



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Πρόγραμμα Μεταπτυχιακών Σπουδών
Ψηφιακά Συστήματα και Υπηρεσίες
Κατεύθυνση: Μεγάλα Δεδομένα και Αναλυτική

Μεταπτυχιακή Διπλωματική Εργασία
Ανάλυση ηχητικών εγγραφών αστικού περιβάλλοντος και αναγνώριση
πλαισίου

Ζαΐρα Φασιανού
Α.Μ. 1710

Επιβλέπων: Ηλίας Μαγκλογιάννης, Καθηγητής

Πειραιάς, 2020

Abstract

At this thesis we tried to apply and evaluate some classification techniques through machine learning algorithms aiming at the automatic recognition of an unknown set of urban sounds. So far, the majority of the researchers have investigate the recognition of the human speech or the categorization of different music types. In this study, we tried to examine how some of the most popular classification algorithms correspond to an urban audio library.

The first step of the process was the creation of an audio library that would include different activities for each of the eight classes studied. Then followed the manipulation of the audio data, such as splitting into the desired lengths of time and further subdividing after applying the frame functions. After labeling every sound data, we convert the audio signal to small dimension vectors of audio features in order to be able to be classified.

The next step was to apply the machine learning algorithms to the processed audio data. The audio attributes were divided into a training and a test set for the estimation of the audio recognition performance. Finally, the machine learning algorithm with the highest accuracy was applied on an unknown set of audio data of varying lengths of time. By this step we tried to investigate its generalization ability.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον κ. Μαγκλογιάννη Ηλία, καθηγητή του τμήματος Ψηφιακών Συστημάτων του Πανεπιστημίου Πειραιώς, για την επίβλεψη και τις συμβουλές του καθ' όλη τη διάρκεια της διπλωματικής μου εργασίας μέχρι την ολοκλήρωσή της.

Περιεχόμενα

| | |
|--|----|
| 1.Εισαγωγή..... | 7 |
| 2.Τεχνολογικό υπόβαθρο - σχετική βιβλιογραφία | 8 |
| 2.1.Ήχος..... | 8 |
| 2.2. Μηχανική Μάθηση για την επεξεργασία και αναγνώριση ήχου | 12 |
| 2.3.Εξόρυξη Δεδομένων..... | 15 |
| 2.4.Ταξινόμηση και Παλινδρόμηση | 16 |
| 2.5.Διασταυρωμένη Επικύρωση – Cross Validation | 17 |
| 2.6.Σχετικές εργασίες..... | 19 |
| 3. Τεχνικές επεξεργασίας και ανάλυσης..... | 26 |
| 3.1.Χαρακτηριστικά ήχου..... | 26 |
| 3.1.1.Χρονικά χαρακτηριστικά ήχου | 27 |
| 3.1.1.1.Ενέργεια – Energy | 27 |
| 3.1.1.2.Ρυθμός μηδενικής διέλευσης - Zero Crossing Rate | 28 |
| 3.1.1.3.Εντροπία ενέργειας - Entropy of energy | 28 |
| 3.1.2.Χαρακτηριστικά συχνότητας ήχου | 29 |
| 3.1.2.1.Φασματικό κέντρο - Spectral Centroid..... | 30 |
| 3.1.2.2.Φασματική διασπορά - Spectral Spread | 30 |
| 3.1.2.3.Φασματική εντροπία - Spectral Entropy | 31 |
| 3.1.2.4.Φασματική ροή - Spectral Flux..... | 31 |
| 3.1.2.5.Κύλιση φάσματος - Spectrum roll off..... | 31 |
| 3.1.2.6.Συντελεστές συχνότητας κλίμακας Mel – MFCCs – Mel frequency cepstral coefficients..... | 32 |
| 3.1.2.7.Διάνυσμα χρώματος - Chroma vector | 33 |
| 3.2.Αλγόριθμοι κατηγοριοποίησης..... | 34 |
| 3.2.1.Κ κοντινότερος γείτονας - KNN – K nearest neighbor | 34 |
| 3.2.2.Μηχανισμοί διανυσμάτων στήριξης - Support vector machines | 37 |
| 3.2.3.Τυχαία δάση - Random forests | 40 |
| 3.2.4.Εξαιρετικά τυχαία δέντρα - Extra Trees - Extremely Randomized Trees..... | 44 |
| 3.2.5.Βαθμωτή ενίσχυση - Gradient boosting..... | 45 |
| 3.3.Μέτρα αξιολόγησης πειραμάτων | 47 |
| 4.Δοκιμές και Αποτελέσματα | 48 |

| | |
|---|----|
| 4.1. Περιγραφή και επεξεργασία των ηχητικών δεδομένων | 48 |
| 4.2.Πείραμα: 8 κλάσεις – παράθυρο 10 sec | 54 |
| 4.3.Πείραμα: 8 κλάσεις – παράθυρο 5 sec | 57 |
| 4.4.Πείραμα: 8 κλάσεις – παράθυρο 25 sec | 59 |
| 4.5.Πείραμα: 4 κλάσεις – παράθυρο10 sec | 63 |
| 5.Συμπεράσματα | 65 |
| Βιβλιογραφία-Πηγές | 66 |

Κατάλογος Πινάκων

| | |
|--|----|
| Πίνακας 1 Πίνακες CM για τη σύγκριση των αλγορίθμων του πρώτου πειράματος | 54 |
| Πίνακας 2 Ακρίβεια ανά κλάση σε τυχαίο σύνολο δεδομένων | 56 |
| Πίνακας 3 Πίνακες CM για τη σύγκριση των αλγορίθμων του δεύτερου πειράματος | 57 |
| Πίνακας 4 Ακρίβεια ανά κλάση σε τυχαίο σύνολο δεδομένων | 59 |
| Πίνακας 5 Πίνακες CM για τη σύγκριση των αλγορίθμων του τρίτου πειράματος..... | 60 |
| Πίνακας 6 Συγκριτική παρουσίαση των μεγεθών accuracy και F_1 όλων των αλγορίθμων για τις χρονικές διάρκειες των 10-5-25 sec για το σύνολο των παραμέτρων | 63 |
| Πίνακας 7 Πίνακες CM για τη σύγκριση των αλγορίθμων του τέταρτου πειράματος | 64 |
| Πίνακας 8 Τα μέτρα απόδοσης ,Precision, Recall και F1, όλων των αλγορίθμων που επιτυγχάνει η βέλτιστη παράμετρος για τις 4 κλάσεις | 64 |
| Πίνακας 9 Ακρίβεια ανά κλάση σε τυχαίο σύνολο δεδομένων | 65 |

Κατάλογος Εικόνων

| | |
|---|----|
| Εικόνα 1 Διάδοση διαμηκών και εγκαρσίων κυμάτων | 10 |
| Εικόνα 2 Τομείς εφαρμογής της εξόρυξης δεδομένων | 15 |
| Εικόνα 3 Στάδια επεξεργασίας δεδομένων | 16 |
| Εικόνα 4 Λάθος εκπαίδευσης και επικύρωσης σε σχέση με τις παραμέτρους | 18 |
| Εικόνα 5 Απεικόνιση της λειτουργίας του KNN αλγορίθμου | 35 |
| Εικόνα 6 Υπερπλάισια σε χώρους χαρακτηριστικών 2 και 3 διαστάσεων | 39 |
| Εικόνα 7 Αριστερά χαμηλή τιμή κανονικοποίησης ενώ δεξιά υψηλή..... | 39 |
| Εικόνα 8 Στην πάνω εικόνα με υψηλό Γάμμα λαμβάνονται υπόψη μόνο τα κοντινά σημεία, ενώ στην κάτω με χαμηλό Γάμμα συνεισφέρουν και τα μακρινά | 40 |
| Εικόνα 9 Το πρώτο «καλό» περιθώριο απέχει εξίσου από τις δύο κλάσεις, αντίθετα το δεύτερο «κακό» περιθώριο βρίσκεται πολύ κοντά στη μπλε κλάση..... | 40 |
| Εικόνα 10 Απεικόνιση του αλγόριθμου Random Forest για αριθμητικά δεδομένα | 41 |
| Εικόνα 11 Δομή του ταξινομητή Random Forest..... | 42 |
| Εικόνα 12 Ψευδοκώδικας του αλγορίθμου Extra Trees | 45 |
| Εικόνα 13 Απεικόνιση του ταξινομητή Gradient Boosting..... | 46 |
| Εικόνα 14 Διαγραμματική απεικόνιση των πέντε ταξινομητών σε σχέση με τις διάφορες παραμέτρους που εξετάστηκαν..... | 55 |
| Εικόνα 15 Μέτρα απόδοσης για την κάθε κλάση έπειτα από την εφαρμογή του ταξινομητή Extra Trees σε αρχεία διάρκειας 10 sec | 56 |
| Εικόνα 16 Απεικόνιση της ακρίβειας (accuracy) και του μέτρου F_1 των πέντε ταξινομητών σε σχέση με τις διάφορες παραμέτρους που εξετάστηκαν | 58 |
| Εικόνα 17 Μέτρα απόδοσης για την κάθε κλάση έπειτα από την εφαρμογή του ταξινομητή Extra Trees σε αρχεία διάρκειας 5 sec | 59 |
| Εικόνα 18 Απεικόνιση της ακρίβειας (accuracy) και του μέτρου F1 των πέντε ταξινομητών σε σχέση με τις διάφορες παραμέτρους που εξετάστηκαν | 61 |

Εικόνα 19 Μέτρα απόδοσης για την κάθε κλάση έπειτα από την εφαρμογή του ταξινομητή Extra Trees σε αρχεία διάρκειας 25 sec 62

1.Εισαγωγή

Ο κόσμος είναι γεμάτος από ήχους. Μετά την όραση, η ακοή είναι η σημαντικότερη αίσθηση, με την οποία αντιλαμβανόμαστε το περιβάλλον. Εκ πρώτης όψεως, όλοι οι ήχοι φαίνεται να είναι διαφορετικοί μεταξύ τους. Εν τούτοις έχουν κάποια όμοια χαρακτηριστικά, όπως το γεγονός ότι μεταφέρουν ενέργεια και κατά συνέπεια μεταδίδουν κάποια πληροφορία από μια περιοχή σ' άλλη, ώστε να επικοινωνούμε με τους άλλους ανθρώπους.

Η παρούσα διπλωματική εργασία πραγματεύεται τη μελέτη και την ανάλυση των ήχων σ' ένα αστικό περιβάλλον. Συγκεκριμένα, ο σκοπός είναι η εφαρμογή και η αξιολόγηση μεθόδων μηχανικής μάθησης πάνω σ' ένα σύνολο περιβαλλοντικών ήχων. Αρκετή έρευνα έχει ήδη διεξαχθεί στον τομέα της ανάλυσης ήχου. Ωστόσο οι περισσότεροι ερευνητές έχουν καταπιαστεί με την αναγνώριση ανθρώπινης ομιλίας είτε με την κατηγοριοποίηση διαφόρων ειδών μουσικής. Στην παρούσα μελέτη θα εξετάσουμε, πως κάποιοι από τους δημοφιλείς αλγορίθμους ταξινόμησης ανταποκρίνονται σε αστικά ηχητικά δεδομένα.

Στο πρώτο στάδιο της διαδικασίας έγινε η συλλογή των δειγμάτων ήχου, με σκοπό τη δημιουργία μιας βιβλιοθήκης, η οποία θα συμπεριλάμβανε πιθανές δραστηριότητες για κάθε μία από τις οχτώ κλάσεις που μελετήθηκαν. Στη συνέχεια, τα ηχητικά δεδομένα καταχωρήθηκαν στην κατάλληλη κλάση, αφού χωρίστηκαν στις επιθυμητές χρονικές διάρκειες και κάθε ηχητικό αρχείο διαιρέθηκε περαιτέρω μετά την εφαρμογή των συναρτήσεων πλαισίου. Επίσης για κάθε ήχο εξάχθηκαν κάποια χαρακτηριστικά, προκειμένου να δημιουργηθούν διανύσματα χαρακτηριστικών συγκεκριμένης διάστασης με τις πληροφορίες, που ήταν απαραίτητες για την ταξινόμηση.

Στο επόμενο στάδιο πραγματοποιήθηκε η εφαρμογή των αλγορίθμων μηχανικής μάθησης στα επεξεργασμένα, πλέον, ηχητικά δεδομένα. Το σύνολο χαρακτηριστικών χωρίστηκε σε σύνολο εκπαίδευσης και σε σύνολο δοκιμής, προκειμένου να αξιολογηθεί η απόδοση του κάθε αλγορίθμου. Επιπλέον, ο αλγόριθμος με την υψηλότερη ακρίβεια δοκιμάστηκε σ' ένα άγνωστο σύνολο ηχητικών δεδομένων διαφορετικής μεταξύ τους χρονικής διάρκειας, για να διερευνηθεί η ικανότητα του να γενικεύει.

Αναλυτικότερα, η οργάνωση της εργασίας είναι η ακόλουθη:

- Κεφάλαιο 1^ο: Το πρώτο κεφάλαιο αποτελεί την εισαγωγή της παρούσας εργασίας
- Κεφάλαιο 2^ο: Στο δεύτερο κεφάλαιο παρουσιάζονται βασικές θεωρητικές έννοιες της μηχανικής μάθησης καθώς και ένα πλήθος σχετικών εργασιών στο πεδίο της ανάλυσης ήχου.
- Κεφάλαιο 3^ο: Το τρίτο κεφάλαιο πραγματεύεται και ορίζει τις τεχνικές επεξεργασίας οι οποίες χρησιμοποιήθηκαν στην παρούσα ανάλυση. Επιπλέον γίνεται η περιγραφή της δημιουργίας του συνόλου ηχητικών δεδομένων.

- Κεφάλαιο 4^ο: Στο τέταρτο κεφάλαιο αναφέρονται οι δοκιμές και τα αποτελέσματα που προέκυψαν από τα πειράματα.
- Κεφάλαιο 5^ο: Το τελευταίο κεφάλαιο περιέχει τα συμπεράσματα της παρούσας ανάλυσης καθώς και κάποιες μελλοντικές κατευθύνσεις με στόχο την βελτίωση των αποτελεσμάτων.

Η βασική συνεισφορά της παρούσας διπλωματικής εργασίας είναι η δημιουργία μιας νέας ηχητικής βιβλιοθήκης και η εύρεση κατάλληλων σχημάτων κατηγοριοποίησης και συνόλου χαρακτηριστικών προκειμένου να επιτύχουν τον καλύτερο διαχωρισμό των κλάσεων.

2.Τεχνολογικό υπόβαθρο - σχετική βιβλιογραφία

2.1.Ήχος

Είναι σχεδόν αδύνατο να φανταστεί κανείς έναν κόσμο χωρίς ήχο. Είναι ίσως το πρώτο πράγμα που νιώθουμε όταν ξυπνάμε το πρωί και πριν κοιμηθούμε το βράδυ. Ο ήχος είναι κυρίαρχος στην καθημερινότητα μας, όταν συνομιλούμε ή όταν ακούμε μουσική. Το περιβάλλον είναι γεμάτο από διαφορετικούς ήχους μερικοί από τους οποίους είναι ευχάριστοι και επιθυμούμε να τους ακούμε (π.χ. η αγαπημένη μας μουσική), ενώ άλλους όχι (π.χ. ο θόρυβος από την κίνηση στον δρόμο ή ένας μονότονος ήχος).

Ένας ήχος μπορεί να είναι είτε απλός/σύνθετος θόρυβος είτε κρότος. Ένας σύνθετος ήχος έχει ένα συγκεκριμένα επαναλαμβανόμενο μοτίβο. Αντίθετα, ένας κρότος είναι ένας ήχος κοφτός με πολύ σύντομη διάρκεια και αντιστοιχεί σε μικρό αριθμό ταλαντώσεων, των οποίων το πλάτος ελαττώνεται απότομα. Αναλυτικότερα, ήχος είναι η ενέργεια που παράγεται, όταν υπάρξει δόνηση. Για παράδειγμα, εάν χτυπήσουμε ένα τύμπανο, η σφιχτή μεμβράνη αρχίζει να δονείται με πολύ μεγάλη ταχύτητα, συχνά η κίνηση είναι τόσο γρήγορη που δεν το βλέπουμε καν, αναγκάζοντας τον αέρα γύρω της να δονείται. Καθώς ο αέρας κινείται, μεταφέρει ενέργεια από το τύμπανο προς όλες τις κατευθύνσεις. Τελικά, ακόμη και ο αέρας μέσα στα αυτιά μας αρχίζει να δονείται. Έτσι, αρχίζουμε να αντιλαμβανόμαστε το δονούμενο τύμπανο ως ήχο. Εν συντομία, υπάρχουν δύο διαφορετικές πτυχές, όσον αναφορά τον ήχο. Μια φυσική διαδικασία που παράγει ηχητική ενέργεια και τη στέλνει δια μέσου του αέρα και υπάρχει και μια ξεχωριστή ψυχολογική διαδικασία που συμβαίνει μέσα στα αυτιά και το μυαλό μας, τα οποία μετατρέπουν την εισερχόμενη ηχητική ενέργεια σε αισθήσεις, τις οποίες ερμηνεύουμε ως θόρυβο, ομιλία και μουσική.

Οι άνθρωποι μπορούν να ακούσουν τα ηχητικά κύματα μόνο ως διακριτούς τόνους, όταν η συχνότητα βρίσκεται μεταξύ περίπου 20 Hz και 20 kHz. Τα ηχητικά κύματα άνω των 20 kHz είναι γνωστά ως υπέρηχοι και δεν είναι αισθητά από τους ανθρώπους. Τα ηχητικά κύματα κάτω από 20 Hz είναι γνωστά ως υπόηχοι. Το εύρος των Hz δεν είναι ίδιο για όλα τα έμβια όντα. Επίσης, το όργανο για την αναπαράσταση των ήχων είναι ο παλμογράφος.

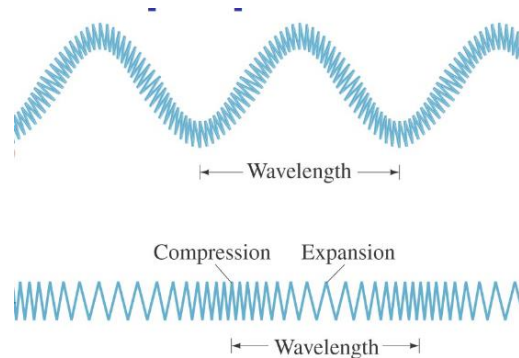
Ο ήχος είναι σαν το φως κατά κάποιο τρόπο, με την έννοια ότι ταξιδεύει από μια συγκεκριμένη πηγή, π.χ. ένα μουσικό όργανο ή ένα θορυβώδες μηχάνημα, ακριβώς όπως και το φως ταξιδεύει από τον ήλιο ή από έναν λαμπτήρα. Αλλά υπάρχουν και κάποιες πολύ σημαντικές διαφορές μεταξύ του φωτός και του ήχου. Γνωρίζουμε ότι το φως μπορεί να ταξιδέψει μέσα στο κενό. Ο ήχος, ωστόσο, δεν μπορεί να ταξιδέψει μέσα στο κενό και πάντα πρέπει να έχει ένα μέσο για να διαδοθεί. Λόγου χάρη ο αέρας, το νερό, το γυαλί ή το μέταλλο είναι μερικά μέσα διάδοσης.

Στην επιστήμη της φυσικής ο αυστηρός ορισμός του ήχου είναι ότι πρόκειται για μια ταλάντωση στην πίεση, την τάση, την μετατόπιση και την ταχύτητα των κινούμενων σωματιδίων που μεταδίδεται σε ένα ελαστικό ή παχύρευστο μέσο. Όπως αναφέρθηκε και παραπάνω, ο ήχος μπορεί να θεωρηθεί ως κίνηση κύματος στον αέρα ή σε άλλα ελαστικά μέσα. Όταν συμβαίνει αυτό, θωρείται ερέθισμα. Τέλος, μπορεί να εξεταστεί ως διέγερση του μηχανισμού ακοής, που έχει ως αποτέλεσμα την αντίληψη του ήχου. Στην περίπτωση αυτή ο ήχος θεωρείται αίσθηση.

Όταν, λοιπόν, ακούμε τον ήχο του ξυπνητηριού, ουσιαστικά ακούμε ενέργεια, η οποία ταξιδεύει. Εκκινεί από κάπου μέσα στο ρολόι, ταξιδεύει στον αέρα και φτάνει λίγο αργότερα στα αυτιά μας. Θυμίζει λίγο τα κύματα που ταξιδεύουν πάνω από τη θάλασσα και ξεκινούν από ένα μέρος, όπου ο άνεμος φυσάει το νερό, ταξιδεύουν πάνω από την επιφάνεια του ωκεανού, και τελικά καταλήγουν στην παραλία. Τα ηχητικά κύματα παράγονται από μια πηγή ήχου, η οποία δημιουργεί κραδασμούς στο περιβάλλον μέσο. Υπάρχουν διάφορα είδη πηγών, οι οποίες προκαλούν τον σχηματισμό των ήχων. Αρχικά, είναι οι ταλαντώσεις. Όταν μια χορδή κιθάρας τεθεί σε παλμική κίνηση, δημιουργεί πυκνώματα και αραιώματα στον περιβάλλοντα αέρα. Αυτό έχει ως συνέπεια τη δημιουργία ενός χαρακτηριστικού ήχου. Επιπρόσθετα, όταν μιλάμε ή τραγουδάμε, οι φωνητικές μας χορδές αυξομειώνουν τα διαστήματα μέσα από τα οποία διέρχεται ο αέρας που βγαίνει από τους πνεύμονες, με αποτέλεσμα να προκαλούνται μεταβολές και τελικά να σχηματίζεται η ομιλία. Κάποιες φορές η πηγή του ήχου μπορεί να είναι θερμική. Για παράδειγμα, ένας ηλεκτρικός σπινθήρας δημιουργεί δυνατό ήχο λόγω της γρήγορης υπερθέρμανσης του αερίου. Τέλος, η πηγή του ήχου μπορεί να είναι υπερηχητική, όπως ο ήχος που προκαλείται από τη γρήγορη εκκίνηση ενός αεροπλάνου. Το ηχητικό κύμα δημιουργείται, καθώς η πηγή δονείται στο μέσο, στο οποίο βρίσκεται και οι δονήσεις αυτές υπό τη μορφή ενέργειας απομακρύνονται με την ταχύτητα του ήχου. Σε σταθερή απόσταση από την πηγή, η πίεση, η ταχύτητα και η μετατόπιση του μέσου διαφοροποιούνται, όσο αλλάζει ο χρόνος.

Υπάρχει, όμως, μια κρίσιμη διαφορά ανάμεσα στα κύματα που κυματίζουν πάνω στη θάλασσα και τα ηχητικά κύματα που φτάνουν στα αυτιά μας. Το νερό μετακινείται προς τα πάνω και προς τα κάτω, χωρίς πραγματικά να μετακινείται οπουδήποτε, καθώς η ενέργεια στο κύμα ταξιδεύει προς τα εμπρός. Κύματα όπως αυτό ονομάζονται εγκάρσια κύματα. Αυτό σημαίνει ακριβώς, ότι το νερό δονείται σε ορθές γωνίες με την κατεύθυνση, στην οποία κινείται το κύμα. Τα ηχητικά κύματα λειτουργούν με εντελώς διαφορετικό τρόπο. Καθώς ένα ηχητικό κύμα κινείται προς τα εμπρός, δημιουργεί ένα εναλλασσόμενο μοτίβο από περιοχές, που έχουν

πιεστεί και άλλες που έχουν τεντωθεί. Με άλλα λόγια, ο ήχος σπρώχνει και τραβά τον αέρα μπροστά και πίσω, ενώ το νερό τον κουνάει πάνω και κάτω. Τα ηχητικά κύματα είναι διαμήκη κύματα, επειδή ο αέρας δονείται κατά την ίδια κατεύθυνση με το κύμα που ταξιδεύει. Τα διαμήκη κύματα, δηλαδή τα κύματα, στα οποία η ταλάντωση του μέσου έχει την ίδια διεύθυνση με εκείνη, στην οποία διαδίδεται ο ήχος, μπορούν επίσης να σχηματιστούν σε μια αέρια στήλη με την κατάλληλη μετακίνηση ενός εμβόλου. Έτσι, προκαλείται μικρή μεταβολή στην πίεση ενός στρώματος αέρα, που γειτονεύει μ' αυτό. Η διάδοση αυτής της μεταβολής στο χώρο αντιστοιχεί με τη διάδοση του ήχου.



Εικόνα 1 Διάδοση διαμηκών και εγκάρσιων κυμάτων

Κατά τη διάδοση τους, τα κύματα μπορούν να ανακλαστούν, να διαθλασθούν ή να εξασθενήσουν. Όταν ένα ξένο σώμα παρεμβάλλεται στην κατεύθυνση των ηχητικών κυμάτων, τότε αυτά υφίστανται ανάκλαση. Αν για παράδειγμα, σταθούμε μπροστά από έναν τοίχο και χτυπήσουμε παλαμάκια, τα κύματα που θα φτάσουν στον τοίχο θα αναγκαστούν να αλλάξουν κατεύθυνση. Αν το εμπόδιο βρίσκεται σε απόσταση μεγαλύτερη των 17 μέτρων, τότε παρατηρούμε, πως ο ήχος που δημιουργήθηκε επαναλαμβάνεται. Αυτό το φαινόμενο ονομάζεται ηχώ. Φυσικά υπάρχει μια καθυστέρηση μεταξύ του ήχου και της ηχώ, επειδή χρειάζεται χρόνος για τον ήχο να φτάσει στον τοίχο και πάλι πίσω. Να αναφερθεί, επίσης, ότι όσο μεγαλύτερη είναι η απόσταση, τόσο μεγαλύτερη είναι και η καθυστέρηση. Επίσης, τα ηχητικά κύματα χάνουν ενέργεια, καθώς ταξιδεύουν. Αυτός είναι ο λόγος για τον οποίο οι ήχοι «ταξιδεύουν λιγότερο καλά» τις ημέρες που έχει κακοκαιρία από εκείνες τις μέρες με καλό καιρό, αφού ο άνεμος διαλύει την ενέργειά τους. Ένας ακόμη παράγοντας που επηρεάζει το πόσο γρήγορα θα εξασθενίσει ένα κύμα είναι το πόσο κολλώδες είναι το μέσο διάδοσης. Για τα περισσότερα μέσα, όπως είναι αέρας ή το νερό, η εξασθένιση του ήχου λόγω κολλώδους μέσου είναι σχεδόν αμελητέα. Ένα ακόμη χαρακτηριστικό των κυμάτων είναι ότι εξαπλώνονται στον χώρο και ταξιδεύοντας κάμπτουν γύρω από γωνίες. Αυτό ονομάζεται περίθλαση κυμάτων. Κάθε κύμα μπορεί να είναι μεγάλο ή μικρό. Τα μεγάλα ηχητικά κύματα έχουν υψηλό πλάτος ή ένταση. Γι' αυτό τα ακούμε σαν πιο δυνατούς ήχους. Εκτός από το πλάτος, αξίζει να σημειωθεί πως το βήμα των ηχητικών κυμάτων ονομάζεται και συχνότητα. Η συχνότητα είναι απλά ο αριθμός των κυμάτων, που παράγεται σ' ένα δευτερόλεπτο.

Πρέπει να αναφερθεί σ' αυτό το σημείο, ότι τα κύματα δεν είναι πανομοιότυπα. Ένα όργανο ή μια ανθρώπινη φωνή παράγει ένα ολόκληρο μείγμα διαφορετικών κυμάτων ταυτόχρονα. Υπάρχει ένα βασικό κύμα μ' ένα ορισμένο πλάτος και ύψος, που ονομάζεται θεμελιώδες και πάνω απ' αυτό υπάρχουν πολλοί υψηλότεροι ήχοι, που ονομάζονται αρμονικοί ή υπερτονικοί. Όλα αυτά τα κύματα προστίθενται μαζί, για να δώσουν ένα μοναδικό σχήμα στο ηχητικό κύμα, που παράγεται κι αυτός είναι ένας λόγος, για τον οποίο ακούγονται διαφορετικά. Ο άλλος λόγος είναι, ότι το πλάτος των κυμάτων από ένα συγκεκριμένο όργανο αλλάζει με ένα μοναδικό τρόπο, καθώς περνούν τα δευτερόλεπτα.

Αφού έχει επισημανθεί, ότι ο ήχος μεταφέρει ενέργεια σ' ένα μοτίβο κυμάτων, μπορεί να παρατηρηθεί, ότι η ταχύτητα του ήχου σημαίνει η ταχύτητα με την οποία κινούνται τα κύματα, δηλαδή η ταχύτητα με την οποία η ενέργεια ταξιδεύει ανάμεσα σε δύο μέρη. Όταν λέμε ότι ένα αεροπλάνο «σπάει το ηχητικό φράγμα», εννοούμε ότι επιταχύνει τόσο γρήγορα, ώστε να ξεπερνά τα απίστευτα μεγάλης έντασης ηχητικά κύματα, που παράγουν οι κινητήρες του, δημιουργώντας ένα φρικτό θόρυβο. Η ταχύτητα του ήχου στον αέρα είναι περίπου 1220 χλμ. / ώρα. Σε σύγκριση με τα φωτεινά κύματα, τα ηχητικά κύματα διαδίδονται εξαιρετικά αργά (περίπου ένα εκατομμύριο φορές πιο αργά). Για παράδειγμα, βλέπουμε τους κεραυνούς πολύ νωρίτερα από ότι τους ακούμε, επειδή τα φωτεινά κύματα φτάνουν άμεσα, ενώ τα ηχητικά κύματα χρειάζονται περίπου 5 δευτερόλεπτα για να καλύψουν 1,6 χιλιόμετρα.

Η διάδοση του ήχου επηρεάζεται από πολλούς συντελεστές. Αν το μέσο διάδοσης κινείται, αυτή η κίνηση μπορεί να αυξήσει ή να μειώσει την ταχύτητα του ηχητικού κύματος ανάλογα με την κατεύθυνση της κίνησης. Για παράδειγμα, αν ο ήχος που διαδίδεται μέσω του ανέμου έχει την ίδια κατεύθυνση με τον άνεμο, τότε η ταχύτητα διάδοσης θα αυξάνεται. Ενώ εάν ο ήχος κι ο άνεμος κινούνται σε αντίθετες κατευθύνσεις, η ταχύτητα του ηχητικού κύματος θα μειωθεί εξαιτίας της αντίθετης ταχύτητας του ανέμου. Επίσης, ο ήχος ταξιδεύει με διαφορετικές ταχύτητες σε στερεά, υγρά και αέρια κι ακόμη κι η ταχύτητά του σ' ένα υλικό μπορεί να αλλάξει. Επιπλέον, διαφέρει ανάλογα με την πυκνότητα του μέσου. Για παράδειγμα, ο ήχος ταξιδεύει περίπου 15 φορές γρηγορότερα στον χάλυβα απ' ότι στον αέρα και περίπου 4 φορές πιο γρήγορα στο νερό απ' ότι στον αέρα. Πιο συγκεκριμένα, υπάρχει μια αναλογία ανάμεσα στην πυκνότητα και την πίεση του μέσου, η οποία καθορίζει την ταχύτητα. Επιπρόσθετα, ο ήχος ταξιδεύει με διαφορετικές ταχύτητες σε διαφορετικά αέρια και μπορεί να φτάσει σε διαφορετικές ταχύτητες, ακόμα και στο ίδιο αέριο. Ταξιδεύει πολύ πιο γρήγορα σε θερμό αέρα κοντά στο έδαφος παρά σε ψυχρό αέρα ψηλότερα. Ακόμη, ταξιδεύει περίπου τρεις φορές ταχύτερα σε αέριο ηλίου απ' ότι στον απλό αέρα, επειδή το ήλιο είναι λιγότερο πυκνό. Αυτός είναι ο λόγος για τον οποίο οι άνθρωποι που αναπνέουν το ήλιο μιλούν με «λεπτή φωνή», αφού τα ηχητικά κύματα, δηλαδή οι φωνές τους, κάνουν τη διαδρομή ταχύτερα, με υψηλότερη συχνότητα. Στην πραγματικότητα, όμως, ο τρόπος που ο ήχος ταξιδεύει μέσω διαφορετικών υλικών είναι πολύ πιο περίπλοκος απ' αυτό και ουσιαστικά θα έπρεπε να συμπεριληφθεί και ο ρόλος των φωτονίων εκτός από την πυκνότητα.

Συχνά επικρατεί η λανθασμένη άποψη ότι τα σωματίδια του μέσου ταξιδεύουν μαζί με το ηχητικό κύμα, το οποίο φυσικά δεν ισχύει. Στη περίπτωση του στερεού υλικού, είναι εύκολο να

το φανταστεί κανείς, αλλά το ίδιο ακριβώς συμβαίνει στα υγρά και τα αέρια. Αυτό πρακτικά σημαίνει, ότι τα σωματίδια στο αέριο ή το υγρό μέσο μεταφέρουν τη δόνηση, αλλά ταυτόχρονα η μέση θέση των σωματιδίων μένει σταθερή, όσο περνά ο χρόνος.

Ο ήχος είναι ένα εξαιρετικά σημαντικό κομμάτι της ζωής στη Γη. Τα περισσότερα ζώα ακούν σε ήχους που σηματοδοτούν τη δυνατότητα φαγητού. Πολλά πλάσματα ανταλλάσσουν, επίσης, σημαντικούς ήχους είτε για να επικοινωνούν με μέλη του ίδιου είδους είτε για να προειδοποιήσουν τους αντιπάλους τους. Οι άνθρωποι έχουν εξελίξει αυτήν την ικανότητα στην ομιλούμενη γλώσσα, ως τρόπο ανταλλαγής πληροφοριών και στη μουσική ως ένα ηχητικό σύστημα για την επικοινωνία των συναισθημάτων.

Έχουμε ακόμη αναπτύξει μια ποικιλία διαφορετικών τεχνολογιών ήχου. Έχουν επινοηθεί μουσικά όργανα, που μπορούν να δημιουργήσουν μια τεράστια ποικιλία διαφορετικών μουσικών ήχων. Επιπλέον, καταγράφονται ήχοι σε δίσκους CD ή σε νεότερες τεχνολογίες, όπως MP3. Στην ιατρική, επιπροσθέτως, χρησιμοποιούνται ήχοι πολύ υψηλής συχνότητας, γνωστούς ως υπερήχους, για τον καθαρισμό χαλασμένων δοντιών μέχρι τη μελέτη ενός μωρού μέσα στη μήτρα μιας μητέρας. Έχουν επίσης διδαχθεί υπολογιστές, που μετατρέπουν τα ηχητικά σήματα σε γραπτό λόγο χρησιμοποιώντας λογισμικό αναγνώρισης φωνής. [30] [31]

2.2. Μηχανική Μάθηση για την επεξεργασία και αναγνώριση ήχου

Η μηχανική μάθηση είναι μια εφαρμογή της τεχνητής νοημοσύνης (Artificial Intelligence), αφού η μάθηση αποτελεί αναπόσπαστη ιδιότητα ενός ευφυούς συστήματος και παρέχει στα συστήματα τη δυνατότητα να μαθαίνουν και να βελτιώνονται αυτόματα από την εμπειρία, χωρίς να έχουν ρητά προγραμματιστεί. Η έρευνα στη μηχανική μάθηση έχει ξεκινήσει από τη δεκαετία του 1940, όταν προτάθηκε το πρώτο μοντέλο μάθησης με βάση το νευρικό σύστημα, γνωστό και ως perceptron (McCulloch and Pitts, 1943). Από τότε και ειδικότερα τα τελευταία 25 χρόνια έχει σημειωθεί μεγάλη πρόοδος.

Η μηχανική μάθηση επικεντρώνεται στην ανάπτυξη προγραμμάτων ηλεκτρονικών υπολογιστών, που μπορούν να έχουν πρόσβαση στα δεδομένα και να τα χρησιμοποιούν, για να εκπαιδεύσουν τον εαυτό τους. Η διαδικασία μάθησης ξεκινά με παρατηρήσεις ή δεδομένα, όπως παραδείγματα, άμεση εμπειρία ή διδασκαλία, προκειμένου να δημιουργηθούν μοτίβα στα δεδομένα και να λάβουμε καλύτερες αποφάσεις στο μέλλον με βάση τα παραδείγματα που παρέχονται. Ο πρωταρχικός στόχος είναι να επιτρέπεται στους υπολογιστές να μαθαίνουν αυτόματα χωρίς την ανθρώπινη παρέμβαση ή βοήθεια και να προσαρμόζουν τις ενέργειες τους ανάλογα. [22]

Η μηχανική μάθηση έχει βρει εφαρμογές σε αρκετές επιστημονικές περιοχές, όπως η μηχανική όραση, η επεξεργασία της γλώσσας, οι μηχανές αναζήτησης παγκόσμιου ιστού, τα ιατρικά πορίσματα, τα οικονομικά ζητήματα, η διάταξη αλληλουχιών DNA στη Βιολογία, η ταυτοποίηση φωνής και η αναγνώριση χαρακτήρων, η ανάπτυξη ευφυών συστημάτων, η ανάπτυξη

αποτελεσματικότερου λογισμικού, η αναγνώριση αντικειμένων στον κλάδο της αστρονομίας, και φυσικά η ρομποτική.

Ένα βασικό πρόβλημα των αλγορίθμων μηχανικής μάθησης είναι η διαμόρφωση και εκπαίδευση τους. Συχνά χρησιμοποιούν στατιστικά στοιχεία, για να βρουν μοτίβα σε μεγάλο πλήθος δεδομένων. Τα δεδομένα αυτά, περιλαμβάνουν πολλά στοιχεία πχ αριθμούς, λέξεις, εικόνες, κλικ ιστοσελίδων. Εάν υπάρχει η δυνατότητα τα δεδομένα να αποθηκευτούν ψηφιακά, τότε μπορούν να τροφοδοτηθούν σ' έναν αλγόριθμο μηχανικής μάθησης. [23]

Οι αλγόριθμοι μηχανικής μάθησης συχνά κατηγοριοποιούνται ως επιτηρούμενοι/εποπτευόμενοι ή μη επιτηρούμενοι/μη εποπτευόμενοι. Οι εποπτευόμενοι αλγόριθμοι μηχανικής μάθησης (Supervised machine learning algorithms) μπορούν να εφαρμόσουν όσα έχουν μάθει στο παρελθόν σε νέα δεδομένα, χρησιμοποιώντας κατηγοριοποιημένα δεδομένα, για να προβλέψουν μελλοντικά γεγονότα. Ξεκινώντας από την ανάλυση ενός γνωστού συνόλου εκπαίδευσης, ο αλγόριθμος εκμάθησης παράγει μια συναγόμενη συνάρτηση, για να κάνει προβλέψεις σχετικά με τις τιμές εξόδου. Το σύστημα είναι σε θέση να παρέχει στόχους για κάθε νέα είσοδο μετά από επαρκή εκπαίδευση. Ο αλγόριθμος εκμάθησης μπορεί επίσης να συγκρίνει το αποτέλεσμα με το σωστό προβλεπόμενο αποτέλεσμα, ώστε να βρει λάθη, για να τροποποιήσει ανάλογα το μοντέλο. Στην εποπτευόμενη μάθηση περιλαμβάνονται δύο κύριες διαδικασίες, η ταξινόμηση και η παλινδρόμηση. Οι δύο αυτές μέθοδοι αναλύονται διεξοδικά σε επόμενη ενότητα. Οι πιο γνωστοί αλγόριθμοι μη επιτηρούμενης μάθησης είναι οι εξής: η γραμμική παλινδρόμηση (Linear Regression), η λογιστική παλινδρόμηση (Logistical Regression), η τυχαία δάση (Random Forest), Gradient Boosted Trees, οι μηχανισμοί διανυσμάτων υποστήριξης (Support Vector Machines -SVM), τα νευρωνικά δίκτυα (Neural Networks, τα δέντρα αποφάσεων (Decision Trees), τα μπευζιανά συστήματα (Naive Bayes) και οι πλησιέστεροι γείτονες (Nearest Neighbor).

Αντίθετα, οι μη εποπτευόμενοι αλγόριθμοι μηχανικής μάθησης (unsupervised machine learning algorithms) εμφανίζονται, όταν οι πληροφορίες, που χρησιμοποιούνται για την εκπαίδευση δεν ταξινομούνται ούτε επισημαίνονται. Το ενδιαφέρον στους μη εποπτευόμενους αλγορίθμους είναι πώς τα συστήματα μπορούν να διαγνώσουν μια συνάρτηση, η οποία περιγράφει την κρυφή δομή των μη επισημασμένων δεδομένων. Το σύστημα δεν υπολογίζει τη σωστή έξοδο, αλλά διερευνά τα δεδομένα και μπορεί να αντλεί συμπεράσματα από σύνολα δεδομένων, για να περιγράψει κρυμμένες δομές από μη επισημασμένα δεδομένα. Οι μη εποπτευόμενοι αλγόριθμοι μάθησης εφαρμόζουν την συσταδοποίηση (clustering) και την μείωση διαστάσεων (Dimensionality reduction) σαν τεχνικές για να περιγράψουν τα δεδομένα. Στη συσταδοποίηση, ένα σύνολο δεδομένων εισόδου πρόκειται να χωριστεί σε ομάδες. Σε αντίθεση με την ταξινόμηση, οι ομάδες δεν είναι γνωστές εκ των προτέρων, καθιστώντας αυτόν τον διαχωρισμό τυπική εργασία της μη επιτηρούμενης μάθησης. Σχετικά με τη μείωση των διαστάσεων είναι γνωστό, ότι υπάρχει πολύ θόρυβος στα εισερχόμενα δεδομένα. Οι αλγόριθμοι μηχανικής μάθησης χρησιμοποιούν τη μείωση των διαστάσεων για την εξάλειψη αυτού του θορύβου, καθώς απομονώνουν τη σχετική πληροφορία. Οι πιο ευρέως χρησιμοποιούμενοι αλγόριθμοι μη εποπτευόμενης μηχανικής μάθησης είναι οι k-means clustering (ομαδοποίηση K μέσου) , t-SNE

(t-Distributed Stochastic Neighbor Embedding), PCA (Principal Component Analysis - ανάλυση βασικών στοιχείων) και Association rules. (Κανόνες συσχέτισης).

Ακόμη, οι ημι-εποπτευόμενοι αλγόριθμοι (Semi-supervised machine learning algorithms) εμπίπτουν κάπου μεταξύ της εποπτευόμενης και της μη εποπτευόμενης μάθησης, δεδομένου ότι χρησιμοποιούν τόσο επισημασμένα όσο και μη επισημασμένα δεδομένα για εκπαίδευση (συνήθως ένα μικρό ποσοστό επισημασμένων και ένα μεγάλο ποσό μη επισημασμένων δεδομένων). Τα συστήματα, που χρησιμοποιούν αυτή τη μέθοδο, είναι σε θέση να βελτιώσουν σημαντικά την ακρίβεια της μάθησης.

Οι αλγόριθμοι ενισχυτικής μάθησης (Reinforcement machine learning algorithms) είναι μια μέθοδος μάθησης που αλληλοεπιδρά με το περιβάλλον της, δημιουργώντας δράσεις και ανακαλύπτοντας λάθη ή ανταμοιβές. Η αναζήτηση σωστού και λάθους και η καθυστερημένη ανταμοιβή είναι τα πιο συναφή χαρακτηριστικά στην ενισχυτική μάθηση. Αυτή η μέθοδος επιτρέπει στους μηχανισμούς μάθησης, να προσδιορίζουν αυτόματα την ιδανική συμπεριφορά μέσα σ' ένα συγκεκριμένο πλαίσιο, προκειμένου να μεγιστοποιήσουν την απόδοσή τους. Ουσιαστικά, αναζητείται κάθε φορά η εύρεση της βέλτιστης συμπεριφοράς ενός μηχανισμού με βάση την ανταμοιβή, που παίρνει σε μια τελική κατάσταση σε κάποιο περιβάλλον, έχοντας ξεκινήσει από μια αρχική κατάσταση και ακολουθώντας μια σειρά από ενέργειες. Στην ενισχυτική μάθηση, δύο είναι οι κύριοι τύποι ανταμοιβών. Ο πρώτος είναι το σήμα της θετικής ανταμοιβής, το οποίο ενθαρρύνει μια συγκεκριμένη ακολουθία δράσεων και ο δεύτερος είναι το αρνητικό σήμα ανταμοιβής, που τιμωρεί για την εκτέλεση ορισμένων δραστηριοτήτων και παροτρύνει τον αλγόριθμο, να διορθωθεί, προκειμένου να σταματήσει να παίρνει ποινές. Ωστόσο, η συνάρτηση του σήματος ανταμοιβής μπορεί να ποικίλει ανάλογα με τη φύση των πληροφοριών. Επομένως, τα σήματα ανταμοιβής μπορούν να ταξινομηθούν περαιτέρω ανάλογα με τις απαιτήσεις. Συνολικά, το σύστημα προσπαθεί να μεγιστοποιήσει τις θετικές ανταμοιβές και να ελαχιστοποιήσει τις αρνητικές. Οι πιο συνηθισμένοι αλγόριθμοι ενισχυτικής μάθησης περιλαμβάνουν τους Q-Learning, Temporal Difference (TD), Monte-Carlo Tree Search (MCTS) και Asynchronous Actor-Critic Agents (A3C). [22][24]

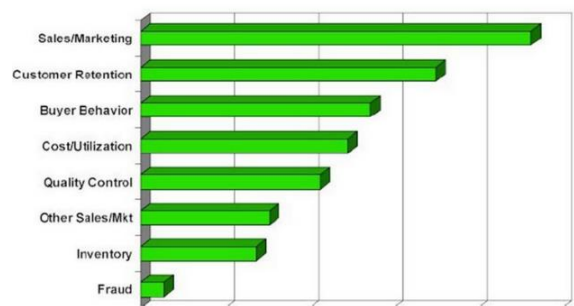
Τα νευρωνικά δίκτυα εμπλέκονται στη περίπτωση, όπου μελετάται η σχέση μεταξύ των ανεξάρτητων και εξαρτημένων μεταβλητών, ακόμη και όταν αυτή η σχέση είναι έντονα μη γραμμική και περίπλοκη, όπως σε προβλέψεις, παλινδρόμηση, μοντελοποίηση δεδομένων, σημαντικότητα μεταβλητών, προσομοίωση συναρτήσεων, μερικές διαφορικές εξισώσεις κλπ. Τα τελευταία χρόνια παρατηρείται έντονη η ενασχόληση ερευνητών και επαγγελματιών με τα νευρωνικά δίκτυα, διότι εφαρμόζονται με επιτυχία σ' ένα ασυνήθιστα ευρύ φάσμα τομέων της επιστήμης και της τεχνολογίας. Η επιτυχία τους πηγάζει από τη θεωρία της ακριβούς προσέγγισης οποιασδήποτε συνεχούς συνάρτησης σε συμπαγή πεδία. Γενικά θεωρείται η επιστημονική περιοχή, η οποία πραγματεύεται τον σχεδιασμό και την ανάπτυξη αλγορίθμων, με την χρήση των οποίων είναι εφικτή η υλοποίηση συστημάτων (π.χ. υπολογιστών) με ικανότητα μάθησης. Ο ορισμός της έννοιας «μάθηση» στη συγκεκριμένη περίπτωση ίσως αποδειχτεί πρόβλημα δυσεπίλυτο. Μάλλον ο μόνος ικανοποιητικός ορισμός είναι ότι μάθηση

είναι οποιαδήποτε αλλαγή σ' ένα σύστημα, που θα του επιτρέψει να είναι αποτελεσματικότερο την επόμενη φορά, που θα επαναλάβει την ίδια εργασία ή κάποια άλλη παρόμοια.

Γενικά, η εκμάθηση μηχανών επιτρέπει την ανάλυση μαζικών ποσοτήτων δεδομένων. Παρόλο που παρέχει ταχύτερα και ακριβέστερα αποτελέσματα, προκειμένου να εντοπίσει επικερδείς ευκαιρίες ή κινδύνους, μπορεί να απαιτήσει επιπλέον χρόνο και πόρους για να εκπαιδευτεί σωστά. Ο συνδυασμός της μηχανικής μάθησης με το AI και τις γνωστικές τεχνολογίες μπορεί να καταστήσει ακόμη πιο αποτελεσματική την επεξεργασία μεγάλου όγκου πληροφοριών [22].

2.3.Εξόρυξη Δεδομένων

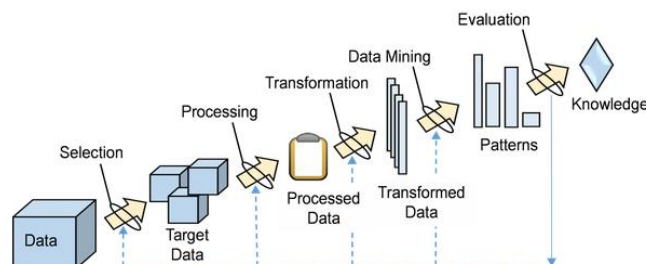
Η εξόρυξη δεδομένων είναι η διαδικασία για την ανεύρεση ενδιαφέρουσας γνώσης από μεγάλο πλήθος δεδομένων (Han και Kamber, 2000). Χρησιμοποιείται ευρέως σε πολλούς τομείς, όπως στο λιανικό εμπόριο, τα οικονομικά, τις τηλεπικοινωνίες και τα κοινωνικά μέσα. Είναι μία περιοχή που έχει μεγάλη συσχέτιση με τη μηχανική μάθηση και προέρχεται από το πεδίο των βάσεων δεδομένων. Πραγματεύεται παρόμοιες μεθόδους, αλλά βασικός άξονας είναι οι εφαρμογές σε μεγάλες βάσεις δεδομένων. Γενικότερα, η εξόρυξη δεδομένων θεωρείται το κεντρικό στοιχείο της διαδικασίας «Ανακάλυψη Γνώσης από Βάσεις Δεδομένων» (Knowledge discovery from Databases - KDD). Είναι ένας διεπιστημονικός τομέας με συμβολές από πολλούς τομείς, όπως είναι η στατιστική, η μηχανική μάθηση, η ανάκτηση πληροφοριών, η αναγνώριση προτύπων και η βιοπληροφορική. [25]



Εικόνα 2 Τομείς εφαρμογής της εξόρυξης δεδομένων

Ο όγκος των διαθέσιμων δεδομένων είναι ένας κρίσιμος παράγοντας, δεδομένου ότι πρόκειται να ανακαλυφθούν μοτίβα σε διαδοχικά ή μη διαδοχικά δεδομένα. Με μια συνολική ματιά, φαίνεται να είναι ένα ολόκληρο χάος. Γι' αυτό η εξόρυξη δεδομένων δεν είναι απλή υπόθεση. Χρειάζεται ένα ορισμένο χρονικό διάστημα και απαιτεί ειδική μεταχείριση. Τα βασικά βήματα της εξόρυξης δεδομένων περιλαμβάνουν τη συλλογή των δεδομένων, τον καθαρισμό, την ανάλυση και τελικά την ερμηνεία τους. Το πρώτο βήμα είναι η συλλογή ορισμένων δεδομένων. Όσο μεγαλύτερο είναι το πλήθος των πληροφοριών τόσο ευκολότερη θα είναι η ανάλυση αργότερα. Πρέπει να διασφαλιστεί, ότι η πηγή των δεδομένων είναι αξιόπιστη. Φυσικά, αφού

υπάρχει ένας μεγάλος όγκος πληροφοριών, πρέπει να κατοχυρωθεί ότι θα παραμείνει μόνο η απαραίτητη πληροφορία και θα αφαιρεθεί οτιδήποτε ανεπιθύμητο. Διαφορετικά, μπορεί να οδηγήσουν σε ψευδή συμπεράσματα. Σειρά έχει η ανάλυση των δεδομένων και η εύρεση των μοτίβων τους. Τέλος, τα δεδομένα που αναλύονται και ερμηνεύονται, ώστε να προκύψουν σημαντικά συμπεράσματα, όπως είναι οι προβλέψεις. [26]



Εικόνα 3 Στάδια επεξεργασίας δεδομένων

Υπάρχουν διάφορα είδη μοντέλων που σχετίζονται με την εξόρυξη δεδομένων. Στην περιγραφική μοντελοποίηση (Descriptive modeling), μελετώνται οι ομοιότητες μεταξύ των συλλεγόμενων δεδομένων και οι λόγοι πίσω απ' αυτές. Αυτό είναι πολύ σημαντικό για την κατασκευή του τελικού συμπεράσματος από το σύνολο δεδομένων. Στην προβλεπτική μοντελοποίηση (Predictive modeling) αναλύονται τα δεδομένα του παρελθόντος για την πρόβλεψη της μελλοντικής συμπεριφοράς. Τα προηγούμενα δεδομένα δίνουν κάποιου είδους καθοδήγηση για τα μελλοντικά συμπεράσματα. Βεβαίως, απαιτείται σωστή επεξεργασία, φιλτράρισμα και σωστή ανάλυση των δεδομένων, για να δημιουργηθούν τέτοια προγνωστικά μοντέλα. [26]

2.4. Ταξινόμηση και Παλινδρόμηση

Όπως αναλυτικά περιεγράφηκε στην παραπάνω ενότητα, ένας εποπτευόμενος αλγόριθμος μηχανικής μάθησης βασίζεται σε δεδομένα εισόδου, που είναι επισήμασμένα με μια κατάλληλη ετικέτα. Χρησιμοποιώντας αυτά τα δεδομένα, εκπαιδεύει μια συνάρτηση, ώστε αυτή να μπορεί με τη σειρά της να κατηγοριοποιεί με επιτυχία κάποια νέα, άγνωστα, μη επισήμασμένα δεδομένα. Ένας επιμέρους διαχωρισμός των αλγορίθμων εποπτευόμενης μάθησης είναι η ταξινόμηση και η παλινδρόμηση, που συνήθως χρησιμοποιούνται στην εξόρυξη δεδομένων, καθώς και σε προβλήματα πρόβλεψης. Η διαφορά μεταξύ των αλγορίθμων παλινδρόμησης και των αλγορίθμων ταξινόμησης προκαλεί αρκετές φορές σύγχυση στους επιστήμονες δεδομένων, διότι η λανθασμένη χρήση τους οδηγεί σε λανθασμένες μεθοδολογίες για την επίλυση των προβλημάτων πρόβλεψης. Η κατανόηση για το αν η μηχανική μάθηση είναι ένα πρόβλημα παλινδρόμησης ή ταξινόμησης είναι το κλειδί για την επιλογή του σωστού αλγορίθμου που θα χρησιμοποιηθεί.

Η ομοιότητα της παλινδρόμησης και της ταξινόμησης είναι, ότι ανήκουν κάτω από την ίδια ομπρέλα της εποπτευόμενης μηχανικής μάθησης, όπως έχει αναφερθεί και παραπάνω. Επίσης

αμφότεροι οι μέθοδοι μοιράζονται την ίδια λογική χρησιμοποιώντας γνωστά σύνολα δεδομένων, που αναφέρονται σαν σύνολα εκπαίδευσης, για να κάνουν προβλέψεις. Στην εποπτευόμενη μάθηση, χρησιμοποιείται ένας αλγόριθμος, για να μάθει τη συνάρτηση χαρτογράφησης από τη μεταβλητή εισόδου (x) στην μεταβλητή εξόδου (y). δηλαδή $y = f(x)$. Ο στόχος ενός τέτοιου προβλήματος είναι να προσεγγίσει όσο γίνεται καλύτερα την συνάρτηση χαρτογράφησης (f), ώστε κάθε φορά που υπάρχει ένα νέο στοιχείο εισόδου (x), η μεταβλητή εξόδου (y) να μπορεί να προβλεφθεί. Η κύρια διαφορά μεταξύ τους είναι, ότι η μεταβλητή εξόδου στην παλινδρόμηση είναι αριθμητική, δηλαδή συνεχής, ενώ στην ταξινόμηση είναι κατηγορηματική, δηλαδή διακριτική.

Στην παλινδρόμηση, οι αλγόριθμοι προσπαθούν να υπολογίσουν τη συνάρτηση χαρτογράφησης (f) από τις μεταβλητές εισόδου (x) σε αριθμητικές ή συνεχείς μεταβλητές εξόδου (y). Σ' αυτήν την περίπτωση, το y είναι μια πραγματική τιμή, η οποία μπορεί να πάρει ακέραιες ή δεκαδικές τιμές. Επομένως, τα προβλήματα παλινδρόμησης είναι συνήθως ποσότητες ή μεγέθη. Η παλινδρόμηση είναι μια διαδικασία εύρεσης ενός μοντέλου ή μιας συνάρτησης για τη διάκριση των δεδομένων σε συνεχείς πραγματικές τιμές. Σύμφωνα με τους μαθηματικούς κανόνες, σ' ένα πρόβλημα παλινδρόμησης, κάποιος προσπαθεί να βρει την προσέγγιση της συνάρτησης με την ελάχιστη απόκλιση σφάλματος. Για παράδειγμα, το ύψος είναι η ανεξάρτητη μεταβλητή και το βάρος είναι η εξαρτημένη μεταβλητή.

Από την άλλη πλευρά, οι αλγόριθμοι ταξινόμησης προσπαθούν να υπολογίσουν τη συνάρτηση χαρτογράφησης (f) από τις μεταβλητές εισόδου (x) σε διακριτές ή κατηγορικές μεταβλητές εξόδου (y). Σ' αυτήν την περίπτωση, το y είναι μια κατηγορία που προβλέπει η συνάρτηση χαρτογράφησης. Εάν παρέχεται με μία ή περισσότερες μεταβλητές εισόδου, ένα μοντέλο ταξινόμησης θα επιχειρήσει να προβλέψει την τιμή ενός ή περισσότερων συμπερασμάτων. Αναλυτικότερα, η ταξινόμηση είναι η διαδικασία εύρεσης ενός μοντέλου ή αλλιώς συνάρτησης, η οποία διαχωρίζει τα δεδομένα σε πολλές κατηγορικές κατηγορίες. Κατά την ταξινόμηση, προσδιορίζεται η ιδιότητα του μέλους στο πρόβλημα, πράγμα που σημαίνει ότι τα δεδομένα ταξινομούνται με διαφορετικές ετικέτες, σύμφωνα με ορισμένες παραμέτρους και κατόπιν οι ετικέτες προβλέπονται για τα νέα δεδομένα εισόδου. [27]

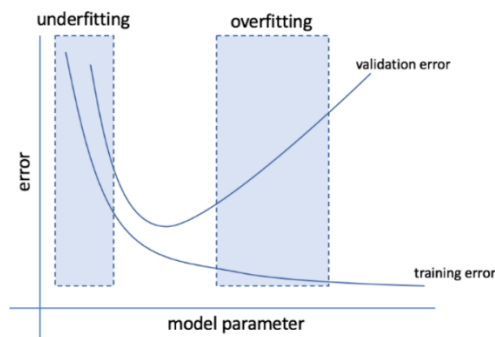
Για την αξιολόγηση των δύο μεθόδων, να αναφερθεί ότι στην περίπτωση της παλινδρόμησης η εκτίμηση γίνεται μέσω του τετραγωνικού σφάλματος ρίζας. Ενώ η ταξινόμηση αξιολογείται με ακρίβεια μέτρησης. Γενικότερα, πρέπει πρωτίστως να εξεταστούν τρεις σημαντικές πτυχές: η ευκολία στην ερμηνεία του αποτελέσματος της μεθόδου, ο απαιτούμενος χρόνος υπολογισμού και η προγνωστική δύναμη του αλγορίθμου.

2.5. Διασταυρωμένη Επικύρωση – Cross Validation

Η τεχνική cross - validation είναι μια στατιστική μέθοδος, που χρησιμοποιείται για την εκτίμηση της ικανότητας των μοντέλων μηχανικής μάθησης. Η επικύρωση είναι μάλλον μία από τις πιο σημαντικές τεχνικές, που χρησιμοποιεί ένας επιστήμονας δεδομένων, αφού υπάρχει πάντα η

ανάγκη να κατακυρωθεί η σταθερότητα του μοντέλου μηχανικής μάθησης και να εξεταστεί το πόσο καλά θα γενικευόταν σε κάποια νέα δεδομένα. Συνήθως χρειαζόμαστε μια διαβεβαίωση, ότι το μοντέλο έχει καταλάβει τα περισσότερα από τα μοτίβα των δεδομένων σωστά και ότι δεν έχει συμπεριλάβει πολύ θόρυβο. Με απλά λόγια, πρέπει να σιγουρευτούμε ότι το μοντέλο έχει το χαμηλότερο συστηματικό σφάλμα (bias) και διακύμανση (variance). [28]

Πρόκειται για μια τεχνική επικύρωσης για την αξιολόγηση του τρόπου γενίκευσης των αποτελεσμάτων μιας στατιστικής ανάλυσης σ' ένα ανεξάρτητο σύνολο δεδομένων. Χρησιμοποιείται κυρίως σε περιπτώσεις, στις οποίες ο στόχος είναι η πρόβλεψη και κάποιος θέλει να υπολογίσει με ακρίβεια πόσο εύστοχο είναι το μοντέλο. Ο στόχος της διασταυρούμενης επικύρωσης είναι ο καθορισμός ενός συνόλου δεδομένων για τη δοκιμή του μοντέλου στη φάση της εκπαίδευσης, το οποίο ονομάζεται σύνολο επικύρωσης, προκειμένου να περιοριστούν προβλήματα, όπως η υπερπροσαρμογή (overfitting), η ελλιπής προσαρμογή (underfitting) και να αποκτηθεί μια εικόνα για το πώς το μοντέλο θα γενικευθεί σ' ένα ανεξάρτητο σύνολο δεδομένων. Είναι σημαντικό τα σύνολα επικύρωσης και εκπαίδευσης να προέρχονται από την ίδια κατανομή. Με τον όρο της ελλιπούς προσαρμογής (underfitting) γίνεται λόγος για τη μη λήψη επαρκών μοτίβων από τα δεδομένα. Το μοντέλο είναι φτωχό απέναντι στο σύνολο εκπαίδευσης και το σύνολο δοκιμών. Αντίθετα, στην περίπτωση της υπερπροσαρμογής (overfitting) εννοούμε τόσο τη λήψη περιττού θορύβου, αλλά και τη καταγραφή μοτίβων, τα οποία δεν μπορούν να γενικευτούν «καλά» σε άγνωστα δεδομένα. Το μοντέλο ανταποκρίνεται εξαιρετικά «καλά» στο σύνολο εκπαίδευσης, αλλά «άσχημα» στο σύνολο δοκιμών.



Εικόνα 4 Λάθος εκπαίδευσης και επικύρωσης σε σχέση με τις παραμέτρους

Υπάρχουν διάφορες τεχνικές, για να εφαρμοστεί η μέθοδος cross validation. Στην πρώτη στρατηγική, τα δεδομένα χωρίζονται σε δύο ομάδες, στο σύνολο εκπαίδευσης και δοκιμής, ώστε το ένα σύνολο να μην αλληλεπικαλύπτει το άλλο. Αυτός είναι ο λόγος, για τον οποίο είναι σημαντικό, να μην υπάρχουν διπλά δείγματα στο σύνολο δεδομένων. Στη συνέχεια, η εκπαίδευση του μοντέλου πραγματοποιείται στο σύνολο εκπαίδευσης, ενώ το σύνολο δοκιμής χρησιμοποιείται για σκοπό επικύρωσης. Αν δεν υπάρχουν αρκετά δεδομένα, για να εκπαιδευτεί το μοντέλο, η κατάργηση ενός μέρους του συνόλου εκπαίδευσης (training set), για να χρησιμοποιηθεί σαν σύνολο επικύρωσης (validation set), θέτει πρόβλημα ελλιπούς

προσαρμογής (underfitting). Ακολούθως, εάν διαθέτουμε περιορισμένα δεδομένα, κινδυνεύουμε να χάσουμε σημαντικά μοτίβα και τάσεις στο σύνολο δεδομένων εκπαίδευσης, γεγονός που αυξάνει την πιθανότητα μεγάλου συστηματικού σφάλματος (bias). Επίσης, τι συμβαίνει στην περίπτωση που ο διαχωρισμός που γίνεται δεν είναι τυχαίος; Για παράδειγμα, αν το σύνολο των δεδομένων έχει μόνο άτομα από ένα συγκεκριμένο κράτος, αμειβόμενους μ' ένα συγκεκριμένο μισθό, μόνο γυναίκες ή μόνο άτομα ορισμένης ηλικίας. Αυτό θα οδηγήσει σε υπερπροσαρμογή (overfitting), καθώς δεν είναι βέβαιο ποια δεδομένα θα καταλήξουν στο σύνολο επικύρωσης. Έτσι το αποτέλεσμα μπορεί να είναι εντελώς διαφορετικό για διαφορετικά σύνολα. Εάν το πλήθος των δεδομένων είναι μεγάλο και τα δείγματα εκπαίδευσης κι ελέγχου ακολουθούν την ίδια κατανομή, τότε αυτό αποφεύγεται και η εν λόγω προσέγγιση είναι αποδεκτή.

Μια ακόμα τεχνική είναι το k cross validation. Κατά τη διάρκεια αυτής της διαδικασίας, λαμβάνεται μια μοναδική παράμετρος που ονομάζεται k , η οποία αναφέρεται στον αριθμό των ομάδων, στις οποίες πρόκειται να χωριστεί ένα δοσμένο σύνολο δεδομένων. Ως εκ τούτου, η διαδικασία συχνά ονομάζεται k -fold cross-validation. Είναι μια δημοφιλής μέθοδος, επειδή είναι απλή στην κατανόησή της και γενικά έχει ως αποτέλεσμα μια λιγότερο μεροληπτική ή λιγότερο αισιόδοξη εκτίμηση της ικανότητας του μοντέλου από άλλες μεθόδους. Αυτή η προσέγγιση περιλαμβάνει την τυχαία διαίρεση του συνόλου των παρατηρήσεων σε ομάδες k , ή πτυχές, περίπου ίσου μεγέθους. Η πρώτη πτυχή αντιμετωπίζεται ως σύνολο επικύρωσης και η μέθοδος εφαρμόζεται στις υπόλοιπες $k - 1$ πτυχές. Στην επικύρωση k Fold, τα δεδομένα χωρίζονται σε υποσύνολα k . Η διαδικασία επικύρωσης επαναλαμβάνεται k φορές, έτσι ώστε κάθε φορά, ένα από τα υποσύνολα k χρησιμοποιείται σαν σύνολο επικύρωσης, ενώ τα υπόλοιπα υποσύνολα $k-1$ σχηματίζουν ένα σύνολο εκπαίδευσης. Το σφάλμα υπολογίζεται κατά μέσον όρο σε όλες τις δοκιμές k , για να υπολογιστεί η συνολική αποτελεσματικότητα (effectiveness) του μοντέλου. Όπως είναι φανερό, κάθε σημείο, δηλαδή κάθε δεδομένο, βρίσκεται σ' ένα σύνολο επικύρωσης ακριβώς μία φορά, ενώ σε ένα σύνολο εκπαίδευσης $k-1$ φορές. Αυτό μειώνει σημαντικά το συστηματικό σφάλμα (bias), καθώς χρησιμοποιούμε τα περισσότερα δεδομένα για προσαρμογή (fitting), αλλά μειώνει σημαντικά και τη διακύμανση (variance), καθώς τα περισσότερα από τα δεδομένα χρησιμοποιούνται και στο σύνολο επικύρωσης. Επίσης, πρόκειται για μια αρκετά χρονοβόρα διαδικασία. [28] [29]

2.6.Σχετικές εργασίες

Πολλές εργασίες έχουν γίνει στο παρελθόν, όσον αφορά την μελέτη και την ανάλυση του ήχου. Επίσης διαφορετικές τεχνικές έχουν χρησιμοποιηθεί, για να προσεγγίσουν τα προβλήματα ταξινόμησης και της εξαγωγής των χαρακτηριστικών, που είναι απαραίτητα σε μια εργασία ανάλυσης ήχου. Παρακάτω παραθέτουμε κάποια από τα έργα διάφορων ερευνητών που ασχολήθηκαν εκτενώς με θέματα ήχου σε διάφορες κατηγορίες.

Αρχικά, στην εργασία τους με τίτλο «Content Based Audio Classification And Retrieval Using Joint Time-Frequency Analysis» οι ερευνητές S. Esmaili, S. Krishnan and K. Raahemifar

παρουσίασαν μια τεχνική ταξινόμησης και ανάκτησης ήχου που εκμεταλλεύεται τη μη στατική συμπεριφορά των μουσικών σημάτων και τη φασματική αλλαγή τους με την πάροδο του χρόνου. Στην τεχνική τους, η ταξινόμηση εκτελέστηκε αναλύοντας τη χρονική συχνότητα και οι ήχοι ταξινομήθηκαν σε 6 ομάδες από διαφορετικά είδη μουσικής, Συγκεκριμένα σε ροκ, κλασική, λαϊκή, τζαζ και ποπ μουσική. Για κάθε μουσικό κομμάτι των 5 δευτερολέπτων, τα χαρακτηριστικά που εξάχθηκαν ήταν η εντροπία, το κέντρο, η αναλογία του κέντρου, το εύρος ζώνης, η αναλογία σιωπής, η αναλογία ενέργειας και η θέση της ελάχιστης και της μέγιστης ενέργειας. Χρησιμοποιώντας μια βάση δεδομένων 143 ηχητικών σημάτων, δημιουργήθηκε ένα σύνολο 10 χαρακτηριστικών χρονικής συχνότητας. Στη συνέχεια, εφαρμόστηκε γραμμική ανάλυση διακρίσεων (LDA) με τη χρήση του λογισμικού SPSS, για να προβλεφθεί η ταξινόμηση των νέων δεδομένων στις κλάσεις. Αυτός ο τύπος ανάλυσης προσπάθησε να βρει έναν γραμμικό συνδυασμό αυτών των εξαγόμενων χαρακτηριστικών, ο οποίος να πετυχαίνει τον καλύτερο διαχωρισμό. Η εν λόγω τεχνική βρίσκει την πρώτη συνάρτηση, που διαχωρίζει τις ομάδες όσο το δυνατόν καλύτερα και στη συνέχεια εντοπίζει περαιτέρω συναρτήσεις, που βελτιώνουν τον διαχωρισμό και δεν σχετίζονται με τις προηγούμενες. Ο αριθμός των συναρτήσεων καθορίζεται από τον αριθμό των χαρακτηριστικών και τον αριθμό των διαθέσιμων ομάδων. Με την χρήση της παραπάνω μεθόδου, της τακτικής γραμμικής διάκρισης, προέκυψε μια ακρίβεια ταξινόμησης περίπου 93%, με τη λαϊκή μουσική να έχει το χαμηλότερο ποσοστό σωστής ταξινόμησης. Τέλος, η μέθοδος επικύρωσης leave one-out που χρησιμοποιήθηκε έδωσε μια ακρίβεια ταξινόμησης 92,3% αποκαλύπτοντας την ισχύ διάκρισης του συνόλου των χαρακτηριστικών.

Μια δεύτερη εργασία με τίτλο «Speech/music classification using PLP and SVM» έγραψε ο R. Thiruvengatanadhan ο οποίος ασχολήθηκε με το πρόβλημα ταξινόμησης ομιλίας / μουσικής, ξεκινώντας από ένα σύνολο χαρακτηριστικών που εξάγονται απευθείας από δεδομένα ήχου. Αρχικά, η εξαγωγή των χαρακτηριστικών έγινε με τη χρήση της μεθόδου PLP, μιας τεχνικής η οποία προτάθηκε από τον Hermansky, όπου τα φασματικά χαρακτηριστικά του ηχητικού σήματος της ομιλίας μετασχηματίστηκαν, ώστε να ταιριάζουν με το ανθρώπινο ακουστικό σύστημα. Μ' αυτήν τη διαδικασία διανύσματα χαρακτηριστικών 9 διαστάσεων υπολογίστηκαν για το κάθε ηχητικό αρχείο wav μορφής. Κατόπιν της διαδικασίας εξαγωγής χαρακτηριστικών, ο αλγόριθμος SVM χρησιμοποιήθηκε ο οποίος εκπαιδεύτηκε σε 100 αρχεία για την κάθε κλάση. Η ταξινόμηση βασίστηκε στην παραγωγή πολλών πλαισίων, των οποίων κάθε δευτερόλεπτο απαρτίστηκε από 100 καρτέ. Η τελική ταξινόμηση ομιλίας / μουσικής βασισμένη στον συγκεκριμένο αλγόριθμο έδωσε μια ακρίβεια της τάξης του 89%.

Επίσης μια ενδιαφέρουσα μελέτη με τίτλο «Hierarchical Automatic Audio Signal Classification» πραγματοποίησαν ο Juan Jose Burred και ο Alexander Lerch, οι οποίοι σχεδίασαν, υλοποίησαν και αξιολόγησαν ένα σύστημα αυτόματης ταξινόμησης ηχητικού σήματος. Τα ηχητικά σήματα ταξινομήθηκαν ανάλογα με τον τύπο τους, μεταξύ τριών τάξεων ομιλίας, 13 μουσικών ειδών και θορύβου υποβάθρου. Κάθε ήχος χωρίστηκε σε πλαίσια τα οποία είχαν μήκος περίπου 10-40 ms και υπολογίστηκε ένα χαρακτηριστικό για κάθε πλαίσιο-παραθύρο. Συνολικά, αξιολογήθηκαν διανύσματα χαρακτηριστικών 90 διαστάσεων. Επιπρόσθετα, ένας KNN και ένας

GMM (Gaussian mixture model) ταξινομητής εξετάστηκαν και προτάθηκε μια ιεραρχική ταξινομική προσέγγιση, η οποία συνίστατο σε μία διαδοχή δέντρων. Μ' αυτήν την προσέγγιση το επίπεδο 1 της ταξινόμησης ανταποκρίνεται σ' ένα πρόβλημα 3 κλάσεων. Έπειτα συνέχισαν με το επίπεδο 2 κ.ο.κ. Αυτό έρχεται σε αντίθεση με την πιο κοινή άμεση προσέγγιση, η οποία εκτελεί μια ενιαία απόφαση και από τις 17 τάξεις. Οι ρυθμοί ταξινόμησης διέφεραν σημαντικά μεταξύ των επιπέδων του δέντρου, δείχνοντας τους ποικίλους βαθμούς δυσκολίας στο διαχωρισμό κάθε υποομάδας κλάσεων ήχου. Τελικά, όπως ήταν αναμενόμενο, οι καλύτερες επιδόσεις επιτεύχθηκαν στα υψηλότερα επίπεδα του δέντρου. Για παράδειγμα, ακρίβεια ίση με 94,59% επιτεύχθηκε στη διαφοροποίηση της ομιλίας, του περιβάλλοντος και της μουσικής, 96,08% στο διαχωρισμό της κλασσικής από τη μη κλασσική μουσική και 81,52% στο διαχωρισμό της μουσικής δωματίου από την μουσική ορχήστρας. Αντίθετα, οι κύριες δυσκολίες προέκυψαν στα πιο συγκεκριμένα είδη, δηλαδή στα χαμηλότερα επίπεδα του δέντρου, ειδικά στην περίπτωση των τεσσάρων τάξεων μουσικής δωματίου, όπου η συνολική ακρίβεια ταξινόμησης ήταν 54,67%. Τα χαμηλότερα επίπεδα του δέντρου, καθώς και ο υψηλός αριθμός κατηγοριών που εξετάστηκαν, κατέστησαν τον ρυθμό ταξινόμησης όλων των κατηγοριών μειωμένο στο 58,71%.

Εξίσου μια αξιολογη εργασία με τίτλο «Audio Feature Extraction And Classification for Urban Sound» πραγματοποιήθηκε από τους Chen Cai, Liu Yuxuan, Sun Haoran, Zhou Moyan, οι οποίοι διερεύνησαν τα χαρακτηριστικά των ήχων του αστικού περιβάλλοντος τα οποία παρουσιάζουν ένα ενδιαφέρον αφού δεν είναι δομημένα και συνήθως συνδυάζονται με έντονο θόρυβο. Έξι χαρακτηριστικά αξιολογήθηκαν, για να αναγνωρίζουν δέκα τύπους αστικών ήχων. Οι δέκα κλάσεις που μελετήθηκαν ήταν το κλιματιστικό, το γαύγισμα του σκύλου, ο κινητήρας, τα παιδιά που παίζουν, η γεώτρηση, ο πυροβολισμός από όπλο, το σφυρί, η μουσική του δρόμου και τα χαρακτηριστικά των ήχων που εξάχθηκαν ήταν ο μετασχηματισμός Fourier (Fast Fourier Transform), η μέση τετραγωνική ρίζα της ενέργειας (Root Mean Squared Energy), οι συντελεστές της συχνότητας Mel (Mel Frequency Cepstral Coefficient), το χρωμόγραμμα (Chromagram), το φάσμα Mel (Mel spectrogram) και η φασματική αντίθεση (Spectral contrast). Στην εργασία εκπαιδεύτηκαν τρία μοντέλα, το συγχρονισμένο Νευρωνικό Δίκτυο (Convolutional Neural Network - CNN), το επαναλαμβανόμενο Νευρωνικό Δίκτυο (Recurrent Neural Network - RNN) και τα διανύσματα στήριξης (Support Vector Machine - SVM). Τελικά η καλύτερη ακρίβεια ήταν 91,1% και επιτεύχθηκε από το μοντέλο RNN με συνδυασμένα χαρακτηριστικά (MFCC, Mel, Contrast, and Chroma).

Ακόμη, στην εργασία τους με τίτλο «Automatic recognition of urban sound sources», οι επιστήμονες Boris Defréville, Pierre Roy, Christophe Rosin και François Pachet μελέτησαν τις εξής 6 κλάσεις του αστικού περιβάλλοντος: αυτοκίνητο, μοτοποδήλατο, λεωφορείο, μοτοσικλέτα, φωνή και πτηνά. Η εργασία επικεντρώθηκε σε δύο αλγορίθμους, στο μοντέλο Gaussian Mixture – GMM και στον K κοντινότερο γείτονα– KNN. Για κάθε 50 ms εκτιμήθηκε το βραχυπρόθεσμο φάσμα. Στη συνέχεια, υπολογίστηκαν οι δέκα πρώτοι συντελεστές Mel (MFCC) για την εκτίμηση του φάσματος κάθε πλαισίου. Ένα τυπικό δείγμα 500 ms αντιπροσωπευόταν από 10 διανύσματα χαρακτηριστικών, δηλ. 100 συντελεστές. Τα αποτελέσματα ήταν

ικανοποιητικά σε σχέση με τέσσερις κατηγορίες ήχου: λεωφορείο (86,8%), μοτοποδήλατο (88,9%), αυτοκίνητο (90,0%) και φωνή (99,0%). Ενώ οι επιδόσεις στις μοτοσικλέτες (72,0%) και τα πτηνά (71,8%) θα ήθελαν κάποια βελτίωση.

Επιπλέον, οι Ivo Draganov και Krasimir Minchev, στην μελέτη τους με τίτλο «Algorithm Selection for Automated Audio Classification based on Content» πρότειναν μια προσέγγιση για την επιλογή ενός αλγορίθμου στην αυτοματοποιημένη ταξινόμηση ήχου με βάση το περιεχόμενο. Οι πειραματικές δοκιμές βασίστηκαν σε μια προσαρμοσμένη βάση δεδομένων που περιλάμβανε 5 ομιλίες και 10 μουσικές εγγραφές σε μη συμπιεσμένη μορφή. Η διάρκεια όλων των ηχογραφημένων ήχων ήταν 1 λεπτό. Τα μουσικά κομμάτια προήλθαν από διαφορετικές χώρες και από διαφορετικά στυλ. Οι ερευνητές επικεντρώθηκαν στη διάκριση των σημάτων ομιλίας και ήχου με βάση την ενέργειά τους. Η ενεργειακή κατανομή των φωνητικών και μουσικών σημάτων αξιολογήθηκε εξετάζοντας τη συχνότητα των μηδενικών διασταυρώσεων (zero crossing), τη βραχυπρόθεσμη ενέργεια (short-term energy) και τη διαμόρφωση 4 Hz (4 Hz modulation) και λαμβάνοντας υπόψη την ελάχιστη ενεργειακή πυκνότητα (Minimum Energy Density -MED), τα πλαίσια χαμηλής ενέργειας (Low Energy Frames-LEF) και την τροποποιημένα χαμηλή αναλογία ενέργειας (Modified Low Energy Ratio -MLER). Η έρευνα αποκάλυψε, ότι όλα τα μουσικά δεδομένα είχαν υψηλότερη κορυφή στις χαμηλές συχνότητες. Τελικά, οδηγήθηκαν στο συμπέρασμα ότι η καλύτερη και πιο ακριβής μέθοδος που ικανοποιεί περισσότερο τη διακριτική συνθήκη και μπορεί να χρησιμοποιηθεί από έναν χρήστη σ' ένα αντικειμενικό περιβάλλον για τον διαχωρισμό της ομιλίας από τη μουσική είναι η μέθοδος 4Hz (ακρίβεια κοντά στο 100%). Στη συνέχεια, ακολουθεί η μέθοδος του ποσοστού μηδενικής διέλευσης (zero crossing rate), αν και δεν είναι πολύ αξιόπιστη, αλλά ο υπολογισμός της είναι πολύ πιο απλός, γεγονός που την καθιστά ιδανική επιλογή, αν ο κύριος στόχος είναι η επιτάχυνση της διαδικασίας αντί για μια ακρίβεια κοντά στο 100%.

Οι Lie Lu, Hao Jiang και HongJiang Zhang ασχολήθηκαν επίσης με την ανάλυση ήχου και στην εργασία τους με τίτλο «A Robust Audio Classification and Segmentation Method», εξέτασαν ένα σύστημα τμηματοποίησης και ταξινόμησης δύο σταδίων, το οποίο κατατάσσει και ταξινομεί μια ροή ήχου σε ομιλία, μουσική, ήχο περιβάλλοντος και σιωπή. Συγκεκριμένα, στο πρώτο στάδιο της ταξινόμησης, τα ηχητικά δεδομένα χωρίστηκαν στις κλάσεις της ομιλίας και της μη ομιλίας, με βάση απλά χαρακτηριστικά, όπως η αναλογία ρυθμού μηδενικής διέλευσης (zero crossing rate ratio), η αναλογία χαμηλής βραχυπρόθεσμης ενέργειας (low shorttime energy ratio), η φασματική ροή (spectrum flux) και η απόσταση LSP (LSP distance). Ο ταξινομητής που χρησιμοποιήθηκε, για να επιτευχθεί ο παραπάνω διαχωρισμός, ήταν ο KNN αλγόριθμος. Στο δεύτερο στάδιο της ταξινόμησης, οι ήχοι της τάξης της μη ομιλίας χωρίστηκαν περαιτέρω σε μουσική, σε περιβαλλοντικούς ήχους και σε σιωπή μ' ένα σύστημα ταξινόμησης που βασίζεται σε κανόνες. Σ' αυτήν τη διαδικασία, προστέθηκαν δύο ακόμη χαρακτηριστικά, η αναλογία του πλαισίου θορύβου (noise frame ratio) και η ζώνη περιοδικότητας (band periodic). Η πειραματική αξιολόγηση έδειξε ότι το προτεινόμενο σύστημα ταξινόμησης ήχου ήταν πολύ αποτελεσματικό με συνολικό ποσοστό ακρίβειας πάνω από 96%.

Μια ακόμα έρευνα σε ηχητικά προβλήματα ήταν η «A multi-device dataset for urban acoustic scene classification» των Annamaria Mesaros, Toni Heittola και Tuomas Virtanen, οι οποίοι μελέτησαν τις εξής κλάσεις αεροδρόμιο, λεωφορείο, μετρό, στάση μετρό, πάρκινγκ, πλατεία, εμπορικό κέντρο, κίνηση, πεζοδρόμιο και τράμ. Για την ανάλυση εφαρμόστηκε το νευρωνικό δίκτυο CNN δύο επιπέδων και για κάθε ηχητικό αρχείο 10 δευτερολέπτων εξάχθηκαν οι ενέργειες Mel. Τα σύνολα χωρίστηκαν σε τρεις κατηγορίες ανάλογα με τη συσκευή που χρησιμοποιήθηκε. Έπειτα, πραγματοποιήθηκε ένας διασταυρωμένος έλεγχος (cross validation) με 30% να είναι το σύνολο επικύρωσης και το confusion matrix έδειξε ότι το σύστημα γενίκευε με μια απόδοση της τάξης του 61% στο σύνολο αξιολόγησης. Επίσης η πιο δύσκολη τάξη να αναγνωριστεί ήταν η πλατεία με την χαμηλότερη απόδοση 33.9% στο σύστημα αξιολόγησης.

Στην εργασία τους με τίτλο «Classification of audio signals using statistical features on time and wavelet transform domains», οι επιστήμονες T. Lambrou, P. Kudumakis, R. Speller, M. Sandler, A. Linney επιδίωξαν να διαπιστώσουν ποια στατιστικά μεγέθη χωρίζουν καλύτερα τα τρία διαφορετικά μουσικά στυλ ροκ, πιάνο και τζαζ. Συνολικά οκτώ στατιστικά χαρακτηριστικά συλλέχθηκαν από κάθε ηχητικό σήμα. Αναλυτικά, αυτά ήταν τα εξής: ο μέσος όρος (Mean), η διακύμανση (Variance), η ασυμμετρία (Skewness), η κύρτωση (Kurtosis), η γωνιακή δεύτερη ροπή (Angular Second Moment), η συσχέτιση (Correlation) και η εντροπία (Entropy). Τέλος, αξιολογήθηκε ο αριθμός των μηδενικών διασταυρώσεων, καθώς υποδεικνύει τη συμπεριφορά του θορύβου σε ένα ηχητικό σήμα. Σε αυτή τη μελέτη εξετάστηκε η απόδοση μερικών ταξινομητών. Αρχικά, μελετήθηκε ο ταξινομητής ελάχιστων αποστάσεων (Minimum Distance Classifier - MDC). Στη συνέχεια, εφαρμόστηκε ο K κοντινότερος γείτονας (k-Nearest Neighbor Distance Classifier - K-NNC). Ο τρίτος ταξινομητής ήταν η ελάχιστη απόσταση τετραγώνων (Least Squares Minimum Distance Classifier - LSMDC) και τελευταίος χρησιμοποιήθηκε ο ταξινομητής του τετραγωνισμού (Quadrature Classifier - QC). Τα αποτελέσματα έδειξαν ότι τα χαρακτηριστικά, που συλλέχθηκαν, έδωσαν πιο βελτιωμένα αποτελέσματα σε σύγκριση με τεχνικές βασισμένες στο κύμα, επιτυγχάνοντας συνολική ακρίβεια ταξινόμησης 91,67% χρησιμοποιώντας είτε τον ταξινομητή ελάχιστης απόστασης (Minimum Distance Classifier) είτε τον ταξινομητή ελάχιστης απόστασης τετραγώνων (Least Squares Minimum Distance Classifier).

Εν συνεχεία, οι ερευνητές Jürgen T. Geiger, Björn Schuller, και Gerhard Rigoll στη μελέτη τους με τίτλο «Large scale audio feature extraction and SVM for acoustic scene classification», περιέγραψαν ένα σύστημα για την ταξινόμηση 10 διαφορετικών κλάσεων των 30 δευτερόλεπτων. Τα χαρακτηριστικά κάθε ηχητικού σήματος εξάχθηκαν από πλαίσια των 25 ms και ομαδοποιήθηκαν στις ακόλουθες κατηγορίες, cepstral, spectral (φασματικά), energy related (ενεργειακά) και voicing (φωνητικά) χαρακτηριστικά. Συνολικά, με το εργαλείο openSMILE ο αριθμός των χαρακτηριστικών που εξάχθηκε ήταν 6669. Στη συνέχεια, μέσω του Weka εφαρμόστηκε ο αλγόριθμος SVM και επιτεύχθηκε ακρίβεια περίπου ίση με 73%. Επίσης η ανάλυση των χαρακτηριστικών που έγινε έδειξε ότι τα φάσματα Mel είναι από τα πιο σημαντικά χαρακτηριστικά και λειτουργούν εξίσου καλά σε σύγκριση με τα MFCC. Επιπλέον, τα υπόλοιπα χρησιμοποιούμενα ενεργειακά και φασματικά χαρακτηριστικά βοήθησαν στην καλύτερη αιχμαλώτιση της πληροφορίας στις ακουστικές σκηνές. Τέλος, παρατηρήθηκε, ότι

ορισμένες ακουστικές σκηνές για παράδειγμα πάρκο, εστιατόριο, μετρό, στάση μετρό δυσκολεύτηκαν να αναγνωριστούν λόγω της μεγάλης μεταβλητότητας της κατηγορίας και της ομοιότητας μεταξύ των διαφόρων κατηγοριών.

Ακολούθως, στην έρευνα των P. Dhanalakshmi, S. Palanivel και V. Ramalingam, «Classification of audio signals using SVM and RBFNN», οι ερευνητές πρότειναν ένα σύστημα αυτόματης ταξινόμησης ήχου, χρησιμοποιώντας τους αλγόριθμους SVM και RBFNN. Οι συντελεστές γραμμικής πρόβλεψης cepstrum (LPC, LPCC) και οι συντελεστές συχνότητας mel χρησιμοποιήθηκαν ως χαρακτηριστικά για την περιγραφή του ακουστικού περιεχομένου. Τα χαρακτηριστικά εξάχθηκαν από 256 μη σιωπηρά ηχητικά δείγματα των 60 sec και χωρίστηκαν σε πλαίσια των 20ms. Τελικά, για κάθε ηχητικό σήμα εξετάστηκαν 39 χαρακτηριστικά. Στη συνέχεια, ένας μη γραμμικός αλγόριθμος μάθησης SVM και το νευρωνικό RBFNN εφαρμόστηκαν, για να προσδιοριστούν οι διάφορες κλάσεις, δηλαδή η μουσική, τα δελτία ειδήσεων, τα αθλητικά νέα, οι διαφήμισης, τα κινούμενα σχέδια και οι ταινίες. Τα πειραματικά αποτελέσματα έδειξαν ότι το ποσοστό της ακρίβειας ήταν 92%. Η απόδοση συγκρίθηκε με το νευρωνικό ακτινωτό δίκτυο (Radial basis function) που έδειξε ακρίβεια 93%. Επίσης, ο ρυθμός ταξινόμησης με τη χρήση LPC και LPCC ήταν ελαφρώς χαμηλότερος από τα 39 διαστάσεων διανύσματα χαρακτηριστικών MFCC.

Στην εργασία τους «Musical Genre Classification of Audio Signals», οι ερευνητές George Tzanetakis και Perry Cook ασχολήθηκαν με την ταξινόμηση δέκα ειδών μουσικής, μπλουζ, country, κλασική, χορευτική, τζαζ, latin, ποπ, RnB, ραπ και ροκ. Τα χαρακτηριστικά τα οποία εξέτασαν ήταν τα μέσα (means) και οι διακυμάνσεις (variances) του κέντρου (centroid), της κύλισης (roll off), της ροής (flux), των μηδενικών διασταυρώσεων (zero crossings), της χαμηλής ενέργειας (low energy), και των συντελεστών MFCC. Τελικά για κάθε ήχο πρόκυψε ένα διάνυσμα 19 χαρακτηριστικών. Έπειτα, για τα χαρακτηριστικά αυτά εξετάστηκαν διάφοροι στατιστικοί αλγόριθμοι, όπως είναι ο KNN. Επίσης ζητήθηκε και από φοιτητές να κρίνουν μετά από ακρόαση μουσικών δειγμάτων 250 ms και μετά από ακρόαση 3 δευτερολέπτων. Παρατηρήθηκε ότι η ακρόαση περισσότερο από 3 δευτερόλεπτα δεν βελτίωσε την απόδοσή τους. Χρησιμοποιώντας τα προτεινόμενα σύνολα χαρακτηριστικών επιτεύχθηκε 61% ακρίβεια στο σύνολο δεδομένων από δέκα μουσικά είδη. Αυτό το αποτέλεσμα συγκρίθηκε με τα αποτελέσματα της ταξινόμησης του μουσικού είδους στον άνθρωπο όπου η ακρίβεια έφτασε στο 53% για τα 250ms.

Στην μελέτη «SVM-based Audio Scene Classification», οι ερευνητές H.Jiang, J.Bai, S.Zhang, B.Xu παρουσίασαν την προσέγγιση, που χρησιμοποιεί ο αλγόριθμος SVM στην ταξινόμηση ηχητικών σημάτων σε μία από τις πέντε κατηγορίες: καθαρή ομιλία, μη καθαρή ομιλία, μουσική, ήχος περιβάλλοντος και σιωπή. Μεταξύ αυτών των κατηγοριών, η μη καθαρή ομιλία μπορεί να χωριστεί περαιτέρω σε ομιλία με μουσική και ομιλία με θόρυβο. Αρχικά, λήφθηκαν υπόψη δεκαέξι είδη χαρακτηριστικών γνωρισμάτων. Συγκεκριμένα, αυτά είναι ο ρυθμός μηδενικής διέλευσης (Zero Crossing Rate), η υψηλή αναλογία του ZCR (High ZCR Ratio), η βραχυπρόθεσμη ενέργεια (Short-Time Energy), η χαμηλή αναλογία STE (LSTER), η μέση τετραγωνική ρίζα (Root Mean Square), η αναλογία σιωπής πλαισίου (Silence Frame Ratio), η κατανομή ενέργειας (Sub-

band Energy Distribution), η φασματική ροή (Spectrum Flux), το φασματικό κέντρο (Spectral Centroid), η φασματική διασπορά (Spectral Spread), η συχνότητα κύλισης (Spectral Rolloff Frequency), η περιοδικότητα υπο-ζωνών (Sub-band Periodicity), ο θόρυβος πλαισίου (Noise Frame Ratio), οι συντελεστές γραμμικού φάσματος (Linear Spectrum Pair), οι συντελεστές Cepstral (Linear Predictive Cepstral Coefficients) και οι συντελεστές συχνότητας Mel (Mel-frequency Cepstral Coefficients). Έπειτα, υπολογίστηκαν τα μέσα και οι τυπικές αποκλίσεις των παραπάνω χαρακτηριστικών. Έτσι η διάσταση του τελικού διανύσματος χαρακτηριστικών, που μελετήθηκε, ήταν 90 για το κάθε ηχητικό σήμα. Τα πειραματικά αποτελέσματα έδειξαν, ότι η μέθοδος SVM μπορεί να επιτύχει 96,61% ποσοστό ακρίβειας χρησιμοποιώντας διαφορετικούς συνδυασμούς χαρακτηριστικών για την ταξινόμηση διαφορετικών ζευγαριών ηχητικών τύπων. Ενώ η ακρίβεια αυξήθηκε στο 96.90% χρησιμοποιώντας ένα ομοιόμορφο σύνολο χαρακτηριστικών.

Επιπροσθέτως, οι Karthikeyan Umapathy, Sridhar Krishnanand και Raveendra K. Rao στην εργασία τους «Audio Signal Feature Extraction and Classification Using Local Discriminant Bases», πρότειναν μια εξαγωγή ακουστικών χαρακτηριστικών και ένα σύστημα ταξινόμησης πολλαπλών ομάδων που επικεντρώνεται στην αναγνώριση συχνοτήτων με την χρήση της τεχνικής των τοπικών διακριτικών βάσεων (LDB). Στο πρώτο επίπεδο, τα ηχητικά σήματα ομαδοποιήθηκαν σε τεχνητούς και φυσικούς ήχους. Κάθε μία από τις ομάδες πρώτου επιπέδου υποδιαίρεθηκε στις ομάδες του δεύτερου επιπέδου, δηλαδή μουσική, αυτοκινητοβιομηχανία, ανθρώπινοι και μη ανθρώπινοι ήχοι. Η τρίτη βαθμίδα διαμορφώθηκε στις τελικές δέκα ομάδες τύμπανα, φλάουτο, πιάνο, αεροσκάφος, ελικόπτερο, άντρες, γυναίκες. Επίσης τα χαρακτηριστικά LDB συνδυάστηκαν με τους συντελεστές MFC. Η τεχνική αυτού του συνδυασμού των χαρακτηριστικών έδειξε να ανταποκρίνεται «πολύ καλά», μιας και η συνολική ακρίβεια ταξινόμησης έφτασε στο 91% για το πρώτο επίπεδο, 99% για το δεύτερο και 95% για το τρίτο επίπεδο.

Ακόμη, οι επιστήμονες Martin F. McKinney και Jeroen Breebaart στο «Features for Audio and Music Classification», αξιολόγησαν τέσσερα σύνολα χαρακτηριστικών ήχου ως προς την ικανότητά τους να ταξινομήσουν πέντε γενικές κλάσεις ήχου και επτά είδη μουσικής. Τα σύνολα χαρακτηριστικών περιλαμβάνουν ιδιότητες χαμηλού επιπέδου σήματος (low-level signal), φασματικούς συντελεστές συχνοτήτων Mel (mel-frequency spectral coefficients) και δύο νέα σύνολα βασισμένα σε αντιληπτικά μοντέλα ακοής. Χρησιμοποιώντας ένα τυπικό Gaussian πλαίσιο για την ταξινόμηση, τα αποτελέσματα δείχνουν ότι η χρονική συμπεριφορά των χαρακτηριστικών είναι σημαντική τόσο για τη μουσική όσο και για την ταξινόμηση του ήχου. Συγκεκριμένα, αποδείχθηκε ότι ενώ η συνολική απόδοση ταξινόμησης των γενικών τάξεων ήχου είναι αρκετά υψηλή ($93 \pm 2\%$), η ταξινόμηση του μουσικού είδους είναι αρκετά χαμηλή ($74 \pm 9\%$). Τέλος, η ταξινόμηση είναι καλύτερη, κατά μέσον όρο, αν βασίζεται σε χαρακτηριστικά από μοντέλα ακουστικής αντίληψης και όχι σε τυπικά χαρακτηριστικά.

Στην εργασία τους με τίτλο «Audio Music Genre Classification Using Different Classifiers and Feature Selection Methods», οι Yusuf Yaslan και Zehra Cataltepe εξέτασαν την απόδοση διαφορετικών ταξινομητών σε διαφορετικά σύνολα ηχητικών χαρακτηριστικών για να

καθορίσουν το είδος ενός συγκεκριμένου μουσικού κομματιού. Το σύνολο δεδομένων που εξετάστηκε περιείχε 1000 κομμάτια μουσικής μήκους 30 δευτερολέπτων το καθένα. Τα 10 είδη μουσικής που ελέγχθηκαν ήταν κλασική, country, ντίσκο, hip-hop, τζαζ, rock, pop και metal. Επίσης για την εξαγωγή των χαρακτηριστικών το λογισμικό MARSYAS χρησιμοποιήθηκε. Τα σύνολα χαρακτηριστικών, που προέκυψαν και μελετήθηκαν, ήταν 6 χαρακτηριστικά με βάση τον ρυθμό, 9 χαρακτηριστικά με βάση το μετασχηματισμό Fourier, 10 χαρακτηριστικά από τους συντελεστές MFCC, 5 χαρακτηριστικά σε σχέση με τα τονικά ύψη και ένα σύνολο που είχε όλα τα παραπάνω 30 χαρακτηριστικά. Επίσης σ' αυτήν την εργασία, πειραματίστηκαν με 6 διαφορετικούς αλγόριθμους, ώστε να επιτευχθεί η μέγιστη ακρίβεια. Τελικά, χρησιμοποιώντας το σύνολο όλων των χαρακτηριστικών και τον συνδυασμό αλγόριθμων PNU(PDC, NBC, UDC), προέκυψε ακρίβεια ίση με 80%.

Καταλήγοντας να αναφερθεί η μελέτη «An Efficient Classification Algorithm For Music Mood Detection In Western and Hindi Music Using Audio Feature Extraction» των Aathreya S. Bhat, Namrata S. Prasad, Amith V.S, και Murali Mohan D., οι οποίοι εξέτασαν συνολικά 8 διαφορετικές διαθέσεις της δυτικής και της ινδικής μουσικής. Αναλυτικότερα, αυτές ήταν η χαρά, ο ενθουσιασμός, η ενέργεια, η μανία, η λύπη, η κατάθλιψη, η ηρεμία και η ικανοποίηση. Συνολικά προέκυψε μια βάση δεδομένων από 100 τραγούδια και σ' αυτήν εξέτασαν τέσσερα σημαντικά χαρακτηριστικά της μουσικής, το τονικό ύψος, την χροιά, την ένταση και τον ρυθμό. Στη συνέχεια πολυεπίπεδα νευρωνικά δίκτυα χρησιμοποιήθηκαν, για να αξιολογήσουν τη διάθεση ενός κομματιού και εκπαιδεύτηκαν χρησιμοποιώντας τον αλγόριθμο Back Propagation. Τελικά, η ανάλυση κατέληξε σ' ένα επίπεδο ακρίβειας έως 94,44%.

3. Τεχνικές επεξεργασίας και ανάλυσης

Στην παρούσα εργασία δημιουργήθηκαν διανύσματα χαρακτηριστικών ήχου στα οποία εν συνεχεία εφαρμόστηκαν αλγόριθμοι μηχανικής μάθησης. Στην παρακάτω ενότητα ακολουθεί μια αναλυτική περιγραφή αυτών.

3.1.Χαρακτηριστικά ήχου

Η εξαγωγή χαρακτηριστικών (Feature Extraction) είναι ένα σημαντικό στάδιο στην ανάλυση του ήχου και ένα βασικό βήμα επεξεργασίας στη μηχανική μάθηση, αφού τα μοντέλα αναγνωρίζουν τα μοτίβα, που χρειάζονται, για να εκπαιδευτούν βασιζόμενοι σε ένα σχετικά μικρό αριθμό χαρακτηριστικών. Τα αρχικά δεδομένα ήχου είναι ογκώδη και ως εκ τούτου, είναι δύσκολο να επεξεργαστούν σε οποιοδήποτε πρόβλημα ανάλυσης. Επομένως, η αρχική αναπαράσταση δεδομένων πρέπει να μετατραπεί σε μια πιο κατάλληλη μορφή, η οποία θα αντιπροσωπεύει τις ιδιότητες των αρχικών ηχητικών σημάτων, μειώνοντας ταυτόχρονα τον όγκο των δεδομένων.

Στις περισσότερες μεθόδους ανάλυσης και επεξεργασίας του ήχου, το σήμα κατατάσσεται αρχικά σε βραχυπρόθεσμα πλαίσια (short-term windows). Αυτή η προσέγγιση χρησιμοποιείται επίσης κατά τη διάρκεια εξαγωγής χαρακτηριστικών. Αρχικά, το ηχητικό σήμα διασπάται σε ενδεχομένως επικαλυπτόμενα πλαίσια και ένα σύνολο χαρακτηριστικών υπολογίζεται στο κάθε πλαίσιο. Μ' αυτόν τον τρόπο παράγονται μοναδικά διανύσματα χαρακτηριστικών για τον κάθε ήχο. Η διάσταση αυτών των διανυσμάτων δεν είναι πάντα σταθερή. Στα περισσότερα προβλήματα ανάλυσης εξάγονται πολλά χαρακτηριστικά από τους διαθέσιμους ήχους, για να προκύψουν τελικά διανύσματα χαρακτηριστικών μεγάλων διαστάσεων. Η εξαγόμενη ακολουθία χαρακτηριστικών υπό μορφή διανυσμάτων χρησιμοποιείται στην ανάλυση των δεδομένων.

Μια άλλη κοινή τεχνική είναι η μεσοπρόθεσμη επεξεργασία της ακολουθίας χαρακτηριστικών. Το ηχητικό σήμα κατανέμεται, αρχικά σε μεσοπρόθεσμα τμήματα (mid-term windows) και στη συνέχεια, για κάθε τμήμα, πραγματοποιείται βραχυπρόθεσμη επεξεργασία. Στη συνέχεια, η ακολουθία των χαρακτηριστικών, χρησιμοποιείται για την επεξεργασία στατιστικών. Στο τέλος, κάθε μεσοπρόθεσμο κομμάτι αντιπροσωπεύεται από ένα σύνολο στατιστικών στοιχείων, που ανταποκρίνονται στις αντίστοιχες ακολουθίες βραχυπρόθεσμων χαρακτηριστικών. Κατά τη διάρκεια του χρόνου, υποθέτουμε ότι τα μεσοπρόθεσμα τμήματα παρουσιάζουν ομοιογένεια σε σχέση με τον τύπο ήχου και συνεπώς, έχει νόημα να προχωρήσουμε με την εξαγωγή στατιστικών στοιχείων στα επιμέρους τμήματα. Στην πράξη, η διάρκεια των μεσοπρόθεσμων πλαισίων (mid-term windows) κυμαίνεται συνήθως ανάμεσα σε 1-10 s.

Κάποιες φορές στο πλαίσιο της κατηγοριοποίησης, είναι επιθυμητό να εξαχθεί ένα μόνο διάνυσμα χαρακτηριστικών ως εκπρόσωπος ολόκληρου του ηχητικού σήματος. Σε τέτοιες περιπτώσεις, αρχικά εξάγονται τα βραχυπρόθεσμα χαρακτηριστικά και τα μεσοπρόθεσμα στατιστικά στοιχεία και υπολογίζονται σε τμήματα κάποιων δευτερολέπτων. Στο τέλος, τα μεσοπρόθεσμα στατιστικά στοιχεία ζυγίζονται μακροπρόθεσμα και προκύπτει ο μέσος όρος τους (long-term average), προκειμένου να παραχθεί μια ενιαία αναπαράσταση διανυσμάτων ολόκληρου του ήχου. Αξίζει να σημειωθεί, ότι ο τελικός μακροπρόθεσμος μέσος όρος των χαρακτηριστικών προβάλλει τα σημαντικότερα χαρακτηριστικά του ήχου. [32]

3.1.1.Χρονικά χαρακτηριστικά ήχου

Γενικά τα χαρακτηριστικά ήχου, που αφορούν τον χρόνο εξάγονται απευθείας από τα δείγματα του ακουστικού σήματος. Χαρακτηριστικά παραδείγματα είναι η βραχυπρόθεσμη ενέργεια και ο ρυθμός μηδενικής διέλευσης. Αυτά τα χαρακτηριστικά προσφέρουν έναν απλό τρόπο ανάλυσης ακουστικών σημάτων, αν και είναι συνήθως απαραίτητος ο συνδυασμός τους με πιο εξελιγμένα γνωρίσματα συχνοτήτων. Παρακάτω ακολουθεί μια αναφορά στα πιο γνωστά από τα χαρακτηριστικά που αφορούν τον χρόνο. [32]

3.1.1.1.Ενέργεια – Energy

Ένα ηχητικό σήμα θεωρείται σαν συνάρτηση του χρόνου. Ο όρος «ενέργεια σήματος» αντιπροσωπεύει την «ισχύ του σήματος». Ουσιαστικά υπάρχουν δύο επιλογές για την μέτρηση της ενέργειας. Η πρώτη είναι ο υπολογισμός της περιοχής κάτω από την απόλυτη τιμή της συνάρτησης και η δεύτερη αφορά τον υπολογισμό της περιοχής κάτω από το τετράγωνο της συνάρτησης. Συνήθως, χρησιμοποιείται η απόλυτη τιμή σε συνδυασμό με το τετράγωνο, ώστε να εξασφαλιστεί η λήψη θετικών τιμών. Αυτή η τακτική ευνοείται λόγω της μαθηματικής της ικανότητας και της ομοιότητάς της με την ευκλείδεια νόρμα, που χρησιμοποιείται στις τεχνικές ανίχνευσης σημάτων. Σύμφωνα με τα παραπάνω, ο τύπος της ισχύς είναι ο ακόλουθος:

$$E_x = \int_{-\infty}^{+\infty} |x(t)|^2 dt$$

Όπου $x(t)$ είναι το ηχητικό σήμα. Σε όρους επεξεργασίας σήματος αυτό ονομάζεται ενέργεια. Ο συγκεκριμένος ορισμός μπορεί να εφαρμοστεί σε οποιοδήποτε σήμα, ανεξάρτητα από το αν διαθέτει πραγματική ενέργεια με τον κλασικό ορισμό της φυσικής ή όχι. Αν το σήμα είναι ένα ηλεκτρικό σήμα, τότε ο παραπάνω ορισμός δίνει τη συνολική ενέργεια του σήματος (σε Joules), που διαχέεται πάνω από μια αντίσταση 1 Ohm. Στη παρούσα εργασία εμείς ασχολούμαστε μόνο με ηχητικά σήματα. Τέλος, πρέπει να αναφερθεί, ότι είναι πολύ σύνηθες η ενέργεια να κανονικοποιείται, έχοντας σαν παρονομαστή το μήκος του πλαισίου (frame length).

3.1.1.2. Ρυθμός μηδενικής διέλευσης - Zero Crossing Rate

Ο ρυθμός μηδενικής διέλευσης είναι ο ρυθμός μεταβολής των σημείων κατά μήκος ενός σήματος, δηλαδή ρυθμός με τον οποίο το σήμα αλλάζει από θετικό σε αρνητικό ή από αρνητικό σε θετικό. Με άλλα λόγια, είναι ο αριθμός των φορών, που ο ήχος μεταβάλλει την τιμή του (από θετικό σε αρνητικό και αντίστροφα), διαιρούμενο με το μήκος του πλαισίου. Αυτό το χαρακτηριστικό έχει χρησιμοποιηθεί σε μεγάλο βαθμό τόσο στην αναγνώριση ομιλίας, όπως και στην ανάκτηση μουσικών πληροφοριών και είναι ένα από τα βασικά χαρακτηριστικά για την ταξινόμηση των ήχων. Ο ρυθμός μηδενικής διέλευσης, σύμφωνα με τα παραπάνω, ορίζεται ως εξής.

$$zcr = \frac{1}{T-1} \sum_{t=1}^{T-1} 1_{\mathbb{R}<0}(s_t s_{t-1})$$

Όπου το s είναι ένα ηχητικό σήμα μήκους T και $1_{\mathbb{R}<0}$ είναι η ταυτοτική συνάρτηση. Το zcr μπορεί να θεωρηθεί σαν ένα μέτρο του θορύβου του ήχου. Συνήθως, εμφανίζει υψηλότερες τιμές, όταν υπάρχει πολύς θόρυβος στον ήχο. Επίσης, για μονοφωνικά ηχητικά σήματα ο ρυθμός μηδενικής διέλευσης μπορεί να χρησιμοποιηθεί ως ένας πρωταρχικός αλγόριθμος ανίχνευσης βήματος.

3.1.1.3. Εντροπία ενέργειας - Entropy of energy

Πρόκειται για την εντροπία των κανονικοποιημένων ενεργειών κάθε υποπλασίου (sub-frame). Μπορεί να ερμηνευτεί ως ένα μέτρο ανίχνευσης απότομων αλλαγών. Για να υπολογιστεί η εντροπία, κάθε βραχυπρόθεσμο πλαίσιο χωρίζεται σε υποπλαίσια K καθορισμένης διάρκειας. Στη συνέχεια, για κάθε υποπλαίσιο j , υπολογίζεται η ενέργεια, όπως περιεγράφηκε παραπάνω και τελικά διαιρείται με τη συνολική ενέργεια $E_{shortframe}$. Η διαδικασία της διαίρεσης είναι μια τυπική διαδικασία και χρησιμεύει ως μέσο για την επεξεργασία της προκύπτουσας ακολουθίας ενεργειακών τιμών υποπλασίου $e_j, j = 1, \dots, K$

ως σειρά από πιθανότητες $e_j = \frac{E_{shortframe j}}{E_{shortframe i}}$

όπου

$$E_{shortframe i} = \sum_{k=1}^K E_{subframe k}$$

Και η εντροπία $H(i)$ της ακολουθίας e_j δίνεται τελικά από την ακόλουθη σχέση

$$H(i) = - \sum_{j=1}^K e_j \log_2(e_j)$$

Οι τιμές της εντροπίας είναι χαμηλότερες, εάν υπάρχουν απότομες αλλαγές στην ενέργεια του πλαισίου. Αυτό συμβαίνει, επειδή, εάν ένα υποπλαίσιο δώσει μια υψηλή ενεργειακή τιμή, τότε μια από τις τελικές πιθανότητες θα είναι μεγάλη και αυτό μειώνει την εντροπία της ακολουθίας e_j . Αυτή η πληροφορία μπορεί να χρησιμοποιηθεί για την ανίχνευση σημαντικών ενεργειακών αλλαγών, όπως στην περίπτωση πυροβολισμών ή εκρήξεων.

3.1.2. Χαρακτηριστικά συχνότητας ήχου

Προκειμένου να ορισθούν τα χαρακτηριστικά της συχνότητας ήχου, θα πρέπει πρώτα να υπολογισθεί ο διακεκριμένος μετασχηματισμός Fourier (Discrete Fourier Transformation). Ο μετασχηματισμός Fourier είναι ευρέως χρησιμοποιούμενος στην ανάλυση του ηχητικού σήματος, διότι παρέχει μια βολική αναπαράσταση της κατανομής της συχνότητας του ήχου. Ας θεωρηθεί, ότι $X_i(k), k = 1, \dots, W_{fL}$ είναι το μέγεθος των DFT συντελεστών του i -οστού ηχητικού πλαισίου. Παρακάτω θα αναλυθεί το πώς υπολογίζονται τα αντίστοιχα φασματικά χαρακτηριστικά βασιζόμενα στους συντελεστές DFT σ' ένα ηχητικό πλαίσιο. Ακόμη, πρέπει να σημειωθεί ότι το W_L είναι ο αριθμός των δειγμάτων σ' ένα βραχυπρόθεσμο πλαίσιο (short-term window). Είναι και ο αριθμός των συντελεστών DFT του συγκεκριμένου πλαισίου. Για την απλότητα των συμβόλων, θα θεωρηθεί ότι W_{fL} είναι ο αριθμός των συντελεστών, που χρησιμοποιούνται στους ακόλουθους υπολογισμούς. Τέλος, να τονιστεί, ότι όλες οι παρακάτω συναρτήσεις φάσματος χρησιμοποιούν σαν είσοδο το μέγεθος των DFT συντελεστών από ένα ηχητικό πλαίσιο και όχι το ίδιο το ηχητικό πλαίσιο. [32]

3.1.2.1. Φασματικό κέντρο - Spectral Centroid

Το φασματικό κέντρο είναι ένα χαρακτηριστικό (feature) που χρησιμοποιείται στην επεξεργασία ψηφιακού σήματος, για να χαρακτηρίσει ένα φάσμα. Υποδεικνύει πού εντοπίζεται το «κέντρο μάζας» του φάσματος και συνδέεται στενά με την εντύπωση της «φωτεινότητας» ενός ήχου. Το φασματικό κέντρο C_i του i -οστού πλαισίου υπολογίζεται ως εξής:

$$C_i = \frac{\sum_{k=1}^{W_{fL}} k X_i(k)}{\sum_{k=1}^{W_{fL}} X_i(k)}$$

Κάποιοι άνθρωποι χρησιμοποιούν το φασματικό κέντρο, για να περιγράψουν τη διάμεσο του φάσματος αντί για το μέσο. Πρόκειται, όμως, για διαφορετικά στατιστικά στοιχεία. Δεδομένου ότι και τα δύο είναι μέτρα φάσματος, σε ορισμένες περιπτώσεις θα εμφανίσουν κάποια ομοιότητα συμπεριφοράς. Αλλά επειδή τα τυπικά ακουστικά φάσματα δεν κατανέμονται κανονικά, τα δύο μέτρα θα δίνουν συχνά διαφορετικές τιμές. Οι υψηλές τιμές του φασματικού κέντρο αντιστοιχούν σε πιο "φωτεινούς" ήχους. Σύμφωνα με πειράματα, η ακολουθία του έχει υψηλές διακυμάνσεις για σήματα ομιλίας, και για τον λόγο αυτό έχει χρησιμοποιηθεί, εκτός των άλλων, σε μεθόδους διαχωρισμού ομιλίας - μουσικής (speech music discrimination).

3.1.2.2. Φασματική διασπορά - Spectral Spread

Η φασματική διασπορά είναι ένα μέτρο της εξάπλωσης του φάσματος σε σχέση με το κέντρο του και τυπικά περιγράφει τη μέση απόκλιση τιμών γύρω από ένα κέντρο, το οποίο συσχετίζεται συνήθως με το εύρος του σήματος. Προκειμένου να υπολογιστεί, πρέπει να γίνει η διαίρεση του φάσματος από το φασματικό κέντρο, σύμφωνα με την ακόλουθη ισότητα.

$$S_i = \sqrt{\frac{\sum_{k=1}^{W_{fL}} (k - C_i)^2 X_i(k)}{\sum_{k=1}^{W_{fL}} X_i(k)}}$$

Τα θορυβώδη σήματα έχουν συνήθως μεγάλη διασπορά, ενώ μεμονωμένοι ήχοι με απομονωμένες κορυφές έχουν ως αποτέλεσμα τη χαμηλή φασματική διασπορά. Παρόμοια με το κέντρο, η διασπορά κανονικοποιείται από την υψηλότερη συχνότητα, ώστε η τιμή της να κυμαίνεται από μηδέν έως ένα. Στην πράξη, η διασπορά φάσματος εφαρμόζεται σε μεθόδους, σύμφωνα με τους οποίους η ενέργεια ενός σήματος, που καταλαμβάνει κάποιο σχετικά περιορισμένο φάσμα συχνοτήτων, κατανέμεται εσκεμμένα σε πολύ μεγαλύτερο φασματικό εύρος, με σκοπό την αύξηση της ασφάλειας των τηλεπικοινωνιών, την αποφυγή υποκλοπών και τη μεγαλύτερη αντοχή στα παράσιτα και τις παρεμβολές. Οι σημαντικότερες μέθοδοι διασποράς φάσματος είναι η διασπορά φάσματος με άμεση ακολουθία (Direct Sequence - DSSS) και η διασπορά φάσματος με εναλλαγή συχνοτήτων (Frequency Hopping - FHSS). Στην πρώτη μέθοδο, οι πληροφορίες κατανέμονται σε πολύ μεγαλύτερο (τουλάχιστον 10 φορές)

εύρος, χρησιμοποιώντας μια ψευδοτυχαία ακολουθία. Σύμφωνα με τη δεύτερη μέθοδο, τα συστήματα μετατόπισης συχνότητας επιτυγχάνουν τα ίδια αποτελέσματα, χρησιμοποιώντας διαφορετική συχνότητα φορέα σε διαφορετικό χρόνο.

3.1.2.3. Φασματική εντροπία - Spectral Entropy

Η φασματική εντροπία, με βάση την εντροπία Shannon στη φυσική, ποσοτικοποιεί την κανονικότητα ή τυχαιότητα του φάσματος ισχύος κατά τη διάρκεια μιας δεδομένης χρονικής περιόδου. Η φασματική εντροπία υπολογίζεται με παρόμοιο τρόπο με την εντροπία της ενέργειας, με τη διαφορά ότι ο υπολογισμός λαμβάνει χώρα στην περιοχή των συχνοτήτων. Αρχικά, το φάσμα του βραχυπρόθεσμου πλαισίου διαιρείται σε L υποπεριοχές. Η ενέργεια E_f της υποζώνης f , $f = 0, \dots, L - 1$ κανονικοποιείται από τη συνολική φασματική ενέργεια, δηλαδή, $n_f = \frac{E_f}{\sum_{f=0}^{L-1} E_f}$. Η εντροπία της κανονικοποιημένης φασματικής ενέργειας n_f υπολογίζεται τελικά σύμφωνα με την εξίσωση:

$$H = - \sum_{f=0}^{L-1} n_f \log_2(n_f)$$

3.1.2.4. Φασματική ροή - Spectral Flux

Η φασματική ροή είναι ένα μέτρο για το πόσο γρήγορα μεταβάλλεται η ισχύς του φάσματος ενός σήματος. Υπολογίζεται συγκρίνοντας την ισχύ σε ένα συγκεκριμένο πλαίσιο (frame) με την ισχύ του φάσματος από το προηγούμενο πλαίσιο. Πιο συγκεκριμένα, συνήθως υπολογίζεται ως η δεύτερη νόρμα, γνωστή και ως Ευκλείδεια απόσταση μεταξύ των δύο κανονικοποιημένων μεγεθών του φάσματος από δύο διαδοχικά βραχυπρόθεσμα πλαίσια. Με τον τρόπο αυτό, η φασματική ροή δεν εξαρτάται από τη συνολική ισχύ, αφού τα φάσματα είναι κανονικοποιημένα. Επίσης, η φασματική ροή μπορεί να χρησιμοποιηθεί για τον προσδιορισμό του χρόνου ενός ηχητικού σήματος, ή για την ανίχνευση της έναρξης του. Για τον υπολογισμό της φασματικής ροής θα χρησιμοποιηθεί ο γενικός τύπος του διακριτού μετασχηματισμού Fourier, ως εξής:

$$FL_{(i,i-1)} = \sum_{k=1}^{W_{fL}} (EN_i(k) - EN_{i-1}(k))^2$$

Όπου, $EN_i(k) = \frac{X_i(k)}{\sum_{l=1}^{W_{fL}} X_i(l)}$

Και $EN_i(k)$ είναι ο k - οστός κανονικοποιημένος συντελεστής DFT- Discrete Fourier transform του i -οστού πλαισίου.

3.1.2.5. Κύλιση φάσματος - Spectrum roll off

Η κύλιση φάσματος είναι η συχνότητα κάτω από ένα καθορισμένο ποσοστό της συνολικής φασματικής ενέργειας, συνήθως 85%. Επιπλέον, αν ο m -οστός DFT συντελεστής ανταποκρίνεται στη συχνότητα κύλισης του i -οστού πλαισίου, τότε ο τύπος για τον υπολογισμό της συχνότητας είναι:

$$\sum_{k=1}^m X_i(k) = C \sum_{k=1}^{W_{fL}} X_i(k)$$

όπου C είναι το ποσοστό το οποίο ορίζει ο κάθε χρήστης. Η συχνότητα της κύλισης φάσματος κανονικοποιείται διαιρώντας το παραπάνω άθροισμα με το W_{fL} , και έτσι τελικά λαμβάνει τιμές μεταξύ 0 και 1.

Αυτό το χαρακτηριστικό μπορεί να χρησιμοποιηθεί για παράδειγμα στην προσέγγιση της μέγιστης ή της ελάχιστης συχνότητας, ρυθμίζοντας το επί τοις εκατό ποσοστό σε μια τιμή κοντά στο 1 ή το 0 αντίστοιχα. Πρόκειται για ένα μέτρο κυρτότητας του φασματικού σχήματος και χρησιμοποιείται για τη διάκριση μεταξύ της φωνής και της αφωνίας στην ομιλία και στη μουσική. Ο άφωνος λόγος έχει υψηλή αναλογία ενέργειας, που περιέχεται στο υψηλής συχνότητας εύρος του φάσματος.

3.1.2.6. Συντελεστές συχνότητας κλίμακας Mel – MFCCs – Mel frequency cepstral coefficients

Τα MFCCs είναι ένα χαρακτηριστικό που χρησιμοποιείται ευρέως στην αυτόματη αναγνώριση ομιλίας και ήχων. Πριν από την εισαγωγή των MFCCs, οι συντελεστές γραμμικής πρόβλεψης (LPCs) και οι γραμμικοί προγνωστικοί συντελεστές (LPCC) ήταν ο κύριος τύπος χαρακτηριστικών για την αυτόματη αναγνώριση ομιλίας (ASR), ειδικά με την χρήση των HMM ταξινομητών.

Στην επεξεργασία ήχου, το MFC είναι μια αναπαράσταση της ενέργειας του φάσματος (power spectrum) ενός ήχου, που βασίζεται σ' ένα γραμμικό μετασχηματισμό συνημίτονου ενός λογάριθμου φάσματος ισχύος σε μια μη γραμμική κλίμακα συχνότητας Mel.

Πιο αναλυτικά, ο τρόπος υπολογισμού των συντελεστών MFCCs είναι ο ακόλουθος: αρχικά, υπολογίζεται ο διακριτός μετασχηματισμός Fourier (DFT) ενός σήματος. Στη συνέχεια, χαρτογραφείται η ισχύς του φάσματος πάνω στην κλίμακα Mel, χρησιμοποιώντας τριγωνικά επικαλυπτόμενα παράθυρα. Έπειτα, λογαριθμείται η ισχύς σε κάθε μία από τις συχνότητες Mel. Τέλος, τα MFCC είναι οι διακριτοί συντελεστές μετασχηματισμού συνημίτονου της λογαριθμικής ισχύος του φάσματος στην κλίμακα Mel.

Δεδομένου ότι εφαρμόζουμε ένα μετασχηματισμό στο ίδιο το φάσμα συχνοτήτων, το φάσμα που προκύπτει δεν είναι ούτε στο πεδίο συχνοτήτων ούτε στο πεδίο του χρόνου. Ανήκει σε ένα νέο πεδίο που ονομάζεται quefrequency. Το λογαριθμισμένο αυτό φάσμα ονομάζεται cepstrum.

Η κλίμακα Mel, είναι μια αντιληπτική κλίμακα τονικών υψών, που κρίνονται από τους ακροατές, ότι είναι ίσα σε απόσταση το ένα απ' τ' άλλο. Το σημείο αναφοράς μεταξύ αυτής της κλίμακας και της μέτρησης της κανονικής συχνότητας ορίζεται από ένα αντιληπτικό βήμα 1000 mels σ' έναν τόνο 1000 Hz, 40 dB (decibel) πάνω από το κατώφλι του ακροατή. Σε πάνω από περίπου 500 Hz, κρίνεται από τους ακροατές ότι όλο και μεγαλύτερα διαστήματα παράγουν ίσες αυξήσεις βήματος. Ως αποτέλεσμα, τέσσερις οκτάβες στην κλίμακα Hertz άνω των 500 Hz καταλαμβάνουν περίπου δύο οκτάβες στην κλίμακα Mel. Η κλίμακα Mel μεγιστοποιεί τη συχνότητα, ώστε να ταιριάζει περισσότερο με το τι μπορεί να ακούσει το ανθρώπινο αυτί. Ο γενικός κανόνας είναι ότι οι άνθρωποι μπορούν πιο εύκολα να αναγνωρίσουν μικρές αλλαγές στην ομιλία σε χαμηλότερες συχνότητες. Η κλίμακα Mel προσπαθεί να καταγράψει τέτοιες διαφορές. Το όνομα Mel προέρχεται από τη λέξη μελωδία, για να δείξει, ότι η κλίμακα βασίζεται σε συγκρίσεις βήματος. Μια συχνότητα, που μετριέται σε Hertz (f) μπορεί να μετατραπεί στην κλίμακα Mel χρησιμοποιώντας τον παρακάτω τύπο:

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

Οι συντελεστές MFCCs είναι ουσιαστικά οι συντελεστές που συνθέτουν συλλογικά ένα MFC. Η διαφορά μεταξύ του cepstrum και του MFC (Mel-frequency cepstrum) είναι ότι στο δεύτερο οι ζώνες συχνοτήτων απέχουν εξίσου στην κλίμακα Mel, το οποίο προσεγγίζει περισσότερο την απόκριση του ανθρώπινου ακουστικού συστήματος από τις γραμμικά διαχωρισμένες ζώνες συχνοτήτων, που χρησιμοποιούνται στο κανονικό cepstrum. Αυτή η στρέβλωση συχνότητας μπορεί να επιτρέψει την καλύτερη αναπαραγωγή του ήχου, όπως στη συμπίεση ήχου.

3.1.2.7. Διάνυσμα χρώματος - Chroma vector

Το διάνυσμα χρώματος είναι μια απεικόνιση (12 στοιχείων) της φασματικής ενέργειας. Υπολογίζεται με την ομαδοποίηση των συντελεστών DFT ενός βραχυπρόθεσμου παραθύρου σε 12 κάδους. Το χρώμα είναι μια ενδιαφέρουσα αναπαράσταση για την μουσική, σύμφωνα με την οποία ολόκληρο το φάσμα προβάλλεται σε 12 κάδους, που αντιπροσωπεύουν τα 12 διαφορετικά ημιτόνια (ή χρώματα) της μουσικής οκτάβας. Κάθε κάδος παράγει το μέσο όρο των μακροπρόθεσμων μεγεθών των αντίστοιχων συντελεστών DFT ως εξής:

$$v_k = \sum_{n \in S_k} \frac{X_i(n)}{N_k}, \quad k \in 0, \dots, 11$$

όπου το S_k είναι ένα υποσύνολο των συχνοτήτων που αντιστοιχούν στους συντελεστές DFT και N_k είναι η πληθικότητα του S_k . Το χρωμοδιάνυσμα v_k , συνήθως υπολογίζεται σε ένα βραχυπρόθεσμο χρονικό διάστημα. Αυτό οδηγεί σ' έναν πίνακα V , ο οποίος στην πραγματικότητα είναι ένας πίνακας αναπαράστασης της ακολουθίας των χρωμο-διανυσμάτων και είναι γνωστός με τον όρο χρωμόγραμμα.

Στη μουσική, το χρώμα ή αλλιώς χρωμόγραμμα είναι ένα χαρακτηριστικό που σχετίζεται στενά με τις δώδεκα διαφορετικές τάξεις τονικού ύψους (απόσταση μεταξύ ημιτονοειδών). Μπορεί να δώσει χρήσιμες μουσικές πληροφορίες για τον ήχο και μπορεί ακόμη και να αποκαλύψει την αντιληπτή μουσική ομοιότητα, που δεν είναι εμφανής στα αρχικά φάσματα.

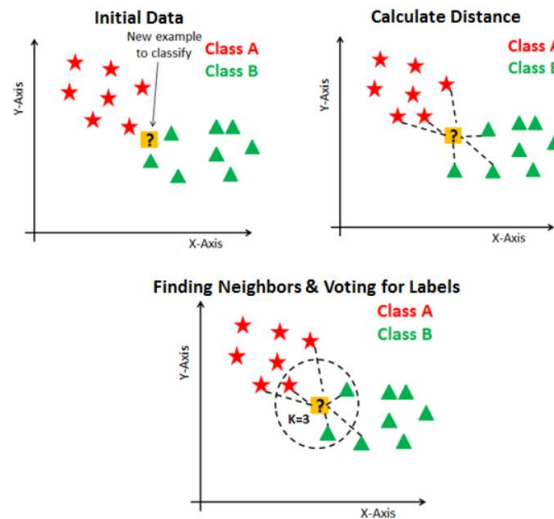
Ο εντοπισμός των διαβαθμίσεων, που διαφέρουν κατά μια οκτάβα, καθιστούν το χρώμα να είναι ένα καθιερωμένο εργαλείο επεξεργασίας και ανάλυσης μουσικών δεδομένων. Για παράδειγμα, η αναγνώριση χορδών βασίζεται σε κάποιο είδος χρωματογραφίας. Επίσης, τα χαρακτηριστικά χρώματος έχουν γίνει πρότυπο στην ευθυγράμμιση της μουσικής και στον συγχρονισμό και στην ανάλυση δομής ήχου. Τέλος, έχει αποδειχθεί ότι αποτελούν ισχυρό εργαλείο στην ταυτοποίηση του τραγουδιού ή στην αναγνώριση και αντιστοίχιση του ήχου.

3.2. Αλγόριθμοι κατηγοριοποίησης

Παρακάτω παρουσιάζονται μερικοί από τους πιο δημοφιλείς αλγορίθμους ταξινόμησης, οι οποίοι χρησιμοποιήθηκαν για μελέτη στην παρούσα εργασία.

3.2.1. Κ κοντινότερος γείτονας - KNN – K nearest neighbor

Η KNN ταξινόμηση είναι μια από τις πιο διαδεδομένες και απλές μεθόδους ταξινόμησης και πρέπει να είναι μια από τις πρώτες επιλογές σε ένα πρόβλημα ταξινόμησης, όταν υπάρχει λίγη ή καθόλου πρότερη γνώση γύρω από την κατανομή των δεδομένων. Ο αλγόριθμος αυτός αναπτύχθηκε από την ανάγκη να πραγματοποιηθεί διακριτή ανάλυση, όταν αξιόπιστες παραμετρικές εκτιμήσεις των πυκνοτήτων πιθανότητας ήταν άγνωστες ή δύσκολο να προσδιοριστούν. Το 1967, διατυπώθηκαν μερικές από τις τυπικές ιδιότητες της KNN μεθόδου, όπως δείχθηκε ότι για $k = 1$ και $n \rightarrow \infty$ το σφάλμα ταξινόμησης οριοθετείται παραπάνω από το διπλάσιο του ποσοστού σφάλματος Bayes. Στη συνέχεια, προέκυψε μια μακρά σειρά από έρευνες, που περιλάμβαναν νέες προσεγγίσεις απόρριψης (Hellman, 1970), βελτιώσεις σε σχέση με το ποσοστό σφάλματος Bayes (Fukunaga & Hostetler, 1975), προσεγγίσεις σταθμισμένης απόστασης (Dudani, 1976, Bailey & Jain, 1978), μέθοδοι υπολογισμού εξομάλυνσης (Bermejo & Cabestany, 2000) και ασαφείς μέθοδοι (Jozwik, 1983, Keller κ.ά., 1985). [13]



Εικόνα 5 Απεικόνιση της λειτουργίας του KNN αλγορίθμου

Ο KNN ταξινομητής βασίζεται συνήθως στην Ευκλείδεια απόσταση μεταξύ δύο συνόλων δεδομένων, του συνόλου δοκιμής και του συνόλου εκπαίδευσης. Συγκεκριμένα, έστω x_i ένα δείγμα εισόδου με p χαρακτηριστικά $(x_{i1}, x_{i2}, \dots, x_{ip})$ και n είναι ο συνολικός αριθμός δειγμάτων εισόδου ($i = 1, 2, \dots, n$) και

p ο συνολικός αριθμός χαρακτηριστικών ($j = 1, 2, \dots, p$). Η ευκλείδεια απόσταση μεταξύ των δειγμάτων x_i και x_l ($l = 1, 2, \dots, n$) ορίζεται ως

$$d(x_i, x_l) = \sqrt{(x_{i1} - x_{l1})^2 + (x_{i2} - x_{l2})^2 + \dots + (x_{ip} - x_{lp})^2}$$

Σύμφωνα με τον αλγόριθμο KNN, ο χρήστης προκαθορίζει την τιμή της σταθερής παραμέτρου K . Ο αλγόριθμος αναζητά μέσα στον διδιάστατο χώρο τα K σημεία-παρατηρήσεις που βρίσκονται πλησιέστερα στη νέα άγνωστη παρατήρηση. Ο κατηγοριοποιητής εκχωρεί τη νέα παρατήρηση στην κλάση που πλειοψηφεί μεταξύ των K πλησιέστερων γειτόνων. [12] Όταν $K = 1$, τότε ο αλγόριθμος παίρνει την απλούστερη μορφή του και είναι γνωστός ως ο κοντινότερος γείτονας (nearest neighbor), δηλαδή το άγνωστο σημείο κατηγοριοποιείται στην κλάση του πιο κοντινού του γείτονα. Υπάρχουν διάφορες προσεγγίσεις, ώστε να βρεθεί ο βέλτιστος αριθμός για το K . Όμως, τυπικά δεν υπάρχει μια δομημένη μέθοδος για τον εντοπισμό του. Πρέπει να το ανακαλυφθεί με διάφορες προσπάθειες μέσα από δοκιμές και λάθη, υποθέτοντας, ότι τα δεδομένα εκπαίδευσης είναι άγνωστα. Η επιλογή πολύ μικρών τιμών για το K μπορεί να είναι θορυβώδης και να έχει μεγάλη επίδραση στο αποτέλεσμα. Από την άλλη βέβαια, μεγαλύτερες τιμές του K θα έχουν χαλαρότερα όρια απόφασης, που σημαίνει χαμηλότερη διακύμανση αλλά και αυξημένο συστηματικό σφάλμα (bias). Επίσης, πρόκειται για μια υπολογιστικά δαπανηρή διαδικασία. Ένας άλλος τρόπος για την διαλογή του K είναι το cross validation. Αρχικά, επιλέγεται ένα μικρό τμήμα από το σύνολο δεδομένων εκπαίδευσης, το οποίο ονομάζεται σύνολο επικύρωσης. Στη συνέχεια, το σύνολο επικύρωσης δοκιμάζεται προκειμένου να

αξιολογηθούν διάφορες πιθανές τιμές του K . Αυτό που θέλουμε να εντοπίσουμε είναι ποια τιμή του K μας δίνει την καλύτερη απόδοση στο σύνολο επικύρωσης. Συνεπώς, μπορούμε να πάρουμε αυτήν την τιμή και να τη χρησιμοποιήσουμε ως την τελική στον αλγόριθμο, ελαχιστοποιώντας το σφάλμα επικύρωσης. Μια ακόμη προσέγγιση είναι επιλέγοντας την τιμή του K να είναι $K = \sqrt{N}$ όπου N ο αριθμός των δειγμάτων στο σύνολο δεδομένων εκπαίδευσης. [15]

Ο KNN αλγόριθμος μπορεί να βελτιωθεί σημαντικά, αν τα σημεία τα οποία είναι πλησιέστερα στη νέα παρατήρηση συνεισφέρουν περισσότερο. Ένας απλός τρόπος να επιτευχθεί αυτό είναι να προστεθούν συντελεστές βαρύτητας στα σημεία. Μία αποτελεσματική ιδέα είναι ότι οι συντελεστές θα μπορούσαν να είναι ίσοι με $\frac{1}{d}$, όπου d η απόσταση του εκάστοτε σημείου από τη νέα παρατήρηση. Ένα σημαντικό μειονέκτημα στον υπολογισμό των μέτρων απόστασης κατευθείαν από το σύνολο εκπαίδευσης είναι στην περίπτωση όπου οι μεταβλητές έχουν διαφορετικές κλίμακες μέτρησης ή υπάρχει ένα μείγμα αριθμητικών και κατηγορικών μεταβλητών. Συγκεκριμένα, οι μεταβλητές με μεγάλο εύρος τιμών επηρεάζουν περισσότερο το αποτέλεσμα από τις μεταβλητές με μικρό εύρος τιμών. Εάν για παράδειγμα οι παρατηρήσεις έχουν δύο γνωρίσματα A και B και το A παίρνει τιμές από 1 έως 1000, ενώ το B παίρνει τιμές από 1 έως 10, τότε το γνώρισμα A επηρεάζει δυσανάλογα την απόσταση σε σχέση με το γνώρισμα B . Το πρόβλημα αυτό μπορεί να αντιμετωπιστεί με την κανονικοποίηση των αριθμητικών τιμών, δηλαδή απλά διαιρώντας τις τιμές των χαρακτηριστικών με την περιοχή τιμών των χαρακτηριστικών ως εξής: [16][14]

$$X_i = \frac{X - \text{Min}}{\text{Max} - \text{Min}}$$

Ένα επιπλέον πρόβλημα είναι ότι ο υπολογισμός της ομοιότητας με βάση την Ευκλείδεια απόσταση υποθέτει την ισότιμη συμμετοχή όλων των γνωρισμάτων, το οποίο γενικώς δεν ισχύει. Το πρόβλημα αυτό αντιμετωπίζεται με τον καθορισμό "βαρών" για την κάθε διάσταση. Ο καθορισμός των βαρών επιτρέπει την αναδιατύπωση του υπολογισμού της απόστασης σύμφωνα με την

$$d(X, Y) = \sqrt{\sum_{i=1}^n w_i (x_i - y_i)^2}$$

όπου w_i είναι το βάρος που αντιστοιχεί στην i -οστή διάσταση. Γενικά, έχουν αναπτυχθεί διάφοροι μέθοδοι για τον υπολογισμό των βαρών. [14]

Στον KNN αλγόριθμο, εκτός από την ευκλείδεια απόσταση, η οποία είναι και η πιο συνηθισμένη, χρησιμοποιούνται ακόμα δύο αποστάσεις. Μια από αυτές είναι η Manhattan

$$\sum_{i=1}^k |x_i - y_i|$$

και η Minkowski

$$\sum_{i=1}^k (|x_i - y_i|)^q$$

Θα πρέπει να σημειωθεί ότι ο περιορισμός και τα τρία μέτρα απόστασης ισχύουν μόνο για συνεχείς αριθμητικές μεταβλητές. Στην περίπτωση κατηγορικών μεταβλητών, κατά κανόνα χρησιμοποιείται η απόσταση Hamming. Η συνάρτηση αυτή επιστρέφει την τιμή 0, εάν οι τιμές του ίδιου κατηγορικού γνωρίσματος δύο διαφορετικών παρατηρήσεων είναι ίδιες, αλλιώς επιστρέφουν την τιμή 1. Πιο συγκεκριμένα, ο τύπος της συνάρτησης είναι ο ακόλουθος: [16]

$$D_H = \sum_{i=1}^k |x_i - y_i|$$
$$\begin{cases} x = y \rightarrow D = 0 \\ x \neq y \rightarrow D = 1 \end{cases}$$

Η μέθοδος K-NN εκτός για κατηγοριοποίηση μπορεί να χρησιμοποιηθεί και για παλινδρόμηση, δηλαδή για πρόβλεψη αριθμητικών τιμών. Για την πρόβλεψη της τιμής της εξαρτημένης μεταβλητής μιας νέας παρατήρησης, ο αλγόριθμος εντοπίζει τις K πλησιέστερες παρατηρήσεις και επιστρέφει ως πρόβλεψη τη μέση τιμή των εξαρτημένων μεταβλητών των επιλεγμένων παρατηρήσεων.

Ένα από τα βασικά πλεονεκτήματα του KNN είναι ότι είναι ιδιαίτερα αποτελεσματικός, όταν υπάρχουν σύνθετες εξαρτήσεις μεταξύ των μεταβλητών. Επίσης, πρόκειται για μια σχετικά απλή μέθοδο και σε ορισμένα προβλήματα πετυχαίνουν υψηλές επιδόσεις κατηγοριοποίησης. Σαφώς δεν πρόκειται για έναν τέλειο αλγόριθμο. Εκτός από τα πλεονεκτήματα διαθέτει και ορισμένα μειονεκτήματα, με το βασικότερο να είναι το γεγονός ότι είναι χρονοβόρος. Ιδιαίτερα στα προβλήματα που ο αριθμός των πιθανών γειτόνων είναι υψηλός. Επιπλέον, κατά τη διάρκεια της κατηγοριοποίησης, γίνονται πολλές συγκρίσεις μεταξύ παρατηρήσεων και απαιτούνται αποτελεσματικές τεχνικές καταλογοποίησης (indexing). Τα τελικά αποτελέσματα μπορούν να επηρεαστούν σημαντικά από τον αριθμό των K γειτόνων. Τέλος, είναι μια μέθοδος ευαίσθητη στα τοπικά χαρακτηριστικά των δεδομένων και στην ύπαρξη μη σημαντικών μεταβλητών εισόδου. [11] [14]

3.2.2. Μηχανισμοί διανυσμάτων στήριξης - Support vector machines

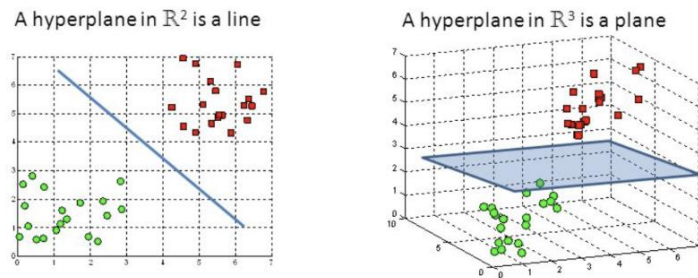
Οι μηχανισμοί διανυσμάτων στήριξης (SVM) χρησιμοποιούνται για να λύσουν ποκίλα προβλήματα του πραγματικού κόσμου. Συγκεκριμένα, είναι χρήσιμα στην κατηγοριοποίηση ενός κειμένου. Επίσης η ταξινόμηση των εικόνων μπορεί να γίνει χρησιμοποιώντας τον ταξινομητή SVM. Τα πειραματικά αποτελέσματα δείχνουν, ότι τα SVM επιτυγχάνουν σημαντικά υψηλότερη ακρίβεια και τα συστήματα κατακερματισμού εικόνας χρησιμοποιούν μια

τροποποιημένη έκδοση του αλγορίθμου. Ακόμη, οι χειρόγραφοι χαρακτήρες μπορούν να αναγνωριστούν χρησιμοποιώντας τους μηχανισμούς διανυσμάτων στήριξης. Ο αλγόριθμος SVM έχει εφαρμοστεί ευρέως στο πεδίο της βιολογίας. Έχουν χρησιμοποιηθεί για την ταξινόμηση πρωτεϊνών και έως το 90% των ενώσεων έχουν ταξινομηθεί σωστά. [18]

Πριν τη δεκαετία του '80, σχεδόν όλοι οι μέθοδοι εκμάθησης βασίζονταν σε γραμμικά προβλήματα αποφάσεων. Στη συνέχεια της δεκαετίας, τα δέντρα αποφάσεων και ο KNN αλγόριθμος επέτρεπαν αποδοτική εκμάθηση σε μη γραμμικά προβλήματα. Το θεωρητικό πλαίσιο ήταν ακόμη σε πολύ χαμηλό επίπεδο και όλα τα πορίσματα επάσχον από τοπικά ελάχιστα. Μέχρι που μια δεκαετία αργότερα χτίστηκαν αλγόριθμοι εκπαίδευσης για μη γραμμικές συναρτήσεις, βασιζόμενοι στην υπολογιστική θεωρία μάθησης. Πλέον τα θεωρήματα ήταν ολοκληρωμένα και χρήσιμες ιδιότητες απέρρεαν από αυτά. Συγχρόνως η διαχωρισιμότητα των μη γραμμικών περιοχών, που χρησιμοποιούν «συναρτήσεις του πυρήνα» έγινε πιο αποδοτική. Επίσης χρησιμοποιήθηκε η τετραγωνική βελτιστοποίηση προβλημάτων, για να αποφευχθούν τα προβλήματα των τοπικών ελαχίστων με νευρωνικά δίκτυα. Ο αλγόριθμος που προέκυψε ήταν ένας αλγόριθμος βελτιστοποίησης, παρά μια ατέρμονη αναζήτηση. [17]

Η ταξινόμηση των δεδομένων είναι ένα πολύ συνηθισμένο έργο στην μηχανική μάθηση. Η βασική ιδέα γύρω από τους μηχανισμούς διανυσμάτων υποστήριξης SVM είναι ακριβώς η ίδια με τα πολλαπλών στρώσεων νευρωνικά δίκτυα. Ένα από τα βασικά χαρακτηριστικά των SVM είναι οι συναρτήσεις πυρήνα (kernel Function), οι οποίες μπορεί να είναι γραμμικές, πολυωνυμικές και εκθετικές. Πρόκειται για συναρτήσεις, οι οποίες παίρνουν σαν είσοδο, χώρους χαμηλών διαστάσεων και τους μετασχηματίζουν σε χώρους υψηλότερων διαστάσεων, δηλαδή μετατρέπουν μη διαχωρίσιμα προβλήματα σε διαχωρίσιμα μέσω εξαιρετικά πολυσύνθετων μετασχηματισμών. Αναλυτικότερα, αρχικά εντοπίζεται το βέλτιστο υπερπλάσιο (hyperplane) για τα γραμμικά διαχωρίσιμα μοτίβα. Έπειτα, γίνεται η επέκταση σε μοτίβα που δεν είναι γραμμικά διαχωρίσιμα με μετασχηματισμούς των αρχικών δεδομένων, ώστε να χαρτογραφηθούν στο νέο χώρο. Αυτό είναι γνωστό σαν το τέχνασμα του πυρήνα (kernel trick) [21]. Ας υποθέσουμε, ότι ορισμένα δεδομένα - σημεία ανήκουν σε μία από τις δύο κλάσεις και ο στόχος είναι να αποφασιστεί σε ποια κατηγορία θα ταξινομηθεί ένα νέο σημείο. Στην περίπτωση του SVM αλγόριθμου, ένα σημειακό δεδομένο θεωρείται σαν ένα p -διάστατο διάνυσμα (μία λίστα από p αριθμούς). Το σημαντικό θέμα είναι αν υπάρχει η δυνατότητα διαχωρισμού του συγκεκριμένου σημείου σ' ένα $(p-1)$ -διάστατο υπερπλάσιο. Στη γεωμετρία με τον όρο υπερπλάσιο (hyperplane) αναφερόμαστε σ' έναν υπόχωρο, του οποίου η διάσταση είναι κατά μία λιγότερη από τον περιβάλλοντα χώρο του. Ουσιαστικά, τα υπερπλάσια είναι όρια απόφασης, που βοηθούν στην ταξινόμηση των νέων σημείων. Τα σημεία που εμπίπτουν σε κάθε πλευρά του υπερπλαισίου μπορούν να αποδοθούν σε διαφορετικές κατηγορίες. Επίσης, η διάσταση του υπερπλαισίου εξαρτάται από τον αριθμό των χαρακτηριστικών (features). Εάν ο αριθμός των χαρακτηριστικών εισόδου είναι 2, τότε το υπερπλάσιο είναι μόνο μια γραμμή. Εάν ο αριθμός των χαρακτηριστικών εισόδου είναι 3, τότε το υπερπλάσιο

γίνεται ένα δισδιάστατο επίπεδο. Είναι δύσκολο να φανταστεί κάποιος τον χώρο, όταν ο αριθμός των χαρακτηριστικών ξεπερνά τα τρία.



Εικόνα 6 Υπερπλάσια σε χώρους χαρακτηριστικών 2 και 3 διαστάσεων

Τα διανύσματα στήριξης είναι σημεία, που είναι πλησιέστερα στο υπερπλάσιο και επηρεάζουν τη θέση και τον προσανατολισμό του ίδιου του υπερπλαισίου. Χρησιμοποιώντας αυτούς τους φορείς στήριξης, μεγιστοποιούμε το περιθώριο (margin) του ταξινομητή. Συνεπώς, η διαγραφή των διανυσμάτων στήριξης θα αλλάξει τη θέση του υπερπλαισίου. Αυτά είναι τα σημεία που μας βοηθούν να κατασκευάσουμε τον SVM. [19] [20]

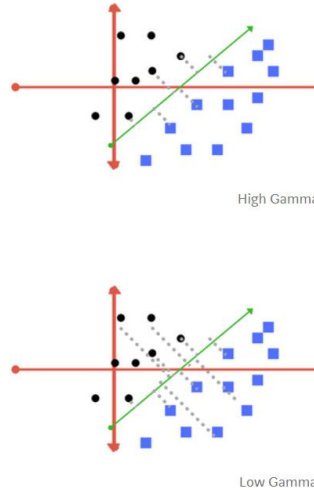
Υπάρχουν αρκετοί ρυθμιστικοί παράγοντες σχετικά με τον αλγόριθμο SVM. Αρχικά, πρέπει να αναφερθεί η παράμετρος κανονικοποίησης, ή αλλιώς παράμετρος C, η οποία ουσιαστικά καθορίζει πόσο θέλουμε να αποφευχθεί η εσφαλμένη ταξινόμηση κάθε δείγματος εκπαίδευσης. Για τις μεγάλες τιμές του C, η βελτιστοποίηση θα επιλέξει ένα υπερπλάσιο μικρότερου περιθωρίου (smaller - margin hyperplane). Αντίθετα, μια πολύ μικρή τιμή του C θα αναγκάσει το βελτιστοποιητή να αναζητήσει ένα μεγαλύτερου περιθωρίου υπερπλάσιο (larger-margin hyperplane), ακόμη και αν αυτό ταξινομεί εσφαλμένα περισσότερα σημεία. Μια καλή επιλογή για το βέλτιστο υπερπλάσιο είναι αυτή που επιτρέπει το μεγαλύτερο διαχωρισμό ή περιθώριο (margin) μεταξύ των δύο τάξεων. Έτσι, επιλέγεται το υπερπλάσιο, ώστε η απόσταση από αυτό στο κοντινότερο σημείο από κάθε πλευρά να μεγιστοποιείται. Ένα τέτοιο υπερπλάσιο ονομάζεται υπερπλάσιο ανώτατου περιθωρίου (maximum -margin hyperplane) και ο γραμμικός ταξινομητής που ορίζει είναι γνωστός ως ταξινομητής μέγιστου περιθωρίου, ή αλλιώς νευρώνας perceptron (νευρώνας αντίληπτρο) βέλτιστης σταθερότητας.



Εικόνα 7 Αριστερά χαμηλή τιμή κανονικοποίησης ενώ δεξιά υψηλή

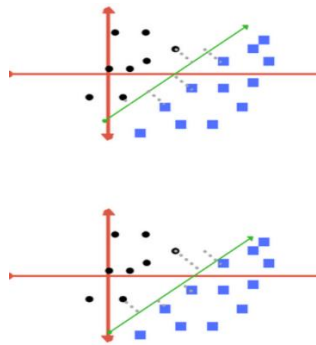
Μια ακόμη παράμετρος σημαντική για τον SVM αλγόριθμο είναι η παράμετρος Γάμμα η οποία καθορίζει πόσο μακριά φτάνει η επιρροή ενός μοναδικού σημείου εκπαίδευσης. Συγκεκριμένα,

για χαμηλό Γάμμα, τα σημεία που είναι μακριά από την πιθανή γραμμή διαχωρισμού χρησιμοποιούνται στον υπολογισμό της γραμμής διαχωρισμού. Αντίθετα, υψηλό γάμμα σημαίνει, ότι τα σημεία που βρίσκονται κοντά στην γραμμή διαχωρισμού είναι αυτά συνεισφέρουν στον υπολογισμό της.



Εικόνα 8 Στην πάνω εικόνα με υψηλό Γάμμα λαμβάνονται υπόψη μόνο τα κοντινά σημεία, ενώ στην κάτω με χαμηλό Γάμμα συνεισφέρουν και τα μακρινά

Η τελευταία παράμετρος και η πιο σημαντική για τον ταξινομητή SVM είναι το περιθώριο (margin). Ένα καλό περιθώριο είναι εκείνο όπου ο διαχωρισμός είναι μεγαλύτερος και για τις δύο τάξεις. Ένα καλό περιθώριο επιτρέπει στα σημεία να είναι στις αντίστοιχες τάξεις τους χωρίς να διασχίζουν την άλλη κατηγορία. [20]

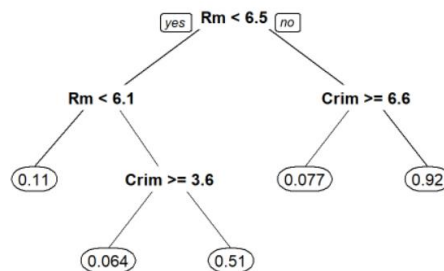


Εικόνα 9 Το πρώτο «καλό» περιθώριο απέχει εξίσου από τις δύο κλάσεις, αντίθετα το δεύτερο «κακό» περιθώριο βρίσκεται πολύ κοντά στη μπλε κλάση

Γενικά, όπως αναφέρθηκε και παραπάνω, όσο μεγαλύτερο είναι το περιθώριο, τόσο μικρότερο είναι το σφάλμα γενίκευσης του ταξινομητή. Η συνάρτηση απώλειας, που βοηθάει να μεγιστοποιηθεί το περιθώριο, είναι η συνάρτητη εύκαπτου συνδέσμου (hinge loss function).

3.2.3. Τυχαία δάση - Random forests

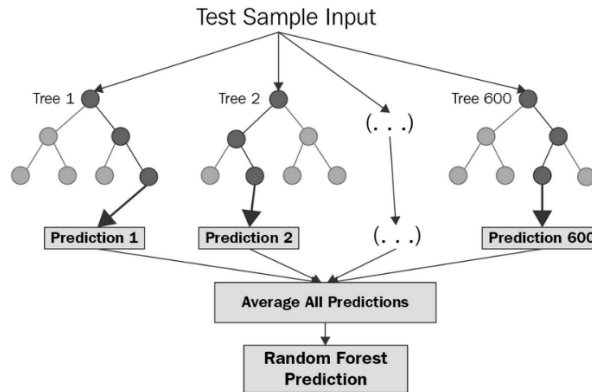
Αρχικά, θα περιγράψουμε τον αλγόριθμο των δέντρων αποφάσεων (decision tree), καθώς είναι ο δομικός λίθος του μοντέλου των τυχαίων δασών (random forest). Ένα δέντρο απόφασης είναι μια δομή που μοιάζει με διάγραμμα ροής στην οποία κάθε εσωτερικός κόμβος αντιπροσωπεύει μια "δοκιμή" σ' ένα χαρακτηριστικό. Κάθε κλάδος αντιπροσωπεύει το αποτέλεσμα της δοκιμής και κάθε κόμβος φύλλων αντιπροσωπεύει μία κλάση (η απόφαση λαμβάνεται στο τέλος αφού υπολογιστούν όλα τα χαρακτηριστικά). Οι διαδρομές από τη ρίζα στα φύλλα ουσιαστικά αντιπροσωπεύουν τους κανόνες ταξινόμησης. Οι μέθοδοι, που βασίζονται σε δέντρα, ενισχύουν τα προγνωστικά μοντέλα με υψηλή ακρίβεια, σταθερότητα και ευκολία ερμηνείας. Σε αντίθεση με τα γραμμικά μοντέλα, χαρτογραφούν τις μη γραμμικές σχέσεις αρκετά καλά. Επίσης, είναι προσαρμόσιμοι στην επίλυση οποιουδήποτε είδους προβλήματος, δηλαδή ταξινόμηση ή παλινδρόμηση. [1]



Εικόνα 10 Απεικόνιση του αλγόριθμου Random Forest για αριθμητικά δεδομένα

Τα τυχαία δάση ή τα τυχαία δάση αποφάσεων είναι μια μέθοδος μάθησης για την ταξινόμηση, και την παλινδρόμηση, τα οποία λειτουργούν κατασκευάζοντας πολλαπλά δέντρα αποφάσεων στο διαθέσιμο χρόνο εκπαίδευσης και στη συνέχεια εξάγουν τη σωστή τάξη (ταξινόμηση) ή τη μέση πρόβλεψη των επιμέρους δένδρων (παλινδρόμηση). Τα τυχαία δάση αποφάσεων διορθώνουν το συνηθισμένο πρόβλημα των δέντρων αποφάσεων, δηλαδή την υπερπροσαρμογή (overfitting) του συνόλου εκπαίδευσης (training set data). Ο πρώτος αλγόριθμος για τα τυχαία δάση αποφάσεων δημιουργήθηκε από τον Tin Kam Ho, χρησιμοποιώντας την τυχαία μέθοδο του υποχώρου, η οποία στη διατύπωση του Ho είναι ένας τρόπος να εφαρμοστεί η προσέγγιση "στοχαστικής διάκρισης" στην ταξινόμηση που προτείνει ο Eugene Kleinberg.

Ο Ho διαπίστωσε, ότι τα δάση των οποίων τα δέντρα χωρίζονται με λοξά υπερπλάσια, μπορούν να αποκτήσουν αρκετά μεγάλη ακρίβεια, καθώς μεγαλώνουν χωρίς να υποφέρουν από υπερπροσαρμογή, αρκεί τα δάση να χωρίζονται τυχαία, ώστε να επηρεάζονται μόνο από επιλεγμένα χαρακτηριστικά. Είναι αξιοσημείωτο, ότι ένας πιο περίπλοκος ταξινομητής, για παράδειγμα ένα μεγαλύτερο δάσος, γίνεται πιο ακριβής σχεδόν μονοτονικά και αυτό είναι σε πλήρη αντίθεση με την κοινή πεποίθηση ότι η πολυπλοκότητα ενός ταξινομητή μπορεί να αυξηθεί μόνο σ' ένα ορισμένο επίπεδο ακρίβειας, πριν μειωθεί εξαιτίας της υπερπροσαρμογής. Η ερμηνεία της αντοχής της δασικής μεθόδου στην υπερπροσαρμογή μπορεί να βρεθεί στη θεωρία των στοχαστικών διακρίσεων του Kleinberg. [3]



Εικόνα 11 Δομή του ταξινομητή Random Forest

Το τυχαίο δάσος, όπως υποδηλώνεται από το όνομά του, αποτελείται από ένα πλήθος ατομικών δέντρων αποφάσεων, που λειτουργούν ως σύνολο. Κάθε μεμονωμένο δέντρο στο τυχαίο δάσος κάνει μια πρόβλεψη κλάσης και η κλάση με τις περισσότερες ψήφους γίνεται η τελική πρόβλεψη του μοντέλου. Η ισχύς του πλήθους είναι η βασική ιδέα πίσω από τα τυχαία δάση. Σύμφωνα με την επιστήμη δεδομένων, ο λόγος που το μοντέλο των τυχαίων δασών λειτουργεί τόσο καλά είναι ότι ένας αρκετά μεγάλος αριθμός ασυσχέτιστων μοντέλων, συγκεκριμένα δέντρων, μπορούν να παράγουν προβλέψεις που είναι ακριβέστερες από οποιαδήποτε μεμονωμένη πρόβλεψη. Η επιτυχία του μοντέλου, όπως αναφέρθηκε παραπάνω, είναι η χαμηλή συσχέτιση μεταξύ των δέντρων. Ο λόγος γι' αυτό το αποτέλεσμα είναι, ότι τα δέντρα προστατεύονται μεταξύ τους από τα ατομικά λάθη, εφόσον δεν κάνουν πάντα λάθος στην ίδια κατεύθυνση. Πιο συγκεκριμένα, ενώ το αποτέλεσμα κάποιων δέντρων είναι λανθασμένο, πολλών άλλων είναι σωστό. Όπως είναι λογικό, η μέθοδος του τυχαίου δάσους πρέπει να διασφαλίσει, ότι η συμπεριφορά κάθε δέντρου δεν είναι αρκετά συσχετισμένη με τη συμπεριφορά οποιουδήποτε άλλου δέντρου στο ίδιο μοντέλο. [2]

Εδώ θα πρέπει να αναφερθούν μερικές σημαντικές τεχνικές. Αρχικά, θα περιγράψουμε την μέθοδο bootstrapping, η οποία είναι μια τεχνική στατιστικής αναδειγματοληψίας, που περιλαμβάνει τυχαία δειγματοληψία ενός συνόλου δεδομένων με αντικατάσταση. Χρησιμοποιείται συχνά ως μέσο ποσοτικοποίησης της αβεβαιότητας που συνδέεται μ' ένα μοντέλο μηχανικής μάθησης. Το bootstrapping είναι εξαιρετικά χρήσιμο, καθώς επιτρέπει τη δημιουργία νέων δειγμάτων από έναν πληθυσμό χωρίς να χρειάζεται να προχωρήσει και να συλλέξει πρόσθετα "στοιχεία εκπαίδευσης". Η διαδικασία είναι να επαναλαμβάνεται η δειγματοληψία των δεδομένων με αντικατάσταση από το αρχικό σύνολο εκπαίδευσης, ώστε να παράγονται πολλαπλά ξεχωριστά υποσύνολα εκπαίδευσης. Αυτά στη συνέχεια χρησιμοποιούνται, για να επιτρέψουν τις μεθόδους μετα-μάθησης (meta-learner), να μειώσουν τη διακύμανση των προβλέψεων, βελτιώνοντας έτσι σημαντικά την απόδοση της πρόβλεψης. Είναι αξιόλογο, να αναφέρουμε, ότι σ' αυτήν την περίπτωση δεν υποδιαιρούμε το σύνολο εκπαίδευσης σε μικρότερα τμήματα, για να εκπαιδεύουμε κάθε δέντρο σε διαφορετικό κομμάτι. Αντίθετα, αν έχουμε ένα δείγμα μεγέθους N , κρατάμε το πλήθος σταθερό και

συνεχίζουμε να εκπαιδεύουμε κάθε δέντρο με σύνολο εκπαίδευσης μεγέθους N . Αλλά αντί για τα αρχικά δεδομένα εκπαίδευσης, παίρνουμε ένα τυχαίο δείγμα μεγέθους N κάνοντας κάθε φορά αντικατάσταση. Για παράδειγμα, εάν υπήρχε μια λίστα, όπου τα δεδομένα εκπαίδευσης ήταν τα ακόλουθα $[1, 2, 3, 4, 5, 6]$, τότε θα μπορούσαμε να δώσουμε σ' ένα από τα δέντρα του δάσους την ακόλουθη λίστα $[1, 2, 2, 3, 6, 6]$. Παρατηρήστε ότι και οι στις δύο περιπτώσεις, πρωτύτερα και κατόπιν της αντικατάστασης, οι δύο λίστες εξακολουθούν να έχουν μήκος 6. Όπως αναφέρθηκε παραπάνω, ένα από τα κύρια μειονεκτήματα των δέντρων αποφάσεων είναι ότι υποφέρουν από εκτιμήσεις μεγάλης διακύμανσης. Αυτό σημαίνει, ότι η προσθήκη ενός μικρού αριθμού πρόσθετων παρατηρήσεων εκπαίδευσης μπορεί να μεταβάλει δραματικά την απόδοση πρόβλεψης ενός ήδη εκπαιδευμένου δέντρου, μολονότι τα δεδομένα εκπαίδευσης δεν θα αλλάξουν σε μεγάλο βαθμό. Αυτό αντιτίθεται μ' έναν εκτιμητή χαμηλής διακύμανσης, όπως είναι η γραμμική παλινδρόμηση, η οποία δεν επηρεάζεται από την προσθήκη επιπλέον δεδομένων, τα οποία είναι σχετικά κοντά με τα υπόλοιπα δεδομένα. Ένας τρόπος για να μετριαστεί αυτό το πρόβλημα είναι μέσω του bagging. Η λογική της μεθόδου αυτής είναι να συνδυαστούν πολλαπλά δέντρα, τα οποία είναι όλα προσαρμοσμένα σε ξεχωριστά δείγματα (bootstrapped samples) και υπολογίζουν τον μέσο όρο των προβλέψεων τους για να μειώσουν τη συνολική διακύμανση αυτών των προβλέψεων. Δηλαδή, εάν ληφθεί ο μέσος όρος των παρατηρήσεων αυτών, η διακύμανση μειώνεται κατά παράγοντα ίσο με τον αριθμό των παρατηρήσεων. Ωστόσο, όταν υπάρχει μόνο ένα σύνολο εκπαίδευσης σημαίνει, ότι είναι δύσκολο να δημιουργηθούν πάρα πολλά ανεξάρτητα σύνολα εκπαίδευσης. Εδώ απαιτείται η τεχνική του bagging. Αν B είναι τα ξεχωριστά δείγματα εκκίνησης (bootstrapped samples), που έχουν δημιουργηθεί με ξεχωριστούς εκτιμητές μοντέλων $\hat{f}^b(x)$, παίρνοντας τον μέσο όρο αυτών έχουμε έναν νέο εκτιμητή χαμηλής πλέον διακύμανσης $\widehat{f}_{avg}(x)$, όπου

$$\widehat{f}_{avg}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^b(x)$$

Ένα από τα βασικά πλεονεκτήματα του bagging είναι ότι δεν κάνει υπερπροσαρμογή το μοντέλο όταν αυξάνεται ο αριθμός των δειγμάτων εκκίνησης B . [4]

Τα τυχαία δάση ακολουθούν μια παρόμοια τεχνική με τη μέθοδο bagging, την τυχαιότητα των χαρακτηριστικών (feature Randomness), στην οποία τυχαία επιλέγεται ένα υποσύνολο από τα p -διάστατα χαρακτηριστικά σε κάθε ξεχωριστό δέντρο που αναπτύσσεται. Αυτό είναι επιθυμητό, ώστε να συμπεριληφθούν όσο το δυνατόν περισσότερα χαρακτηριστικά (features), προκειμένου να συγκεντρωθούν περισσότερες πληροφορίες για το μοντέλο. Ωστόσο, αυτό έχει ως στόχο τη σκόπιμη αποφυγή πολύ ισχυρών προγνωστικών χαρακτηριστικών που οδηγούν σε παρόμοιους διαχωρισμούς στα δέντρα. Σε ένα κανονικό δέντρο αποφάσεων, κατά τη διάρκεια του διαχωρισμού ενός κόμβου, κάθε πιθανό χαρακτηριστικό εξετάζεται, ώστε τελικά να επιλεγεί αυτό που παράγει τον μεγαλύτερο διαχωρισμό μεταξύ των παρατηρήσεων στον αριστερό κόμβο έναντι εκείνων που βρίσκονται στον δεξιό κόμβο. Ο αλγόριθμος των τυχαίων δέντρων ακολουθεί μια εντελώς αντίθετη προσέγγιση. Συγκεκριμένα, κάθε δέντρο επιλέγει μόνο από ένα τυχαίο υποσύνολο χαρακτηριστικών. Αυτό οδηγεί σε ακόμη περισσότερες παραλλαγές μεταξύ των δέντρων του μοντέλου και τελικά προκύπτει χαμηλότερη συσχέτιση

μεταξύ των δέντρων και μεγαλύτερη διαφοροποίηση. Έτσι, στο τυχαίο δάσος, καταλήγουμε σε δέντρα που όχι μόνο εκπαιδεύονται σε διαφορετικά σύνολα δεδομένων, αλλά και χρησιμοποιούν διαφορετικά χαρακτηριστικά κατά τη λήψη αποφάσεων.

3.2.4.Εξαιρετικά τυχαία δέντρα - Extra Trees - Extremely Randomized Trees

Ένα μοναδικό δέντρο αποφάσεων, όπως αναφέρθηκε παραπάνω, παρουσιάζει συνήθως υψηλή απόκλιση, καθώς υπερπροσαρμόζει (overfit) τα δεδομένα, από τα οποία εκπαιδεύεται. Αυτό πρακτικά συμβαίνει, γιατί μαθαίνει από ένα μόνο μονοπάτι των αποφάσεων. Λογικό επακόλουθο είναι ότι οι προβλέψεις από ένα δέντρο απόφασης συνήθως δεν είναι ακριβείς για την κατηγοροποίηση των νέων δεδομένων. Μια εξέλιξη της παραπάνω μεθόδου είναι το τυχαίο δάσος, όπως αναλυτικά έχει περιγραφεί, το οποίο παρουσιάζει μεσαία απόκλιση και εισάγει την έννοια της τυχειότητας. [5]

Τα Εξαιρετικά Τυχαία Δέντρα (Extremely Randomized Trees) έχουν γενικά χαμηλή απόκλιση. Ο κύριος στόχος τους είναι η περαιτέρω τυχειοποίηση της δομής των δέντρων, όπου η επιλογή του βέλτιστου σημείου κοπής είναι υπεύθυνη για την διακύμανση του επαγόμενου δένδρου. Μοιάζουν πολύ με τα τυχαία δάση, καθώς οικοδομούν πολλαπλά δέντρα (n estimators) και κόβουν κόμβους χρησιμοποιώντας τυχαία υποσύνολα χαρακτηριστικών, αλλά έχουν δύο βασικές διαφορές. Πρώτον τα εξαιρετικά τυχαία δάση δεν ακολουθούν την τεχνική bootstrapping για τις παρατηρήσεις, που σημαίνει ότι ο αλγόριθμος χρησιμοποιεί δείγματα χωρίς αντικατάσταση. Δεύτερον οι κόμβοι χωρίζονται βασιζόμενοι σε τυχαίους διαχωρισμούς μεταξύ ενός τυχαίου υποσυνόλου των χαρακτηριστικών που επιλέγονται στον κάθε κόμβο. Στα Εξαιρετικά Τυχαία Δέντρα, η τυχειότητα δεν προέρχεται από το bootstrapping των δεδομένων, αλλά πηγάζει από τυχαίους διαχωρισμούς όλων των παρατηρήσεων και ο αλγόριθμος προσπαθεί αντί να βρει ένα βέλτιστο σημείο κοπής για καθένα από τα τυχαία επιλεγμένα χαρακτηριστικά K κάθε κόμβου, να επιλέξει ένα σημείο κοπής τυχαία.[6]

Αυτή η μέθοδος οδηγεί συχνά σε αυξημένη ακρίβεια χάρη στην εξομάλυνση και ταυτοχρόνως μειώνει σημαντικά την υπολογιστική επιβάρυνση που συνδέεται με τον προσδιορισμό του βέλτιστου σημείου κοπής των δέντρων στα τυχαία δάση.

Από στατιστικής απόψεως, η απόρριψη της ιδέας του bootstrapping οδηγεί σε ένα πλεονέκτημα όσο αναφορά το συστηματικό σφάλμα (bias), ενώ η τυχειοποίηση των σημείων τομής έχει συχνά εξαιρετικό αποτέλεσμα στην μείωση της διακύμανσης. Αυτή η μέθοδος έχει αποφέρει προηγμένα αποτελέσματα σε αρκετά μεγάλα πολύπλοκα προβλήματα. [5]

Table 1 Extra-Trees splitting algorithm (for numerical attributes)

Split_a_node(S)

Input: the local learning subset S corresponding to the node we want to split

Output: a split $[a < a_c]$ or nothing

- If **Stop_split**(S) is TRUE then return nothing.
- Otherwise select K attributes $\{a_1, \dots, a_K\}$ among all non constant (in S) candidate attributes;
- Draw K splits $\{s_1, \dots, s_K\}$, where $s_i = \text{Pick_a_random_split}(S, a_i), \forall i = 1, \dots, K$;
- Return a split s_* such that $\text{Score}(s_*, S) = \max_{i=1, \dots, K} \text{Score}(s_i, S)$.

Pick_a_random_split(S, a)

Inputs: a subset S and an attribute a

Output: a split

- Let a_{\max}^S and a_{\min}^S denote the maximal and minimal value of a in S ;
- Draw a random cut-point a_c uniformly in $[a_{\min}^S, a_{\max}^S]$;
- Return the split $[a < a_c]$.

Stop_split(S)

Input: a subset S

Output: a boolean

- If $|S| < n_{\min}$, then return TRUE;
- If all attributes are constant in S , then return TRUE;
- If the output is constant in S , then return TRUE;
- Otherwise, return FALSE.

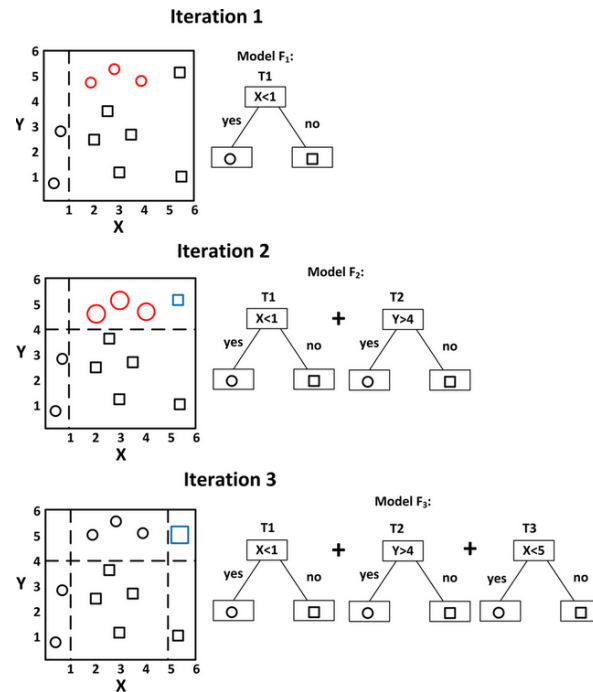
Εικόνα 12 Ψευδοκώδικας του αλγορίθμου Extra Trees

3.2.5.Βαθμωτή ενίσχυση - Gradient boosting

Η τεχνική της ενίσχυσης (Boosting) προέκυψε από την ιδέα, πως ένας αδύναμος ταξινομητής μπορεί να τροποποιηθεί, για να γίνει καλύτερος. Είναι ,δηλαδή, μια μέθοδος μετατροπής των αδύναμων ταξινομητών σε ισχυρούς. Μια αδύναμη υπόθεση ή ένας ανίσχυρος ταξινομητής ορίζεται ως εκείνος του οποίου η απόδοση είναι τουλάχιστον ελαφρώς καλύτερη από μια τυχαία επιλογή. Με την τεχνική της ενίσχυσης (boosting), κάθε νέο δέντρο είναι μια προσαρμογή σε μια τροποποιημένη έκδοση του αρχικού συνόλου δεδομένων. [10]

Ο αλγόριθμος gradient boosting μπορεί να αποσαφηνιστεί εύκολα, αφού πρώτα αναφερθούμε στον αλγόριθμο AdaBoost. Ο αλγόριθμος AdaBoost αρχίζει με την εκπαίδευση ενός δέντρου αποφάσεων, στο οποίο σε κάθε παρατήρηση αποδίδεται ίσο βάρος, δηλαδή όλες οι παρατηρήσεις είναι όμοια ζυγισμένες. Μετά την αξιολόγηση του πρώτου δένδρου, αυξάνουμε τα βάρη εκείνων των παρατηρήσεων που ήταν δύσκολο να ταξινομηθούν σωστά και μειώνουμε τα βάρη για τις παρατηρήσεις, οι οποίες ευχερώς ταξινομήθηκαν. Κατά συνέπεια, το δεύτερο δέντρο που δημιουργείται αναπτύσσεται σ' αυτά τα νέα σταθμισμένα δεδομένα. Εδώ, ο στόχος είναι να βελτιωθούν οι προβλέψεις του πρώτου δέντρου. Το νέο μοντέλο, επομένως, είναι «Δέντρο 1 + Δέντρο 2». Κατόπιν, υπολογίζεται το σφάλμα ταξινόμησης απ' αυτό το νέο μοντέλο δύο δέντρων και δημιουργείται ένα τρίτο δέντρο για την πρόβλεψη των αναθεωρημένων υπολοίπων (residuals). Η εν λόγω διαδικασία επαναλαμβάνεται για ένα συγκεκριμένο αριθμό επαναλήψεων. Τα μεταγενέστερα δέντρα βοηθούν στην ταξινόμηση των παρατηρήσεων, που δεν είναι σωστά ταξινομημένες από τα προηγούμενα δέντρα. Οι προβλέψεις του τελικού

μοντέλου είναι, συνεπώς, το σταθμισμένο άθροισμα των προβλέψεων, που προέκυψαν από τα προηγούμενα μοντέλα (δέντρα).



Εικόνα 13 Απεικόνιση του ταξινομητή Gradient Boosting

Σχετικά με τον αλγόριθμο Gradient Boosting, να αναφέρουμε ότι λειτουργεί εκπαιδευοντας πολλά μοντέλα με βαθμιαίο, προσθετικό και διαδοχικό τρόπο. Αρχικά, όπως στην περίπτωση του AdaBoost, χρησιμοποιούνται πολύ σύντομα δέντρα απόφασης, που έχουν μόνο ένα μοναδικό διαχωρισμό, το οποίο ονομάζεται decision stump. Η μεγάλη διαφορά μεταξύ του AdaBoost και του Gradient Boosting είναι ο τρόπος με τον οποίο οι δύο αλγόριθμοι εντοπίζουν τις ατέλειες των αδύναμων ταξινομητών (στη συγκεκριμένη περίπτωση των δέντρων αποφάσεων). Ενώ το μοντέλο AdaBoost αναγνωρίζει τις αδυναμίες χρησιμοποιώντας υψηλού βάρους δεδομένα, το μοντέλο gradient boosting κάνει το ίδιο τοποθετώντας βαθμίδες (gradients) στη συνάρτηση απώλειας (loss function). Έτσι, πίσω από τον αλγόριθμο GB βρίσκεται η επαναλαμβανόμενη μόχλευση των μοτίβων στα υπόλοιπα (residuals), η ενίσχυση ενός μοντέλου με αδύναμες προβλέψεις και τελικά η βελτίωση του. Μόλις φτάσουμε σ' ένα στάδιο, όπου τα υπόλοιπα (residuals) δεν έχουν κανένα πρότυπο, που θα μπορούσε να διαμορφωθεί, η μοντελοποίηση σταματάει, διαφορετικά ίσως προκύψει υπερπροσαρμογή (overfitting) των δεδομένων. Αλγοριθμικά, ελαχιστοποιούμε τη συνάρτηση απώλειας, ώστε να φθάσει στις ελάχιστες τιμές της. [9]

Η συνάρτηση απώλειας είναι ένα μέτρο που υποδεικνύει πόσο καλοί είναι οι συντελεστές (coefficients) του μοντέλου στην προσαρμογή των υποκείμενων δεδομένων [7]. Για να κατανοήσουμε τη συνάρτηση απώλειας, πρέπει να εξετάσουμε το τι προσπαθούμε να βελτιστοποιήσουμε. Για παράδειγμα, αν προσπαθούμε να προβλέψουμε τις τιμές πώλησης ακινήτων, τότε η συνάρτηση απώλειας θα βασίζεται στο σφάλμα μεταξύ των πραγματικών και

των προβλεπόμενων τιμών των κατοικιών. Γενικά, η συνάρτηση απώλειας που χρησιμοποιείται εξαρτάται από τον τύπο του προβλήματος που επιλύεται. Ο αλγόριθμος υποστηρίζει πολλές τυπικές συναρτήσεις απώλειας, αλλά και ο χρήστης μπορεί να ορίσει τη δική του. Για παράδειγμα, στην παλινδρόμηση μπορεί να χρησιμοποιηθεί το τετραγωνικό σφάλμα σαν συνάρτηση απώλειας, ενώ για τα προβλήματα ταξινόμησης μπορεί να εξεταστεί η λογαριθμική απώλεια (logarithmic loss). Ένα πλεονέκτημα του gradient boosting είναι ότι δεν χρειάζεται να προκύψει ένας νέος αλγόριθμος ενίσχυσης για κάθε συνάρτηση απώλειας, που μπορεί να θέλει να χρησιμοποιήσει ο χρήστης, αν αυτός πρόκειται για ένα αρκετά γενικό πλαίσιο που είναι ικανό να χειριστεί οποιαδήποτε διαφοροποιημένη συνάρτηση απώλειας. [8] Ένα λοιπόν από τα μεγαλύτερα κίνητρα για τη χρήση της του αλγορίθμου gradient boosting είναι ότι επιτρέπει τη βελτιστοποίηση μιας συνάρτησης κόστους που καθορίζεται από το χρήστη, αντί για μια συγκεκριμένη συνάρτηση απώλειας που συνήθως προσφέρει λιγότερο έλεγχο και δεν ανταποκρίνεται ουσιαστικά στις εφαρμογές του πραγματικού κόσμου.

3.3. Μέτρα αξιολόγησης πειραμάτων

Ένας τρόπος για να ελέγξουμε το πόσο καλά ο αλγόριθμος έμαθε το σύνολο εκπαίδευσης είναι να υπολογίσουμε το σφάλμα εκπαίδευσης, το οποίο είναι το ποσοστό των διανυσμάτων του συνόλου εκπαίδευσης, που παραμένουν λάθος ταξινομημένα στο τέλος του σταδίου της εκπαίδευσης. Ιδανικά, το σφάλμα εκπαίδευσης θα πρέπει να είναι ίσο με το μηδέν για να θεωρηθεί ο αλγόριθμος τέλειος. Ωστόσο, το ζητούμενο είναι ο ταξινομητής να είναι ικανός να γενικεύσει. Αυτό σημαίνει ότι θέλουμε ο αλγόριθμος να μπορεί να ανταπεξέλθει ικανοποιητικά σε δεδομένα εκτός αυτών του συνόλου εκπαίδευσης. Όπως λοιπόν προκύπτει, ένα μηδενικό σφάλμα εκπαίδευσης δεν εγγυάται γενίκευση και ίσως κρύβει το πρόβλημα της υπερπροσαρμογής (overfitting).

Ένα σημαντικό εργαλείο για την ανάλυση της απόδοσης των αλγορίθμων είναι ο πίνακας confusion matrix, ο οποίος ομαδοποιεί τα αποτελέσματα ταξινόμησης σ' έναν πίνακα και βοηθά στην κατανόηση των σφαλμάτων, που συμβαίνουν κατά τη διάρκεια της δοκιμής και της εκπαίδευσης των δεδομένων. Ο confusion matrix CM είναι ένας πίνακας $N_c \times N_c$, του οποίου οι σειρές αναφέρονται στην αληθινή (ground truth) κλάση, ενώ οι στήλες στην προβλεπόμενη κλάση του συνόλου δεδομένων. Μ' άλλα λόγια, κάθε στοιχείο, $CM(i, j)$ αντιπροσωπεύει τον αριθμό των δειγμάτων της κλάσης i που είχαν αποδοθεί στην κλάση j , σύμφωνα με την επιλεγμένη μέθοδο ταξινόμησης. Επομένως, η διαγώνιος του confusion matrix καταγράφει τις σωστές αποφάσεις ταξινόμησης ($i = j$). [32]

Ο confusion matrix που χρησιμοποιείται στην παρούσα εργασία κανονικοποιείται με την διαίρεση κάθε στοιχείου του CM με το άθροισμα των στοιχείων της αντίστοιχης σειράς, δηλαδή από τον (αληθινό) πληθυσμό της τάξης που έχει χαρτογραφηθεί στη συγκεκριμένη σειρά. Ουσιαστικά, όλες οι κλάσεις θεωρούνται ίσου μεγέθους και το σύνολο δεδομένων ισορροπημένο. Επίσης, εκτός από τον confusion matrix, εξετάζουμε την ακρίβεια (accuracy),

την ανάκληση (recall) και την ακρίβεια (precision) για κάθε μία από τις οχτώ κλάσεις καθώς και το μέτρο F_1 .

Η συνολική ακρίβεια (accuracy) του ταξινομητή, ορίζεται ως το ποσοστό των δειγμάτων του συνόλου δεδομένων που έχουν ταξινομηθεί σωστά. Συγκεκριμένα, η συνολική ακρίβεια μπορεί να υπολογιστεί διαιρώντας το άθροισμα των διαγώνιων στοιχείων (αριθμός των σωστά ταξινομημένων δειγμάτων) με το συνολικό άθροισμα των στοιχείων ενός CM (συνολικός αριθμός δειγμάτων στο σύνολο δεδομένων). Προφανώς, η ποσότητα 1-Accuracy δείχνει το συνολικό σφάλμα ταξινόμησης. Αναλυτικότερα ο τύπος της ακρίβειας είναι:

$$Acc = \frac{\text{number of correct predictions}}{\text{total number of predictions}}$$

Εκτός από τη συνολική ακρίβεια, υπάρχουν και κάποια άλλα μεγέθη που περιγράφουν πόσο καλά ο αλγόριθμος ταξινόμησης ανταποκρίνεται σε κάθε κλάση. Ένα απ' αυτά είναι η ανάκληση (recall), η οποία ορίζεται σαν την αναλογία των δεδομένων που προβλέφθηκαν σωστά ότι ανήκουν σε μια κλάση προς τα δεδομένα που ανήκουν όντως σ' αυτήν την κλάση. Συγκεκριμένα, η ανάκληση δίνεται από:

$$Re = \frac{\text{True Positive}}{\text{Predicted Results}}$$

Επιπλέον, ένα μέτρο απόδοσης είναι το precision, που ορίζεται ως η αναλογία των δεδομένων που προβλέφθηκαν σωστά ότι ανήκουν σε μια κλάση προς τα δεδομένα που ταξινομήθηκαν σε αυτήν την κλάση. Ο τύπος για την ακρίβεια precision είναι ο ακόλουθος:

$$Pr = \frac{\text{True Positive}}{\text{Actual Results}}$$

Πρέπει να σημειωθεί ότι, εάν όλες οι κλάσεις περιέχουν τον ίδιο αριθμό δειγμάτων, δηλαδή εάν όλες οι κλάσεις είναι εκ των προτέρων ισοδύναμες, τότε τα παραπάνω τρία μέτρα απόδοσης μπορούν να υπολογιστούν από οποιαδήποτε μορφή του confusion matrix, ανεξάρτητα από την κανονικοποίηση. Τέλος, ένα ευρέως χρησιμοποιούμενο μέτρο απόδοσης, που συνδυάζει τις τιμές ακριβείας (precision) και ανάκλησης (recall) είναι το μέτρο F_1 , το οποίο υπολογίζεται ως εξής:

$$F_1 = \frac{2RePr}{Re + Pr}$$

4. Δοκιμές και Αποτελέσματα

4.1. Περιγραφή και επεξεργασία των ηχητικών δεδομένων

Για την παρούσα εργασία χρησιμοποιήθηκε η βιβλιοθήκη pyAudioAnalysis, η οποία είναι μια ανοιχτή Python βιβλιοθήκη που παρέχει ένα ευρύ φάσμα συναρτήσεων σχετιζόμενων με τον ήχο, εστιάζοντας σε θέματα εξαγωγής χαρακτηριστικών (feature extraction), ταξινόμησης (classification), τμηματοποίησης (segmentation) και οπτικοποίησης (visualization). Για τη σωστή εφαρμογή και εκτέλεση των συναρτήσεων της βιβλιοθήκης απαραίτητη ήταν η εγκατάσταση, εκτός της pyAudioAnalysis, των ακόλουθων βιβλιοθηκών: Numpy, Matplotlib, Scipy, SKlearn, hmmlearn, Simplejson, eyeD3 και pydub. Το εκτελέσιμο γράφτηκε στο Jupyter 5.0.0 της Anaconda σε Python version 3.

Όσον αφορά τα ηχητικά δεδομένα, πρέπει να αναφερθεί, ότι δεν χρησιμοποιήθηκε ένα έτοιμο σύνολο δεδομένων, αντ'αυτού δημιουργήθηκε ένα σύνολο μελέτης εξ ολοκλήρου από την αρχή με την χρήση διαθέσιμων δεδομένων από το διαδίκτυο. Οι δύο ιστότοποι από τους οποίους έγινε η λήψη των διαθέσιμων αρχείων είναι το Zapsplat και το Freesound. Αρχικά, συλλέχθηκαν διάφοροι τύποι ηχητικών αρχείων, όπως mp3, wav και aiff. Στη συνέχεια, μετασχηματίστηκαν όλα τα διαθέσιμα αρχεία στην ίδια μορφή wav, δηλαδή όλοι οι τελικοί ήχοι που μελετήθηκαν ήταν σε κυματομορφή (Waveform Audio File Format).

Επιπλέον, τα δεδομένα, όπως είναι αναμενόμενο, είχαν στιγμές σιωπών και φυσικά δεν είχαν όλα την ίδια χρονική διάρκεια. Ήταν, επομένως, απαραίτητο να απομακρυνθούν οι σιωπές από τα διαθέσιμα αρχεία. Σαν threshold σιωπών θεωρήσαμε τα 20Hz. Η συχνότητα αυτή επιλέχθηκε σαν το κατώτατο όριο, αφού το ανθρώπινο αυτί είναι ευαίσθητο σε ηχητικό φάσμα μεταξύ 20 και 20,000 Hz. Βέβαια σπάνια μπορεί να νιώσουμε ήχους χαμηλότερους από 20 Hz, όπως για παράδειγμα έναν σεισμό. Πρέπει εδώ να σημειωθεί, ότι όσον αφορά τις σιωπές ορίστηκε και ένα χρονικό διάστημα για τη διάρκεια τους. Για να θεωρηθεί μια σιωπή ως αληθινή, το χρονικό κατώφλι ορίστηκε σαν 25ms, το οποίο είναι το μισό μήκος κύματος ενός ήχου συχνότητας 20Hz. Εάν σκοπεύαμε να αφαιρούσαμε σιωπές μικρότερες απ' αυτό, τότε υπήρχε ο κίνδυνος να χάσουμε τον μη σιωπηλό ήχο σ' ένα χαμηλό σημείο του κύματος του. Επίσης, αυτό το ελάχιστο δεν ισχύει για σιωπές στην αρχή και στο τέλος του αρχείου. Αυτές αναγνωρίστηκαν ως σιωπή ανεξάρτητα από το μήκος. Πρακτικά σημαίνει, ότι κάθε φορά που η συχνότητα του κύματος ήταν χαμηλότερη των 20 Hz και διαρκούσε για περισσότερο από 25 msec, τότε το διάστημα σιωπής απομακρυνόταν από το αρχείο.

Στη συνέχεια, τα αρχεία του συνόλου εκπαίδευσης έπρεπε να έχουν ακριβώς την ίδια χρονική διάρκεια. Τα διαθέσιμα ηχητικά δεδομένα κλαδεύτηκαν στα 5 sec, 10 sec και 25 sec και πειράματα πραγματοποιήθηκαν και για τις τρεις χρονικές διάρκειες. Αυτό, όπως είναι λογικό, έγινε μετά την απομάκρυνση των σιωπών, μιας και τα εναπομείναντα αρχεία είχαν μικρότερη διάρκεια από τα αρχικά. Κάποιοι από τους ήχους που λήφθηκαν από τους προαναφερθέντες ιστότοπους είχαν διάρκεια λεπτών, ενώ κάποιων άλλων κλάσεων τα αρχεία είχαν πολύ μικρή χρονική διάρκεια. Σ' αυτήν την περίπτωση αντί του κλαδεύματος του αρχικού ήχου, επιλέχθηκε η ένωση, ώστε τελικά να προκύψει ηχητικό δείγμα ίσου μεγέθους με τα υπόλοιπα.

Το δεύτερο μέρος της διαδικασίας ήταν η κατασκευή του συνόλου δεδομένων εκπαίδευσης, ένα σύνολο από διανύσματα χαρακτηριστικών, των οποίων οι ετικέτες κάθε κλάσης είναι ήδη

γνωστές. Το σύνολο δεδομένων εκπαίδευσης είναι το προϊόν μιας χρονοβόρας διαδικασίας, η οποία περιλαμβάνει τη συγκέντρωση των ηχητικών δειγμάτων, τις δοκιμές ακρόασης, τη χειροκίνητη επισήμανση (tagging) της κάθε κλάσης και την οργάνωση των ηχητικών δεδομένων σε συγκεκριμένες δομές, προκειμένου να διευκολυνθεί η ανταλλαγή δεδομένων. Γενικά, η ποιότητα του συνόλου εκπαίδευσης καθορίζει την ποιότητα του κατηγοριοποιητή. Ουσιαστικά, τα δεδομένα εκπαίδευσης υψηλής ποιότητας είναι απαραίτητα στα διάφορα προβλήματα κατηγοριοποίησης. [32] Κατόπιν αυτής της διαδικασίας, δημιουργήθηκαν συνολικά οχτώ κλάσεις με 100 αρχεία η κάθε μια, για να εξυπηρετήσουν τις πειραματικές δοκιμές, που θα αναφερθούν παρακάτω.

Η πρώτη κλάση είναι η παραλία, η οποία περιλαμβάνει δεδομένα, όπως κύματα που καταλήγουν σε άμμο ή σε πέτρες, παραλίες σε διαφορετικές στιγμές του 24ώρου και με διαφορετικές συνθήκες (με δυνατό άνεμο, με άπνοια, πρωινές ή νυχτερινές ώρες), ήχους από τζιτζίκια, γλαροπούλια, πυροτεχνήματα, παιδιά που παίζουν. Επίσης, ακούγονται δραστηριότητες, όπως βήματα ανθρώπων σε αμμώδη ή πετρώδη παραλία, surfers στη σανίδα, κάποιοι που κάνουν beach party ή μπάνιο σε υπαίθρια ντουζιέρα. Ακόμη, περιέχεται ο ήχος του νερού και η θάλασσα από διαφορετικές αποστάσεις (ενδεικτικά, 40m , 100m).

Η δεύτερη κλάση είναι οι δημόσιες τουαλέτες, στην οποία περιέχονται ηχητικά δεδομένα όπως το καζανάκι, η ανοιχτή βρύση που τρέχει ή που στάζει νερό, μηχάνημα για το στέγνωμα των βρεγμένων χεριών, το καπάκι από τον κάδο απορριμμάτων που ανοίγει και κλείνει. Ακόμη, συμπεριλαμβάνονται ήχοι, όπως το χαρτί τουαλέτας που ξετυλίγεται, βούρτσα τουαλέτας, χαλασμένο καζανάκι, διάφορα είδη πόρτας πχ. ξύλινη, μεταλλική η οποία ανοίγει και κλείνει, χαρτί που κόβεται από το ρολό, καθάρισμα δοντιών, το πόμολο και το κλείδωμα μιας πόρτας. Επίσης, σ' αυτήν την κλάση, τοποθετήθηκαν ηχητικά δεδομένα από ανεμιστήρα καθαρισμού χώρου, άνοιγμα του παράθυρου μπάνιου και του νερού στους σωλήνες.

Η τρίτη κλάση είναι ο ζωολογικός κήπος. Σ' αυτήν περιέχονται ήχοι, όπως είδη πτηνών (παπαγάλοι, πάπιες, κότες, ωδικά πουλιά). Επίσης συμπεριλαμβάνεται ένα πλήθος θηλαστικών, όπως αρκούδες, ελέφαντες, λιοντάρια, μαϊμούδες, ρινόκεροι, ιπποπόταμοι, λύκοι, άλογα, αγριογούρουνα, τίγρεις, ελάφια, αγριόγατες. Μεταξύ των άγριων ζώων υπάρχουν και ηχητικά δεδομένα από πιγκουίνους, νυχτερίδες, και φίδια. Τέλος, μέσα στην κλάση υπάρχουν δείγματα από ομιλίες ανθρώπων μέσα σε ζωολογικό κήπο.

Η τέταρτη κλάση είναι η εκκλησία, η οποία περιλαμβάνει ήχους από είδη καμπανών και χτύπους κατά τη διάρκεια διαφορετικών θρησκευτικών περιστάσεων όπως είναι για παράδειγμα ένας γάμος ή μια κηδεία. Επίσης ακούγονται μελωδικά εκκλησιαστικά όργανα, τα τραγούδια της εκκλησιαστικής χορωδίας και χτύποι καμπανών ηχογραφημένοι από το εσωτερικό και από το εξωτερικό της εκκλησίας. Επιπλέον, υπάρχουν ήχοι του ρολογιού της εκκλησίας και εκκλησιαστικοί ψαλμοί.

Η πέμπτη κλάση είναι τα εστιατόρια-bars. Αρχικά, προορίζονταν για δύο εντελώς ανεξάρτητες κλάσεις, αλλά οι ήχοι των δύο αυτών κλάσεων ήταν αρκετά παρόμοιοι και έτσι αποφασίστηκε

η σύζευξή τους. Στην εν λόγω κλάση περιέχονται ήχοι από ένα Βαυαρικό μαγαζί, διάφορα είδη μουσικής, όπως κουβανέζικη, πιάνο, jazz, lounge κ.α. Επίσης ακούγονται άνθρωποι που χορεύουν, τραγουδούν και χειροκροτούν μετά το τέλος της μουσικής. Ακόμη, περιέχονται ήχοι από σπάσιμο και μάζεμα πιάτων με σκούπα, μηχανής που χτυπάει τον καφέ, τράβηγμα τραπεζίων. Τέλος, συμπεριλαμβάνονται δεδομένα από ήχο μπιλιάρδου σε μια rub και ανθρώπους να τρώνε.

Η έκτη κλάση είναι το σπίτι, όπου τα ηχητικά δεδομένα είναι σχετικά με την οργάνωση των βάζων, την επιδιόρθωση με σφυρί, το χτένισμα των μαλλιών, το βούρτσισμα των δοντιών με ηλεκτρική και απλή οδοντόβουρτσα. Ακόμη, συμπεριλαμβάνονται ο ήχος από θερμαντήρα νερού, ξύλα που καίγονται στο τζάκι, καρέκλες με ροδάκια που σύρονται, βήματα, κλιματισμός, το χτύπημα του τηλεφώνου. Επιπροσθέτως, περιέχονται αρχεία με τράβηγμα επίπλων, κρεβάτια που τρίζουν, ανέβασμα ανθρώπου σε σκάλα αλουμινίου καθώς και με δραστηριότητες στην κουζίνα (μαγείρεμα, χρονοδιακόπτης φαγητού, πιάτα και μαχαίρια), αλλά και στο νοικοκυριό γενικότερα (ξεσκόνισμα, καθάρισμα τζαμιών, ηλεκτρικές συσκευές). Τέλος, στη συγκεκριμένη κλάση υπάρχουν και δεδομένα από συναγερμό πυρκαγιάς, ξυπνητήρι και βήματα σε ξύλινο δάπεδο.

Η έβδομη κλάση είναι ο παιδότοπος/παιδική χαρά, η οποία καταλέγει ήχους από παιδική χαρά κατά τη διάρκεια διαφορετικών εποχών και τοποθεσιών (π.χ. χειμώνας-καλοκαίρι, χωριό-πόλη). Ακούγονται παιδιά να παίζουν ποδόσφαιρο, να κάνουν τραμπολίνο, να γελάνε, να ουρλιάζουν πάνω σε ένα πειρατικό πλοίο και γενικά να μαλώνουν και να παίζουν. Επίσης, περιέχονται ήχοι από γονιούς που παίζουν με τα παιδιά τους και από παιδική χαρά σε ζωολογικό κήπο, αλλά και σε προαύλιο σχολικού κτηρίου.

Η όγδοη και τελευταία κλάση είναι το αεροδρόμιο. Στη συγκεκριμένη κλάση, περιέχονται ήχοι από αεροδρόμια διαφορετικών χωρών, ανακοινώσεις από υπάλληλους αεροδρομίου, αλλά και από αεροσυνοδούς (π.χ. για απαγόρευση καπνίσματος ή για κάλεσμα αργοπορημένων επιβατών). Επίσης, υπάρχουν δεδομένα από αεροδρόμια με πολύ κόσμο και φασαρία, καθώς και με λιγότερους ανθρώπους, άρα και πιο ήσυχα. Ακούγονται ακόμα ομιλίες απόμνων από διάφορα σημεία του αεροδρομίου, όπως από την αίθουσα αναμονής, ήχοι από βαλίτσες που σέρνονται και φυσικά αεροπλάνα που προσγειώνονται και απογειώνονται. Στις περισσότερες εφαρμογές, το ηχητικό σήμα αναλύεται με τη βραχυπρόθεσμη τεχνική (short-term ή short-time), σύμφωνα με την οποία ο ήχος χωρίζεται σε πιθανά αλληλεπικαλυπτόμενα βραχυπρόθεσμα παράθυρα (frames). Ο κύριος λόγος που υιοθετείται μια τέτοια τεχνική είναι ότι συνήθως τα ηχητικά σήματα είναι μη στάσιμα από τη φύση τους, δηλαδή οι ιδιότητές τους ποικίλλουν με την πάροδο του χρόνου. Δεν έχει, λοιπόν, νόημα να υπολογιστεί, για παράδειγμα, η μέση ένταση των δειγμάτων μιας ολόκληρης καταγραφής, επειδή η προκύπτουσα τιμή θα επηρεαζόταν από τα πιο έντονα δείγματα. Αντίθετα, θα ήταν πιο χρήσιμο η εγγραφή να σπάσει σε σύντομα τμήματα και να υπολογιστεί μία τιμή έντασης ανά τμήμα. Αυτή είναι και η κύρια ιδέα πίσω από τη βραχυπρόθεσμη επεξεργασία. Τα παραπάνω υπογραμμίζουν το γεγονός ότι οι σημαντικές παράμετροι στην τεχνική της βραχυπρόθεσμης επεξεργασίας είναι το μήκος του βραχυπρόθεσμου παραθύρου (short term window length) και

το μέγεθος του βραχυπρόθεσμου βήματος (short term step). Συνήθως, το μήκος παραθύρου/πλαϊσίου κυμαίνεται από 10 ms έως 50 ms. Από την άλλη πλευρά, το βήμα παραθύρου ελέγχει το βαθμό επικάλυψης μεταξύ διαδοχικών πλαϊσίων. Εάν, για παράδειγμα, η επιθυμητή επικάλυψη είναι 75% και το μήκος παραθύρου είναι 40ms, τότε το βήμα παραθύρου πρέπει να είναι 10ms. [32] Στα προβλήματα κατηγοριοποίησης συνηθίζεται το βραχυπρόθεσμο μήκος (short term window) και βήμα παραθύρου (short term step) να ταυτίζονται. Συγκεκριμένα, στην παρούσα εργασία τα βραχυπρόθεσμα μεγέθη ισούνται με 50 ms.

Εκτός από την βραχυπρόθεσμη διαδικασία σημαντικό είναι και το στάδιο της μεσοπρόθεσμης (midterm) ανάλυσης. Έπειτα από αυτή τη διαδικασία παράγεται μια ακολουθία από μεσοπρόθεσμα διανύσματα χαρακτηριστικών για κάθε ηχητικό αρχείο, ανάλογα με τη διάρκεια του αρχείου και το μήκος του μεσοπρόθεσμου παραθύρου (mid-term window). Στην παρούσα ανάλυση το μεσοπρόθεσμο μήκος (midterm window) και το μεσοπρόθεσμο βήμα (midterm step) ορίστηκε το 1sec. Ωστόσο, το επιθυμητό είναι σε κάθε ηχητικό σήμα να αντιστοιχεί ένα μόνο διάνυσμα χαρακτηριστικών, όπως ακολουθήθηκε και στις υπόλοιπες εργασίες ανάλυσης ήχου. Εδώ, λοιπόν, λαμβάνει χώρα ένα βήμα μακροπρόθεσμου ζυγίσματος (long-term average) που εκτελείται στα εξαγόμενα μεσοπρόθεσμα διανύσματα χαρακτηριστικών. Μ' αυτόν τον τρόπο, τελικά, παράγεται ένα διάνυσμα χαρακτηριστικών ανά ηχητικό αρχείο.

Η εξαγωγή χαρακτηριστικών, όπως περιεγράφηκε σε προηγούμενο κεφάλαιο, μπορεί να θεωρηθεί σαν μια πυραμίδα στρωμάτων. Στο χαμηλότερο επίπεδο, οι τεχνικές εξόρυξης βραχυπρόθεσμων χαρακτηριστικών δημιουργούν αλληλουχίες διανυσμάτων χαρακτηριστικών από ηχητικά δεδομένα. Στο επόμενο υψηλότερο επίπεδο υπολογίζεται ένας αριθμός στατιστικών για την κάθε αλληλουχία βραχυπρόθεσμων χαρακτηριστικών. Στη συνέχεια, τα εξαγόμενα στατιστικά στοιχεία μπορούν να ομαδοποιηθούν για να δημιουργηθεί ένα μόνο διάνυσμα χαρακτηριστικών. Αν D οι ακολουθίες χαρακτηριστικών που έχουν εξαχθεί και S τα στατιστικά στοιχεία που έχουν υπολογιστεί για κάθε ακολουθία, τότε το προκύπτον διάνυσμα χαρακτηριστικών, έστω x , είναι $L = D \times S$ -διαστάσεων. Στόχος είναι να υπολογιστεί η ετικέτα κλάσης y ενός τμήματος ήχου που αντιπροσωπεύεται από το διάνυσμα χαρακτηριστικών x . [32] Για την ανάλυση που πραγματεύεται η παρούσα εργασία δημιουργήθηκε για κάθε ηχητικό σήμα διάνυσμα διάστασης 136 χαρακτηριστικών. Τα χαρακτηριστικά που μελετήθηκαν είναι όσα είχαν αναφερθεί στην ενότητα των χαρακτηριστικών με την επιπλέον προσθήκη των στατιστικών τους στοιχείων. Τελικά, η κάθε κλάση που προέκυψε διαθέσιμη για πειραματισμό περιείχε 100 διανύσματα (ηχητικά σήματα), με 136 χαρακτηριστικά το καθένα (διάσταση διανύσματος).

Επίσης, απαραίτητο είναι τα διανύσματα χαρακτηριστικών του συνόλου εκπαίδευσης να κανονικοποιηθούν, πριν τελικά χρησιμοποιηθούν από τον ταξινομητή. Μια λύση σ' αυτό το ζήτημα είναι η χρήση της γραμμικής μεθόδου, όπου κανονικοποιείται το j -οστό χαρακτηριστικό $j = 1, \dots, L$ ως εξής:

$$\hat{v}_i(j) = \frac{v_i(j) - \mu(j)}{\sigma(j)}, \quad i = 1, \dots, M, \quad j = 1, \dots, L$$

όπου M είναι ο αριθμός των δειγμάτων εκπαίδευσης, L είναι η διάσταση του χώρου των χαρακτηριστικών, $\mu(j)$ είναι η μέση τιμή του j -οστού χαρακτηριστικού και $\sigma(j)$ η αντίστοιχη τυπική απόκλιση. [32] Οι παράμετροι κανονικοποίησης (μέση τιμή και τυπική απόκλιση) χρησιμοποιούνται για την ομαλοποίηση των δεδομένων εκπαίδευσης και κάθε συνόλου δοκιμής, πριν αυτό ταξινομηθεί.

Μετά τη δημιουργία του συνόλου δεδομένων κατάρτισης, εκπαιδεύτηκαν πέντε αλγόριθμοι ταξινόμησης, οι kNN, Support Vector Machines, Random forests, Extra trees και Gradient boosting. Για κάθε έναν από τους παραπάνω αλγορίθμους, προσπαθήσαμε να βρούμε την βέλτιστη παράμετρο, που μεγιστοποιεί την απόδοση του ταξινομητή, καθώς και ποιος από τους πέντε αλγορίθμους λειτουργεί καλύτερα στο διαθέσιμο σύνολο ηχητικών δεδομένων. Πιο συγκεκριμένα, οι παράμετροι που εξετάστηκαν για τους προαναφερθέντες ταξινομητές είναι η παράμετρος C του περιθωρίου μεταξύ των τιμών [0.001, 0.01, 0.5, 1.0, 5.0, 10.0, 20.0] για τον αλγόριθμο SVM, ο αριθμός των κοντινότερων γειτόνων K ανάμεσα στους αριθμούς [1, 3, 5, 7, 9, 11, 13, 15] για τον ταξινομητή KNN, ο αριθμός των δέντρων μεταξύ των τιμών [10, 25, 50, 100, 200, 500] για τους αλγορίθμους random forest και extra trees και τέλος ο αριθμός των σταδίων ενίσχυσης από τις τιμές [10, 25, 50, 100, 200, 500] για τον ταξινομητή gradient boosting. Επίσης, τα αρχικά ηχητικά δεδομένα μετατράπηκαν σε αρχεία μορφής arff, για να είναι διαχειρίσιμα από τους αλγορίθμους και να έχουν το ίδιο όνομα με το μοντέλο. Στα αρχεία αυτά αποθηκεύεται ολόκληρο το σύνολο των διανυσμάτων χαρακτηριστικών (feature vectors) και οι αντίστοιχες ετικέτες των κλάσεων.

Αφού ο ταξινομητής εκπαιδευτεί κατάλληλα, μπορεί να χρησιμοποιηθεί για την ταξινόμηση ηχητικών αρχείων, που ανήκουν στο σύνολο δοκιμής. Το σύνολο δοκιμής περιλαμβάνει διανύσματα χαρακτηριστικών, των οποίων η κατηγορία πρέπει να προσδιοριστεί από τον ταξινομητή. Αφού γίνει η πρόβλεψη, οι αποφάσεις ταξινόμησης συγκρίνονται με τις πραγματικές κλάσεις των δεδομένων. Είναι σημαντικό να σημειωθεί ότι τα σύνολα εκπαίδευσης και δοκιμής δεν πρέπει να επικαλύπτονται, δηλαδή να περιέχουν κοινά δεδομένα, γιατί σ' αυτήν την περίπτωση η αξιολόγηση του ταξινομητή θα είναι άκυρη.

Στην παρούσα εργασία πραγματοποιήθηκαν 100 πειράματα επικύρωσης για την κάθε παράμετρο του κάθε ταξινομητή. Πρόκειται για μια repeated random sub-sampling validation (hold-out) μέθοδο επικύρωσης σύμφωνα με την οποία σε κάθε επανάληψη το σύνολο δεδομένων χωρίζεται τυχαία σε δύο υποομάδες (τυχαία δειγματοληψία) και γίνεται η επανάληψη της μεθόδου $k=100$ φορές. Η απόδοση της ταξινόμησης προκύπτει με τον μέσο όρο των hold-out επαναλήψεων. Η μέθοδος αποτελεί μια βελτίωση της κλασσικής μεθόδου hold-out, ώστε να αποφευχθεί το overfitting, λόγω της τυχαίας φύσης της διαδικασίας χωρισμού των δύο συνόλων. Φυσικά υπάρχει ακόμα ο κίνδυνος κάποια δεδομένα να μην συμπεριληφθούν ποτέ στο σύνολο εκπαίδευσης αλλά ούτε και στο σύνολο δοκιμής. Εν κατακλείδι, στην επιλογή της τελικής παραμέτρου για τον κάθε αλγόριθμο, αποφασίσαμε

μεταξύ των μέτρων απόδοσης να επιλέγεται η παράμετρος εκείνη που επιτυγχάνει την καλύτερη ακρίβεια (accuracy).

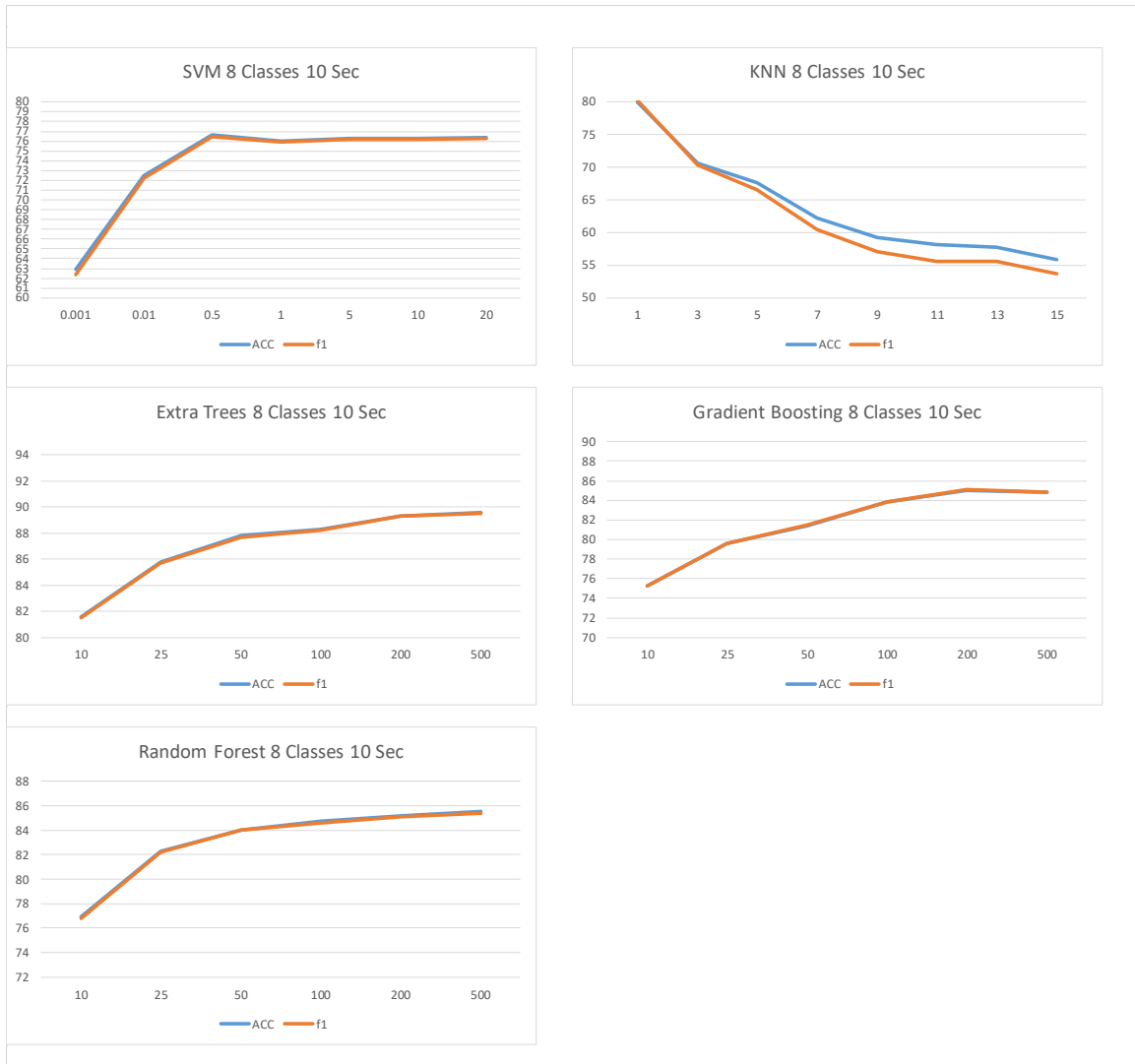
4.2.Πείραμα: 8 κλάσεις – παράθυρο 10 sec

Η πρώτη δοκιμή που εκτελέστηκε συμπεριλάμβανε και τις 8 κλάσεις. Τα 100 αρχεία της κάθε κλάσης είχαν χρονική διάρκεια 10 δευτερολέπτων το καθένα. Οι πίνακες *CM* του πρώτου πειράματος για τους 5 αλγόριθμους ήταν οι εξής:

| Confusion Matrix SVM 8 Classes 10 Sec | | | | | | | | | Confusion Matrix Extra Trees 8 Classes 10 Sec | | | | | | | | |
|---|---------|----------------|-------|--------|-------|------------|----------------|------|---|---------|----------------|-------|--------|-------|------------|----------------|-------|
| | Airport | Bar Restaurant | Beach | Church | House | Playground | Public Toilets | Zoo | | Airport | Bar Restaurant | Beach | Church | House | Playground | Public Toilets | Zoo |
| Airport | 9.45 | 0.76 | 1.55 | 0.16 | 0.05 | 0.03 | 0.19 | 0.31 | Airport | 10.74 | 0.81 | 0.15 | 0 | 0 | 0.01 | 0.38 | 0.41 |
| Bar Restaurant | 1.46 | 9.9 | 0.28 | 0.11 | 0 | 0.36 | 0.01 | 0.38 | Bar Restaurant | 0.3 | 11.94 | 0.06 | 0 | 0.01 | 0.18 | 0 | 0.01 |
| Beach | 1.3 | 0.38 | 9.16 | 0 | 0.43 | 0.29 | 0.83 | 0.13 | Beach | 1.1 | 0.33 | 9.69 | 0 | 0.89 | 0.33 | 0.18 | 0 |
| Church | 0.09 | 0 | 0.09 | 12.19 | 0 | 0.01 | 0.13 | 0 | Church | 0 | 0 | 0.11 | 12.31 | 0 | 0.01 | 0.05 | 0.01 |
| House | 0.06 | 0.04 | 1.04 | 0 | 9.05 | 0.05 | 1.06 | 1.2 | House | 0 | 0 | 0.33 | 0 | 11.72 | 0 | 0.3 | 0.15 |
| Playground | 0.11 | 0.46 | 0.06 | 0.05 | 0.05 | 11.24 | 0 | 0.53 | Playground | 0.15 | 0.15 | 0 | 0 | 0.03 | 11.87 | 0 | 0.3 |
| Public Toilets | 0.88 | 0.2 | 1.25 | 0.16 | 1.19 | 0 | 8.17 | 0.65 | Public Toilets | 0 | 0 | 0.55 | 0 | 0.74 | 0 | 10.95 | 0.26 |
| Zoo | 1.11 | 0.65 | 0.43 | 0.01 | 1.14 | 1.09 | 0.61 | 7.46 | Zoo | 0.61 | 0.5 | 0.23 | 0.03 | 0.15 | 0.45 | 0.16 | 10.37 |
| Confusion Matrix KNN 8 Classes 10 Sec | | | | | | | | | Confusion Matrix Gradient Boosting 8 Classes 10 Sec | | | | | | | | |
| | Airport | Bar Restaurant | Beach | Church | House | Playground | Public Toilets | Zoo | | Airport | Bar Restaurant | Beach | Church | House | Playground | Public Toilets | Zoo |
| Airport | 9.74 | 0.09 | 1.8 | 0 | 0 | 0.56 | 0.13 | 0.19 | Airport | 10.31 | 0.75 | 0.59 | 0 | 0 | 0.05 | 0.2 | 0.6 |
| Bar Restaurant | 1.71 | 8.24 | 1.21 | 0 | 0 | 1.19 | 0 | 0.15 | Bar Restaurant | 0.71 | 10.82 | 0.34 | 0 | 0.01 | 0.2 | 0.1 | 0.31 |
| Beach | 0.79 | 0 | 10.91 | 0 | 0.53 | 0.13 | 0.15 | 0 | Beach | 0.94 | 0.08 | 9.74 | 0 | 0.65 | 0.36 | 0.49 | 0.25 |
| Church | 0.91 | 0 | 0.33 | 11.25 | 0.01 | 0 | 0 | 0 | Church | 0.01 | 0 | 0.39 | 11.89 | 0 | 0 | 0.06 | 0.15 |
| House | 0.43 | 0.01 | 0.86 | 0 | 9.76 | 0.24 | 0.79 | 0.41 | House | 0.09 | 0.01 | 0.61 | 0.03 | 10.79 | 0.04 | 0.43 | 0.51 |
| Playground | 0.24 | 0.3 | 0.03 | 0 | 0.04 | 11.75 | 0 | 0.15 | Playground | 0.28 | 0.09 | 0.01 | 0 | 0 | 11.44 | 0 | 0.69 |
| Public Toilets | 0.51 | 0.01 | 0.91 | 0 | 0.41 | 0 | 10.42 | 0.23 | Public Toilets | 0.31 | 0 | 0.56 | 0 | 0.89 | 0.01 | 10.01 | 0.71 |
| Zoo | 1.3 | 0.74 | 0.88 | 0.18 | 0.36 | 0.74 | 0.36 | 7.95 | Zoo | 0.41 | 0.2 | 0.16 | 0.19 | 0.36 | 0.78 | 0.36 | 10.04 |
| Confusion Matrix Random Forest 8 Classes 10 Sec | | | | | | | | | | | | | | | | | |
| | Airport | Bar Restaurant | Beach | Church | House | Playground | Public Toilets | Zoo | | | | | | | | | |
| Airport | 10.1 | 1.21 | 0.35 | 0 | 0 | 0 | 0.44 | 0.4 | | | | | | | | | |
| Bar Restaurant | 0.4 | 11.49 | 0.38 | 0 | 0 | 0.14 | 0 | 0.1 | | | | | | | | | |
| Beach | 1.11 | 0.38 | 9.42 | 0 | 1.03 | 0.21 | 0.33 | 0.03 | | | | | | | | | |
| Church | 0 | 0 | 0.15 | 12.25 | 0 | 0 | 0.1 | 0 | | | | | | | | | |
| House | 0 | 0 | 0.39 | 0 | 10.84 | 0.03 | 0.71 | 0.54 | | | | | | | | | |
| Playground | 0.09 | 0.33 | 0 | 0 | 0 | 11.61 | 0 | 0.48 | | | | | | | | | |
| Public Toilets | 0.06 | 0 | 0.78 | 0 | 1.14 | 0 | 10.3 | 0.23 | | | | | | | | | |
| Zoo | 0.44 | 0.61 | 0.43 | 0 | 0.53 | 0.84 | 0.2 | 9.46 | | | | | | | | | |

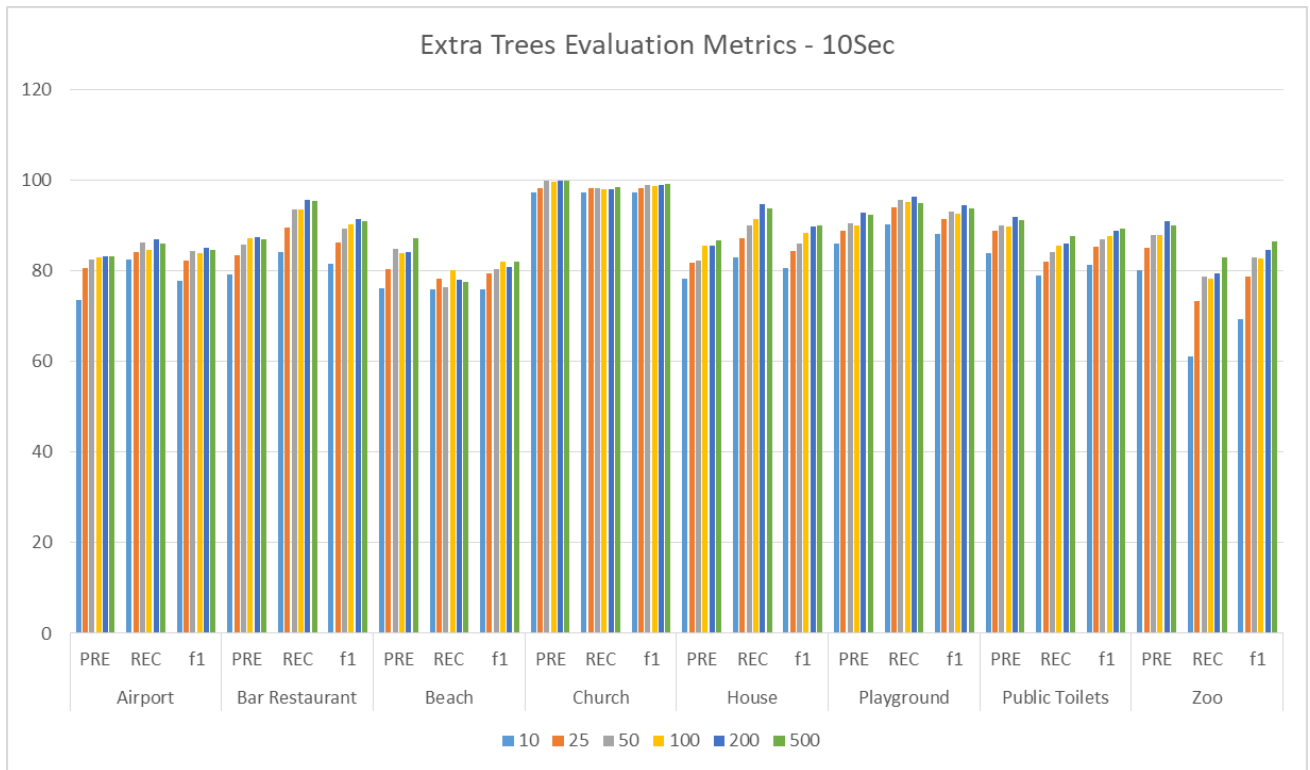
Πίνακας 1 Πίνακες *CM* για τη σύγκριση των αλγορίθμων του πρώτου πειράματος

Επίσης παρακάτω φαίνεται μια διαγραμματική απεικόνιση των μεγεθών F_1 και Acc σε σχέση με τις διάφορες παραμέτρους που εξετάστηκαν για τον κάθε αλγόριθμο. Στον άξονα των y βλέπουμε τις διαφορετικές παραμέτρους ενώ στον άξονα x τα evaluation metrics Accuracy και F_1 .



Εικόνα 14 Διαγραμματική απεικόνιση των πέντε ταξινομητών σε σχέση με τις διάφορες παραμέτρους που εξετάστηκαν

Είναι φανερό ότι ο ταξινομητής που πετυχαίνει την καλύτερη ακρίβεια είναι ο Extra Trees, για πλήθος δέντρων ίσο με 500. Συγκεκριμένα, η ακρίβεια που προκύπτει είναι 89.59% και το μέγεθος F1 φτάνει στο 89.5%. Επιπλέον, στο ακόλουθο γράφημα αναπαρίστανται τα μεγέθη F_1 , $recall$ και $precision$ όλων των κλάσεων για τον αλγόριθμο Extra Trees σε σχέση με τις διάφορες παραμέτρους που εξετάστηκαν.



Εικόνα 15 Μέτρα απόδοσης για την κάθε κλάση έπειτα από την εφαρμογή του ταξινομητή Extra Trees σε αρχεία διάρκειας 10 sec

Παρατηρούμε ότι η κλάση «Εκκλησία» έχει επιτύχει τα υψηλότερα ποσοστά για όλα τα μεγέθη αξιολόγησης και για όλες τις παραμέτρους. Έπειτα, θελήσαμε να εξετάσουμε πως συμπεριφέρεται ο αλγόριθμος extra trees σε ένα εντελώς άγνωστο δείγμα 104 ηχητικών δεδομένων διαφόρων χρονικών διαρκειών το καθένα. Τα αρχεία μετατράπηκαν σε μορφή wav και ο επιλεγμένος ταξινομητής εφαρμόστηκε σ' αυτά, προκειμένου να ληφθεί η απόφαση για το ποια ετικέτα χαρακτηρίζει το καθένα του συνόλου.

Τα αποτελέσματα των σωστά ταξινομημένων ήχων είναι τα ακόλουθα:

| Extra Trees - 10 sec | Accuracy Per Class |
|----------------------|--------------------|
| Airport | 76.9% |
| Bar-Restaurant | 61.5% |
| Beach | 84.6% |
| Church | 92.3% |
| House | 76.9% |
| Playground | 84.6% |
| Public Toilet | 69.2% |
| Zoo | 76.9% |

Πίνακας 2 Ακρίβεια ανά κλάση σε τυχαίο σύνολο δεδομένων

Βλέπουμε ότι τα χαμηλότερα ποσοστά ακρίβειας παρατηρούνται για τις κλάσεις «Εστιατόριο» και «Δημόσια Τουαλέτα».

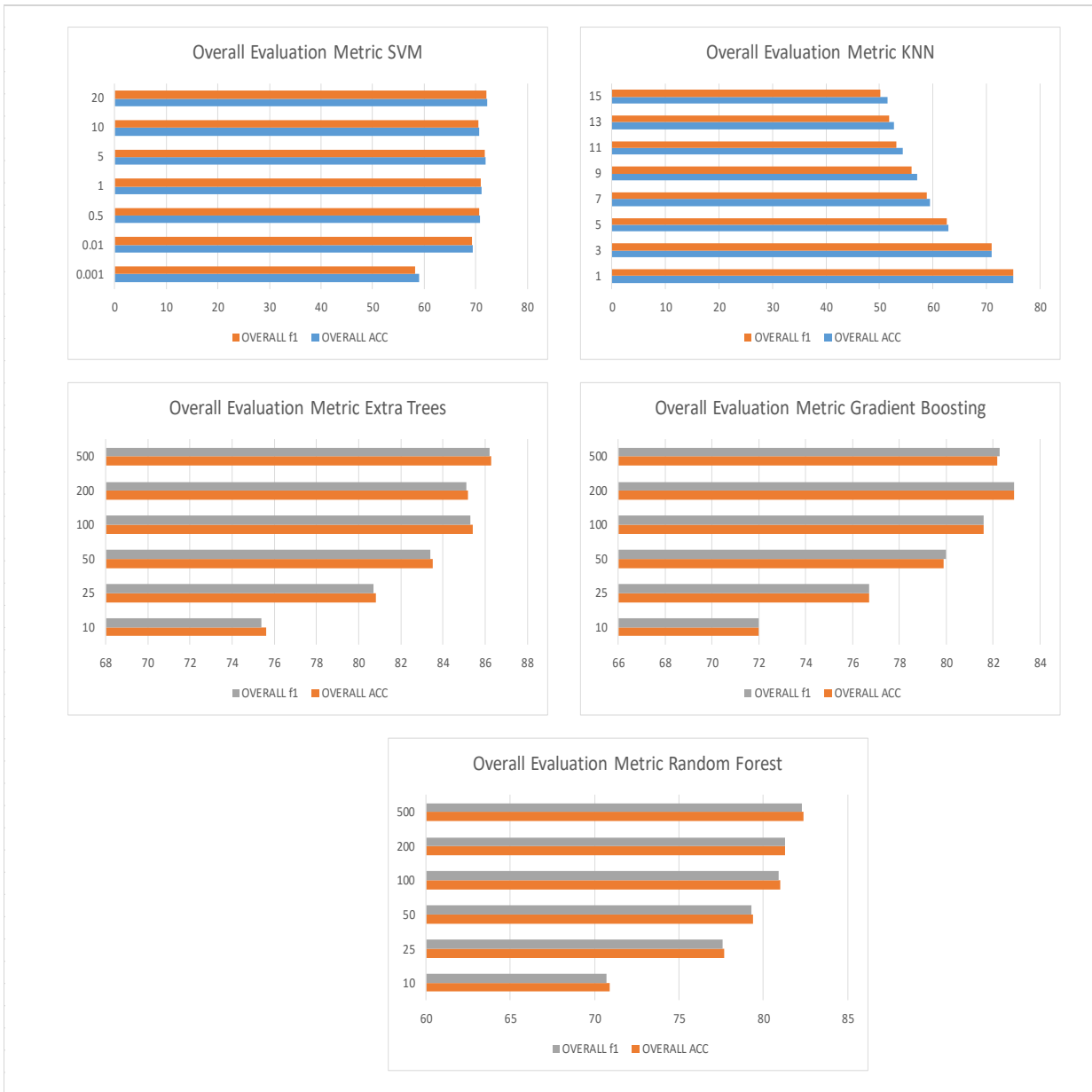
4.3.Πείραμα: 8 κλάσεις – παράθυρο 5 sec

Το δεύτερο πείραμα ακολουθεί τα ίδια βήματα με το προηγούμενο, με τη διαφορά ότι τα αρχεία που χρησιμοποιήθηκαν για την εκπαίδευση των πέντε αλγορίθμων είχαν χρονική διάρκεια 5 sec, μειωμένη, δηλαδή κατά το ήμισυ σε σχέση με τα ηχητικά σήματα της προηγούμενης δοκιμασίας. Οι πίνακες *CM* που προέκυψαν για το σύνολο των ταξινομητών ήταν οι εξής:

| Confusion Matrix SVM 8 Classes 5 Sec | | | | | | | | | Confusion Matrix Gradient boosting 8 Classes 5 Sec | | | | | | | | |
|--|---------|----------------|-------|--------|-------|------------|----------------|------|--|---------|----------------|-------|--------|-------|------------|----------------|------|
| | Airport | Bar Restaurant | Beach | Church | House | Playground | Public Toilets | Zoo | | Airport | Bar Restaurant | Beach | Church | House | Playground | Public Toilets | Zoo |
| Airport | 9.34 | 0.7 | 1.24 | 0.06 | 0.31 | 0.14 | 0.14 | 0.58 | Airport | 10.09 | 0.74 | 0.86 | 0 | 0.01 | 0.16 | 0.13 | 0.51 |
| Bar Restaurant | 1.43 | 9.62 | 0.28 | 0 | 0.01 | 0.64 | 0.06 | 0.46 | Bar Restaurant | 0.96 | 10.01 | 0.25 | 0.15 | 0 | 0.41 | 0.13 | 0.59 |
| Beach | 1.28 | 0.48 | 9.02 | 0 | 0.73 | 0.26 | 0.71 | 0.03 | Beach | 0.84 | 0.03 | 9.94 | 0 | 0.44 | 0.51 | 0.61 | 0.14 |
| Church | 0.16 | 0 | 0.24 | 12.06 | 0 | 0.03 | 0.01 | 0 | Church | 0.21 | 0.06 | 0.08 | 11.91 | 0.01 | 0 | 0.15 | 0.08 |
| House | 0.08 | 0.19 | 1.05 | 0 | 8.87 | 0.04 | 1.25 | 1.03 | House | 0 | 0 | 0.48 | 0 | 10.65 | 0.01 | 0.73 | 0.64 |
| Playground | 0.44 | 0.71 | 0.21 | 0 | 0.15 | 9.7 | 0 | 1.29 | Playground | 0.19 | 0.35 | 0.01 | 0 | 0.1 | 11.27 | 0.25 | 0.33 |
| Public Toilets | 1.33 | 0.29 | 1.84 | 0.13 | 1.23 | 0.06 | 6.97 | 0.66 | Public Toilets | 0.34 | 0.11 | 0.79 | 0.01 | 0.63 | 0.11 | 9.49 | 1.03 |
| Zoo | 1.21 | 0.9 | 0.59 | 0.1 | 1.15 | 1.33 | 0.6 | 6.62 | Zoo | 0.1 | 0.53 | 0.11 | 0.13 | 0.89 | 0.68 | 0.51 | 9.56 |
| Confusion Matrix KNN 8 Classes 5 Sec | | | | | | | | | Confusion Matrix Random Forest 8 Classes 5 Sec | | | | | | | | |
| | Airport | Bar Restaurant | Beach | Church | House | Playground | Public Toilets | Zoo | | Airport | Bar Restaurant | Beach | Church | House | Playground | Public Toilets | Zoo |
| Airport | 8.82 | 0.84 | 1.59 | 0 | 0 | 0.71 | 0.53 | 0.01 | Airport | 9.49 | 1.19 | 0.74 | 0 | 0 | 0.2 | 0.13 | 0.76 |
| Bar Restaurant | 1.69 | 8.24 | 0.53 | 0.13 | 0 | 1.8 | 0 | 0.13 | Bar Restaurant | 0.56 | 10.71 | 0.43 | 0 | 0 | 0.49 | 0.09 | 0.23 |
| Beach | 0.65 | 0.01 | 11.09 | 0 | 0.03 | 0.04 | 0.45 | 0.24 | Beach | 1.03 | 0.41 | 9.54 | 0 | 0.81 | 0.25 | 0.4 | 0.06 |
| Church | 0.7 | 0.01 | 0.5 | 11.01 | 0 | 0.13 | 0.01 | 0.14 | Church | 0.03 | 0 | 0.08 | 12.21 | 0 | 0 | 0.04 | 0.15 |
| House | 0.08 | 0 | 0.76 | 0 | 9.81 | 0.06 | 1.15 | 0.64 | House | 0.05 | 0 | 0.61 | 0 | 10.11 | 0 | 0.78 | 0.95 |
| Playground | 0.43 | 0.15 | 0.51 | 0 | 0.01 | 10.86 | 0.13 | 0.41 | Playground | 0.19 | 0.29 | 0 | 0 | 0.06 | 11.65 | 0 | 0.31 |
| Public Toilets | 0.23 | 0.31 | 2.01 | 0 | 0.69 | 0.23 | 8.4 | 0.64 | Public Toilets | 0.35 | 0 | 0.7 | 0 | 1.16 | 0 | 10 | 0.29 |
| Zoo | 1.59 | 0.59 | 0.89 | 0.36 | 0.59 | 1.19 | 0.44 | 6.86 | Zoo | 0.41 | 0.89 | 0.2 | 0.1 | 0.75 | 0.85 | 0.61 | 8.69 |
| Confusion Matrix Extra Trees 8 Classes 5 Sec | | | | | | | | | | | | | | | | | |
| | Airport | Bar Restaurant | Beach | Church | House | Playground | Public Toilets | Zoo | | | | | | | | | |
| Airport | 10.45 | 0.51 | 0.55 | 0 | 0.05 | 0.2 | 0.14 | 0.6 | | | | | | | | | |
| Bar Restaurant | 0.54 | 11.37 | 0.11 | 0 | 0 | 0.3 | 0.11 | 0.06 | | | | | | | | | |
| Beach | 0.76 | 0.19 | 10.26 | 0 | 0.58 | 0.35 | 0.31 | 0.05 | | | | | | | | | |
| Church | 0 | 0 | 0.13 | 12.22 | 0 | 0 | 0.03 | 0.13 | | | | | | | | | |
| House | 0.09 | 0 | 0.53 | 0 | 11.04 | 0 | 0.56 | 0.29 | | | | | | | | | |
| Playground | 0.14 | 0.15 | 0 | 0 | 0.18 | 11.85 | 0 | 0.19 | | | | | | | | | |
| Public Toilets | 0.28 | 0 | 1.19 | 0 | 0.95 | 0 | 9.77 | 0.31 | | | | | | | | | |
| Zoo | 0.33 | 0.56 | 0.26 | 0.11 | 0.55 | 0.98 | 0.43 | 9.29 | | | | | | | | | |

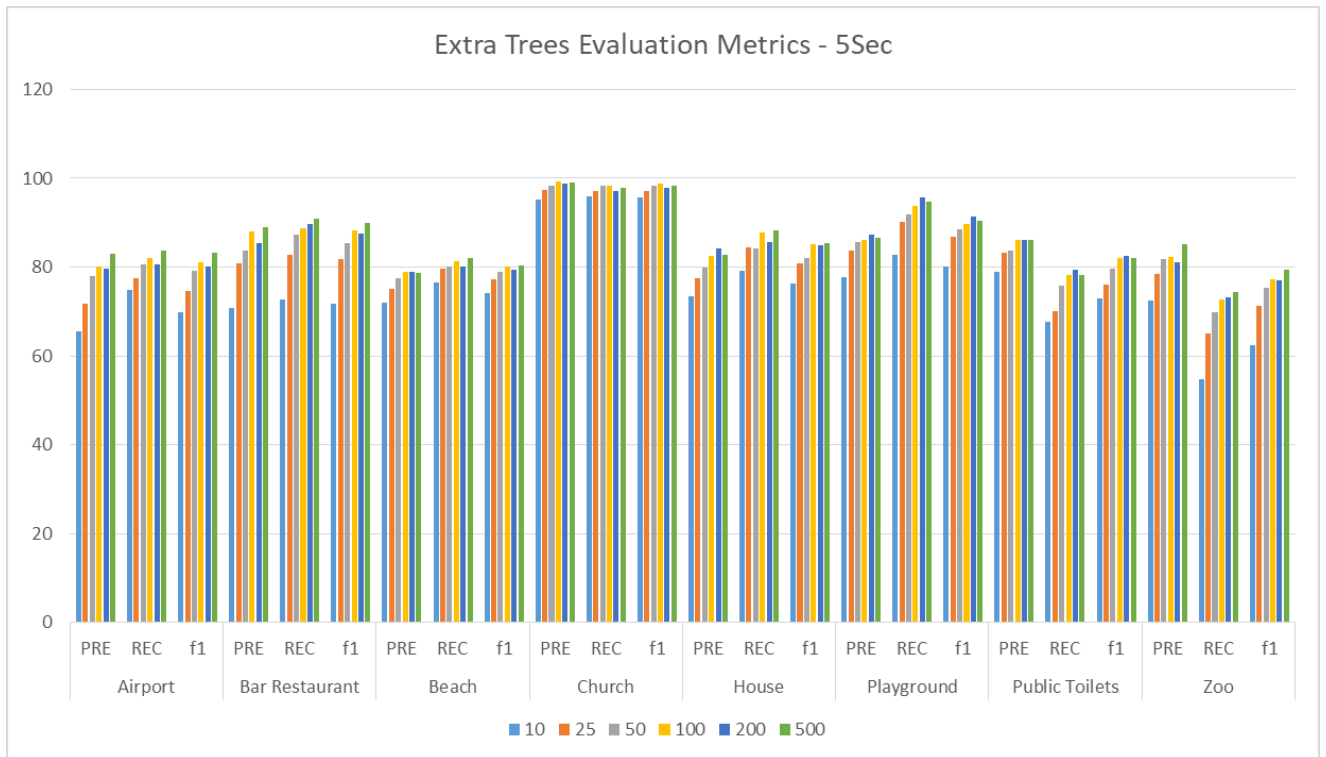
Πίνακας 3 Πίνακες *CM* για τη σύγκριση των αλγορίθμων του δεύτερου πειράματος

Όπως φαίνεται παρακάτω, πάλι ο αλγόριθμος Extra Forest δίνει την καλύτερη ακρίβεια Accuracy και F_1 για την παράμετρο των 500 δέντρων. Σε σχέση, όμως, με το προηγούμενο πείραμα η ακρίβεια είναι ελαφρώς χαμηλότερη –κατά 3.34 μονάδες. Συγκεκριμένα, το accuracy του ταξινομητή για το συγκεκριμένο πείραμα ισούται με 86.3% έναντι του 89.6% που ήταν το αντίστοιχο της προηγούμενης δοκιμασίας.



Εικόνα 16 Απεικόνιση της ακρίβειας (accuracy) και του μέτρου F_1 των πέντε ταξινομητών σε σχέση με τις διάφορες παραμέτρους που εξετάστηκαν

Επίσης στο γράφημα που ακολουθεί φαίνονται τα μεγέθη F_1 , *recall* και *precision* όλων των κλάσεων για τον αλγόριθμο Extra Trees σε σχέση με τις διάφορες παραμέτρους που εξετάστηκαν.



Εικόνα 17 Μέτρα απόδοσης για την κάθε κλάση έπειτα από την εφαρμογή του ταξινομητή Extra Trees σε αρχεία διάρκειας 5 sec

Είναι φανερό ότι η «Εκκλησία» συνεχίζει να δίνει τα υψηλότερα ποσοστά ακρίβειας σε σχέση με τις υπόλοιπες κλάσεις.

Ακολούθως εξετάσαμε κατά πόσο ο παραπάνω αλγόριθμος έχει την ικανότητα να γενικεύει στο ίδιο σύνολο δεδομένων που εξετάστηκε στην προηγούμενη δοκιμή. Τα αποτελέσματα της ακρίβειας ανά κλάση είναι τα ακόλουθα.

| Extra Trees - 5 sec | Accuracy Per Class |
|---------------------|--------------------|
| Airport | 92.3% |
| Bar-Restaurant | 61.5% |
| Beach | 84.6% |
| Church | 84.6% |
| House | 69.2% |
| Playground | 84.6% |
| Public Toilet | 84.6% |
| Zoo | 69.2% |

Πίνακας 4 Ακρίβεια ανά κλάση σε τυχαίο σύνολο δεδομένων

Παρατηρούμε ότι η «Παραλία» η «Εκκλησία» και ο «Παιδότοπος» συνεχίζουν να έχουν υψηλή ακρίβεια σε σχέση με το πρώτο πείραμα.

4.4.Πείραμα: 8 κλάσεις – παράθυρο 25 sec

Η τρίτη δοκιμή που πραγματοποιήθηκε στην παρούσα εργασία είναι να εκπαιδευτούν αλγόριθμοι σε αρχεία μεγέθους 25sec δηλαδή μεγαλύτερης διάρκειας από τα προηγούμενα δύο πειράματα. Οι πίνακες CM για τους αλγόριθμους που εξετάστηκαν είναι οι ακόλουθοι:

| Confusion Matrix SVM 8 Classes 25 Sec | | | | | | | | | Confusion Matrix Gradient Boosting 8 Classes 25 Sec | | | | | | | | |
|---|---------|----------------|-------|--------|-------|------------|----------------|------|---|---------|----------------|-------|--------|-------|------------|----------------|------|
| | Airport | Bar Restaurant | Beach | Church | House | Playground | Public Toilets | Zoo | | Airport | Bar Restaurant | Beach | Church | House | Playground | Public Toilets | Zoo |
| Airport | 10.79 | 0.53 | 0.6 | 0 | 0.04 | 0 | 0.33 | 0.23 | Airport | 10.69 | 0.81 | 0.2 | 0.03 | 0.11 | 0.01 | 0.3 | 0.35 |
| Bar Restaurant | 1.64 | 9.31 | 0.01 | 0.25 | 0.05 | 0.86 | 0.06 | 0.31 | Bar Restaurant | 0.55 | 10.45 | 0.01 | 0.19 | 0.04 | 0.91 | 0 | 0.35 |
| Beach | 0.96 | 0 | 9.29 | 0 | 0.65 | 0.44 | 1.1 | 0.06 | Beach | 0.66 | 0 | 10.04 | 0 | 0.49 | 0.56 | 0.58 | 0.18 |
| Church | 0.15 | 0.36 | 0 | 11.96 | 0 | 0.01 | 0 | 0.01 | Church | 0.41 | 0.21 | 0.13 | 11.52 | 0 | 0 | 0.15 | 0.08 |
| House | 0.01 | 0.18 | 0.84 | 0 | 9.92 | 0.01 | 0.45 | 1.09 | House | 0.05 | 0.09 | 0.13 | 0 | 11.5 | 0 | 0.2 | 0.54 |
| Playground | 0.1 | 0.99 | 0.18 | 0 | 0.14 | 10.4 | 0 | 0.7 | Playground | 0.04 | 0.46 | 0.13 | 0 | 0 | 11.75 | 0 | 0.13 |
| Public Toilets | 0.76 | 0.16 | 0.75 | 0.1 | 0.68 | 0.04 | 9.92 | 0.09 | Public Toilets | 0.43 | 0.01 | 1 | 0.01 | 0.5 | 0.04 | 10.17 | 0.34 |
| Zoo | 0.68 | 0.61 | 0.51 | 0.35 | 1.91 | 1.29 | 0.15 | 7 | Zoo | 0.54 | 0.34 | 0.08 | 0.24 | 1.18 | 0.7 | 0.03 | 9.41 |
| Confusion Matrix KNN 8 Classes 25 Sec | | | | | | | | | Confusion Matrix Gradient Boosting 8 Classes 25 Sec | | | | | | | | |
| | Airport | Bar Restaurant | Beach | Church | House | Playground | Public Toilets | Zoo | | Airport | Bar Restaurant | Beach | Church | House | Playground | Public Toilets | Zoo |
| Airport | 11.35 | 0.26 | 0.4 | 0 | 0.01 | 0.2 | 0.14 | 0.14 | Airport | 10.94 | 0.93 | 0.06 | 0 | 0 | 0.14 | 0.26 | 0.18 |
| Bar Restaurant | 2.35 | 7.41 | 0.81 | 0.03 | 0.18 | 1.28 | 0.14 | 0.31 | Bar Restaurant | 0.54 | 10.37 | 0.06 | 0.2 | 0.14 | 0.85 | 0 | 0.34 |
| Beach | 0.88 | 0 | 11.32 | 0 | 0.24 | 0.01 | 0.03 | 0.03 | Beach | 1.43 | 0.01 | 9.1 | 0 | 1.01 | 0.58 | 0.36 | 0.01 |
| Church | 0.91 | 0.25 | 0.25 | 10.77 | 0.03 | 0.06 | 0.23 | 0 | Church | 0.33 | 0.41 | 0.04 | 11.39 | 0 | 0.05 | 0.28 | 0.01 |
| House | 0.2 | 0 | 0.35 | 0 | 10.97 | 0.11 | 0.26 | 0.6 | House | 0.09 | 0.14 | 0.26 | 0 | 11.41 | 0 | 0.24 | 0.36 |
| Playground | 0.24 | 0.14 | 1.34 | 0 | 0.13 | 10.55 | 0 | 0.11 | Playground | 0.03 | 0.44 | 0 | 0.03 | 0 | 11.87 | 0 | 0.14 |
| Public Toilets | 0.69 | 0.03 | 1.67 | 0.23 | 0.46 | 0.14 | 9.17 | 0.11 | Public Toilets | 0.5 | 0 | 0.6 | 0 | 0.78 | 0 | 10.57 | 0.05 |
| Zoo | 0.81 | 0.25 | 0.68 | 0.18 | 1.43 | 1.26 | 0.31 | 7.59 | Zoo | 0.65 | 0.36 | 0 | 0.09 | 1.09 | 0.64 | 0.05 | 9.62 |
| Confusion Matrix Extra Trees 8 Classes 25 Sec | | | | | | | | | | | | | | | | | |
| | Airport | Bar Restaurant | Beach | Church | House | Playground | Public Toilets | Zoo | | | | | | | | | |
| Airport | 11.15 | 0.88 | 0.14 | 0 | 0 | 0.03 | 0.14 | 0.18 | | | | | | | | | |
| Bar Restaurant | 0.46 | 10.91 | 0.1 | 0.13 | 0.14 | 0.59 | 0 | 0.18 | | | | | | | | | |
| Beach | 0.86 | 0 | 9.92 | 0 | 0.64 | 0.61 | 0.46 | 0 | | | | | | | | | |
| Church | 0.31 | 0.33 | 0.08 | 11.42 | 0 | 0 | 0.33 | 0.04 | | | | | | | | | |
| House | 0.01 | 0.03 | 0.05 | 0 | 11.65 | 0 | 0.43 | 0.34 | | | | | | | | | |
| Playground | 0.01 | 0.2 | 0 | 0 | 0 | 12.05 | 0 | 0.24 | | | | | | | | | |
| Public Toilets | 0.68 | 0 | 0.46 | 0 | 0.5 | 0 | 10.8 | 0.06 | | | | | | | | | |
| Zoo | 0.55 | 0.45 | 0.08 | 0.06 | 0.91 | 0.44 | 0.25 | 9.76 | | | | | | | | | |

Πίνακες 5 Πίνακες CM για τη σύγκριση των αλγορίθμων του τρίτου πειράματος

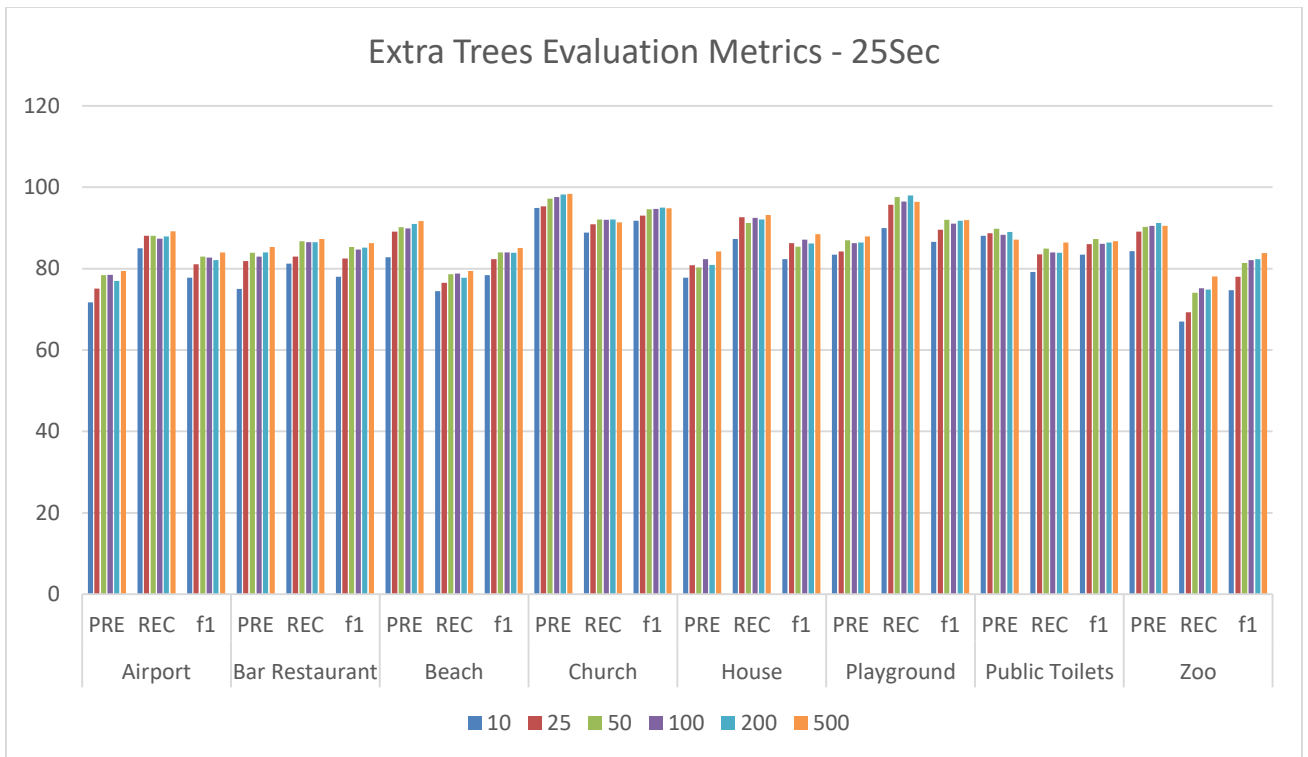
Ακόμη παρακάτω φαίνεται μια διαγραμματική απεικόνιση όλων των παραμέτρων που εξετάστηκαν για τον κάθε αλγόριθμο και είναι φανερό ποιοι είναι οι παράμετροι που ρυθμίζουν κατάλληλα τους ταξινομητές ώστε να πετυχαίνουν την μέγιστη ακρίβεια.



Εικόνα 18 Απεικόνιση της ακρίβειας (accuracy) και του μέτρου F1 των πέντε ταξινομητών σε σχέση με τις διάφορες παραμέτρους που εξετάστηκαν

Ο αλγόριθμος Extra Trees συνεχίζει να δίνει την μέγιστη ακρίβεια κατά τη διάρκεια του σταδίου της επικύρωσης. Συγκεκριμένα, η ακρίβεια (accuracy) για τον συγκεκριμένο αλγόριθμο και για παράμετρο 500 δέντρα φτάνει στο 87.66%.

Στο παρακάτω διάγραμμα απεικονίζεται η απόδοση των μεγεθών F_1 , *recall* και *precision* για τον αλγόριθμο Extra Trees σε κάθε κλάση εξετάζοντας όλες τις παραμέτρους.



Εικόνα 19 Μέτρα απόδοσης για την κάθε κλάση έπειτα από την εφαρμογή του ταξινομητή Extra Trees σε αρχεία διάρκειας 25 sec

Η «Εκκλησία» και ο «Παιδότοπος» φαίνεται να έχουν την καλύτερη απόδοση ανάμεσα στις υπόλοιπες κλάσεις για τον συγκεκριμένο αλγόριθμο.

Επιπρόσθετα, στον παρακάτω πίνακα φαίνεται μία σύγκριση ανάμεσα στα τρία πειράματα που διεξάχθηκαν παραπάνω μεταξύ όλων των αλγορίθμων και με τον συνδυασμό όλων των παραμέτρων.

| Parametre | OVERALL SVM 10 sec | | OVERALL SVM 5 sec | | OVERALL SVM 25 sec | |
|-----------|--------------------|------|-------------------|------|--------------------|------|
| | ACC | f1 | ACC | f1 | ACC | f1 |
| 0.001 | 62.9 | 62.4 | 59.1 | 58.3 | 67.7 | 66.5 |
| 0.01 | 72.5 | 72.2 | 69.4 | 69.3 | 78.2 | 78 |
| 0.5 | 76.6 | 76.5 | 70.8 | 70.7 | 78.6 | 78.4 |
| 1 | 76 | 75.9 | 71.1 | 71 | 78.4 | 78.3 |
| 5 | 76.3 | 76.2 | 71.9 | 71.7 | 78.4 | 78.3 |
| 10 | 76.3 | 76.2 | 70.6 | 70.5 | 78.6 | 78.4 |
| 20 | 76.4 | 76.3 | 72.2 | 72.1 | 77.5 | 77.3 |

| Parametre | OVERALL KNN 10 sec | | OVERALL KNN 5 sec | | OVERALL KNN 25 sec | |
|-----------|--------------------|------|-------------------|------|--------------------|------|
| | ACC | f1 | ACC | f1 | ACC | f1 |
| 1 | 80 | 80.2 | 75.1 | 75.1 | 79.1 | 79 |
| 3 | 70.6 | 70.4 | 71 | 71 | 70.8 | 71.1 |
| 5 | 67.7 | 66.5 | 62.9 | 62.6 | 68.7 | 68.5 |
| 7 | 62.2 | 60.5 | 59.4 | 58.8 | 68.6 | 68.3 |
| 9 | 59.2 | 57.1 | 57 | 56.1 | 64.5 | 63.6 |
| 11 | 58.1 | 55.6 | 54.4 | 53.2 | 61.6 | 60.1 |
| 13 | 57.8 | 55.6 | 52.8 | 51.9 | 58.9 | 57.1 |
| 15 | 55.9 | 53.7 | 51.6 | 50.2 | 56.5 | 54.1 |

| Parametre | OVERALL Extra Trees 10 | | OVERALL Extra Trees 5 | | OVERALL Extra Trees 25 | |
|-----------|------------------------|------|-----------------------|------|------------------------|------|
| | ACC | f1 | ACC | f1 | ACC | f1 |
| 10 | 81.6 | 81.5 | 75.6 | 75.4 | 81.6 | 81.6 |
| 25 | 85.8 | 85.7 | 80.8 | 80.7 | 84.9 | 84.8 |
| 50 | 87.8 | 87.7 | 83.5 | 83.4 | 86.7 | 86.6 |
| 100 | 88.3 | 88.2 | 85.4 | 85.3 | 86.6 | 86.6 |
| 200 | 89.3 | 89.3 | 85.2 | 85.1 | 86.6 | 86.6 |
| 500 | 89.6 | 89.5 | 86.3 | 86.2 | 87.7 | 87.7 |

| Parametre | OVERALL Gradient Boosting 10 sec | | OVERALL Gradient Boosting 5 sec | | OVERALL Gradient Boosting 25 sec | |
|-----------|----------------------------------|------|---------------------------------|------|----------------------------------|------|
| | ACC | f1 | ACC | f1 | ACC | f1 |
| 10 | 75.3 | 75.3 | 72 | 72 | 76.2 | 76.2 |
| 25 | 79.6 | 79.6 | 76.7 | 76.7 | 80.9 | 80.8 |
| 50 | 81.4 | 81.5 | 79.9 | 80 | 82.4 | 82.3 |
| 100 | 83.8 | 83.8 | 81.6 | 81.6 | 84 | 83.9 |
| 200 | 85 | 85.1 | 82.9 | 82.9 | 85.5 | 85.5 |
| 500 | 84.8 | 84.8 | 82.2 | 82.3 | 85 | 85 |

| Parametre | OVERALL Random Forest 10 sec | | OVERALL Random Forest 5 sec | | OVERALL Random Forest 25 sec | |
|-----------|------------------------------|------|-----------------------------|------|------------------------------|------|
| | ACC | f1 | ACC | f1 | ACC | f1 |
| 10 | 76.9 | 76.8 | 70.9 | 70.7 | 78.1 | 78.1 |
| 25 | 82.3 | 82.2 | 77.7 | 77.6 | 82.8 | 82.7 |
| 50 | 84 | 84 | 79.4 | 79.3 | 83.1 | 83 |
| 100 | 84.7 | 84.6 | 81 | 80.9 | 84.2 | 84.2 |
| 200 | 85.2 | 85.1 | 81.3 | 81.3 | 83.6 | 83.6 |
| 500 | 85.5 | 85.4 | 82.4 | 82.3 | 85.3 | 85.3 |

Πίνακας 6 Συγκριτική παρουσίαση των μεγεθών accuracy και F_1 όλων των αλγορίθμων για τις χρονικές διάρκειες των 10-5-25 sec για το σύνολο των παραμέτρων

4.5.Πείραμα: 4 κλάσεις – παράθυρο 10 sec

Το τελευταίο πείραμα της παρούσας εργασίας είναι να εξεταστούν οι μισές κλάσεις από αυτές των αρχικών πειραμάτων. Συγκεκριμένα οι κλάσεις που μελετήθηκαν σε αυτή τη δοκιμασία ήταν οι εξής: «Παραλία», «Εκκλησία», «Παιδότοπος» και «Σπίτι». Οι αλγόριθμοι εκπαιδεύτηκαν σε αρχεία χρονικής διάρκειας 10sec, με 100 διαθέσιμα αρχεία για την κάθε κλάση. Οι πίνακες CM που προέκυψαν για τους πέντε ταξινομητές φαίνονται παρακάτω.

| Confusion Matrix SVM 4 Classes 10 Sec | | | | | Confusion Matrix Extra Trees 4 Classes 10 Sec | | | | |
|---|-------|--------|-------|------------|---|-------|--------|-------|------------|
| | Beach | Church | House | Playground | | Beach | Church | House | Playground |
| Beach | 23.2 | 0 | 1.15 | 0.65 | Beach | 22.1 | 0 | 1.88 | 1.03 |
| Church | 0.25 | 24.72 | 0 | 0.03 | Church | 0.38 | 24.62 | 0 | 0 |
| House | 2.35 | 0.08 | 21.52 | 1.05 | House | 0.75 | 0 | 24.25 | 0 |
| Playground | 0.95 | 0 | 0.33 | 23.72 | Playground | 0 | 0 | 0.2 | 24.8 |
| Confusion Matrix KNN 4 Classes 10 Sec | | | | | Confusion Matrix Gradient Boosting 4 Classes 10 Sec | | | | |
| | Beach | Church | House | Playground | | Beach | Church | House | Playground |
| Beach | 23.82 | 0 | 1.03 | 0.15 | Beach | 22.4 | 0.33 | 1.43 | 0.85 |
| Church | 1.1 | 23.87 | 0.03 | 0 | Church | 0.55 | 24.45 | 0 | 0 |
| House | 2.68 | 0 | 22.02 | 0.3 | House | 2.38 | 0.08 | 22.45 | 0.1 |
| Playground | 0 | 0 | 0.08 | 24.92 | Playground | 0.1 | 0 | 0.35 | 24.55 |
| Confusion Matrix Random Forest 4 Classes 10 Sec | | | | | | | | | |
| | Beach | Church | House | Playground | | | | | |
| Beach | 21.17 | 0 | 2.55 | 1.28 | | | | | |
| Church | 0.58 | 24.42 | 0 | 0 | | | | | |
| House | 1.55 | 0 | 23.37 | 0.08 | | | | | |
| Playground | 0 | 0 | 0.3 | 24.7 | | | | | |

Πίνακας 7 Πίνακες CM για τη σύγκριση των αλγορίθμων του τέταρτου πειράματος

Στο συγκεκριμένο πείραμα όλοι οι αλγόριθμοι πετυχαίνουν ακρίβεια πάνω από 90% κατά το στάδιο της επικύρωσης. Αναλυτικότερα στον παρακάτω πίνακα παρουσιάζονται τα μεγέθη F_1 , $recall$ και $precision$ για κάθε κλάση για την βέλτιστη παράμετρο του κάθε αλγορίθμου καθώς και η συνολική ακρίβεια (accuracy) όλων των ταξινομητών κατά τη διαδικασία της επικύρωσης.

| SVM 4 Classes | | | | | | | | | | | | | | |
|-----------------------------|-------|------|------|--------|------|------|-------|------|------|------------|------|------|---------|------|
| | Beach | | | Church | | | House | | | Playground | | | OVERALL | |
| Parametre | PRE | REC | f1 | PRE | REC | f1 | PRE | REC | f1 | PRE | REC | f1 | ACC | f1 |
| 0.5 | 86.7 | 92.8 | 89.7 | 99.7 | 98.9 | 99.3 | 93.6 | 86.1 | 89.7 | 93.2 | 94.9 | 94.1 | 93.2 | 93.2 |
| KNN 4 Classes | | | | | | | | | | | | | | |
| | Beach | | | Church | | | House | | | Playground | | | OVERALL | |
| Parametre | PRE | REC | f1 | PRE | REC | f1 | PRE | REC | f1 | PRE | REC | f1 | ACC | f1 |
| 1 | 86.3 | 95.3 | 90.6 | 100 | 95.5 | 97.7 | 95.1 | 88.1 | 91.5 | 98.2 | 99.7 | 99 | 94.6 | 94.7 |
| Extra Trees 4 Classes | | | | | | | | | | | | | | |
| | Beach | | | Church | | | House | | | Playground | | | OVERALL | |
| Parametre | PRE | REC | f1 | PRE | REC | f1 | PRE | REC | f1 | PRE | REC | f1 | ACC | f1 |
| 200 | 95.2 | 88.4 | 91.7 | 100 | 98.5 | 99.2 | 92.1 | 97 | 94.5 | 96 | 99.2 | 97.6 | 95.8 | 95.7 |
| Gradient Boosting 4 Classes | | | | | | | | | | | | | | |
| | Beach | | | Church | | | House | | | Playground | | | OVERALL | |
| Parametre | PRE | REC | f1 | PRE | REC | f1 | PRE | REC | f1 | PRE | REC | f1 | ACC | f1 |
| 500 | 88.1 | 89.6 | 88.8 | 98.4 | 97.8 | 98.1 | 92.7 | 89.8 | 91.2 | 96.3 | 98.2 | 97.2 | 93.8 | 93.8 |
| Random Forest 4 Classes | | | | | | | | | | | | | | |
| | Beach | | | Church | | | House | | | Playground | | | OVERALL | |
| Parametre | PRE | REC | f1 | PRE | REC | f1 | PRE | REC | f1 | PRE | REC | f1 | ACC | f1 |
| 200 | 90.9 | 84.7 | 87.7 | 100 | 97.7 | 98.8 | 89.1 | 93.5 | 91.3 | 94.8 | 98.8 | 96.8 | 93.7 | 93.6 |

Πίνακας 8 Τα μέτρα απόδοσης, Precision, Recall και F1, όλων των αλγορίθμων που επιτυγχάνει η βέλτιστη παράμετρος για τις 4 κλάσεις

Είναι φανερό ότι και στις τέσσερις κλάσεις τα μέτρα απόδοσης είναι υψηλά για την επιλεγμένη κάθε φορά παράμετρο. Μετά την διαδικασία της επικύρωσης επιλέξαμε τον αλγόριθμο Extra Trees για να δοκιμαστεί σε ένα άγνωστο σύνολο 52 ηχητικών αρχείων. Τα αποτελέσματα ήταν τα εξής για την κάθε κλάση:

| Extra Trees - 4 Cl | Accuracy Per Class |
|--------------------|--------------------|
| Beach | 92.3% |
| Church | 92.3% |
| House | 76.9% |
| Playground | 100.0% |

Πίνακας 9 Ακρίβεια ανά κλάση σε τυχαίο σύνολο δεδομένων

Τα αποτελέσματα είναι σαφώς βελτιωμένα σε σύγκριση με τα προηγούμενα πειράματα στα οποία μελετήθηκαν και οι οχτώ κλάσεις μαζί. Στη συγκεκριμένη δοκιμή η συνολική ακρίβεια του πειράματος ξεπερνά το 90%.

5. Συμπεράσματα

Στην παρούσα εργασία, μελετήθηκαν μερικοί γνωστοί αλγόριθμοι πάνω σε ηχητικά δεδομένα. Χρησιμοποιήθηκε μια κατασκευασμένη από εμάς βάση δεδομένων, στην οποία συμπεριλάβαμε όλους τους πιθανούς ήχους που τυχόν μπορεί να περιέχονται στην κάθε κατηγορία/κλάση. Σ' όλα τα πειράματα, ο πιο αποδοτικός αλγόριθμος φάνηκε να είναι ο ταξινομητής Extra Trees με μία ιδανική παράμετρο πεντακοσίων δέντρων. Αν και κατά τη διάρκεια της επικύρωσης τα ποσοστά ακρίβειας ήταν αρκετά υψηλά, κατά το στάδιο της εφαρμογής του αλγορίθμου σ' ένα άγνωστο σύνολο ηχητικών δεδομένων κάποιες από τις κλάσεις πέτυχαν χαμηλή ακρίβεια. Ίσως αυτό συνέβη λόγω της μεγάλης μεταβλητότητας της κατηγορίας και της ομοιότητας μεταξύ των διαφόρων κλάσεων. Κατά τη διάρκεια της επικύρωσης, παρατηρήθηκε ότι οι αλγόριθμοι που εκπαιδεύτηκαν στα ηχητικά δεδομένα διάρκειας 5sec, είχαν την χαμηλότερη απόδοση. Όσον αφορά τον αλγόριθμο SVM, φαίνεται να πετυχαίνει μεγαλύτερη ακρίβεια στην περίπτωση που τα αρχεία του συνόλου εκπαίδευσης είχαν διάρκεια 25sec. Επίσης, οι ταξινομητές Gradient Boosting, Random Forest και KNN μένουν σχεδόν ίδιοι τόσο όταν τα δεδομένα εκπαίδευσης είναι διάρκειας 10sec όσο και όταν είναι 25 sec. Τέλος, ο αλγόριθμος Extra Trees επιτυγχάνει μεγαλύτερη ακρίβεια για τα δεδομένα εκπαίδευσης, που διαρκούν 10 sec. Η κλάση «Εκκλησία» παρουσιάζει τα υψηλότερα ποσοστά ακρίβειας ταξινόμησης σε όλα τα πειράματα, ενώ η κλάση «Εστιατόρια» βρίσκεται ανάμεσα στα χαμηλότερα ποσοστά.

Στο τελευταίο πείραμα εξετάστηκαν ,μετά από αφαίρεση, λιγότερες κλάσεις για να περιοριστεί η ομοιότητα στο περιεχόμενο τους. Ο αλγόριθμος Extra Tress εκπαιδεύτηκε σε τέσσερις κλάσεις και τρεις απ' αυτές προσέγγισαν μεγάλα ποσοστά ακρίβειας. Γενικά, η ακρίβεια του πειράματος ,τόσο κατά την επικύρωση όσο και κατά την εκτέλεση στο άγνωστο δείγμα, ήταν αρκετά υψηλή.

Ασφαλώς υπάρχει περιθώριο βελτίωσης στην παραπάνω έρευνα. Αρχικά, θα ήταν χρήσιμος ο εμπλουτισμός της βάσης δεδομένων με ηχητικά δεδομένα, έτσι θα συμπεριληφθούν ακόμα περισσότερες δραστηριότητες σε κάθε κλάση, για να ισχυροποιηθεί η γενίκευση του κάθε αλγορίθμου. Αυτός είναι ίσως και ένας τρόπος να αυξηθεί η ακρίβεια στις κλάσεις με τα

χαμηλότερα ποσοστά. Επίσης μια πολύ ενδιαφέρουσα προσέγγιση, η οποία έχει ακολουθηθεί από διάφορους ερευνητές είναι να μελετηθούν διαφορετικά χαρακτηριστικά ήχου απ' αυτά της παρούσας εργασίας, όπως και να εξεταστούν εναλλακτικοί αλγόριθμοι μηχανικής μάθησης. Ακόμη, η εφαρμογή των αλγορίθμων σε έναν συνδυασμό ηχητικών και οπτικών χαρακτηριστικών ίσως πετύχει μεγαλύτερη ακρίβεια. Τέλος η αυτόματη κατηγοριοποίηση περιβαλλοντικών ήχων σε πραγματικό χρόνο, θα μπορούσε να χρησιμοποιηθεί ως εφαρμογή σε κινητές συσκευές.

Βιβλιογραφία-Πηγές

- [1] "Understanding Random Forest", Tony Yiu, Jun. 2019
- [2] "Random Forests", Leo Breiman and Adele Cutler
- [3] "Random Decision Forests", Ho, Tin Kam, 1995
- [4] "An Intuitive Explanation of Random Forest and Extra Trees Classifiers", Frank Ceballos, Jul. 2019
- [5] "Extremely randomized trees", Pierre Geurts, Damien Ernst, Louis Wehenkel, Mar. 2006
- [6] "Extra Trees Classifier- How does Extra Trees Classifier reduce the risk of overfitting?" Naman Bhandari, Oct. 2018
- [7] "Understanding Gradient Boosting Machines", Harshdeep Singh, Nov. 2018
- [8] "A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning", Jason Brownlee, Sep. 2016
- [9] "Gradient Boosting from scratch", Prince Grover, Dec. 2017
- [10] "A Quick Guide to Boosting in ML", Jocelyn D'Souza, Mar. 2018
- [11] "Machine Learning Basics with the K-Nearest Neighbors Algorithm", Onel Harrison, Sep. 2018
- [12] "Εισαγωγή στην υπολογιστική νοημοσύνη", Β. Καμπουρλάζος, Γ. Παπακώστας, 2015
- [13] "Biodiesel: Quality, Emissions and By-Products", Gisela Montero, Margarita Stoytcheva
- [14] "Επιχειρηματική Ευφυΐα και Εξόρυξη Δεδομένων", Ε.Γ.Κύρκος, 2015
- [15] "A Simple Introduction to K-Nearest Neighbors Algorithm", Dhilip Subramanian, Dhilip Subramanian, Jun. 2019
- [16] "An Introduction to Data Science", Dr.Saed Sayad

- [17] "Support Vector Machines notes", R.Berwick, MIT University
- [18] "Support Vector Machines-Applications", Wikipedia
- [19] "Support Vector Machine — Introduction to Machine Learning Algorithms", Rohith Gandhi, Jun. 2018
- [20] "Chapter 2: SVM (Support Vector Machine) — Theory", Savan Patel, May 2017
- [21] "Understanding Support Vector Machine algorithm from examples", Sunil Ray, Sep. 2017
- [22] "What is Machine Learning? A definition", Expert System, Mar. 2017
- [23] "What is machine learning?" Karen Hao, Nov. 2018
- [24] "4 Types of Machine Learning Algorithms", Volodymyr Bilyk
- [25] "R and Data Mining", Yangchang Zhao, 2013
- [26] "Data Mining in Brief", Sidath Asiri, Dec. 2017
- [27] "Regression Versus Classification Machine Learning: What's the Difference?", Michael J. Garbade. Aug. 2018
- [28] "Cross-Validation", Georgios Drakos, Aug. 2018
- [29] "Why and how to Cross Validate a Model?", Sanjay.M, Nov. 2018
- [30] "Sound", Chris Woodford, Feb. 2019
- [31] "Φυσική του Ήχου", Αραβαντινός, ΤΕΙ Αθήνας
- [32] "Introduction to AUDIO ANALYSIS", T.Giannakopoulos, A.Pikrakis