

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



- ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ -

«Κατηγοριοποίηση της Μουσικής σε είδη με τη χρήση
Αρχιτεκτονικών Βαθιάς Μάθησης»

Ειρήνη Φωκά

ΑΜ:Π2013083

Επιβλέπων: Κάτια-Λήδα Κερμανίδου

Οκτώβριος 2020

Πρόλογος

Η παρούσα εργασία έχει θέμα την κατηγοριοποίηση της μουσικής σε είδη με τη χρήση Αρχιτεκτονικών Βαθιάς Μάθησης. Με την έννοια κατηγοριοποίηση της μουσικής εννοούμε την ταξινόμηση των μουσικών σημάτων σε μοναδικές κατηγορίες, που βασίζονται σε υπολογιστικές αναλύσεις αναπαραστάσεων των μουσικών χαρακτηριστικών. Έχουν γίνει πολλές προσπάθειες κατηγοριοποίησης μουσικής με χειρωνακτικό τρόπο, όμως η μέθοδος αυτή αποτελεί μια ακριβή, χρονοβόρα και υποκειμενική διαδικασία σε τέτοιο βαθμό που απαιτείται η αυτοματοποίησή της. Η αυτόματη κατηγοριοποίηση της μουσικής αποτελεί βασικό στοιχείο της ανάκτησης μουσικής πληροφορίας και έχει κερδίσει μεγάλη βαρύτητα ως τρόπος δόμησης και οργάνωσης του αυξανόμενου μεγάλου αριθμού μουσικών αρχείων που είναι διαθέσιμα ψηφιακά στο διαδίκτυο.

Ένας τρόπος για την αυτόματη κατηγοριοποίηση της μουσικής είναι η χρήση μοντέλων βαθιάς μάθησης, τα οποία εκπαιδεύονται με τη χρήση μεγάλων επισημασμένων συνόλων δεδομένων και αρχιτεκτονικές νευρωνικών δικτύων που «μαθαίνουν» τα χαρακτηριστικά των δεδομένων αυτόματα, χωρίς τη δημιουργία ανάγκης της χειροκίνητης εξαγωγής τους.

Στο πρώτο κεφάλαιο γίνεται μια εισαγωγή στο ρόλο της μουσικής από την αρχαιότητα έως σήμερα, στην ανάγκη για την κατηγοριοποίησή της, στα είδη κατηγοριοποίησης και στα χαρακτηριστικά της μουσικής που χρησιμοποιούνται. Έπειτα, εισάγονται βασικοί ορισμοί όπως η Τεχνητή Νοημοσύνη, η Μηχανική Μάθηση, τα Νευρωνικά δίκτυα και η Βαθιά Μάθηση. Αναφέρονται προσπάθειες κατηγοριοποίησης της μουσικής που έχουν γίνει μέχρι σήμερα. Στη συνέχεια, διεξάγονται πειράματα με το εργαλείο WEKA, με τη χρήση του συνόλου δεδομένων «Greek Audio Dataset», αλγόριθμο ταξινόμησης τον Deep Learning Mlp Classifier του WEKA, και μεθόδους ταξινόμησης cross-validation και train-test split. Τέλος, γίνονται συγκρίσεις μεταξύ ανεπεξέργαστων και επεξεργασμένων δεδομένων, μεταξύ χαρακτηριστικών του ήχου, και μεταξύ δεδομένων ήχου και δεδομένων στίχων και αναλύονται τα αποτελέσματα.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά την επιβλέπουσα καθηγήτριά μου, κ.Κάτια-Λήδα Κερμανίδου για την υποστήριξή της, την κατανόηση και την εμπιστοσύνη της.

Επίσης, θα ήθελα να ευχαριστήσω τον κ. Δήμο Μακρή για την παραχώρηση του συνόλου δεδομένων «Greek Audio Dataset» και τις πολύτιμες συμβουλές του.

Περιεχόμενα

1. Εισαγωγή	- 1 -
1.1 Λίγα λόγια για τη μουσική	- 1 -
1.2 Κατηγοριοποίηση της μουσικής	- 2 -
1.2.1 Κατηγοριοποίηση με βάση τη διάθεση	- 3 -
1.2.2 Κατηγοριοποίηση με βάση το μουσικό είδος	- 5 -
1.3 Χαρακτηριστικά της μουσικής για την κατηγοριοποίησή της	- 6 -
2. Βαθιά Μάθηση(Deep learning)	- 9 -
2.1 Τεχνητή νοημοσύνη	- 9 -
2.2 Μηχανική μάθηση	- 10 -
2.3 Βαθιά μηχανική μάθηση και Νευρωνικά δίκτυα	- 11 -
2.4 Συναρτήσεις ενεργοποίησης	- 14 -
2.5 Τοπολογίες νευρωνικών δικτύων.....	- 15 -
2.6 Στάδια λειτουργίας νευρωνικών δικτύων.....	- 17 -
3. Προσπάθειες κατηγοριοποίησης της μουσικής.....	- 18 -
3.1 Πειράματα κατηγοριοποίησης μουσικής σε είδη	- 18 -
3.2 Ταξινόμηση με βάση τη διάθεση.....	- 24 -
3.3 Ταξινόμηση με <i>deep learning</i>	- 25 -
4. Πειράματα με το εργαλείο Meka.....	- 31 -
4.1 <i>Greek Audio Dataset (GAD)</i>	- 31 -
4.2 Λίγα λόγια για τα εργαλεία Weka και Meka	- 33 -
4.3 Προεπεξεργασία δεδομένων	- 34 -
4.4 Το πακέτο βαθιάς μάθησης ‘WekaDeeplearning4j’	- 37 -
4.5 Ταξινόμηση πολλαπλών ετικετών	- 39 -
4.5.1 Μέθοδοι Αξιολόγησης.....	- 41 -
4.5.2 Μετρικές αξιολόγησης	- 42 -
5. Αποτελέσματα πειραμάτων	- 44 -
5.1 Σύγκριση επεξεργασμένων-μη επεξεργασμένων δεδομένων	- 44 -
5.1.1 Σε <i>audio</i> δεδομένα	- 44 -
5.1.2 Σε δεδομένα στίχων.....	- 48 -
5.2 Σύγκριση χαρακτηριστικών του ήχου	- 50 -
5.3 Σύγκριση δεδομένων Audio με δεδομένα στίχων	- 52 -
6 Συμπεράσματα και μελλοντικές προεκτάσεις	- 57 -
Βιβλιογραφία	- 59 -

1. Εισαγωγή

1.1 Λίγα λόγια για τη μουσική

Η μουσική είναι μια τέχνη που εξελίχθηκε παράλληλα με την ιστορία του ανθρώπου. Η σημασία της μουσικής στην αρχαία Ελλάδα ήταν η αδιάλυτη ενότητα λόγου και ήχου. Αποτελούσε μέσο πνευματικής ωρίμανσης και συνώνυμά της ήταν η παιδεία, η δύναμη και η πνευματική καλλιέργεια. Αρχαίοι φιλόσοφοι ασχολήθηκαν με την τέχνη της μουσικής. Για τον Πυθαγόρα η μουσική ήταν μια μαθηματική επιστήμη, η ουσία της οποίας ήταν ο αριθμός, και η ομορφιά της η έκφραση των αρμονικών σχέσεων των αριθμών. Μάλιστα ήταν ο πρώτος που έθεσε τις βάσεις της επιστήμης της μουσικής με μια επιστημονικά θεμελιωμένη θεωρία. Ανακάλυψε τη σχέση ανάμεσα στο μήκος των χορδών και το τονικό ύψος που δίνουν. Για τον Πλάτωνα η μουσική δίνει ψυχή στο σύμπαν, φτερά στη σκέψη, απογειώνει τη φαντασία, χαρίζει χαρά στη λύπη και ζωή στα πάντα. Διατύπωσε πως όταν αλλάζουν οι νόμοι της μουσικής, αλλάζουν και οι νόμοι της πολιτείας.

Σήμερα θα λέγαμε πως η μουσική είναι ένας τρόπος με τον οποίο οι άνθρωποι μπορούν να εκφράζονται, να επικοινωνούν, να μοιράζονται, να επινοούν, να δημιουργούν. Είναι μια γλώσσα, και γλώσσα δεν εννοούμε μόνο τη γλώσσα που μιλάμε, αλλά κάθε τρόπο οργανωμένης έκφρασης και επικοινωνίας. Οι άνθρωποι χρησιμοποιούν τη γλώσσα της μουσικής για να οργανώσουν τον ήχο με τέτοιο τρόπο ώστε να εκφράζονται καλλιτεχνικά και να κοινωνούν τα συναισθήματά τους, τις ελπίδες και τα οράματά τους.

Όπως όλες οι γλώσσες, έτσι και η μουσική διαμορφώνεται και μεταβάλλεται με την πάροδο του χρόνου, μέσα από ιστορικές και κοινωνικές συνθήκες. Δεν είναι ένα αυτόνομο φαινόμενο που υπακούει σε αυστηρούς κανόνες, αλλά συνυφαίνεται με την καθημερινότητα των ανθρώπων, τις ιδέες και τις πεποιθήσεις που χαρακτηρίζουν κάθε ιστορική περίοδο. Με αυτόν τον τρόπο στην πορεία των χρόνων εμφανίζονται διαφορετικά μουσικά είδη και μορφές και μεταβάλλεται το περιεχόμενο των διαφορετικών μουσικών ρευμάτων και στυλ που παρατηρούνται στη διάρκεια της ιστορίας.

1.2 Κατηγοριοποίηση της μουσικής

Από την αρχή του 21^{ου} αιώνα ο κόσμος βρίσκεται σε μια συνεχή ανάπτυξη της “online” μουσικής πληροφορίας. Τα ψηφιακά σύνολα δεδομένων γίνονται ολοένα και πιο δημοφιλή, τόσο στα επαγγελματικά αποθετήρια, όσο και στις προσωπικές συλλογές. Με τη συνεχή εξέλιξη των υπηρεσιών του διαδικτύου αναμένεται ακόμα μεγαλύτερη αύξηση του αριθμού των ανθρώπων που ασχολούνται με τις ακουστικές βιβλιοθήκες[1]. Η αύξηση αυτή των μουσικών συνόλων δεδομένων στο διαδίκτυο και στα συστήματα πολυμέσων έχουν δημιουργήσει την ανάγκη για την υπόσταση εφαρμογών ανάκτησης μουσικής πληροφορίας και για την ανάλυση των μουσικών συνόλων.

Η Ανάκτηση Μουσικής Πληροφορίας αφορά το ερευνητικό πεδίο που στοχεύει στην ανάπτυξη μεθόδων και τεχνικών όσο αφορά την εξόρυξη γνώσης από τη μουσική πληροφορία. Είναι ένα πεδίο ιδιαίτερα διεπιστημονικό και εμπεριέχει πεδία όπως η μουσικολογία, η ψυχολογία, η επεξεργασία σήματος, η ανάκτηση πληροφορίας, η μηχανική μάθηση και η διάδραση ανθρώπου- υπολογιστή(Καρύδης Ιωάννης,2015).

Ένα θέμα ανάκτησης μουσικής πληροφορίας είναι το πρόβλημα της κατηγοριοποίησης της μουσικής. Με την έννοια κατηγοριοποίηση της μουσικής εννοούμε την ταξινόμηση των μουσικών σημάτων σε μοναδικές κατηγορίες, που βασίζονται σε υπολογιστικές αναλύσεις αναπαραστάσεων των μουσικών χαρακτηριστικών.

Έχουν γίνει πολλές προσπάθειες κατηγοριοποίησης μουσικής με χειρωνακτικό τρόπο, όμως η μέθοδος αυτή αποτελεί μια ακριβή, χρονοβόρα και υποκειμενική διαδικασία σε τέτοιο βαθμό που απαιτείται η αυτοματοποίησή της. Η αυτόματη κατηγοριοποίηση της μουσικής αποτελεί βασικό στοιχείο της ανάκτησης μουσικής πληροφορίας και έχει κερδίσει μεγάλη βαρύτητα ως τρόπος δόμησης και οργάνωσης του αυξανόμενου μεγάλου αριθμού μουσικών αρχείων που είναι διαθέσιμα ψηφιακά στο διαδίκτυο[3].

Η διαδικασία κατηγοριοποίησης της μουσικής χωρίζεται σε δύο σκέλη. Αρχικά γίνεται εξαγωγή χαρακτηριστικών από το μουσικό σήμα, τα οποία περιγράφουν το μουσικό περιεχόμενο και παρέχεται ένα πλαίσιο για την ανάπτυξη και την αξιολόγησή τους. Παραδείγματα στα οποία μπορούν να χρησιμοποιηθούν τέτοια χαρακτηριστικά είναι η ανάκτηση ομοιότητας, διαχωρισμού και η ηχητική αποτύπωση, δίνοντας βάση στα θεμέλια των τεχνικών ηχητικής ανάλυσης. Η πλειοψηφία τέτοιων τεχνικών αναλύσεων χρησιμοποιεί αριθμητικά χαρακτηριστικά που προσπαθούν να αντλήσουν πληροφορίες από το κάθε μουσικό κομμάτι που προσδιορίζουν. Στη συνέχεια σχεδιάζεται ένας

μηχανισμός- αλγόριθμος ή μαθηματικό μοντέλο- ο οποίος αναγνωρίζει τις ετικέτες από την αναπαράσταση των μουσικών ήχων με βάση τα χαρακτηριστικά που εξήχθησαν στο πρώτο βήμα.

Είναι πιθανό στο σύντομο μέλλον όλη η μουσική που έχει καταγραφεί στη ανθρώπινη ιστορία να είναι διαθέσιμη στο διαδίκτυο. Η αυτόματη μουσική ανάλυση θα είναι μια από τις εμπορικές υπηρεσίες της κατανομής μουσικού περιεχομένου που θα χρησιμοποιείται για να προσεγγίσει πελάτες (Tzanetakis, 2002).

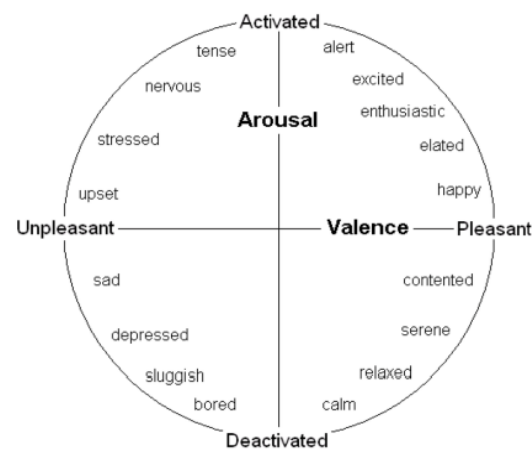
Αρκετοί ερευνητές έχουν ασχοληθεί με την επίλυση του προβλήματος. Η κατηγοριοποίηση γίνεται κυρίως με γνώμονα το μουσικό είδος αλλά και τις διαθέσεις που προκαλεί η μουσική.

1.2.1 Κατηγοριοποίηση με βάση τη διάθεση

Η συναισθηματική διάσταση της μουσικής συχνά αναφέρεται ως συναίσθημα ή διάθεση που αυτή προκαλεί. Η διάθεση ενός τραγουδιού εκφράζεται μέσω των χαρακτηριστικών που εξάγονται από κάθε μουσικό κομμάτι. Εξαιτίας της φυσικής υποκειμενικότητας της μουσικής αντίληψης δεν υπάρχουν γενικευμένα αποδεκτά πρότυπα για κατηγορίες διαθέσεων. Επιπλέον, στην επιστημονική κοινότητα Ανάκτησης Μουσικής Πληροφορίας υπάρχουν θεμελιώδη ζητήματα που παραμένουν άλυτα, όπως η έλλειψη γενικής συναίνεσης που αφορούν στις κατηγορίες διαθέσεων, δυσχεραίνοντας τη σύγκριση των προσεγγίσεων ταξινόμησής τους, και η μη εδραίωση συγκεκριμένης επιστημονικής ορολογίας. Μουσικοί ψυχολόγοι έχουν δημιουργήσει πολλά διαφορετικά μοντέλα, τα οποία έχουν κατακριθεί για έλλειψη απόδοσης σημασίας στο μουσικό περιεχόμενο [5].

Παρόλα αυτά η συναισθηματική διάσταση της μουσικής έχει αναγνωριστεί ως σημαντικός παράγοντας στην οργάνωση και την επεξεργασία της μουσικής πληροφορίας (Hu, Xiao, 2010). Ψυχολογικές έρευνες [7] έχουν δείξει πως μέρος της σημασιολογικής πληροφορίας υπάρχει αποκλειστικά στους στίχους ενός μουσικού κομματιού. Σύμφωνα με τους Aucouturier, Julien και Pachet (2003), πολλά χαρακτηριστικά υψηλού επιπέδου, όπως τα σημασιολογικά δε γίνονται διακριτά χρησιμοποιώντας αποκλειστικά τεχνικές με φασματογραφήματα. Λόγω της συναισθηματικής πληροφορίας που περιέχεται στους στίχους πολλοί ερευνητές [8] ξεκίνησαν να τους χρησιμοποιούν ως επέκταση του ήχου και έχει ήδη αναφερθεί σημαντική βελτίωση στα αποτελέσματα κατηγοριοποίησης σε είδη.

Ένα χρήσιμο εργαλείο για την κατηγοριοποίηση της μουσικής σε είδη είναι το μοντέλο του *Thayer(1989)* με το οποίο γίνεται μια χαρτογράφηση των συναισθημάτων. Στο μοντέλο υπάρχουν δύο διαστάσεις, valence και arousal οι οποίες διαιρούν το δισδιάστατο συναισθηματικό πλάνο σε τέσσερα μέρη, όπως φαίνεται στην **Εικόνα 1**. Η διάσταση arousal αντιστοιχεί στη ενέργεια, δηλαδή στο συναίσθημα που προκαλείται από ένα μουσικό κομμάτι και η διάσταση valence αντιστοιχεί στην ένταση του συναισθήματος.



Εικόνα 1. Το δισδιάστατο συναισθηματικό πλάνο του *Thayer*(από [31])

1.2.2 Κατηγοριοποίηση με βάση το μουσικό είδος

Το μεγάλο μέγεθος ψηφιακής μουσικής δημιουργεί τη ανάγκη για αποτελεσματικούς τρόπους περιήγησης, οργάνωσης και δυναμικής ενημέρωσης των συλλογών. Η συσχέτιση μουσικών κομματιών με είδη είναι καίριας σημασίας για τη βοήθεια των χρηστών να εντοπίζουν αυτό που αναζητούν.

Τα μουσικά είδη είναι κατηγορίες οι οποίες δημιουργούνται από τη σύνθετη αλληλεπίδραση κουλτούρων και καλλιτεχνών, αλλά και πολιτισμικών, εμπορικών και ιστορικών παραγόντων, για να χαρακτηρίσουν τις ομοιότητες μεταξύ μουσικών ή συνθέσεων και να οργανώσουν τις μουσικές συλλογές. Αποτελούν τα πιο δημοφιλή μεταδεδομένα για την περιγραφή του μουσικού περιεχομένου.

Υπάρχουν πολλές οπτικές σχετικά με το τι είναι μουσικό είδος και τις προσεγγίσεις κατηγοριοποίησής τους. Ο *Begstra(2006)* παρουσιάζει δυο έννοιες του είδους που θέτουν τις παραμέτρους που το περιγράφουν ως μηχανισμούς για μουσική σύσταση. Χωρίζει τα είδη σε παραδοσιακά, δηλαδή αυτά που έχουν προκύψει από τις δισκογραφικές βιομηχανίες, και σε συλλογικά, που προσδιορίζονται από online σύνολα δεδομένων με μεγάλο αριθμό ακροατών που εφαρμόζουν τις δικές τους ετικέτες. Το μουσικό είδος είναι η πιο ευρέως διαδεδομένη περιγραφή του μουσικού περιεχομένου, παρόλο που δεν είναι σαφώς καθορισμένο, καθώς οι ετικέτες του ενδέχεται να βασίζονται είτε σε στοιχεία που αφορούν την τέχνη, είτε σε πολιτισμικούς και εμπορικούς παράγοντες (*Aucouturier et al.,2003*). Τα μουσικά είδη προσδιορίζονται από ειδικούς, αλλά και από ερασιτέχνες όπως οι απλοί χρήστες. Οι *Li, Tao & Tzanetakis το 2003*, αναφέρουν ως μουσικό είδος προσδιορισμούς που έχουν δημιουργήσει οι άνθρωποι με σκοπό να δομηθεί το αχανές σύμπαν της μουσικής. Οι *Pachet και Cazaly(2000)* εισήγαγαν μια ιεραρχία ειδών τα οποία διαχωρίζονται από ενορχηστρωτικές και ιστορικές παραμέτρους. Όλα τα παραπάνω αποτελούν εναλλακτικούς τρόπους της συμβατικής ερμηνείας του είδους που ενσαρκώνει όρους όπως jazz, rock και pop.

Τα όρια μεταξύ των ειδών είναι χαώδη και παρά την εξέλιξη στη κατηγοριοποίηση της μουσικής δεν έχουν οριστεί πλήρως αν υπάρχουν συγκεκριμένα χαρακτηριστικά που καθορίζουν το είδος ή αν αυτό εξαρτάται από την υποκειμενική αντίληψη του ακροατή[11]. Ρίχνοντας μια ματιά σε συγκεκριμένα και συχνά χρησιμοποιούμενα είδη έχει παρατηρηθεί πόσα διαφορετικά κριτήρια που καθορίζουν το είδος μπορούν να υπάρχουν. Για

παράδειγμα η Ινδική μουσική ορίζεται γεωγραφικά ενώ η Μπαρόκ μουσική σχετίζεται με μια ιστορική περίοδο που περιλάμβανε ένα μεγάλο εύρος από στυλ[14].

Οι δυσκολίες στο διαχωρισμό των ειδών είναι λογικό να υπάρχουν γιατί η μουσική είναι μια τέχνη που εξελίσσεται και οι ερμηνευτές και συνθέτες έχουν επηρεαστεί από μουσικές που ανήκουν σε πολλά διαφορετικά είδη.

Παρά τα ασαφή και πολλές φορές επικαλυπτόμενα όρια, έχει παρατηρηθεί πως τα μουσικά σήματα –ψηφιακά ή αναλογικά- που ανήκουν σε ένα είδος μοιράζονται κοινά χαρακτηριστικά σχετικά με την ενορχήστρωση, τη ρυθμική δομή και το τονικό περιεχόμενο[12]. Το γεγονός αυτό υποδεικνύει ότι είναι εφικτή η υλοποίηση της αυτόματης κατηγοριοποίησης της μουσικής σε είδη.

1.3 Χαρακτηριστικά της μουσικής για την κατηγοριοποίησή της

Οι άνθρωποι ταξινομούν ή επισημειώνουν τη μουσική με βάση ποικίλα χαρακτηριστικά που εξάγονται από τα μουσικά σήματα. Η διαδικασία αυτή αποτελεί το πρώτο βήμα για την κατηγοριοποίηση της μουσικής. Τα μουσικά κομμάτια που ανήκουν στο ίδιο είδος μοιράζονται κοινά χαρακτηριστικά όσο αφορά τα μουσικά όργανα με τα οποία έχει γίνει η σύνθεσή τους, τους ρυθμούς, το τονικό ύψος, την ομοιότητα σε στίχους, καθώς και τις διαθέσεις που προκαλούν. Ένα παράδειγμα είναι οι συγκεκριμένες ετικέτες διαθέσεων που εφαρμόζονται σε συγκεκριμένα είδη, λόγου χάριν η ετικέτα «θυμός» αντιστοιχεί συνήθως σε punk μουσική, η ετικέτα «χαρά» αντιστοιχεί πολύ συχνά σε παιδικά τραγούδια. Ο βαρύς ήχος μια ηλεκτρικής κιθάρας με έντονη παραμόρφωση σε συνδυασμό με μια φωνή που γρυλίζει είναι μια καλή ένδειξη πως ή μουσική που ακούγεται ανήκει στην κατηγορία «μέταλ».

Η ουσία της μουσικής κατηγοριοποίησης είναι η μετάφραση των χαρακτηριστικών αυτών σε αριθμητικές αναπαραστάσεις, που μπορούν να αντιληφθούν και να επεξεργαστούν οι υπολογιστικές μηχανές. Η διαδικασία αυτή αποτελείται συνήθως από υπολογιστικά βήματα που αρχικά μετατρέπουν το ηχητικό περιεχόμενο σε αναπαραστάσεις συναρτήσεων του χρόνου ή της συχνότητας. Στη συνέχεια εξάγονται χαρακτηριστικά που διαφοροποιούνται εμφανώς το ένα από το άλλο και τέλος συγκεντρώνονται τα αποτελέσματα και

επαναλαμβάνεται η διαδικασία μέχρις ότου να μπορεί να οριστεί η κατάλληλη κατηγορία στην οποία μπορεί να ταξινομηθεί η κάθε μουσική[15]. Τα χαρακτηριστικά πρέπει να είναι περιεκτικά, δηλαδή να αναπαριστούν τη μουσική με βέλτιστο τρόπο, συμπαγή, καταλαμβάνοντας μικρό αποθηκευτικό χώρο, και αποτελεσματικά, χωρίς να απαιτούν πολλούς υπολογισμούς για την εξαγωγή τους[3]. Οι έρευνες ανάκτησης μουσικής πληροφορίας έχουν την τάση να επικεντρώνονται σε ηχητικές(mp3), συμβολικές(παρτιτούρες, MIDI) ή πολιτισμικές πηγές πληροφορίας(web data, ετικέτες χρηστών, άλλες έρευνες) για την εξαγωγή χαρακτηριστικών.

Στην περίπτωση εξαγωγής χαρακτηριστικών από ακουστικά σήματα ,τα χαρακτηριστικά συνήθως σχετίζονται με τις κύριες διαστάσεις της μουσικής, όπως η μελωδία, η αρμονία, το τονικό ύψος, ο ρυθμός και η χροιά του ήχου.

Τονικό ύψος ονομάζεται η συχνότητα που εκπέμπει ο κάθε ήχος και καθορίζει πόσο «ψηλός» ή «χαμηλός» ακούγεται αυτός. Τα χαρακτηριστικά τονικού περιεχομένου περιγράφουν πληροφορίες σχετικά με τη μελωδία και την αρμονία των μουσικών σημάτων και η ανάκτησή τους πραγματοποιείται με ποικίλες τεχνικές εντοπισμού τονικού ύψους.

Ο ρυθμός είναι η περιοδική εναλλαγή των τονισμών των ήχων και των παύσεων στο χρόνο. Τα χαρακτηριστικά ρυθμικού περιεχομένου περιγράφουν την κίνηση των μουσικών σημάτων συναρτήσει του χρόνου και περιέχουν πληροφορίες για την ομαλότητα του ρυθμού, το μπιτ, το τέμπο και τη ρυθμική αγωγή.

Το ηχόχρωμα είναι βασικό χαρακτηριστικό του ήχου και αναφέρεται στις ταλαντώσεις της θεμελιώδους συχνότητας και των αρμονικών της, που εμπεριέχονται σε ένα ηχητικό κύμα . Τα ηχοχρωματικά χαρακτηριστικά χρησιμοποιούνται για να διαφοροποιήσουν σύνολα ήχων, οι οποίοι είναι πιθανόν να έχουν όμοιο ρυθμικό και τονικό περιεχόμενο. Τα χαρακτηριστικά αυτά χρησιμοποιήθηκαν αρχικά για την αναγνώριση ομιλίας.

Η μουσική πληροφορία μπορεί να αναπαρασταθεί με μεγαλύτερη ακρίβεια και υψηλότερου επιπέδου αναπαραστάσεις. Οι συμβολικές μορφές αποτελούν ένα ισχυρό εργαλείο αναπαράστασης μουσικών κομματιών με τη χρήση υπολογιστών. Δίνουν πληροφορίες παρόμοιες με αυτές που δίνονται από μια παρτιτούρα, συμπεριλαμβάνοντας πληροφορίες για τους τόνους, τις διάρκειες, τις εντάσεις και τη χροιά για κάθε μουσικό γεγονός του κομματιού. Επίσης δίνονται και διάφορες καθολικές πληροφορίες, όπως το τέμπο και το τονικό κλειδί της σύνθεσης.

Από τα συμβολικά δεδομένα μπορούν να εξαχθούν υψηλού επιπέδου μουσικές έννοιες που δύσκολα μπορούν να αποδοθούν από τύπους δεδομένων που χρησιμοποιούνται συνήθως, όπως τα ηχητικά δεδομένα. Το μόνο αρνητικό στοιχείο είναι ότι οι ακριβείς

συμβολικές αναπαραστάσεις είναι σπάνια διαθέσιμες, ειδικά όταν πρόκειται για νέα μουσικά κομμάτια.

Τέλος, σημαντικό ρόλο διαδραματίζουν και τα χαρακτηριστικά που προκύπτουν από τη συναισθηματική διάσταση της μουσικής. Οι στίχοι αποτελούν μια πλούσια πηγή πληροφοριών για τη διάθεση ενός τραγουδιού. Πολλοί τύποι κειμενικών χαρακτηριστικών μπορούν να εξαχθούν από αυτούς, όπως η συχνότητα ενός όρου στους στίχους ενός τραγουδιού, tf-idf βάρη και Boolean μεταβλητές που υποδεικνύουν παρουσία ή απουσία ενός όρου στους στίχους, με τη βοήθεια του επιστημονικού πεδίου της γλωσσικής τεχνολογίας.

2. Βαθιά Μάθηση(Deep learning)

2.1 Τεχνητή νοημοσύνη

Η ιστορία της Τεχνητής Νοημοσύνης ξεκινά από την αρχαιότητα με μύθους και ιστορίες να παρουσιάζουν τεχνητά όντα προικισμένα με σοφία και συναισθήματα. Κατά το δεύτερο παγκόσμιο πόλεμο ο Alan Turing ανακάλυψε το γνωστό «τεστ της μίμησης», θέτοντας το μέτρο ευφυΐας μια νοήμονος μηχανής. Ο όρος «Τεχνητή Νοημοσύνη» χρησιμοποιήθηκε επίσης για πρώτη φορά το 1956 στο Dartmouth College, σε συνέδριο ερευνητών για τη μελέτη δυνατοτήτων χρήσης των υπολογιστών ως προς την προσομοίωση της ανθρώπινης νοημοσύνης. Σήμερα ο όρος είναι αρκετά διαδεδομένος και αποτελεί αναπόσπαστο κομμάτι της επιστήμης των υπολογιστών, λόγω του αυξημένου όγκου δεδομένων, των προηγμένων αλγορίθμων και των βελτιώσεων στην ισχύ των υπολογιστών και την αποθήκευση των δεδομένων.

Οι ορισμοί σχετικά με την Τεχνητή Νοημοσύνη αλλάζουν με την πάροδο του χρόνου, εξαιτίας της ραγδαίας εξέλιξης της προόδου της. Οι πιο πρόσφατοι ορισμοί αναφέρονται στη μίμηση της ανθρώπινης συμπεριφοράς και συνοψίζονται στον εξής ορισμό: Τεχνητή Νοημοσύνη είναι ο κλάδος της επιστήμης των υπολογιστών που ασχολείται με το σχεδιασμό ευφυών υπολογιστικών συστημάτων, και την εξέλιξή τους ως προς την ικανότητά τους να επιδεικνύουν διαδικασίες ανθρώπινης λογικής, όπως η μάθηση, η συλλογιστική, η επίλυση προβλημάτων, η κατανόηση φυσικής γλώσσας και η αυτοδιόρθωση.

Σημαντικοί ερευνητικοί χώροι της Τεχνητής Νοημοσύνης είναι η αναπαράσταση γνώσης, η μηχανική όραση, μηχανισμοί εξαγωγής συμπερασμάτων και επίλυσης προβλημάτων, η ρομποτική, ευφυείς υπηρεσίες διαδικτύου και σημασιολογικού ιστού και η μηχανική μάθηση.

2.2 Μηχανική μάθηση

Η μάθηση είναι ένα σύνθετο εσωτερικό βιολογικό και πνευματικό φαινόμενο που έχει μελετηθεί από διάφορους κλάδους της επιστήμης, όπως ψυχολογία, παιδαγωγική, φυσιολογία, ιατρική και βιολογία. Ως βιολογική διαδικασία, η μάθηση παρατηρείται στους ανθρώπους και τα ζώα και είναι αποτέλεσμα ατέρμονης άσκησης και επανάληψης που επηρεάζεται από νευροφυσιολογικούς μηχανισμούς. Δύο βασικές ιδιότητες που συνδέονται με την έννοια της μάθησης είναι η ικανότητα απόκτησής της κατά την αλληλεπίδραση με το περιβάλλον, και η ικανότητα βελτίωσης του τρόπου εκτέλεσης μιας ενέργειας μέσω της επανάληψης. Οι διαδικασίες της μάθησης δεν έχουν ερμηνευτεί και κατανοηθεί πλήρως με παραδεκτό τρόπο από τα πεδία όλων των επιστημών που τη μελετούν, λόγω της ποικιλομορφίας τους.

Ο άνθρωπος κατανοεί το περιβάλλον με την παρατήρησή του και τη δημιουργία μιας απλοποιημένης εκδοχής του. Η αφαιρετική αυτή εκδοχή ονομάζεται μοντέλο και η δημιουργία αυτού επαγωγική μάθηση. Επιπλέον ο άνθρωπος έχει την ικανότητα οργάνωσης και συσχέτισης των εμπειριών του, με αποτέλεσμα τη δημιουργία νέων δομών που ονομάζονται πρότυπα[16].

Η δημιουργία μοντέλων ή προτύπων από ένα σύνολο δεδομένων, από ένα υπολογιστικό σύστημα, ονομάζεται μηχανική μάθηση. Η μηχανική μάθηση έχει ως σκοπό τη δημιουργία μηχανών ικανών να μαθαίνουν και να βελτιώνουν την απόδοση τους κατά την εκτέλεση συγκεκριμένων εργασιών, μέσω της αξιοποίησης προηγούμενης γνώσης και εμπειρίας, χωρίς να υπάρχει ανάγκη για εκ νέου προγραμματισμό τους.

Έχουν διατυπωθεί πολλοί ορισμοί για τη μηχανική μάθηση. Ο Herbert A. Simon το 1983 διατύπωσε πως η μάθηση σηματοδοτεί προσαρμοστικές αλλαγές σε ένα σύστημα, με την έννοια ότι αυτές του επιτρέπουν να κάνει την ίδια εργασία, ή άλλες εργασίες της ίδιας κατηγορίας, πιο αποδοτικά και αποτελεσματικά την επόμενη φορά. Ο Tom Mitchell το 2006 έδωσε τον εξής ορισμό για τη μηχανική μάθηση: « Ένα πρόγραμμα υπολογιστή λέμε ότι μαθαίνει από την εμπειρία E ως προς κάποια κλάση εργασιών T και μέτρο απόδοσης P , αν η απόδοσή του σε εργασίες από το T , όπως μετρείται από το P , βελτιώνεται μέσω της εμπειρίας E ». Τέλος σύμφωνα με τους Witten και Frank(2011) κάτι μαθαίνει όταν αλλάζει την συμπεριφορά του κατά τέτοιο τρόπο ώστε να αποδίδει καλύτερα στο μέλλον.

Στη μηχανική μάθηση αναπτύσσονται τρεις βασικές κατηγορίες μάθησης, ανάλογες με τους τρόπους μάθησης των ανθρώπων: η επιβλεπόμενη μάθηση, η μη- επιβλεπόμενη μάθηση και η ενισχυτική μάθηση.

Στην επιβλεπόμενη μάθηση ή μάθηση με παραδείγματα το σύστημα μαθαίνει επαγωγικά μια συνάρτηση που ονομάζεται συνάρτηση στόχος και αποτελεί έκφραση του μοντέλου που περιγράφει τα δεδομένα. Η συνάρτηση στόχος χρησιμοποιείται για την πρόβλεψη των τιμών μεταβλητών εξόδου(εξαρτημένες μεταβλητές), βάσει των τιμών των μεταβλητών εισόδου(ανεξάρτητες μεταβλητές). Χρησιμοποιείται σε προβλήματα ταξινόμησης και παρεμβολής.

Στη μη-επιβλεπόμενη μάθηση η τιμή της συνάρτησης στόχου των παραδειγμάτων δεν είναι γνωστή. Το σύστημα κατασκευάζει ένα μοντέλο για κάποιο σύνολο εισόδων υπό μορφή παρατηρήσεων χωρίς να γνωρίζει τις επιθυμητές εξόδους. Εφόσον το σύνολο εισόδου δεν είναι επισημασμένο δεν υπάρχει σφάλμα ή σήμα ανταμοιβής για την αξιολόγηση των πιθανών αποτελεσμάτων και η απόδοση μετριέται υποκειμενικά. Χρησιμοποιείται σε προβλήματα ανάλυσης συσχετισμών και ομαδοποίησης.

Στην ενισχυτική μάθηση το σύστημα μαθαίνει μια στρατηγική ενεργειών μέσα από άμεση αλληλεπίδραση με το περιβάλλον. Χρησιμοποιείται κυρίως σε προβλήματα σχεδιασμού, όπως για παράδειγμα ο έλεγχος κίνησης ρομπότ και η βελτιστοποίηση εργασιών σε εργοστασιακούς χώρους.

Πολλά μοντέλα μηχανικής μάθησης που έχουν σχεδιαστεί βρίσκουν εφαρμογή σε πολλά επιστημονικά πεδία που απαιτούν υπολογιστικές εργασίες, όπου ο σχεδιασμός και ο προγραμματισμός ρητών αλγορίθμων με ικανοποιητική απόδοση είναι δύσκολος ή ακόμη και αδύνατος. Σημαντικές κατηγορίες συστημάτων μηχανικής μάθησης είναι: τα δέντρα αποφάσεων, οι μηχανές διανυσμάτων υποστήριξης, τα δίκτυα Bayes, οι γενετικοί αλγόριθμοι, ο επαγωγικός λογικός προγραμματισμός, τα συστήματα συσταδοποίησης, η αναπαραστατική μάθηση και τα τεχνητά νευρωνικά δίκτυα.

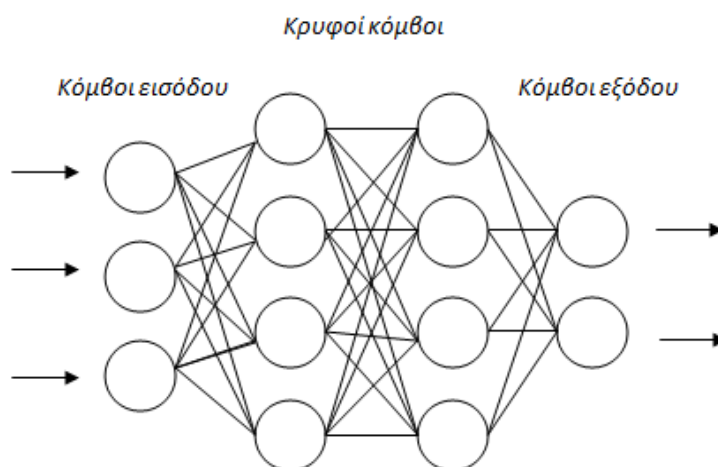
2.3 Βαθιά μηχανική μάθηση και Νευρωνικά δίκτυα

Η βαθιά μάθηση(deep learning), γνωστή και ως βαθιά δομημένη μάθηση(deep structured learning) ή ιεραρχική μάθηση(hierarchical learning) είναι μέρος μια ευρύτερης κατηγορίας

των μεθόδων μηχανικής μάθησης που μυούν τους υπολογιστές στη φυσική για τους ανθρώπους διαδικασία της μάθησης με παραδείγματα. Τα μοντέλα εκπαιδεύονται με τη χρήση μεγάλων επισημασμένων συνόλων δεδομένων και αρχιτεκτονικές νευρωνικών δικτύων που «μαθαίνουν» τα χαρακτηριστικά των δεδομένων αυτόματα, χωρίς τη δημιουργία ανάγκης της χειροκίνητης εξαγωγής τους. Οι αλγόριθμοι βαθιάς μάθησης χρησιμοποιούν πολλαπλά επίπεδα για να εξάγουν σταδιακά χαρακτηριστικά υψηλότερου επιπέδου από μια ακατέργαστη είσοδο. Ο όρος «βαθιά» αναφέρεται στα κρυμμένα αυτά επίπεδα.

Τα τεχνητά νευρωνικά δίκτυα είναι συστήματα επεξεργασίας δεδομένων που αποτελούνται από ένα πλήθος απλών μονάδων επεξεργασίας, τους τεχνητούς νευρώνες, οι οποίοι είναι οργανωμένοι σε δομές παρόμοιες με αυτές τους ανθρώπινου εγκεφάλου. Σε αναλογία με το βιολογικό νευρώνα, η βασική δομική μονάδα του τεχνητού νευρωνικού δικτύου είναι ο τεχνητός νευρώνας, που συχνά αναφέρεται ως κόμβος.

Μπορούμε να ξεχωρίσουμε τρεις τύπους κόμβων στα νευρωνικά συστήματα: τους κόμβους εισόδου, τους κόμβους εξόδου και τους κρυμμένους κόμβους (**Εικόνα 2.**). Οι κόμβοι εισόδου λαμβάνουν δεδομένα από το εξωτερικό περιβάλλον του νευρωνικού δικτύου και αποτελούν το επίπεδο εισόδου του νευρωνικού δικτύου. Οι κόμβοι εξόδου εξάγουν τα δεδομένα από το νευρωνικό δίκτυο και απαρτίζουν το επίπεδο εξόδου. Τέλος, οι είσοδοι και οι έξοδοι των κρυμμένων κόμβων παραμένουν μέσα στο δίκτυο. Οι κρυμμένοι κόμβοι αποτελούν τα κρυμμένα(εσωτερικά)επίπεδα του νευρωνικού δικτύου(μπορεί να είναι ένα ή περισσότερα).



Εικόνα 2. Είδη κόμβων

Ο νευρώνας λοιπόν, δέχεται εισόδους από το περιβάλλον ή από άλλους νευρώνες και υπολογίζει μια έξοδο την οποία μεταδίδει στους άλλους νευρώνες με τους οποίους συνδέεται. Το σύστημα των νευρώνων πραγματοποιεί εγγενώς παράλληλες διεργασίες, καθώς πολλοί κόμβοι διεξάγουν ταυτόχρονα τους υπολογισμούς τους(Εικόνα 3).

Μια δεύτερη διαδικασία που εκτελείται στο νευρώνα είναι η προσαρμογή των βαρών, ανάλογα με την τοπολογία του δικτύου και τους κανόνες εκπαίδευσης. Κάθε σύνδεση ή αλλιώς σύναψη χαρακτηρίζεται από ένα βάρος w , η τιμή του οποίου καθορίζει το βαθμό αλληλεπίδρασης κάθε ζεύγους νευρώνων. Η συνεισφορά του σήματος είναι ανάλογη με το πόσο μικρό ή μεγάλο είναι το βάρος w . Ένα θετικό βάρος συνεπάγεται με τη διέγερση ενός νευρώνα, ενώ ένα αρνητικό με την αναστολή του. Ο τεχνητός νευρώνας έχει μια εξωτερική σύνδεση b^k σταθερού βάρους w^0 που ονομάζεται πόλωση και η τιμή του οποίου είναι πάντα 1. Το σταθερό βάρος καθορίζει την αύξηση ή τη μείωση της τιμής της εισόδου του δικτύου στη συνάρτηση ενεργοποίησης, ανάλογα με το αν είναι θετικό ή αρνητικό. Κάθε σήμα x_i εισόδου πολλαπλασιάζεται με το αντίστοιχο βάρος. Στη συνέχεια όλα τα γινόμενα αθροίζονται μαζί με το βάρος w^0 (πόλωση) και το άθροισμα αυτό αποτελεί τη συνολική είσοδο του νευρώνα.

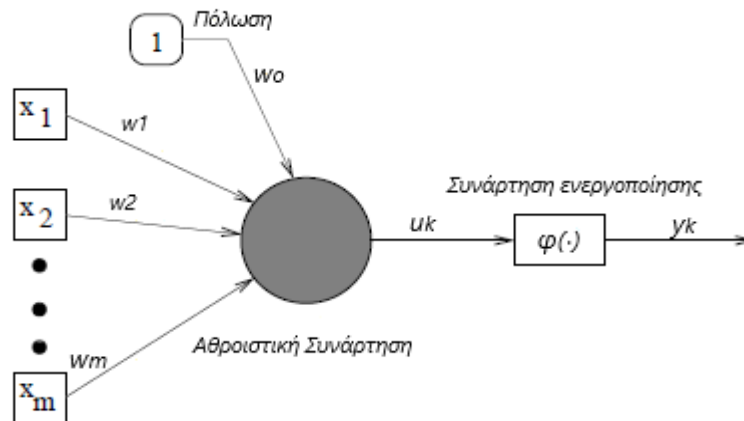
Η συνολική είσοδος μετασχηματίζεται με τη συνάρτηση ενεργοποίησης $\phi(u)$, από την οποία προκύπτει η τελική έξοδος του νευρώνα που θα διαδοθεί στο νευρωνικό δίκτυο προς την έξοδο.

Με μαθηματικούς όρους μπορούμε να περιγράψουμε ένα νευρώνα από το παρακάτω ζεύγος εξισώσεων:

$$u_k = \sum_{i=1}^m x^i * w^{ki} + w^0$$

$$y_k = \phi(u_k)$$

όπου x^i οι m είσοδοι του νευρώνα, w^{ki} τα βάρη των συνάψεων του νευρώνα k , w^0 η πόλωση του νευρώνα, u η συνολική είσοδος του νευρώνα και $\phi()$ η συνάρτηση ενεργοποίησης.



Εικόνα 3. Μοντέλο νευρώνα

2.4 Συναρτήσεις ενεργοποίησης

Τρεις βασικοί τύποι συναρτήσεων είναι η συνάρτηση με κατώφλι (Threshold function), η τμηματική γραμμική συνάρτηση (Piecewise-Linear function) και η σιγμοειδής συνάρτηση ενεργοποίησης (**Εικόνα 4**).

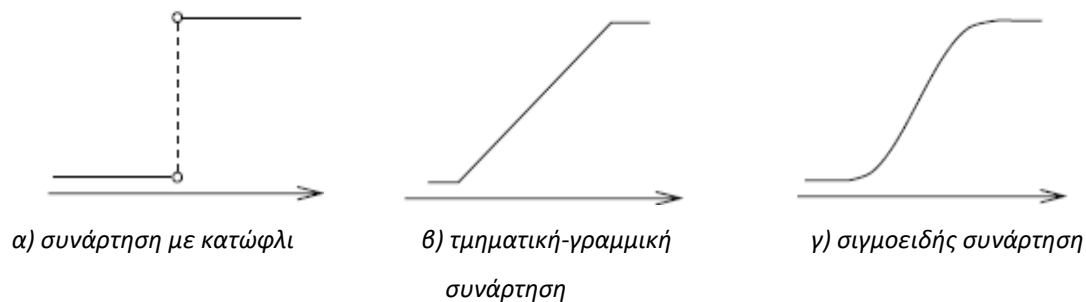
Στη συνάρτηση με κατώφλι η έξοδος του νευρώνα παίρνει την τιμή 1 εάν η δυνατότητα ενεργοποίησης του νευρώνα είναι θετική, διαφορετικά παίρνει την τιμή 0. Η συνάρτηση αυτή χρησιμοποιείται για τη δημιουργία νευρώνων που χωρίζουν τις εισόδους σε δυο διαφορετικές κατηγορίες.

Η τμηματική γραμμική συνάρτηση η έξοδος παίρνει τις τιμές: 1 αν το $u \geq \frac{1}{2}$, u αν $\frac{1}{2} > u > -\frac{1}{2}$ και 0 αν $u \leq -\frac{1}{2}$.

Ως επί το πλείστον για την κατασκευή νευρωνικών δικτύων χρησιμοποιείται η σιγμοειδής συνάρτηση ενεργοποίησης. Σε αντίθεση με τη συνάρτηση με κατώφλι που μπορεί να πάρει τις τιμές 0 ή 1, η σιγμοειδής συνάρτηση μπορεί να πάρει όλες τις τιμές από το [0,1].

$$\varphi(u) = \frac{1}{1 + \exp(-au)}$$

όπου α, η παράμετρος κλίσης της σιγμοειδούς συνάρτησης.



Εικόνα 4. Οι τρεις βασικοί τύποι συναρτήσεων ενεργοποίησης(από[30])

2.5 Τοπολογίες νευρωνικών δικτύων

Η τοπολογία ενός νευρωνικού δικτύου ορίζεται με βάση τον τρόπο σύνδεσης των κόμβων του και τη διάδοση των δεδομένων. Οι δύο κύριες κατηγορίες τοπολογιών είναι τα δίκτυα εμπρόσθιας τροφοδότησης(Feed-forward networks) και τα αναδρομικά δίκτυα(Reccurent networks).

Τα δίκτυα εμπρόσθιας τροφοδότησης χωρίζονται σε δύο κατηγορίες: Τα δίκτυα εμπρόσθιας τροφοδότησης ενός επιπέδου και τα δίκτυα εμπρόσθιας τροφοδότησης πολλαπλών επιπέδων.

Στα δίκτυα ενός επιπέδου υπάρχει ένα επίπεδο με κόμβους εισόδου και ένα επίπεδο με κόμβους εξόδου. Με τον όρο «ένα επίπεδο» εννοείται το επίπεδο εξόδου, διότι σε αυτό πραγματοποιούνται υπολογισμοί. Στα δίκτυα πολλαπλών επιπέδων υπάρχει ένα ή περισσότερα κρυφά επίπεδα. Τα σήματα εξόδου των νευρώνων κάθε κρυφού επιπέδου γίνονται σήματα εισόδου για τους νευρώνες του επόμενου επιπέδου.

Και στις δύο περιπτώσεις η ροή των δεδομένων διέρχεται από τους κόμβους εισόδου με κατεύθυνση τους κόμβους εξόδου. Κάθε κόμβος ενός επιπέδου μπορεί να συνδεθεί με κάθε κόμβο του επόμενου επιπέδου. Σε ένα πλήρως διασυνδεδεμένο δίκτυο εμπρόσθιας τροφοδότησης ο κάθε κόμβος ενός επιπέδου συνδέεται με όλους τους κόμβους στο επόμενο επίπεδο. Δεν υπάρχει καμία σύνδεση μεταξύ των νευρώνων του ίδιου επιπέδου και καμία ανάδραση μεταξύ των επιπέδων του δικτύου.

Τα αναδρομικά νευρωνικά δίκτυα είναι αυτά που η δομή τους μοιάζει περισσότερο με τη λειτουργία του ανθρώπινου εγκεφάλου, όπου οι διεργασίες της αντίληψης, της λήψης αποφάσεων και της εκμάθησης είναι μη γραμμικές. Το κύριο χαρακτηριστικό τους που τα διαφοροποιεί από τα δίκτυα εμπρόσθιας τροφοδότησης είναι ότι περιέχουν τουλάχιστον μια ανάδραση ανάμεσα στους κόμβους του ίδιου επιπέδου ή κόμβους διαφορετικών επιπέδων. Άλλο βασικό χαρακτηριστικό των αναδρομικών δικτύων είναι η αντίληψη του χρόνου, για παράδειγμα σε εφαρμογές με δυναμικά συστήματα μπορούν να χρησιμοποιηθούν για να προβλέψουν την έξοδο της στιγμής $t+1$ έχοντας δεδομένα της στιγμής t ή και προηγούμενων. Κάθε χρονική στιγμή οι κόμβοι έχουν μια τιμή ενεργοποίησης. Τέλος, σημαντικό χαρακτηριστικό είναι η εισαγωγή της μνήμης στα νευρωνικά δίκτυα. Τα αναδρομικά δίκτυα κατηγοριοποιούνται σε πλήρως αναδρομικά δίκτυα, στα οποία επιτρέπεται η σύνδεση όλων των νευρώνων μεταξύ τους και στα τοπικά αναδρομικά δίκτυα.

2.6 Στάδια λειτουργίας νευρωνικών δικτύων

Η λειτουργία των νευρωνικών δικτύων χωρίζεται σε δύο στάδια: τη μάθηση ή εκπαίδευση(learning) και την ανάκληση(recall). Μάθηση είναι η διαδικασία της προσαρμογής των τιμών των βαρών του δικτύου με τη χρήση κανόνων ή αλλιώς αλγορίθμων εκμάθησης, με σκοπό δοθέντος ενός συγκεκριμένου διανύσματος εισόδου στο δίκτυο, να παράγεται ανάλογο διάνυσμα εξόδου. Κατά τη διαδικασία της εκπαίδευσης οι συναπτικές συνδέσεις μεταξύ των νευρώνων ενισχύονται ή αποδυναμώνονται ώστε συνολικά το δίκτυο να επιτύχει το στόχο της επίλυσης του προβλήματος. Το είδος της μάθησης καθορίζεται με βάση τον τρόπο που πραγματοποιούνται οι αλλαγές των παραμέτρων. Όπως αναφέρθηκε και στο κεφάλαιο 2.2 υπάρχουν τρεις βασικές κατηγορίες μάθησης: η επιβλεπόμενη μάθηση, η μάθηση χωρίς επίβλεψη και η ενισχυτική μάθηση.

Όταν το δίκτυο φτάσει στην επιθυμητή απόδοση το στάδιο εκμάθησης τερματίζεται και ξεκινά η διαδικασία της ανάκλησης. Ανάκληση ονομάζεται η διαδικασία υπολογισμού ενός διανύσματος εξόδου, για συγκεκριμένο διάνυσμα εισόδου και τιμές βαρών.

3. Προσπάθειες κατηγοριοποίησης της μουσικής

3.1 Πειράματα κατηγοριοποίησης μουσικής σε είδη

Πρωτοπόροι στην κατηγοριοποίηση είδους μουσικής είναι οι Tzanetakis και Cook(2002), οι οποίοι χρησιμοποίησαν τρεις διαφορετικές συλλογές δεδομένων - υποενότητες της βάσης δεδομένων GZTAN που δημιούργησαν οι ίδιοι. Εφάρμοσαν μια ταξινόμηση λόγου που αναγνωρίζει ανδρική φωνή, γυναικεία φωνή καθώς και ομιλία από αθλητικές εκπομπές. Για κάθε μουσική κατηγορία χρησιμοποιήθηκαν για εκπαίδευση 100 αντιπροσωπευτικά δείγματα ακουστικών αρχείων, διάρκειας 30 δευτερολέπτων έκαστο, οποία ανακτήθηκαν από ραδιοφωνικούς σταθμούς, CD και συμπιεσμένα MP3. Τα δείγματα αποθηκεύτηκαν σε αρχεία 22 050 Hz -16-bit- mono. Τα μουσικά είδη διακρίθηκαν στις εξής κατηγορίες.

Η γενική κατηγορία 'Genres dataset' αποτελείται από 10 κλάσεις: κλασσική, country, disco, hip hop, jazz, rock blues, reggae, pop και metal. Η υποκατηγορία 'Classical dataset' αποτελείται από 4 κλάσεις: χορωδία, ορχήστρα, κουαρτέτο, πιάνο, κουαρτέτο εγχόρδων. Τέλος η υποκατηγορία 'Jazz dataset' αποτελείται από 6 κλάσεις: big band, cool, fusion, piano, quartet και swing.

Το 30διάστατο διάνυσμα χαρακτηριστικών απαρτίζεται από ένα συνδυασμό χαρακτηριστικών: 19 χαρακτηριστικά με βάση το ηχόχρωμα[9 FFT και 10 MFCC], 6 χαρακτηριστικά ρυθμικού περιεχομένου και 5 χαρακτηριστικά τονικού ύψους.

Μετά από δοκιμές όλων των πιθανών συνδυασμών εξήχθησαν τα εξής συμπεράσματα: Τα χαρακτηριστικά τα οποία δε βασίζονται στο ηχόχρωμα όπως το ιστόγραμμα τονικού ύψους και το ιστόγραμμα ρυθμού απέδωσαν χειρότερα από τα χαρακτηριστικά ηχοχρώματος (FFT, MFCC) σε όλες τις περιπτώσεις. Τα χαρακτηριστικά ρυθμικού και τονικού περιεχομένου διαδραματίζουν λιγότερο σημαντικό ρόλο στην ταξινόμηση 'Jazz και classical dataset' σε σύγκριση με το 'Genre dataset'.

Για λόγους ταξινόμησης χρησιμοποιήθηκε ένας αριθμός από πρότυπα στατιστικά δείγματα ταξινομητών αναγνώρισης (SPR). Οι πιο σημαντικοί από τους οποίους εξήχθησαν αξιοσημείωτα συμπεράσματα είναι: ο ταξινομητής simple Gaussian (GS), ο ταξινομητής Gaussian mixture model (GMM) (K), όπου K συγκεκριμένος αριθμός από τις πολυδιάστατες Γκαουσιανές κατανομές και ο αλγόριθμος K-means όπου το K καθορίζει τον αριθμό των

πλησιέστερων γειτόνων. Ο **πίνακας 1** απεικονίζει τα πιο σημαντικά αποτελέσματα από κάθε σύνολο δεδομένων για τη μέση ακρίβεια και την τυπική απόκλιση.

	Genres(10)	Classical (4)	Jazz (6)
GS	59% ± 4	77% ± 6	61% ± 8
GMM(2)	60% ± 4	81% ± 5	66% ± 7
GMM(3)	61% ± 4	88% ± 4	68% ± 7
GMM(5)	61% ± 4	88% ± 5	59% ± 6
KNN(3)	60% ± 4	78% ± 6	58% ± 7
KNN(5)	56% ± 3	78% ± 6	56% ± 6

Πίνακας 1. Σημαντικότερα αποτελέσματα κάθε συνόλου δεδομένων.

Έκτοτε, οι ερευνητές έχουν αφοσιωθεί στην αύξηση του αριθμού των χαρακτηριστικών.

Το μικρό έργο που έχει παραχθεί όσο αφορά την αυτόματη ταξινόμηση μουσικών ειδών, καθώς και οι σχετικά χαμηλές αναφερόμενες τιμές της ακρίβειας οδήγησαν σε νέες μεθόδους εξαγωγής χαρακτηριστικών.

Τα ιστογράμματα συντελεστών κυματιδίων Daubechies (DWCHs) που εισήγαγαν οι Tao Li, Mitsunori Ogihara and Qi Li το 2003, παράλληλα με τη χρήση μηχανών διανυσμάτων υποστήριξης (SVMs) στα πειράματα κατηγοριοποίησης, βελτίωσαν την ακρίβεια της κατηγοριοποίησης μουσικών ειδών. Τα DWCHs είναι βασισμένα σε ιστογράμματα κυματιδίων με σκοπό να συλλέξουν ταυτόχρονα την προέλευση και πληροφορίες για τα μουσικά σήματα. Αποτελούνται από 28 σύνολα, 4 για κάθε ένα από 7 εύρη συχνότητας.

Για τα πειράματά τους χρησιμοποίησαν δυο σύνολα δεδομένων. Το πρώτο σύνολο A από τη βάση δεδομένων GZTAN των Tzanetakis και Cook, η οποία όπως αναφέρθηκε παραπάνω περιέχει 10 ταξινομημένα είδη μουσικής με 100 αποσπάσματα για κάθε ένα από αυτά. Το

σύνολο B αποτελείται από 756 ηχητικά αρχεία, τα οποία διαχωρίζονται σε 5 είδη. Το σύνολο δεδομένων κατασκευάστηκε από τους συγγραφείς από τη συλλογή CD με 189 μουσικά άλμπουμ του Mitsunori Ogiwara.

Ακολουθείται το ίδιο μοτίβο με τη συλλογή GZTAN με ηχητικά αρχεία διάρκειας 30 δευτερολέπτων που αποθηκεύονται σε αρχεία 22 050 Hz – 16-bit - mono. Η κατανομή των ειδών μουσικής είναι η εξής: Ατμοσφαιρική (109), Κλασσική (164), Fusion (136), Jazz (251), και Rock (96).

Για τη δημιουργία του πειράματος χρησιμοποίησαν κάθε χαρακτηριστικό ξεχωριστά και έπειτα κάθε πιθανό συνδυασμό αυτών, με σκοπό να συγκρίνουν την ακρίβεια που εξάγεται από κάθε πιθανό συνδυασμό με την ακρίβεια που εξάγεται από τα προτεινόμενα DWCHs. Για μεθόδους κατηγοριοποίησης χρησιμοποίησαν τους αλγορίθμους: GMM (Gaussian Mixture Models) με 3 Γκαουσιανούς συνδυασμούς για τη μοντελοποίηση κάθε είδους, τον αλγόριθμο k-NN πλησιέστερων γειτόνων, όπου k=5, τον LDA (Linear Discriminant Analysis) και τμηματικές επεκτάσεις των μηχανών διανυσμάτων υποστήριξης (SVM).

Στον **πίνακα 2** εμφανίζεται η ακρίβεια για κάθε πιθανό συνδυασμό και έπειτα το συνδυασμό των DWCHs με το μετασχηματισμό Fourier και τους συντελεστές συχνότητας Cepstrum. Οι τιμές της ακρίβειας είναι υπολογισμένες με διασταυρούμενη επικύρωση 10 τμημάτων (ten-fold cross-validation). Οι αριθμοί στις παρενθέσεις δηλώνουν την τυπική απόκλιση.

Τα αποτελέσματα επιβεβαιώνουν ότι κάθε ένα από τα τέσσερα σύνολα χαρακτηριστικών περιέχει χρήσιμες πληροφορίες για τον χαρακτηρισμό των μουσικών σημάτων. Επίσης χρήσιμα συμπεράσματα προέκυψαν από τις μίξεις συνδυασμών ρυθμικών, ηχοχρωματικών και τονικών χαρακτηριστικών. Ο συνδυασμός FFT με MFCC λειτουργεί καλύτερα από οποιονδήποτε συνδυασμό με δύο χαρακτηριστικά. Μόνο όταν ένα τέταρτο χαρακτηριστικό προστίθεται στον παραπάνω συνδυασμό τα αποτελέσματα γίνονται καλύτερα – με τη χρήση των DWCHs έχει επιτευχθεί ακρίβεια 78.5%. Αυτή είναι μια αξιοσημείωτη βελτίωση σχετικά με την ακρίβεια 61% που προέκυψε από τους Tzanetakis και Cook[4]. Στον **πίνακα 2** φαίνονται τα αποτελέσματα από τις μεθόδους που χρησιμοποιήθηκαν στο σύνολο δεδομένων A, με κάθε δυνατό συνδυασμό των χαρακτηριστικών.

Χαρακτηριστικά	Μέθοδοι					
	SVM1	SVM2	MPSVM	GMM	LDA	KNN
DWCHs	74.9 (4.97)	78.5 (4.07)	68.3 (4.34)	63.5 (4.72)	71.3 (6.10)	62.1 (4.54)
Beat+FFT+MFCC+Pitch	70.8 (5.39)	71.9 (5.09)	66.2 (5.23)	61.4 (3.87)	69.4 (6.93)	61.3 (4.85)
Beat+FFT+MFCC	71.2 (4.98)	72.1 (4.68)	64.6 (4.16)	60.8 (3.25)	70.2 (6.61)	62.3 (4.03)
Beat+FFT+Pitch	65.1 (4.27)	67.2 (3.79)	56.0 (4.67)	53.3 (3.82)	61.1 (6.53)	51.8 (2.94)
Beat+MFCC+Pitch	64.3 (4.24)	63.7 (4.27)	47.8 (3.82)	50.4 (2.22)	61.7 (7.23)	54.0 (3.30)
FFT+MFCC+Pitch	70.9 (6.22)	72.2 (3.90)	64.9 (5.06)	59.6 (2.22)	69.9 (6.76)	61.0 (5.40)
Beat+FFT	61.7 (5.12)	62.6 (4.83)	50.8 (5.16)	48.3 (2.82)	56.0 (6.73)	48.8 (5.07)
Beat+MFCC	60.4 (3.19)	60.2 (4.84)	53.5 (4.45)	47.7 (2.24)	59.6 (4.03)	50.5 (4.53)
Beat+Pitch	42.7 (5.37)	41.1 (4.68)	35.6 (4.27)	34.0 (2.69)	36.9 (4.38)	35.7 (3.59)
FFT+MFCC	70.5 (5.98)	71.8 (4.83)	63.6 (4.71)	59.1 (3.20)	66.8 (6.77)	62.1 (7.12)
FFT+Pitch	64.0 (5.16)	68.2 (3.79)	55.1 (5.82)	53.7 (3.15)	60.0 (6.68)	53.8 (4.73)
MFCC+Pitch	60.6 (4.54)	64.4 (4.37)	53.3 (2.95)	48.2 (2.71)	59.4 (4.50)	54.7 (3.50)
Beat	26.5 (3.30)	21.5 (2.71)	22.1 (3.04)	22.1 (1.91)	24.9 (2.99)	22.8 (5.12)
FFT	61.2 (6.74)	61.8 (3.39)	50.6 (5.76)	47.9 (4.91)	56.5 (6.9)	52.6 (3.81)
MFCC	58.4 (3.31)	58.1 (4.72)	49.4 (2.27)	46.4 (3.09)	55.5 (3.57)	53.7 (4.11)
Pitch	36.6 (2.95)	33.6 (3.23)	29.9 (3.76)	25.8 (3.02)	30.7 (2.79)	33.3 (3.20)

Πίνακας 2. Αποτελέσματα μεθόδων συνόλου δεδομένων Α

Στο δεύτερο σύνολο δεδομένων η ακρίβεια ήταν 74.2%, ξανά από τις SVMs. Τα ίδια αποτελέσματα προκύπτουν και από το σύνολο δεδομένων A, με την ίδια επιλογή χαρακτηριστικών. Η βάση αυτή δημιουργήθηκε με μικρό χειρισμό και υπήρξε σύγχυση μεταξύ των ειδών. Για παράδειγμα το είδος «Ατμοσφαιρική μουσική» είναι πολύ ασαφές και μπορεί να καλύψει μουσικό εύρος από την κλασσική μουσική αλλά και από τη τζαζ.

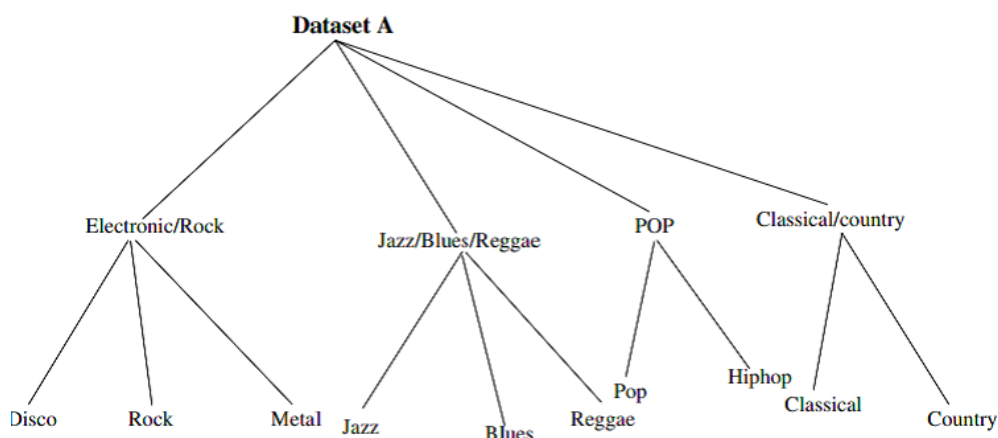
Η κατηγοριοποίηση της μουσικής σε είδη συνέχισε να εξελίσσεται με την είσοδο νέων χαρακτηριστικών και προσεγγίσεων. Η χρήση των MIDI μαζί με το χαρακτηριστικό audio είναι ένας από τους παραπάνω τρόπους [6]. Γενικά το ακουστικό σήμα μπορεί να αναπαρασταθεί με δύο τρόπους. Ο πρώτος είναι η συμβολική αναπαράσταση βασισμένη σε μουσικές παρτιτούρες. Παραδείγματα αυτής της αναπαράστασης είναι το MIDI και το humdrum -σύνολο εργαλείων για την ανάλυση της μουσικής- στα οποία για κάθε νότα, το τονικό ύψος, η διάρκεια, ο χρόνος έναρξης/λήξης και οι δυναμικές είναι αποθηκευμένα στο αρχείο. Ο δεύτερος τρόπος είναι βασισμένος στα ακουστικά σήματα, καταγράφοντας την ένταση -ενεργητικότητα του ακουστικού σήματος ως συνάρτηση του χρόνου κάνοντας δειγματοληψία σε συγκεκριμένη συχνότητα. Μπορεί να είναι σε συμπιεσμένη ή σε ασυμπιεστη μορφή, όπως χρησιμοποιείται από τα συστήματα εξαγωγής ηχητικών χαρακτηριστικών. Χρησιμοποιήθηκαν οι ταξινομητές NCD (normalized compression distance), ο οποίος χρησιμοποιήθηκε για μέτρο ομοιότητας μεταξύ αρχείων αναπαράστασης μουσικής και ο αλγόριθμος k-πλησιέστερων γειτόνων με $k=10$. Εντούτοις, η ανάγκη για την ύπαρξη σωστών MIDI αρχείων, όπου το καθένα θα ανταποκρίνεται στο αυθεντικό ακουστικό αρχείο περιπλέκει την κατάσταση και μέχρι τώρα δεν είναι διαθέσιμο κάποιο αξιολογούμενο αποτέλεσμα.

Οι Thomas Lidy και Andreas Rauber [1] παρατηρώντας ότι τα ρυθμικά χαρακτηριστικά έχουν χαμηλή απόδοση για την κατηγοριοποίηση σε μουσικά είδη, πρότειναν δύο νέα σύνολα χαρακτηριστικών: το Statistical Spectrum Descriptor και το Rhythm Histogram Features. Η αξιολόγηση των συνόλων αυτών έγινε με τη χρήση τριών διαφορετικών ακουστικών συλλογών, με σκοπό να αποσπαστούν πληροφορίες από διαφορετικά μουσικά είδη και να εντοπιστούν προβλήματα που αφορούν συγκεκριμένους τύπους ήχου. Οι συλλογές που χρησιμοποιήθηκαν είναι: η GTZAN και δύο από τις συλλογές που χρησιμοποιήθηκαν στο διαγωνισμό ρυθμικής κατηγοριοποίησης ISMIR 2004. Η πρώτη περιέχει 698 αποσπάσματα από 8 είδη μουσικής για χορούς ζευγαριών και η δεύτερη 1458 ολόκληρα τραγούδια ανισομερώς κατανεμημένα σε 6 είδη. Τα πειράματα διεξήχθησαν με ταξινομητές SVMs με διασταυρούμενη επαλήθευση 10 τμημάτων. Τα αποτελέσματα από τη GTZAN έδειξαν

74.9% ακρίβεια, όπου αποτελεί βελτίωση 16.4% σε σχέση με την περιγραφή χαρακτηριστικών ρυθμικών μοτίβων που χρησιμοποιήθηκε στα πειράματα Tzanetakis και Cook. Στις επόμενες δύο συλλογές τα αποτελέσματα ήταν παρεμφερή.

Μια διαφορετική προσέγγιση είναι η εφαρμογή κατηγοριοποίησης ειδών με ταξινομίες. Οι Tao Li και Mitsunori Ogiwara [3] αποδεικνύουν ότι η χρήση ταξινομιών βελτιώνει τη γενική απόδοση κατηγοριοποίησης. Οι δομές ταξινομιών αναγνωρίζουν τη σχέση εξάρτησης μεταξύ των ειδών και αναπτύσσουν τη χρηστικότητα αναζητώντας το βαθμό επιτυχίας και της ικανοποίησης των χρηστών. Χρησιμοποιώντας ταξινομίες επιτρέπεται να γίνονται περισσότερο αποδεκτά τα λάθη κατηγοριοποίησης σε σχέση με την περίπτωση της απλής ταξινόμησης. Για τη δημιουργία του πειράματος χρησιμοποιήθηκαν η βάση δεδομένων GTZAN και ένα δεύτερο σύνολο δεδομένων που δημιουργήθηκε από τους ερευνητές για τη διεκπεραίωση του πειράματος, το οποίο περιέχει 756 τραγούδια από 5 είδη.

Η χρήση ταξινομιών βασίζεται σε μια προσέγγιση επιπέδων, όπου τα είδη διατάσσονται σε ένα δέντρο ιεραρχίας δυο επιπέδων και ο ταξινομητής εκπαιδεύεται για κάθε εσωτερικό κόμβο. Στο πρώτο επίπεδο ταξινόμησης πραγματοποιείται διαχωρισμός ανάμεσα στις ομάδες που έχουν περιγραφεί στο αρχικό επίπεδο. Στο δεύτερο επίπεδο δημιουργούνται ταξινομητές για κάθε ομάδα κλάσεων του πρώτου επιπέδου. Το σύνολο χαρακτηριστικών που χρησιμοποιήθηκαν είναι ένας συνδυασμός του FFT, των MFCC και των DWCH. Οι SVMs χρησιμοποιήθηκαν για το σκοπό της ταξινόμησης. Τα αποτελέσματα απέδωσαν 72% για την απλή ταξινόμηση, 72.7 για την ιεραρχική ταξινόμηση και 75% για το πρώτο επίπεδο ομάδων. Παρατηρήθηκε μικρή βελτίωση στην ακρίβεια. Στην εικόνα φαίνεται η δομή της ταξινομίας για τη βάση GTZAN.



Εικόνα 5. Δομή ταξινομίας για τη βάση GTZAN (από [3])

Οι James Bergstra κ.α (2006) παρουσίασαν έναν αλγόριθμο που πραγματοποιεί διαχωρισμό κομματιών ανάλογα το μουσικό είδος και τον καλλιτέχνη. Για την επιλογή ακουστικών χαρακτηριστικών χρησιμοποιήθηκε η μέθοδος μηχανικής μάθησης ADABOOST. Ο αλγόριθμος απέδωσε ακρίβεια 83% στη βάση GZTAN.

Οι ερευνητές Panagakis και Cotropoulos[21] πρότειναν ένα ισχυρό πλαίσιο ταξινόμησης σε μουσικά είδη το οποίο σε αρχικό στάδιο μειώνει την διαστατικότητα της απεικόνισης του εγκεφάλου κατά τη διάρκεια ακουστικών ερεθισμάτων. Έπειτα, μοντελοποιεί κάθε αναπαράσταση σε ένα άθροισμα συντελεστών ταξινομημένων ατόμων, από τα οποία προκύπτουν οι εκπαιδευμένες απεικονίσεις των γνωστών μουσικών ειδών, αν υποθέσουμε πως οι απεικονίσεις των μουσικών εγγραφών του ίδιου είδους μοιάζουν αρκετά. Η ταξινόμηση σε είδη μουσικής στη βάση δεδομένων GZTAN ήταν εντυπωσιακά υψηλή αγγίζοντας ακρίβεια 93.7%.

Οι Scaringella, Zoia, και Mlynek[14] ανέφεραν μια προσέγγιση μη επιβλεπόμενης μάθησης για να ομαδοποιήσουν δεδομένα, έτσι ώστε η ταξινόμηση να προκύπτει από τα ίδια τα δεδομένα δίνοντας βάση σε αντικειμενικές μετρήσεις ομοιότητας. Με αυτόν τον τρόπο απέφυγαν τους περιορισμούς μιας καθορισμένης ταξινομίας, στην οποία ενδεχομένως να υπάρχουν ασάφειες και αφηρημένες έννοιες. Σε αυτή την προσέγγιση ένα ακουστικό αρχείο αναπαρίσταται από ένα σύνολο χαρακτηριστικών και ένα μέτρο ομοιότητας χρησιμοποιείται για τη σύγκριση. Τα μέτρα ομοιότητας μπορεί να είναι απλά όπως η ευκλείδεια απόσταση, αλλά και πιο σύνθετα, όπως τα κρυμμένα μοντέλα Markov ή Γκαουσιανά μοντέλα διασποράς.

3.2 Ταξινόμηση με βάση τη διάθεση

Η διάθεση και τα συναισθήματα που προκαλεί η μουσική αποτελεί έναν διαφορετικό τρόπο ταξινόμησης. Για να γίνει εφικτή μια τέτοια κατηγοριοποίηση απαιτείται η μοντελοποίηση της διάθεσης και μια διαδικασία χαρτογράφησης της. Ένας τρόπος που

χρησιμοποιείται συχνά είναι το μοντέλο του Thayer(Κεφ 1.2.1). Για να κατηγοριοποιηθεί η μουσική ανάλογα με διαθέσεις που προκαλεί, με επιστημονική αξιολόγηση είναι απαραίτητη μια επισημασμένη βάση δεδομένων με αληθείς ετικέτες, παρόλο που η μουσική δε κατηγοριοποιείται αποκλειστικά σε ένα και μόνο είδος. Μόνο το ανθρώπινο αυτί μπορεί να έχει απόλυτη κρίση. Όμως οι κριτικές στη μουσική είναι εκ των πραγμάτων υποκειμενικές και η συμφωνία μεταξύ εκτιμητών επιτυγχάνεται δύσκολα. Ως αποτέλεσμα οι σχολιασμοί για τις διαθέσεις πραγματοποιούνται κυρίως χειρωνακτικά από τρεις ή περισσότερους σχολιαστές, που ιδανικά αντιπροσωπεύουν ψυχολόγους και μουσικολόγους.

Η ταξινόμηση με βάση τη διάθεση πραγματοποιούνται με τη χρήση στίχων, ακουστικού αρχείου ή συνδυασμό και των δυο. Οι στίχοι περιέχουν πλούσια πληροφορία από την οποία μπορούν να εξαχθούν πολλοί τύποι χαρακτηριστικών. Τα γλωσσολογικά χαρακτηριστικά που έχουν εξαχθεί από τους στίχους συνήθως περιλαμβάνουν συλλογές από πακέτα λέξεων, όπου τα κείμενα αντιμετωπίζονται σαν αταξινόμητες λέξεις, μαζί με τη συχνότητά τους. Περιγραφικές λέξεις με ή χωρίς αποκοπή κατάληξης, επισημάνσεις με μέρη του λόγου και λειτουργικές λέξεις είναι τα πιο κοινά γλωσσολογικά χαρακτηριστικά.

Προηγούμενες έρευνες έχουν αποδείξει ότι οι λέξεις μόνες τους δεν είναι τόσο αποτελεσματικές όσο το όσο ο ήχος στην ταξινόμηση μουσικής. Οι Yang και Lee [22] ήταν οι πρώτοι που πρότειναν τον συνδυασμό στίχων και αριθμού ακουστικών χαρακτηριστικών με σκοπό να μεγιστοποιήσουν την ακρίβεια της ταξινόμησης. Ακόμα έχει πραγματοποιηθεί έρευνα, η οποία έχει βελτιώσει την απόδοση χρησιμοποιώντας υβριδικές προσεγγίσεις με ήχο και στίχους, καθώς και το συνδυασμό ήχου, στίχων και μουσικής σημειογραφίας.

3.3 Ταξινόμηση με *deep learning*

Οι Honglak Lee, Yan Largman, Peter Pham και Andrew Y. Ng[23] εφάρμοσαν Συνελικτικά Βαθιά Δίκτυα Πεποιθήσεων (CDBN) σε μη κατηγοριοποιημένα ακουστικά δεδομένα, όπως ομιλία και ήχο, και αξιολόγησαν τις αναπαραστάσεις μάθησης χαρακτηριστικών σε διαφορετικές περιπτώσεις ταξινόμησης των ακουστικών αρχείων. Ως βασική δομή των CDBN χρησιμοποίησαν τις Περιορισμένες Μηχανές Boltzman(CRBMs).

Για να εφαρμόσουν τα CDBM στα ακουστικά δεδομένα μετατρέπουν σήματα συναρτήσει του χρόνου σε φασματογραφήματα. Λόγω των πολλών διαστάσεων των

φασματογραφημάτων, εφάρμοσαν σε αυτά τη μέθοδο Ανάλυσης Κύριων Συνιστωσών με την οποία επαναπροσδιορίζονται οι συντεταγμένες των δεδομένων ,αφού έχει γίνει προεπεξεργασία των δεδομένων με τη μέθοδο Whitening, η οποία αφαιρεί τα πλεονάζοντα στοιχεία από αυτά.

Έγινε ταξινόμηση με βάση το είδος των τραγουδιών και τον καλλιτέχνη. Η ταξινόμηση σε είδη έγινε με 5 διαφορετικούς τρόπους , από τους οποίους καλύτερη απόδοση είχε το πρώτο επίπεδο χαρακτηριστικών CDBN. **Πίνακας 3.**

Αντίστοιχες διαδικασίες έγιναν για την αναγνώριση καλλιτέχνη με 4 διαφορετικούς τρόπους, όπου καλύτερη απόδοση είχε ο συνδυασμός πρώτου και δεύτερου επιπέδου χαρακτηριστικών CDBN.

Train examples	RAW	MFCC	CDBN L1	CDBN L2	CDBN L1 + L2
1	51.6%	54%	66.1%	62.5%	64.3%
2	57%	62.1%	69.7%	67.9%	69.5%
3	59.7%	65.3%	70%	66.7%	69.5%
5	65.8%	68.3%	73.1%	69.2%	72.7%

Πίνακας 3 : Ακρίβεια για τους 5 τρόπους κατηγοριοποίησης σε είδη

Οι Philippe Hamel και Douglas Eck[24] παρουσίασαν ένα σύστημα αυτόματης εξαγωγής κατάλληλων χαρακτηριστικών για συγκεκριμένα ζητούμενα από ακουστικά αρχεία. Το σύστημα απαρτίζεται από ένα Βαθύ Δίκτυο Πεποιθήσεων σε ακουστικά αρχεία όπου έχει εφαρμοστεί Διακριτός Μετασχηματισμός Fourier. Οι ενεργοποιήσεις του εκπαιδευμένου δικτύου χρησιμοποιήθηκαν ως είσοδοι σε μη γραμμικές μηχανές διανυσμάτων υποστήριξης με σκοπό την εκπαίδευση των χαρακτηριστικών για να λύνουν προβλήματα κατηγοριοποίησης μουσικής σε είδη.

Τα βαθιά δίκτυα πεποιθήσεων αποτελούνται από πολλά επίπεδα Περιορισμένων Μηχανών Boltzman. Για να αξιοποιήσουν με τον καλύτερο τρόπο τις μηχανές αυτές χρησιμοποίησαν τον αλγόριθμο ‘Contrastive Divergence’.

Χρησιμοποιήθηκαν δύο σύνολα ακουστικών δεδομένων . Το σύνολο Tzanetakis χρησιμοποιήθηκε για την αναγνώριση ειδών μουσικής . Τα αρχεία χωρίστηκαν σε μικρά πλαίσια 46.44ms και για κάθε πλαίσιο υπολογίστηκε Διακριτός Μετασχηματισμός Fourier.

Το σύνολο Majorminer, το οποίο περιέχει περισσότερες από 300 ετικέτες για 2500 αποσπάσματα ,χρησιμοποιήθηκε για τον έλεγχο των εκπαιδευμένων χαρακτηριστικών. Στη συνέχεια έγινε επεξεργασία του δικτύου ώστε να επιλεχθούν οι καλύτεροι παράμετροι , που απέδωσαν ακρίβεια 73.7%.

Χρησιμοποίησαν συνδυασμούς των εκπαιδευμένων χαρακτηριστικών για κατηγοριοποίηση και επισημείωση του δεύτερου συνόλου δεδομένων, ώστε να αποσπάσουν τη μεγαλύτερη ακρίβεια(84.3%), η οποία ήταν το ίδιο ικανοποιητική με άλλες προσεγγίσεις εξαγωγής χαρακτηριστικών από ακουστικό σήμα.

Οι Costa, Oliviera και Silla Jr [25] σύγκριναν αποτελέσματα που αποκτήθηκαν από συνελκτικά νευρωνικά δίκτυα (CNN) , με αποτελέσματα που αποκτήθηκαν χρησιμοποιώντας ταξινομητές χειρωνακτικών χαρακτηριστικών, όπως ο προσδιοριστής Local Binary Pattern(LBP) , η παραλλαγή του RLBP , τρεις τύποι ακουστικών χαρακτηριστικών και μηχανές διανυσμάτων υποστήριξης (SVM). Επιπλέον έκαναν πειράματα συνδυάζοντας τα παραπάνω αποτελέσματα για να εκτιμήσουν τη συμπληρωματικότητα μεταξύ των δυο διαφορετικών αναπαραστάσεων για προβλήματα ταξινόμησης μουσικής.

Τα πειράματά τους έγιναν σε τρία μουσικά σύνολα δεδομένων με ξεχωριστά χαρακτηριστικά: το σύνολο δεδομένων ISMIR 2004, μια συλλογή μουσικής της Λατινικής Αμερικής και μια συλλογή ηχογραφήσεων έθνικ Αφρικανικής μουσικής.

Δημιούργησαν φασματογραφήματα από τα περιεχόμενα του μουσικού σήματος για να πραγματοποιήσουν μια οπτικοποιημένη εξαγωγή χαρακτηριστικών, χρησιμοποιώντας 30s από κάθε μουσικό κομμάτι.

Τα αποτελέσματα έδειξαν πως τα CNN μπορούν να συγκριθούν ευνοϊκά με άλλους ταξινομητές. Στο αφρικανικό σύνολο τα CNN υπερέβησαν αποτελέσματα προηγούμενων ερευνών, στο σύνολο Λατινικής μουσικής ο συνδυασμός CNN και RLBP απέδωσαν ακρίβεια 92%, μεγαλύτερη από κάθε άλλη μέχρι τότε, και τέλος στο σύνολο ISMIR ενώ τα CNN δεν ξεπέρασαν σε απόδοση τις μέχρι τότε έρευνες, απέδωσαν καλύτερα από άλλους ταξινομητές που είναι βασισμένοι αποκλειστικά σε άλλα είδη χαρακτηριστικών.

Οι Chiyuan Zhang, Georgios Evangelopoulos, Stephen Voinea, Lorenzo Rosasco, Tomaso Poggio[27] , το 2014, πρότειναν μια πολυεπίπεδη ιεραρχική αρχιτεκτονική για μια βαθιά αναπαράσταση, αμετάβλητη σε μετασχηματισμούς και σταθερές, και εξαγωγή χαρακτηριστικών σε πλαίσια υψηλής ανάλυσης . Απέδειξαν πως μια βαθιά αναπαράσταση

αμετάβλητη σε τυπικούς μετασχηματισμούς βελτιώνει την κατηγοριοποίηση της μουσικής και πως η μη επιβλεπόμενη μάθηση είναι εφαρμόσιμη χρησιμοποιώντας αποθηκευμένα πρότυπα και μετασχηματισμούς τους. Για τα πειράματά τους χρησιμοποίησαν το σύνολο ακουστικών αρχείων GTZAN.

Ρύθμισαν και πρόσθεσαν διαφορετικά επίπεδα για να επιβάλλουν σταθερότητα σε ακουστικές αλλαγές, όπως η παραμόρφωση, τοπικές μεταφορές και μετατοπίσεις τονικού ύψους.

Στο βασικό επίπεδο δημιούργησαν ένα λογαριθμικό φασματογράφημα χρησιμοποιώντας βραχύχρονο μετασχηματισμό Fourier για δομές ακουστικών σημάτων με μεγάλο εύρος. Η εκτίμηση σφάλματος φια το πρώτο επίπεδο ήταν 35.5%. Πρόσθεσαν ένα δεύτερο επίπεδο για να συγκεντρώσουν τις προβολές αλλοιωμένων προτύπων στα υψηλότερο επίπεδο του φασματογραφήματος. Σε αυτό το επίπεδο παρατηρήθηκε απόκλιση σφάλματος 22%. Στο τρίτο επίπεδο αντιμετώπισαν τις τοπικές μεταφορές του ήχου χρησιμοποιώντας τη διαδικασία max- pooling για 8 γειτονικά πλαίσια με απόκλιση σφάλματος 16.5%. Στο τέταρτο επίπεδο εφάρμοσαν προβολή και συγκέντρωση των προτύπων μετατόπισης τονικού ύψους. Η απόκλιση σφάλματος ήταν 18%, βελτιωμένη σε σχέση με τα προηγούμενα επίπεδα και πολύ μικρότερη σε σχέση με τους MFCC(67%).

Οι Keunwoo Choi, Gyorgy Fazekas, Mark Sandler[26], το 2016, πρότειναν έναν αλγόριθμο αυτόματης επισημείωσης βασισμένο σε ένα Βαθύ Πλήρως Συνελικτικό Δίκτυο(FCN). Τα FCN είναι βαθιά συνελικτικά δίκτυα τα οποία αποτελούνται αποκλειστικά από συνελικτικά επίπεδα, χωρίς κανένα πλήρως συνδεδεμένο επίπεδο. Η αρχιτεκτονική του δικτύου περιέχει 4 συνελικτικά επίπεδα και 4 συγκεντρωτικά επίπεδα που χρησιμοποιούν max- pooling. Το δίκτυο έχει ως είσοδο ένα φασματογράφημα σε κλίμακα mel και ως έξοδο προβλέπει ένα 50- διάστατο διάνυσμα ετικετών. Η αρχιτεκτονική επεκτάθηκε με τρία ακόμα βαθύτερα επίπεδα(FCN-{5,6,7}).

Χρησιμοποιήθηκαν δυο σύνολα δεδομένων για να αξιολογηθεί το προτεινόμενο σύστημα. Στα πειράματα υπολογίστηκαν οι καμπύλες ROC και το εμβαδό κάτω από αυτές(AUC) της αρχιτεκτονικής με διαφορετικές πολυπλοκότητες και τύπους εισόδου.

Το σύνολο δεδομένων MagnaTagATune το οποίο χρησιμοποιήθηκε στο πρώτο πείραμα περιέχει 25.856 αποσπάσματα των 29.1s σε μορφή mp3 με 185 ετικέτες, από τις οποίες χρησιμοποιήθηκαν οι καλύτερες 50 που περιέχουν είδη, όργανα και διαθέσεις. Αξιολογήθηκαν τρεις αρχιτεκτονικές (FCN-{3,4,5}), καθώς και ο βραχύχρονος μετασχηματισμός Fourier, MFCC και φασματογραφήματα σε κλίμακα mel ως είσοδος στο

τέταρτο επίπεδο. Τα αποτελέσματα έδειξαν πως ο προτεινόμενος αλγόριθμος ανταγωνίζεται ικανοποιητικά τις άλλες προσεγγίσεις, παρ' όλα αυτά υπήρξαν μικρές διαφορές μεταξύ των αλγορίθμων. Παρατηρήθηκε ότι η πρόσθεση επιπέδων δεν οδηγεί πάντα σε βελτιωμένη απόδοση.

Το δεύτερο σύνολο δεδομένων Million Song Dataset , το οποίο είναι εννέα φορές μεγαλύτερο από το προηγούμενο σύνολο, χρησιμοποιήθηκε στο δεύτερο πείραμα με ετικέτες από το last.fm και αξιολογήθηκαν οι αποδόσεις της αρχιτεκτονικής μεταβάλλοντας τον αριθμό των επιπέδων(FCN-{3,4,5,6,7}). Σε αντίθεση με το πρώτο πείραμα αποδείχθηκε ότι βαθύτερα μοντέλα ξεπερνούν σε απόδοση την αρχιτεκτονική τεσσάρων επιπέδων, δείχνοντας ότι οι βαθύτερες δομές ωφελούν περισσότερο από τα επαρκή δεδομένα.

Οι Sergio Oramas, Oriol Nieto, Francesco Barbieri, Xavier Serra[28], το 2018, ασχολήθηκαν με την κατηγοριοποίηση της μουσικής με πολλαπλές και λεπτομερείς ετικέτες χαρακτηρισμών, με μια προσέγγιση βαθιάς μάθησης, χρησιμοποιώντας τρεις διαφορετικούς τρόπους: ακουστικά αρχεία, κείμενο και εικόνα.

Παρουσίασαν το νέο σύνολο δεδομένων MuMu, με 31.000 άλμπουμ ταξινομημένα σε ένα ή περισσότερα από 250 είδη, το οποίο συνδυάζει πληροφορίες από το σύνολο δεδομένων κριτικών του Amazon και το Million Song Dataset.

Αφού έγινε επεξεργασία σήματος στα μουσικά κομμάτια του συνόλου, χρησιμοποιήθηκαν ως είσοδος σε ένα Συνελικτικό Νευρωνικό Δίκτυο με τέσσερα συνελικτικά επίπεδα, στο οποίο έγιναν πειράματα με διαφορετικούς αριθμούς και μεγέθη φίλτρων, και τροποποίηση παραμέτρων εξόδου.

Ενσωμάτωσαν όλες τις κριτικές από το σύνολο MuMu που αναφέρονται σε κάθε άλμπουμ, σε ένα ενιαίο κείμενο, στα οποία εφάρμοσαν ένα Μοντέλο Διανυσματικού Χώρου(VSM) για να δημιουργηθεί ένα διάνυσμα χαρακτηριστικών για κάθε ένα από αυτά. Ένα δίκτυο εμπρόσθιας τροφοδοσίας με δύο πυκνά επίπεδα των 2048 νευρώνων και μια Διορθωμένη Γραμμική Μονάδα (ReLU) έπειτα από κάθε επίπεδο, εκπαιδεύτηκαν για να προβλέπουν χαρακτηριστικά ειδών.

Κάθε άλμπουμ στο σύνολο έχει μια εικόνα εξωφύλλου. Για να πραγματοποιηθεί κατηγοριοποίηση σε είδη από τις εικόνες χρησιμοποίησαν ένα Υπολειμματικό Νευρωνικό Δίκτυο που αποτελείται από 101 επίπεδα(Res Net-101).

Τα διανύσματα που προέκυψαν από κάθε διαφορετικό τρόπο συνδέθηκαν σε ένα ενιαίο διάνυσμα που χρησιμοποιήθηκε ως είσοδος σε ένα πολυστρωματικό Perceptron(MLP), όπου η είσοδος είναι ευθέως συνδεδεμένη με την έξοδο.

Η προσέγγιση με βάση το κείμενο ξεπέρασε σε απόδοση τους άλλους δύο τρόπους. Η κατηγοριοποίηση με εικόνες είχε χαμηλή απόδοση αλλά βελτίωσε τα αποτελέσματα στη συνδυαστική προσέγγιση. Τέλος ο συνδυασμός των τριών τρόπων απέδωσε τα καλύτερα αποτελέσματα.

Οι Shaleen Bengani, S. Vadivel και J. Angel Arul Jothi[29], το 2019, παρουσίασαν ένα σύνολο μεθόδων βασισμένων στη βαθιά μάθηση για αυτόματη επισημείωση μουσικών χαρακτηριστικών, με υψηλή απόδοση και λιγότερες απαιτήσεις σε υπολογιστικό χρόνο και μνήμη σε σχέση με προηγούμενες προσπάθειες.

Η προτεινόμενη μέθοδος είναι ένα σύνολο τριών συνελκτικών νευρωνικών δικτύων. Τα δύο πρώτα μοντέλα έχουν ως είσοδο φασματογραφήματα mel. Το πρώτο αποτελείται από μονοδιάστατα συνελκτικά επίπεδα, ενώ το δεύτερο από δισδιάστατα. Το τρίτο μοντέλο δέχεται ως είσοδο ακατέργαστα ακουστικά σήματα. Για κάθε μοντέλο ,και για κάθε συνδυασμό μοντέλων υπολογίστηκε το εμβαδό AUC.

Για τα πειράματά τους χρησιμοποίησαν το επισημασμένο σύνολο δεδομένων MagnaTagATune χωρισμένο σε 16 σύνολα, με τα οποία έγιναν συνδυασμοί. Χρησιμοποιήθηκαν μόνο οι 50 ετικέτες με τη μεγαλύτερη συχνότητα εμφάνισης.

Αποδείχθηκε ότι τα δύο πρώτα μοντέλα και το τρίτο μαθαίνουν διαφορετικά είδη ακουστικών χαρακτηριστικών. Ο συνδυασμός των μοντέλων απέδωσε αρκετά ικανοποιητικά αποτελέσματα, ισότιμα με αυτά των μεθόδων με τα καλύτερα αποτελέσματα ως τότε.

4. Πειράματα με το εργαλείο Meka

4.1 Greek Audio Dataset (GAD)

Η βάση ακουστικών αρχείων που χρησιμοποιήθηκε για τα πειράματα είναι μια δωρεάν διαθέσιμη συλλογή χαρακτηριστικών και μεταδεδομένων ήχου που δημιουργήθηκε από τους Δήμο Μακρή, Κάτια Λήδα Κερμανίδου και Ιωάννη Καρύδη. Περιέχει 1000 τραγούδια, 1000 txt αρχεία με στίχους, 1000 υπερσυνδέσμους στο YouTube ,277 ξεχωριστούς καλλιτέχνες και 8 είδη. Η συλλογή δεν περιέχει ακουστικά αρχεία λόγω πνευματικών δικαιωμάτων, αλλά περιέχει σημαντικά χαρακτηριστικά που έχουν εξαχθεί από αυτά, όπως στίχους και σχολιασμούς σχετικά με το είδος και τη διάθεση για κάθε ακουστικού κομματιού. Για κάθε κομμάτι υπάρχει διαθέσιμος υπερσύνδεσμος με ακουστικό περιεχόμενο στο Youtube.

Η ακουστική συλλογή καλύπτει ολόκληρο το εύρος της ελληνικής μουσικής από την παραδοσιακή μέχρι τη μοντέρνα. Οι στίχοι που περιλαμβάνονται στη συλλογή ανακτήθηκαν από ποικίλες πηγές όπως stixoi.info και χρησιμοποιήθηκαν για ακαδημαϊκούς και επιστημονικούς σκοπούς. Χρησιμοποιούνται μόνο για πληροφορίες σχετικά με τη διάθεση του κάθε κομματιού.

Η επισημείωση των κομματιών έγινε χειρωνακτικά με ετικέτες που περιγράφουν το είδος και τη διάθεση των κομματιών, από φοιτητές του Μουσικού τμήματος Ιονίου Πανεπιστημίου. Για την κατηγοριοποίηση των κομματιών ως προς τη διάθεση χρησιμοποιήθηκε το μοντέλο του Thayer, το οποίο περιέχει συνολικά 16 ταξινομίες διαθέσεων. Ως μέτρο εγκυρότητας των σχολιασμών χρησιμοποιήθηκε η σταθερή συμφωνία των σχολιαστών.

Τα είδη στα οποία κατηγοριοποιήθηκαν τα κομμάτια είναι: Λαϊκό, Ρεμπέτικο, Έντεχνο, Ροκ, Μοντέρνο Λαϊκό, Ποπ, Εναλλακτικό και Hip Hop. Η κατηγοριοποίησή τους έγινε με

επισημείωση μονής ετικέτας, δηλαδή ένταξη του κάθε κομματιού σε μία μοναδική κατηγορία και με επισημείωση πολλαπλών ετικετών, δηλαδή ένταξη του κάθε κομματιού σε μία ή και περισσότερες κατηγορίες.

Χρησιμοποιήθηκαν χαρακτηριστικά με βάση τον ήχο, τους στίχους και τα συμβολικά δεδομένα(MIDI, παρτιτούρες). Τα χαρακτηριστικά ήχου χωρίζονται σε τρεις κατηγορίες με βάση: το ηχόχρωμα, τη ρυθμική αγωγή και το τονικό ύψος. Η κάθε κατηγορία αποτελείται από σύνολα χαρακτηριστικών.

Τα ηχοχρωματικά χαρακτηριστικά χρησιμοποιήθηκαν για τη διαφοροποίηση της μίξης του ήχου με βάση τα ορχηστρικά μέρη των κομματιών, στις περιπτώσεις όπου τα μελωδικά και τα τονικά στοιχεία ήταν όμοια. Χωρίστηκαν σε τέσσερις βασικές κατηγορίες: FFT, MFCC, Spectrum και Method of Moments και στο σύνολό τους είναι 251.

Τα ρυθμικά χαρακτηριστικά χρησιμοποιήθηκαν για να περιγράψουν την κανονικότητα του ρυθμού, το beat, το τέμπο και τη χρονική υπογραφή. Στο σύνολο είναι 73 και χωρίστηκαν σε δύο βασικές κατηγορίες: Χαρακτηριστικά Beat και συχνότητας και Ιστόγραμμα Beat.

Τέλος, τα χαρακτηριστικά τονικού ύψους χρησιμοποιούνται για πολλούς διαφορετικούς σκοπούς, όπως η επισημείωση συγχορδιών, ανάλυση δομής και μουσικός συγχρονισμός και βασίζονται σε τεχνικές εντοπισμού του τονικού ύψους. Το χαρακτηριστικό που χρησιμοποιήθηκε είναι το Linear Predictive Coding. Προσθέτοντας παράγωγα χαρακτηριστικά το σύνολο τους ήταν 18.

Η εξαγωγή των ακουστικών χαρακτηριστικών έγινε ως επί το πλείστον με το βασισμένο σε Java πρόγραμμα εξαγωγής χαρακτηριστικών jAudio, όπου διαθέτει 27 ξεχωριστά χαρακτηριστικά. Επιπλέον χρησιμοποιήθηκε το εργαλείο AudioFeatureExtraction το οποίο παρουσιάστηκε από την ερευνητική ομάδα ανάκτησης μουσικής πληροφορίας του Πανεπιστημίου Τεχνολογίας της Βιέννης για την εξαγωγή του χαρακτηριστικού Beat Histogram από το σύνολο ρυθμικών χαρακτηριστικών, αφού έδινε καλύτερα αποτελέσματα.

Το σύνολο δεδομένων είναι διαθέσιμο σε μορφή HDF5 και σε μορφή CSV που είναι συμβατή με το εργαλείο Weka.

4.2 Λίγα λόγια για τα εργαλεία Weka και Meka

Για τη διεξαγωγή των πειραμάτων χρησιμοποιήθηκαν τα εργαλεία WEKA και MEKA.

Το WEKA [33] είναι ένα εργαλείο που παρέχει την εφαρμογή αλγορίθμων μάθησης σε σύνολα δεδομένων. Διαθέτει πολλές μεθόδους για την επίλυση προβλημάτων εξόρυξης δεδομένων, όπως η ταξινόμηση, η ομαδοποίηση, ή επιλογή χαρακτηριστικών, καθώς και εργαλεία για την προεπεξεργασία των δεδομένων που ονομάζονται φίλτρα.

Το WEKA υποστηρίζει ταξινόμηση μονής ετικέτας(single-label classification), δηλαδή κάθε τραγούδι της βάσης δεδομένων μπορεί να ταξινομηθεί σε μια και μόνο κατηγορία. Στα πειράματα της παρούσας εργασίας πραγματοποιείται ταξινόμηση πολλαπλών ετικετών(multi-label classification), διότι κάθε τραγούδι μπορεί να ανήκει σε παραπάνω από μια κατηγορία. Για το λόγο αυτό το WEKA χρησιμοποιήθηκε μόνο για την προεπεξεργασία των δεδομένων, ενώ για τη διεξαγωγή των πειραμάτων έγινε χρήση του εργαλείου MEKA.

Το MEKA αποτελεί προέκταση του WEKA και χρησιμοποιεί τη μορφή αρχείου ARFF του WEKA. Υποστηρίζει την ταξινόμηση πολλαπλών ετικετών, δηλαδή κάθε τραγούδι μπορεί να ταξινομηθεί σε παραπάνω από μία κατηγορίες. Χρησιμοποιεί πολλαπλές κλάσεις, μία για κάθε χαρακτηρισμό, αντί για μια κλάση που χρησιμοποιεί το WEKA. Ο αριθμός των χαρακτηριστικών ορίζεται είτε με $-c$ είτε με $-C$, τα οποία αναγράφονται στην κεφαλίδα του αρχείου ARFF. Το MEKA χρησιμοποιεί εσωτερικά αναφορές στους δείκτες των κλάσεων για να υποδηλώσει τον αριθμό των χαρακτηριστικών. Σε περίπτωση που ο αριθμός των χαρακτηριστικών τείνει να διαφοροποιείται σε κάθε σύνολο, το MEKA επιτρέπει την αποθήκευση της επιλογής στο ARFF αρχείο με την έκφραση @relation 'όνομα αρχείου', όπου η άνω κάτω τελεία (:) χρησιμοποιείται για να διαχωρίσει το όνομα του συνόλου από τις επιλογές. Το παράδειγμα που ακολουθεί περιγράφει την κεφαλίδα ενός ARFF αρχείου για ταξινόμηση πολλαπλών ετικετών, με μεταβλητές τα πρώτα οχτώ χαρακτηριστικά.

@relation 'arff_όνομα_συνόλου_δεδομένων: -C 8'

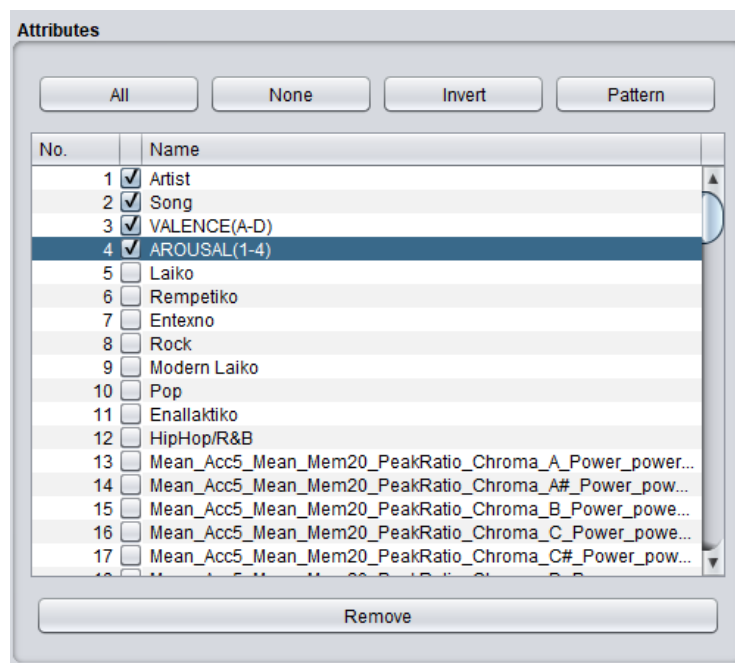
Οι παραπάνω παράμετροι λειτουργούν εφόσον τα χαρακτηριστικά που θα χρησιμοποιηθούν ως κλάσεις βρίσκονται στην αρχή του συνόλου δεδομένων. Σε περίπτωση που βρίσκονται στο τέλος, η χρήση του μείον (-) πριν τον αριθμό των κλάσεων υποδηλώνει τον αριθμό των χαρακτηριστικών που θα χρησιμοποιηθούν ως κλάσεις(@relation

'arff_όνομα_συνόλου_δεδομένων: -C- 8'). Σε διαφορετική περίπτωση απαιτείται η εφαρμογή φίλτρου μετατόπισης χαρακτηριστικών.

4.3 Προεπεξεργασία δεδομένων

Η μορφή των αρχείων του συνόλου δεδομένων (CSV) είναι συμβατή με την είσοδο που δέχονται τα εργαλεία WEKA και MEKA, οπότε δε χρειάστηκε κάποια αλλαγή.

Για την προεπεξεργασία του συνόλου δεδομένων χρησιμοποιήθηκε το εργαλείο WEKA, όπου αφαιρέθηκαν οι περιττές κλάσεις, δηλαδή αυτές που δεν μας δίνουν καμία πληροφορία και δεν επηρεάζουν την ταξινόμηση σε είδη. Οι κλάσεις αυτές είναι: 'Artist', 'Song', 'Valence' και 'Arousal'.



Εικόνα 6. Διαγραφή των περιττών κλάσεων 'Artist', 'Song', 'Valence' και 'Arousal'

Ένα σύνολο δεδομένων ορίζεται ανομοιόμορφο όταν οι κατηγορίες ταξινόμησης δεν αναπαρίστανται κατά προσέγγιση ισομερώς. Στο Greek Audio Dataset οι τέσσερις από τις οχτώ κατηγορίες είναι ανομοιόμορφες. Για παράδειγμα 195 τραγούδια ανήκουν στην κλάση ‘Ρεμπέτικο’, ενώ 1205 όχι. Η ανομοιομορφία αυτή ενδέχεται να οδηγήσει σε μη έγκυρη ταξινόμηση. Για το λόγο αυτό στα δεδομένα εφαρμόστηκε το φίλτρο ‘SMOTE’ που υπάρχει ενσωματωμένο στο WEKA.

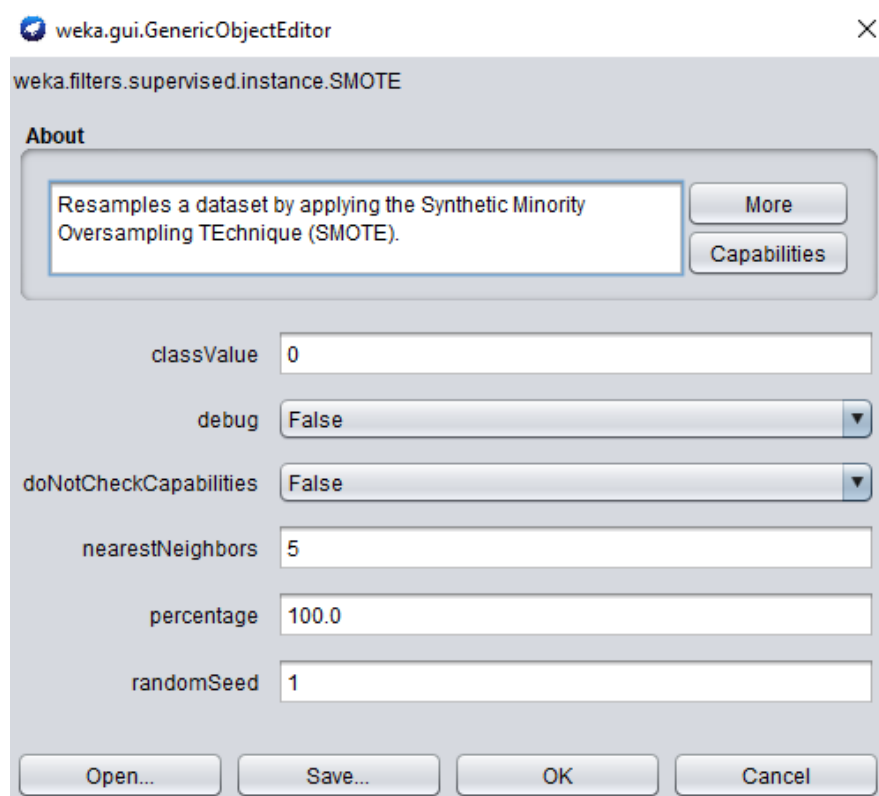
Η τεχνική SMOTE(Synthetic Minority Oversampling Technique) είναι μια προσέγγιση oversampling, κατά την οποία δημιουργούνται συνθετικά παραδείγματα, τα οποία προστίθενται στη μειονοτική κλάση[32]. Σε κάθε δείγμα προστίθενται συνθετικά παραδείγματα που προκύπτουν από κάποιους ή όλους τους k πλησιέστερους γείτονες της μειονοτικής κλάσης. Ανάλογα με το ποσοστό oversampling που απαιτείται κάποιοι γείτονες από τους k πλησιέστερους επιλέγονται τυχαία.

Στα πειράματα επιλέγονται οι πέντε πλησιέστεροι γείτονες. Αυτό σημαίνει πως αν απαιτείται oversampling 200%, οι δύο από τους πέντε πλησιέστερους γείτονες θα χρησιμοποιηθούν ως συνθετικά παραδείγματα.

Η εφαρμογή του φίλτρου SMOTE στο WEKA είναι μια απλή διαδικασία. Στην καρτέλα της προεπεξεργασίας των δεδομένων η επιλογή SMOTE βρίσκεται στα supervised filters για instances. Για τα πειράματα εφαρμόστηκε το φίλτρο 4 φορές σε κάθε CSV αρχείο στις κλάσεις για κάθε μειονοτική κλάση είδους. SMOTE εφαρμόστηκε στις κλάσεις: ‘Ρεμπέτικο’, ‘Ποπ’, ‘Εναλλακτικό’ και ‘HipHop R&B’, με ποσοστά oversampling 125%, 125%, 90% και 150% αντίστοιχα.

Είδη	Πριν	Μετά
Ρεμπέτικο	188	438
Ποπ	195	482
Εναλλακτικό	265	568
HipHop R&B	59	147

Πίνακας 4. Οι μειονοτικές κλάσεις πριν και μετά την προεπεξεργασία



Εικόνα 7. Ρυθμίσεις του SMOTE: Επιλογή πλησιέστερων γειτόνων και ποσοστό oversampling

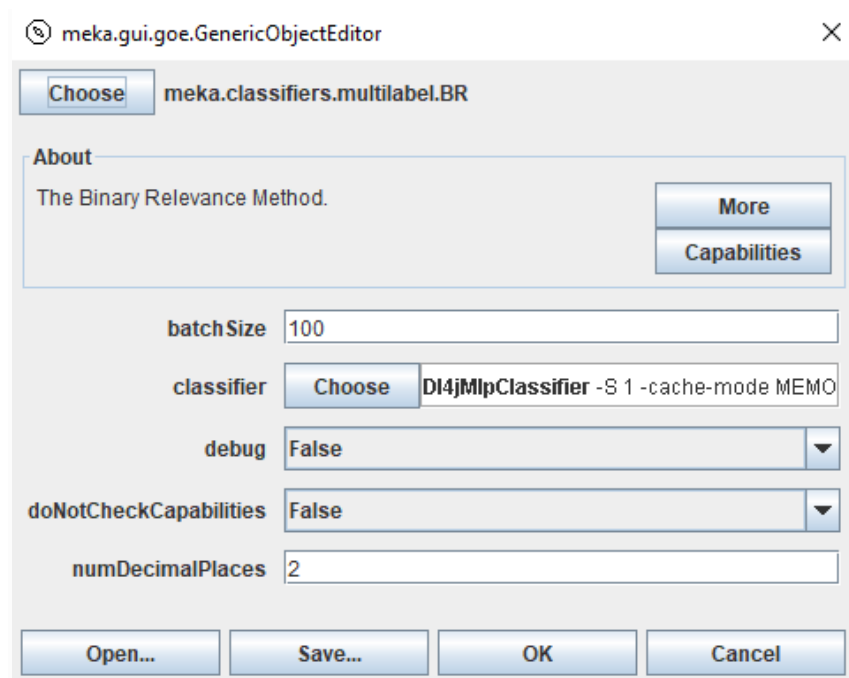
4.4 Το πακέτο βαθιάς μάθησης 'WekaDeeplearning4j'

Το WEKADeeplearning4j είναι ένα πακέτο βαθιάς μάθησης προσαρμοσμένο στο εργαλείο WEKA. Έχει αναπτυχθεί για να ενσωματώσει στο WEKA τις πιο σύγχρονες τεχνικές βαθιάς μάθησης. Τα πιο γνωστά περιβάλλοντα βαθιάς μάθησης απαιτούν από το χρήστη τη σύνταξη κώδικα για να εφαρμόσουν τις τεχνικές και τους αλγορίθμους της. Το πακέτο WEKADeeplearning4j επιτρέπει την πρόσβαση στη βαθιά μάθηση, μέσω ενός γραφικού περιβάλλοντος χρήστη(GUI). Το σύστημα υποστήριξης του πακέτου WEKADeeplearning4j παρέχεται από τη βιβλιοθήκη Deeplearning4j της Java. Περιλαμβάνει υποστήριξη μονάδας επεξεργασίας γραφικών (GPU) και επιτρέπει την εκπαίδευση και τον έλεγχο μοντέλων βαθύων νευρωνικών δικτύων μέσω του GUI, όπως τα συνελκτικά και τα ανατροφοδοτούμενα νευρωνικά δίκτυα. Επιπλέον, παρέχει λειτουργίες προ-επεξεργασίας δεδομένων κειμένου και εικόνας. Το πακέτο WekaDeeplearning4j υποστηρίζει πλήρως συνδεδεμένα δίκτυα τροφοδοτούμενα προς τα εμπρός, συνελκτικά και ανατροφοδοτούμενα δίκτυα. Παρέχει λειτουργία φόρτωσης δεδομένων, τα οποία μπορεί να είναι συμβατικά δεδομένα σε μορφή πίνακα, εικόνα, κείμενο ή ομαδοποιημένα δεδομένα.

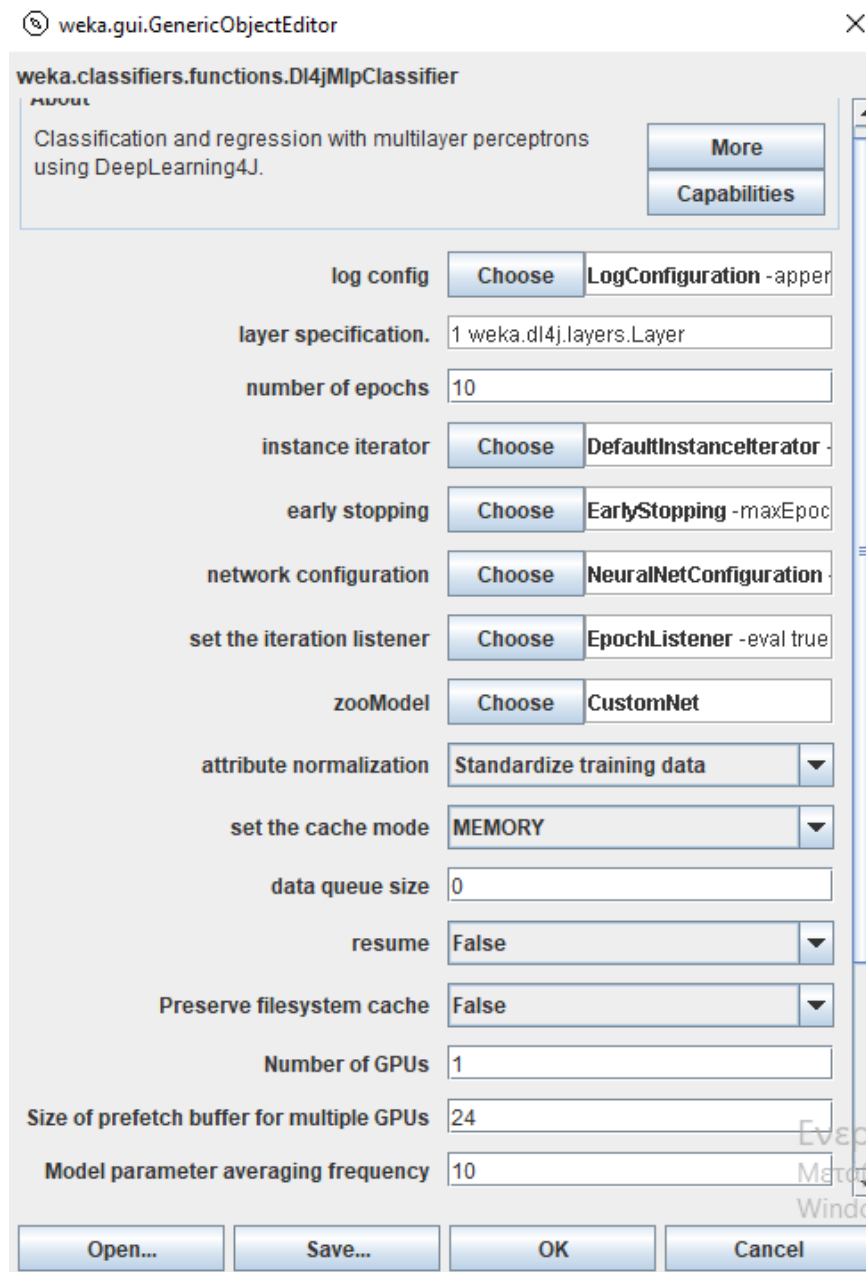
Όπως προαναφέρθηκε, κύριος στόχος του εργαλείου είναι να δημιουργηθεί πρόσβαση στις τεχνικές deep learning δίχως τη χρήση κώδικα. Ο χρήστης για να διεξάγει πειράματα μέσω GUI πρέπει αρχικά να φορτώσει τα δεδομένα στο εργαλείο με τη μορφή ARFF. Στη συνέχεια ακολουθεί η ρύθμιση της αρχιτεκτονικής του δικτύου και η επιλογή του πρωτοκόλλου που θα εφαρμοστεί στο πείραμα. Τέλος, το μόνο που απομένει είναι να τρέξει το πείραμα.

Το WEKADeeplearning4j δημιουργήθηκε ως πακέτο του WEKA, και με αυτόν τον τρόπο τα μοντέλα Deep Learning είναι διαθέσιμα σε ολόκληρο το περιβάλλον του WEKA, καθώς και της προέκτασής του MEKA. Το κυριότερο μοντέλο που είναι διαθέσιμο ονομάζεται DI4jMlpClassifier και επιτρέπει τη δημιουργία αυθαίρετων νευρωνικών δικτύων τροφοδοτούμενων προς τα εμπρός, όπως τα συνελκτικά δίκτυα.

Για να επιλέξω τον DI4jMlpClassifier στο MEKA, αφού έχω φορτώσει τα δεδομένα μου επιλέγω την καρτέλα "Classify" και πατώντας πάνω στον επιλεγμένο ταξινομητή εμφανίζονται επιλογές των διαθέσιμων αλγορίθμων. Έπειτα επιλέγοντας επάνω στον αλγόριθμο DI4jMlpClassifier εμφανίζονται οι παράμετροι που μπορούν να τροποποιηθούν. Στο πείραμα της παρούσας εργασίας θα διατηρηθούν οι προκαθορισμένες ρυθμίσεις.



Εικόνα 8.Επιλογή του *Dl4jMlpClassifier*



Εικόνα 9. Ρυθμίσεις παραμέτρων του αλγορίθμου *DL4jMlpClassifier*

4.5 Ταξινόμηση πολλαπλών ετικετών

Όπως προαναφέρθηκε για την ταξινόμηση του Greek Audio Dataset χρησιμοποιήθηκε το εργαλείο MEKA , το οποίο επιτρέπει την ταξινόμηση πολλαπλών ετικετών. Αυτό το είδος ταξινόμησης μέχρι πρόσφατα χρησιμοποιούταν κυρίως για κατηγοριοποίηση κειμένου.

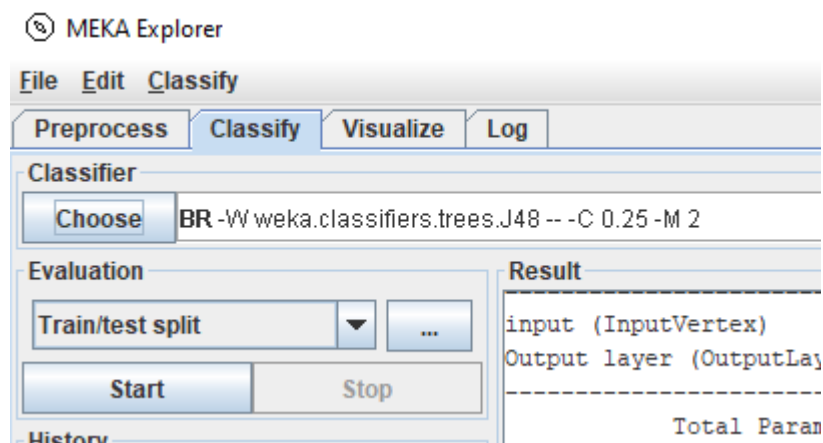
Πλέον, τα μοντέλα κατηγοριοποίησης πολλαπλών ετικετών χρησιμοποιούνται σε πιο σύγχρονες εφαρμογές, όπως η ταξινόμηση πρωτεϊνικών δομών, η μουσική κατηγοριοποίηση και η σημασιολογική ταξινόμηση[1].

Στην ταξινόμηση πολλαπλών ετικετών πρέπει να καθορίζεται το είδος ή τα είδη στα οποία ανήκει κάθε ένα από τα παραδείγματα(τραγούδια) του συνόλου δεδομένων που ταξινομείται.

Υπάρχουν δυο μέθοδοι ταξινόμησης πολλαπλών ετικετών: η μέθοδος μετασχηματισμού προβλήματος και η μέθοδος προσαρμογής αλγορίθμου. Η μέθοδος μετασχηματισμού προβλήματος μετασχηματίζει ένα πρόβλημα ταξινόμησης πολλαπλών ετικετών σε ένα ή περισσότερα προβλήματα ταξινόμησης μονής ετικέτας ή σε προβλήματα παλινδρόμησης. Η μέθοδος προσαρμογής αλγορίθμου χρησιμοποιούν επέκταση συγκεκριμένων αλγορίθμων για να χειριστούν απευθείας τα δεδομένα πολλαπλών ετικετών[34].

Στο παρόν πείραμα θα χρησιμοποιηθεί η μέθοδος μετασχηματισμού προβλήματος Binary Relevance(BR) [35], την οποία έχει το MEKA ως προεπιλογή. Η μέθοδος αυτή μετασχηματίζει κάθε πρόβλημα πολλαπλών ετικετών σε ένα δυαδικό πρόβλημα για κάθε ετικέτα. Έτσι για κάθε ετικέτα το δυαδικό μοντέλο εκπαιδεύεται να προβλέπει κάθε παράδειγμα το οποίο ανήκει στην ετικέτα αυτή.

Για να χρησιμοποιήσω τον ταξινομητή BR στο περιβάλλον του MEKA επιλέγω “choose”(classifier) και διαλέγω τον αντίστοιχο ταξινομητή.



Εικόνα 10. Επιλογή ταξινομητή BR

Το σύνολο δεδομένων Greek Audio Dataset περιέχει 8 δυαδικές κλάσεις, η κάθε μια από τις οποίες αναπαριστά ένα μουσικό είδος με δύο πιθανές τιμές: «ναι», στην περίπτωση που το εκάστοτε τραγούδι ανήκει στη συγκεκριμένη κλάση και «όχι», στην περίπτωση που δεν ανήκει.

4.5.1 Μέθοδοι Αξιολόγησης

Στα πειράματα χρησιμοποιούνται δύο μέθοδοι αξιολόγησης: η μέθοδος “Holdout” ή “Train/Test Split” και η μέθοδος “Cross Validation”(Διασταυρωμένη Επικύρωση).

Στη μέθοδο “Holdout” τα δεδομένα χωρίζονται σε δύο σύνολα, που ονομάζονται «Σύνολο εκπαίδευσης» και «Σύνολο ελέγχου». Το σύνολο εκπαίδευσης είναι το υποσύνολο του συνόλου δεδομένων, στο οποίο εφαρμόζεται ο αλγόριθμος μηχανικής μάθησης και «μαθαίνει» τις σχέσεις μεταξύ των χαρακτηριστικών και των ετικετών τους. Το σύνολο ελέγχου χρησιμοποιείται για να διαπιστωθεί η απόδοση του μοντέλου σε δεδομένα που δεν έχει «ξαναδεί». Στα πειράματα το ποσοστό του συνόλου που χρησιμοποιείται για εκπαίδευση είναι 64%.

“Cross Validation” είναι η μέθοδος όπου το σύνολο δεδομένων χωρίζεται σε k υποσύνολα. Ένα υποσύνολο k χρησιμοποιείται για τον έλεγχο του μοντέλου(test set), ενώ τα υπόλοιπα $k-1$ σύνολα χρησιμοποιούνται για την εκπαίδευση του μοντέλου(training set). Η διαδικασία επαναλαμβάνεται έως ότου όλα τα υποσύνολα k να έχουν χρησιμοποιηθεί ως σύνολα ελέγχου. Το τελικό αποτέλεσμα θα είναι ο μέσος όρος των επιμέρους αποτελεσμάτων.

Για παράδειγμα για διασταυρωμένη επικύρωση 5 υποσυνόλων, το σύνολο δεδομένων θα χωριστεί σε 5 τμήματα και το μοντέλο θα εκπαιδευτεί και θα ελεγχθεί 5 φορές, ώστε κάθε τμήμα να χρησιμοποιηθεί ως test set, όπως φαίνεται στον παρακάτω πίνακα(πίνακας 5).

Στα πειράματα θα χρησιμοποιηθεί Cross-Validation 10 τμημάτων.

Διαίρεση 1	τμήμα 1	τμήμα 2	τμήμα 3	τμήμα 4	τμήμα 5	Αποτελέσματα 1
Διαίρεση 2	τμήμα 1	τμήμα 2	τμήμα 3	τμήμα 4	τμήμα 5	Αποτελέσματα 2
Διαίρεση 3	τμήμα 1	τμήμα 2	τμήμα 3	τμήμα 4	τμήμα 5	Αποτελέσματα 3
Διαίρεση 4	τμήμα 1	τμήμα 2	τμήμα 3	τμήμα 4	τμήμα 5	Αποτελέσματα 4
Διαίρεση 5	τμήμα 1	τμήμα 2	τμήμα 3	τμήμα 4	τμήμα 5	Αποτελέσματα 5

Πίνακας 5. Με κόκκινο επισημαίνονται τα σύνολα που χρησιμοποιούνται ως *test set* σε κάθε κύκλο και με μπλε τα σύνολα που χρησιμοποιούνται ως *train set*.

4.5.2 Μετρικές αξιολόγησης

Αφού το μοντέλο έχει εκπαιδευτεί και δοκιμαστεί αξιολογείται η απόδοσή του. Για το σκοπό αυτό χρησιμοποιούνται μετρικές αξιολόγησης. Στην ταξινόμηση πολλαπλών ετικετών οι προβλέψεις για ένα παράδειγμα είναι ένα σύνολο από ετικέτες. Συνεπώς οι προβλέψεις μπορεί να είναι πλήρως σωστές, μερικώς σωστές και πλήρως λάθος. Οι πιο διαδεδομένες μετρικές αξιολόγησης είναι οι “Exact match” και “Hamming Loss” και “Accuracy” (ακρίβεια) [33], [34], [35] για κάθε ετικέτα. Επίσης θα καταγραφούν και οι τιμές των μετρικών F1 και της καμπύλης AUC-ROC.

Η μετρική “Exact match” είναι πολύ αυστηρή και η τιμή της δηλώνει το ποσοστό του δείγματος που έχει ταξινομηθεί σωστά σε όλες τις ετικέτες. Όπως αναφέρθηκε παραπάνω, πολλές φορές, στα προβλήματα ταξινόμησης πολλαπλών ετικετών, τα αποτελέσματα είναι μερικώς θετικά, δηλαδή δεν έχουν ταξινομηθεί όλα τα παραδείγματα στις σωστές ετικέτες. Το μειονέκτημα της μετρικής αυτής είναι ότι αγνοεί τα μερικώς θετικά αποτελέσματα.

Το “Hamming Loss” δηλώνει το ποσοστό των ετικετών που έχουν προβλεφθεί λάθος. Είναι το κλάσμα με αριθμητή τις λανθασμένες προβλέψεις ετικετών και παρονομαστή τον αριθμό των συνολικών ετικετών. Όσο μικρότερη είναι η τιμή του, τόσο καλύτερη απόδοση έχει ο ταξινομητής.

Για την σαφέστερη περιγραφή των υπολοίπων μετρικών θα χρησιμοποιηθούν οι εξής παράμετροι[37]: TP(True Positive) και TN(True Negative) είναι οι ετικέτες που έχουν προβλεφθεί σωστά και είναι είτε αληθώς θετικές, είτε αληθώς αρνητικές και FP(False

Positive) και FN(False Negative) είναι οι ετικέτες που έχουν ταξινομηθεί λάθος και είναι είτε ψευδώς θετικές, είτε ψευδώς αρνητικές.

Η ακρίβεια(Accuracy)= $\frac{TP+TN}{TP+TN+FN+TN}$ αποτελεί μια πολύ καλή μετρική αξιολόγησης όταν τα δεδομένα είναι ομοιόμορφα, δηλαδή όταν οι κατηγορίες ταξινόμησης αναπαρίστανται ισομερώς. Η τιμή της ακρίβειας για κάθε παράδειγμα ορίζει το ποσοστό των σωστών ταξινομημένων ετικετών προς τον αριθμό των συνολικών ετικετών του εκάστοτε παραδείγματος. Η γενική ακρίβεια είναι ο μέσος όρος της ακρίβειας όλων των παραδειγμάτων. Στα πειράματα της παρούσας εργασίας θα σημειωθεί η ακρίβεια για κάθε παράδειγμα.

Η τιμή της μετρικής $F1 = \frac{2 \times (Recall \times Precision)}{Recall + Precision}$ δηλώνει τον αρμονικό μέσο του precision (ακρίβεια)= $\frac{TP}{TP+FP}$ και του recall(ανάκληση) = $\frac{TP}{TP+FN}$, όπου precision είναι η αναλογία των θετικών ετικετών που έχουν προβλεφθεί σωστά, προς το σύνολο των ετικετών που έχουν προβλεφθεί θετικές και ανάκληση είναι η αναλογία των θετικών ετικετών που έχουν προβλεφθεί σωστά, προς το σύνολο των ετικετών που είναι πραγματικά θετικές. Το F1 είναι πολλές φορές περισσότερο χρήσιμο από το Accuracy, ειδικά όταν τα δεδομένα μας είναι κατανεμημένα άνισα.

Η καμπύλη ROC (Receiver Operating Characteristics)[36] είναι μια καμπύλη πιθανοτήτων σε δισδιάστατο χώρο που αναπαριστά το βαθμό διακριτότητας μεταξύ των κλάσεων μιας ετικέτας. Στον άξονα x απεικονίζεται η ανάκληση(recall), ενώ στον άξονα y το

specificity = $\frac{TN}{TN+FP}$, δηλαδή η αναλογία των ετικετών που έχουν προβλεφθεί αληθώς αρνητικές προς τον πραγματικό αριθμό των αρνητικών ετικετών. Όσο μεγαλύτερο είναι το εμβαδό κάτω από την καμπύλη AUC(Area Under the Curve), με μέγιστο τον αριθμό 1, τόσο καλύτερη είναι η απόδοση του μοντέλου στο να διαχωρίζει τις θετικές από τις αρνητικές κλάσεις.

5. Αποτελέσματα πειραμάτων

Τα CSV αρχεία του Greek Dataset χωρίζονται σε τρία σύνολα με βάση τα χαρακτηριστικά του ήχου(Audio), τους στίχους(Lyrics) και τα συμβολικά δεδομένα(Symbolic). Για το πείραμα χρησιμοποιήθηκαν τα δύο πρώτα σύνολα.

Το πρώτο σύνολο αφορά τα χαρακτηριστικά του ήχου και περιέχει τρία υποσύνολα: “Pitch Chroma”, “Rhythm” και “Timbral”. Το υποσύνολο “Pitch Chroma” περιέχει ένα CSV αρχείο που αφορά τα χαρακτηριστικά τονικού ύψους των τραγουδιών. Το υποσύνολο “Rhythm” περιέχει ένα CSV αρχείο με ρυθμικά χαρακτηριστικά και τέλος, το υποσύνολο “Timbral” περιέχει δύο CSV αρχεία με ηχοχρωματικά χαρακτηριστικά. Το δεύτερο σύνολο περιέχει 10 CSV αρχεία που περιέχουν tf-idf βάρη.

Πραγματοποιήθηκαν πειράματα με τον αλγόριθμο deeplearning4j και με δύο μεθόδους αξιολόγησης, τη μέθοδο “Cross Validation” και “Train-test split” για κάθε CSV αρχείο ξεχωριστά και στη συνέχεια ακολούθησαν συγκρίσεις των αποτελεσμάτων.

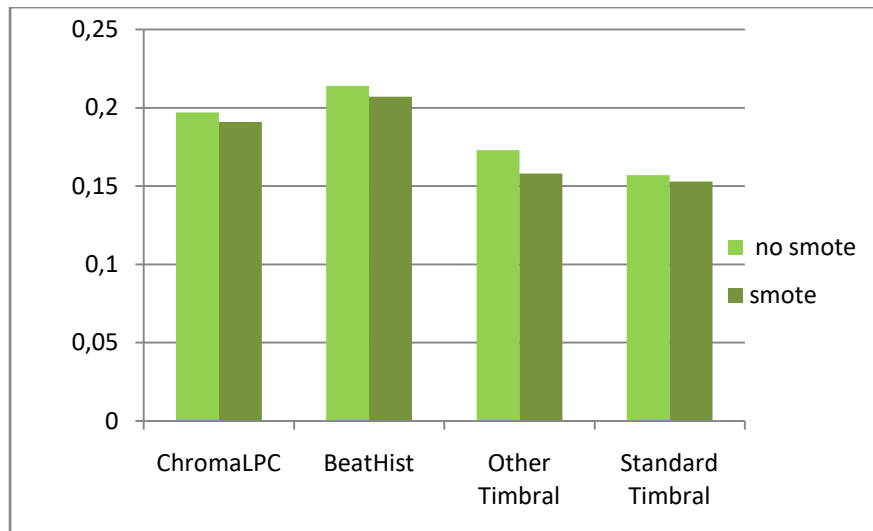
5.1 Σύγκριση επεξεργασμένων-μη επεξεργασμένων δεδομένων

5.1.1 Σε audio δεδομένα

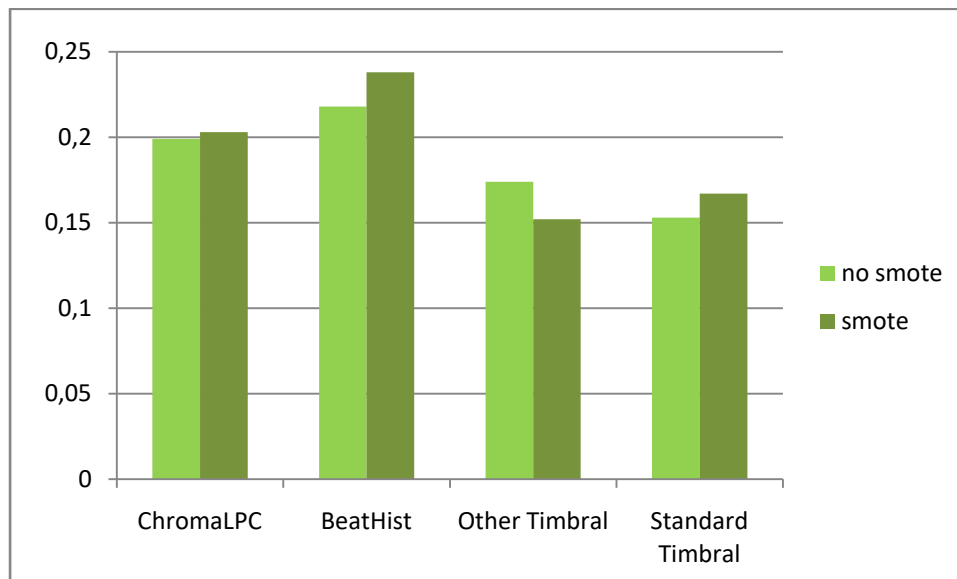
Αρχικά πραγματοποιήθηκε σύγκριση μεταξύ των αρχικών δεδομένων και των δεδομένων μετά την επεξεργασία τους με την τεχνική SMOTE. Στα παρακάτω γραφήματα φαίνονται τα αποτελέσματα.

Στα δεδομένα που δεν έχει εφαρμοστεί η τεχνική SMOTE, τα αποτελέσματα της μετρικής απόδοσης Hamming Loss που προέκυψαν από cross-validation έχουν υψηλότερες τιμές από τα ανεπεξέργαστα δεδομένα. Αυτό σημαίνει πως το ποσοστό ετικετών που έχουν προβλεφθεί λάθος είναι μεγαλύτερο στην πρώτη περίπτωση.

Στα αποτελέσματα που προέκυψαν από τη μέθοδο train-test split τα χαρακτηριστικά “Chromal LPC” και “Standard Timbral” έχουν καλύτερη απόδοση ανεπεξέργαστα, ενώ τα “Beat Hist” και “Other Timbral” αποδίδουν καλύτερα επεξεργασμένα.

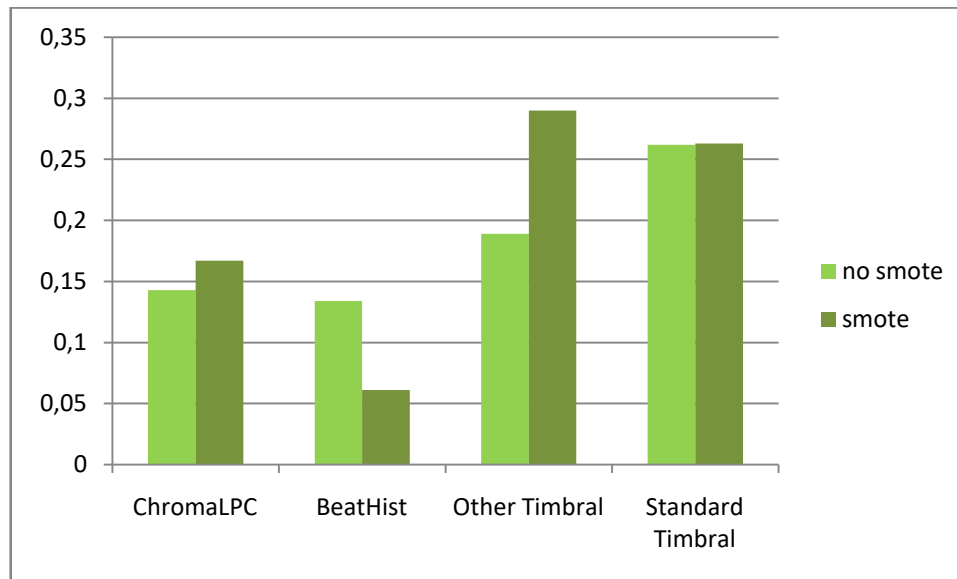


Γράφημα 1. Hamming loss σε audio δεδομένα με cross-validation



Γράφημα 2. Hamming loss σε audio δεδομένα με Test-train split

Τα αποτελέσματα της μετρικής Exact Match έχουν καλύτερη απόδοση στα δεδομένα που έχει εφαρμοστεί SMOTE και με τις δύο μεθόδους αξιολόγησης cross-validation και train-split. Μεγαλύτερο ποσοστό του δείγματος έχει ταξινομηθεί σωστά σε όλες τις ετικέτες, σε σχέση με τα ανεπεξέργαστα δεδομένα.



Γράφημα 3. Exact Match με train-test split

Στα αποτελέσματα της μετρικής F1, συνολικά, ανά παράδειγμα και ανά ετικέτα, τα επεξεργασμένα δεδομένα έχουν ελαφρώς καλύτερη απόδοση από τα ανεπεξέργαστα με τη μέθοδο cross-validation. Με τη μέθοδο train-test split η απόδοση είναι καλύτερη στα ανεπεξέργαστα δεδομένα. Όπως παρουσιάζεται στα **γραφήματα 5** και **6** οι διαφορές είναι αμελητέες.

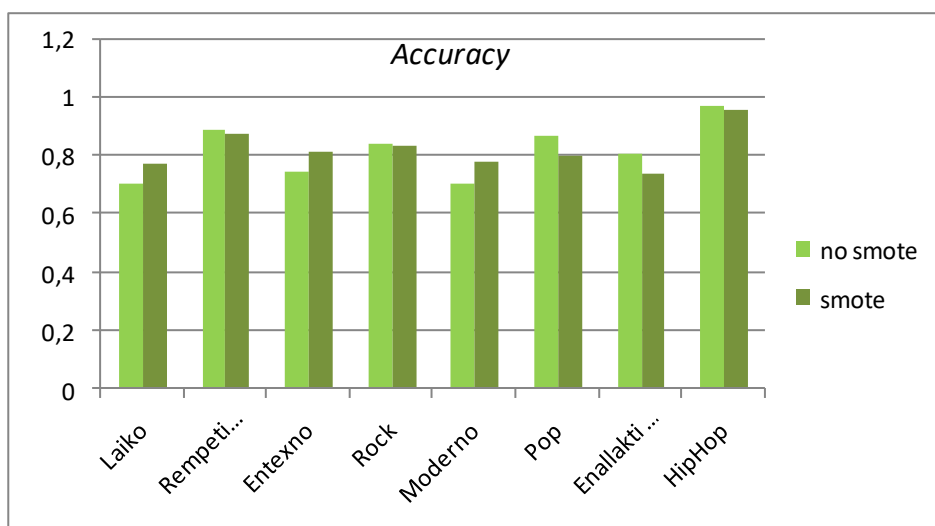
Audio	F1	F1	F1 example	F1 example	F1 label	F1 label
ChromaLPC	0,625	0,62	0,58	0,568	0,442	0,452
BeatHist	0,556	0,528	0,519	0,495	0,262	0,23
Other Timbral	0,696	0,722	0,651	0,675	0,583	0,66
Standard Timbral	0,717	0,72	0,66	0,665	0,616	0,647
Average	0,6485	0,6475	0,6025	0,60075	0,47575	0,49725

Γράφημα 4. F1, F1 example και F1 label για ανεπεξέργαστα(πράσινο) και επεξεργασμένα(χακί) δεδομένα με cross-validation.

Audio	F1	F1	F1 example	F1 example	F1 label	F1 label
ChromaLPC	0,664	0,646	0,628	0,608	0,486	0,51
BeatHist	0,63	0,581	0,598	0,547	0,364	0,4
Other Timbral	0,705	0,74	0,66	0,696	0,589	0,671
Standard Timbral	0,74	0,724	0,698	0,68	0,64	0,668
Average	0,68475	0,67275	0,646	0,63275	0,51975	0,56225

Γράφημα 5. F1, F1 example και F1 label για ανεπεξέργαστα(πράσινο) και επεξεργασμένα(χακί) δεδομένα με train-test split.

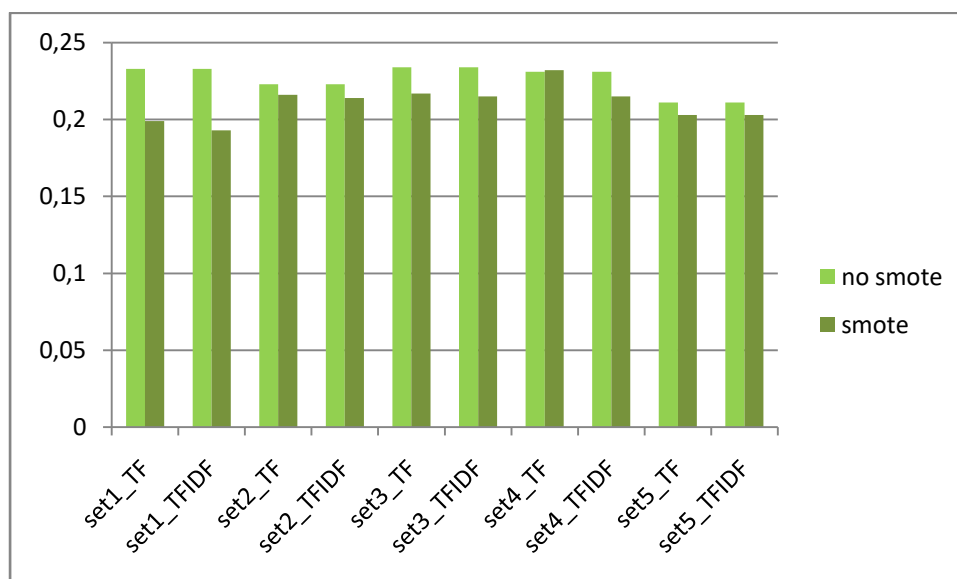
Στα επεξεργασμένα δεδομένα, τα αποτελέσματα που προέκυψαν και από τις δύο μεθόδους και αφορούν την ακρίβεια έχουν καλύτερη απόδοση στα μουσικά είδη «Λαϊκό», «Έντεχνο» και «Μοντέρνο», ενώ στα υπόλοιπα είδη καλύτερη απόδοση έχουν τα ανεπεξέργαστα δεδομένα.



Γράφημα 6. Ακρίβεια για κάθε ετικέτα με cross-validation

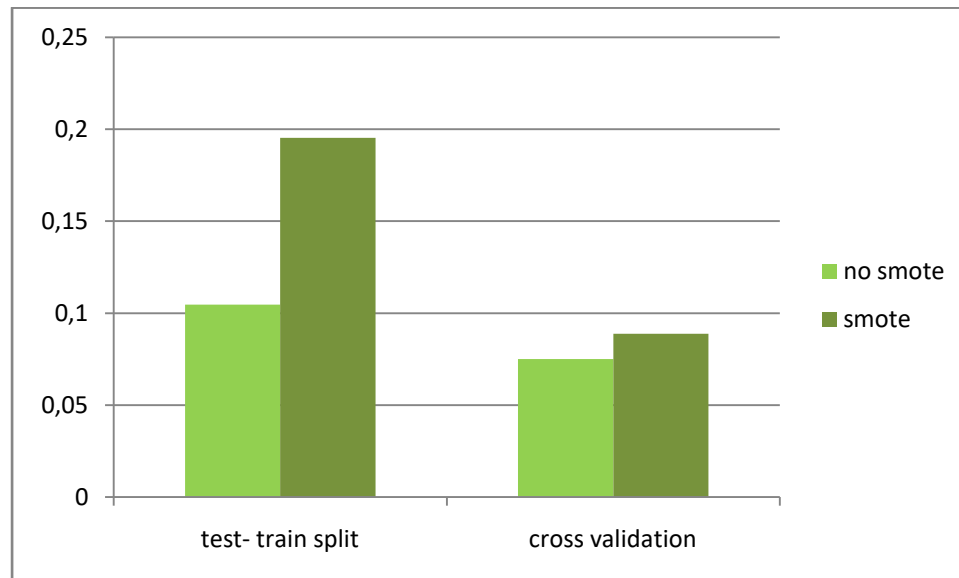
5.1.2 Σε δεδομένα στίχων

Παρόμοια είναι τα αποτελέσματα από τα δεδομένα στίχων. Οι τιμές των αποτελεσμάτων της μετρικής Hamming loss των ανεπεξέργαστων δεδομένων στίχων είναι συνολικά μεγαλύτερες από αυτές των επεξεργασμένων δεδομένων και με τις δύο μεθόδους.



Γράφημα 7. Hamming Loss με cross-validation

Οι τιμές των αποτελεσμάτων της μετρικής Exact match είναι μεγαλύτερες στα δεδομένα που έχει εφαρμοστεί smote και με τις δύο μεθόδους.



Γράφημα 8. Exact match με train-test split και cross-validation

Οι τιμές των F1, F1 example και F1 στα δεδομένα των στίχων είναι συνολικά καλύτερες στα επεξεργασμένα δεδομένα και με τις δύο μεθόδους.

Τέλος , οι τιμές της ακρίβειας είναι μεγαλύτερες στις ετικέτες «Λαϊκό», «Έντεχνο» και «Μοντέρνο» στα επεξεργασμένα δεδομένα και με τις δύο μεθόδους ταξινόμησης.

Lyrics	F1	F1	F1 example	F1 example	F1 label	F1 label
set1_TF	0,589	0,644	0,544	0,599	0,442	0,556
set1_TFIDF	0,589	0,664	0,544	0,622	0,442	0,577
set2_TF	0,553	0,533	0,512	0,49	0,281	0,298
set2_TFIDF	0,553	0,533	0,512	0,488	0,281	0,29
set3_TF	0,545	0,536	0,509	0,492	0,283	0,329
set3_TFIDF	0,545	0,536	0,509	0,49	0,283	0,313
set4_TF	0,485	0,485	0,457	0,457	0,169	0,191
set4_TFIDF	0,485	0,493	0,457	0,443	0,169	0,232
set5_TF	0,589	0,581	0,54	0,529	0,267	0,401
set5_TFIDF	0,589	0,586	0,54	0,534	0,367	0,508
Average	0,5522	0,5591	0,5124	0,5144	0,2984	0,3695

Πίνακας 6. F1, F1 example, F1 label με cross-validation

Lyrics	F1	F1	F1 example	F1 example	F1 label	F1 label
set1_TF	0,598	0,639	0,561	0,6	0,442	0,557
set1_TFIDF	0,598	0,656	0,561	0,619	0,442	0,566
set2_TF	0,606	0,592	0,575	0,553	0,363	0,442
set2_TFIDF	0,606	0,585	0,575	0,543	0,363	0,441
set3_TF	0,569	0,575	0,538	0,535	0,362	0,434
set3_TFIDF	0,569	0,571	0,538	0,535	0,362	0,413
set4_TF	0,599	0,542	0,576	0,526	0,336	0,332
set4_TFIDF	0,598	0,553	0,575	0,511	0,336	0,35
set5_TF	0,625	0,617	0,589	0,576	0,442	0,5
set5_TFIDF	0,625	0,605	0,589	0,566	0,442	0,481
Average	0,5993	0,5935	0,5677	0,5564	0,389	0,4516

Πίνακας 7. F1, F1 example, F1 label με train-test split

Για τις υπόλοιπες συγκρίσεις στη συνέχεια θα χρησιμοποιηθούν τα επεξεργασμένα δεδομένα, καθώς τα αποτελέσματά τους συνολικά ήταν καλύτερα από τα ανεπεξέργαστα δεδομένα.

5.2 Σύγκριση χαρακτηριστικών του ήχου

Τα κύρια χαρακτηριστικά του ήχου είναι ο ρυθμός, το τονικό ύψος και το ηχόχρωμα. Η επόμενη σύγκριση που έγινε μετά από τα πειράματα ήταν αυτή μεταξύ των χαρακτηριστικών αυτών. Τα 4 χαρακτηριστικά του Greek Dataset που συγκρίνονται είναι το “ChromaLPC” χαρακτηριστικό τονικού ύψους, το “ BeatHist” ρυθμικό χαρακτηριστικό και τα “ Other Timbral” και “ Standard Timbral” ηχοχρωματικά χαρακτηριστικά .

Καλύτερη απόδοση φαίνεται να έχουν τα χαρακτηριστικά “ Other Timbral” και “ Standard Timbral” και με τις δύο μεθόδους κατηγοριοποίησης. Παρουσιάζουν τις μικρότερες τιμές για τη μετρική απόδοση Hamming Loss και τις μεγαλύτερες τιμές για τις μετρικές Exact match, F1 και AUC-ROC , ενώ χαμηλότερη απόδοση παρουσιάζει το ρυθμικό χαρακτηριστικό “ BeatHist” , με τις υψηλότερες τιμές στη μετρική Hamming Loss και τις χαμηλότερες στις υπόλοιπες μετρικές και με τις δύο μεθόδους αξιολόγησης.

	Hamming Loss	Exact match	F1	F1 example	F1 label	AUC-ROC
<i>ChromaLPC</i>	0,191/0,203	0,146/0,167	0,62/0,646	0,568/0,608	0,452/0,51	0,788/0,785
<i>BeatHist</i>	0,207/0,238	0,03/0,061	0,528/0,851	0,495/0,547	0,23/0,4	0,659/0,673
<i>Other Timbral</i>	0,158/0,152	0,255/0,29	0,722/0,74	0,675/0,696	0,66/0,671	0,867/0,877
<i>Standard</i>	0,153/0,167	0,255/0,263	0,72/0,724	0,665/0,68	0,647/0,668	0,875/0,871

Πίνακας 8. Μετρικές αξιολόγησης με cross-validation/train-test split με επισημασμένες τις καλύτερες αποδόσεις.

Καλύτερη απόδοση στην ακρίβεια έχει το χαρακτηριστικό “ Other Timbral” στα είδη «Λαϊκό», «Ροκ», και «Εναλλακτικό» , και το χαρακτηριστικό “ Standard Timbral” στα είδη «Ρεμπέτικο», «Μοντέρνο» και «Ποπ» και με τις δύο μεθόδους. Στο είδος «Έντεχνο» καλύτερη απόδοση έχει το χαρακτηριστικό “ Standard Timbral” με cross-validation και το χαρακτηριστικό “ Other Timbral” με train-test split, ενώ στο είδος Hip/Hop αποδίδει καλύτερα το “ Other Timbral” με cross-validation και το “ Standard Timbral” με train-test split.

Χειρότερη απόδοση έχει το χαρακτηριστικό “Beat Hist” στα είδη «Λαϊκό», «Ρεμπέτικο», «Έντεχνο», «Ροκ»(με ελάχιστη διαφορά από τα υπόλοιπα) και «Hip/Hop», το χαρακτηριστικό “ChromaLPC” στα είδη «Μοντέρνο» και «Ποπ», και το χαρακτηριστικό “Beat Hist” στ είδος “Hip/Hop”. Όλα τα παραπάνω ισχύουν και για τις δύο μεθόδους αξιολόγησης.

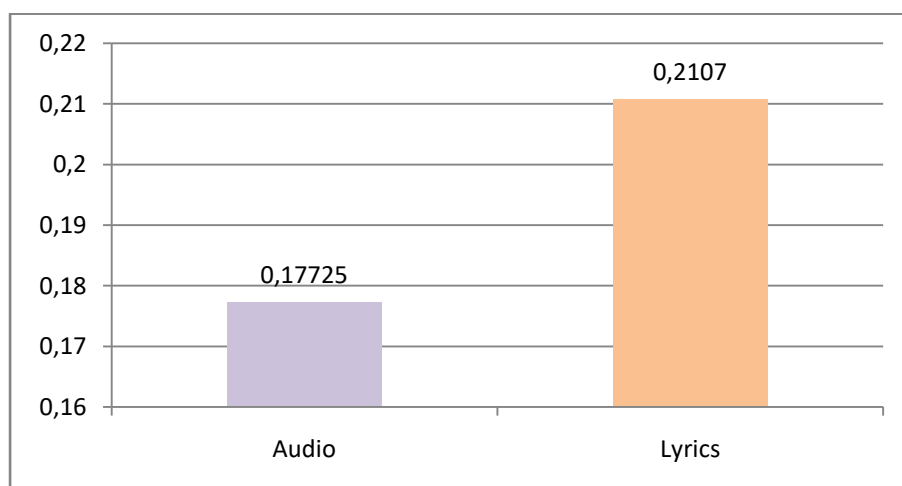
	Laiko	Rempetiko	Entexno	Rock	Moderno	Pop	Enallaktiko	HipHop
<i>ChromaLPC</i>	0,763/0,743	0,876/0,872	0,802/0,781	0,81/0,809	0,761/0,738	0,781/0,751	0,738/0,73	0,943/0,951
<i>BeatHist</i>	0,675/0,59	0,8/0,735	0,798/0,794	0,81/0,803	0,777/0,745	0,8/0,789	0,745/0,718	0,94/0,92
<i>OtherTimbral</i>	0,827/0,836	0,899/0,883	0,823/0,829	0,862/0,878	0,78/0,781	0,785/0,816	0,791/0,807	0,969/0,953
<i>Standard</i>	0,806/0,793	0,918/0,931	0,836/0,803	0,853/0,84	0,793/0,793	0,833/0,835	0,66/0,749	0,972/0,967

Πίνακας 9. Ακρίβεια με cross-validation/train-test split με επισημασμένες τις καλύτερες αποδόσεις.

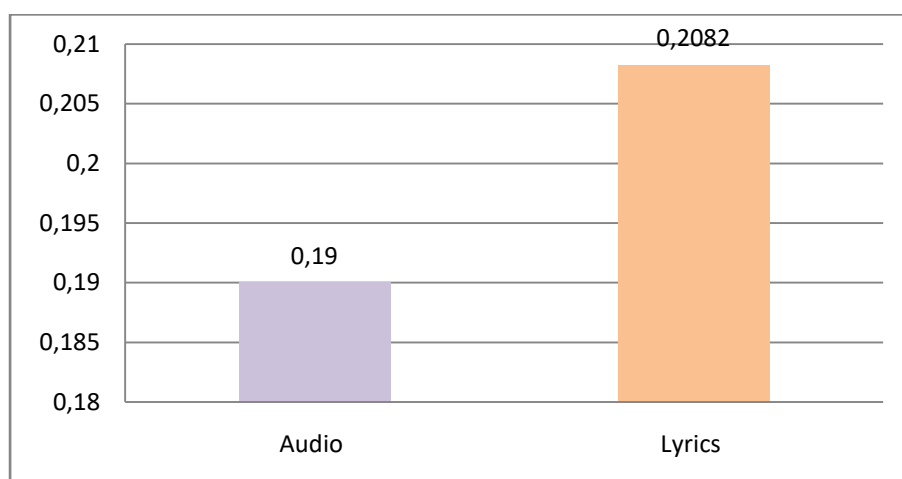
5.3 Σύγκριση δεδομένων Audio με δεδομένα στίχων

Στη συνέχεια έγινε σύγκριση μεταξύ των χαρακτηριστικών του ήχου και των χαρακτηριστικών των στίχων και με τις δύο μεθόδους αξιολόγησης cross-validation και train-test split.

Οι τιμές των αποτελεσμάτων της μετρικής Hamming Loss είναι χαμηλότερες στα δεδομένα ήχου και με τις δύο μεθόδους αξιολόγησης. Συνεπώς, ποσοστό που έχει προβλεφθεί λάθος είναι μεγαλύτερο στα δεδομένα των στίχων.

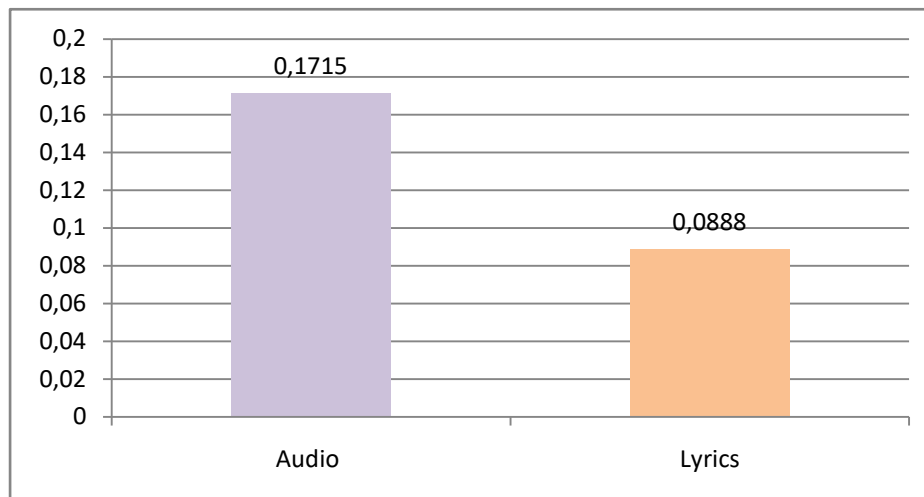


Γράφημα 9. Hamming Loss Audio και Lyrics με cross-validation.

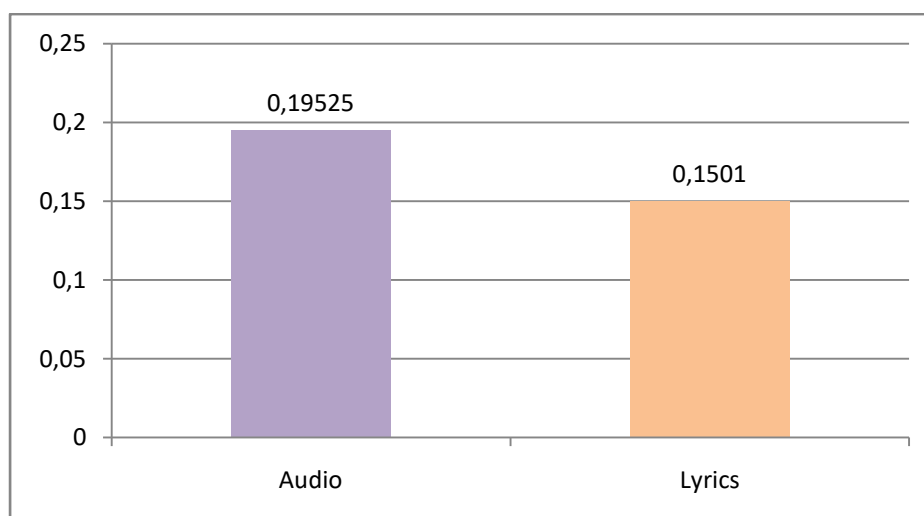


Γράφημα 10. Hamming Loss Audio και Lyrics με train-test split.

Οι τιμές της μετρικής Exact match είναι μεγαλύτερες κατά 0,08 μονάδες με cross-validation και 0,04 μονάδες με train-test split. Το ποσοστό δείγματος που έχει ταξινομηθεί σωστά σε όλες τις ετικέτες είναι μεγαλύτερο στα δεδομένα ήχου.

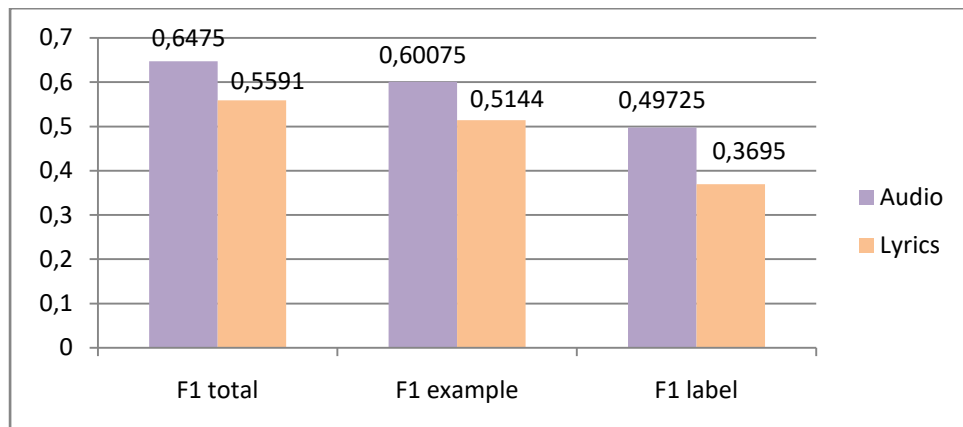


Γράφημα 11. *Exact Match Audio και Lyrics με cross-validation*

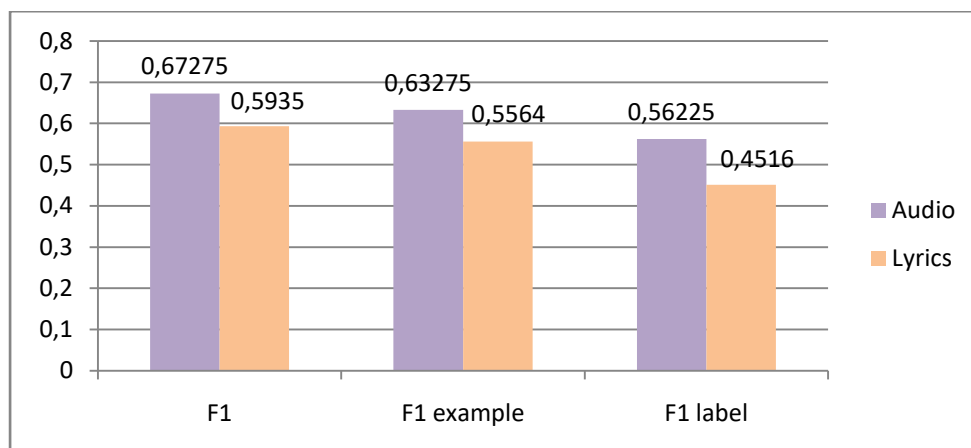


Γράφημα 12. *Exact Match Audio και Lyrics με train-test split*

Οι τιμές της μετρικής F1(αρμονικός μέσος precision και recall), F1 για κάθε παράδειγμα και F1 για κάθε ετικέτα είναι μεγαλύτερες στα δεδομένα ήχου κατά περίπου 0,08 μονάδες και με τις δύο μεθόδους αξιολόγησης.

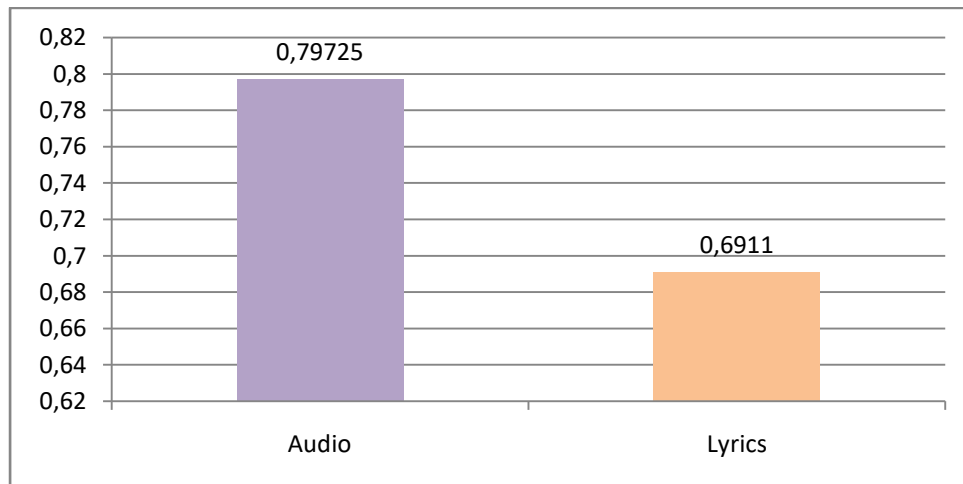


Γράφημα 13. F1, F1 example, F1 label με cross-validation.

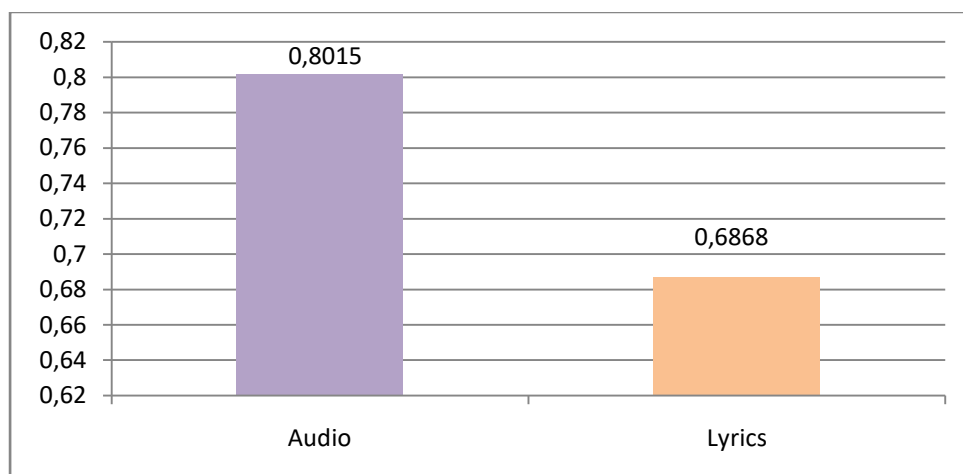


Γράφημα 14. F1, F1 example, F1 label με train-test split.

Οι τιμές της καμπύλης AUC-ROC είναι μεγαλύτερες στα δεδομένα του ήχου κατά 0,10 μονάδες κατά προσέγγιση. Συνεπώς ο βαθμός διακριτότητας μεταξύ των κλάσεων είναι μεγαλύτερος στα δεδομένα του ήχου.



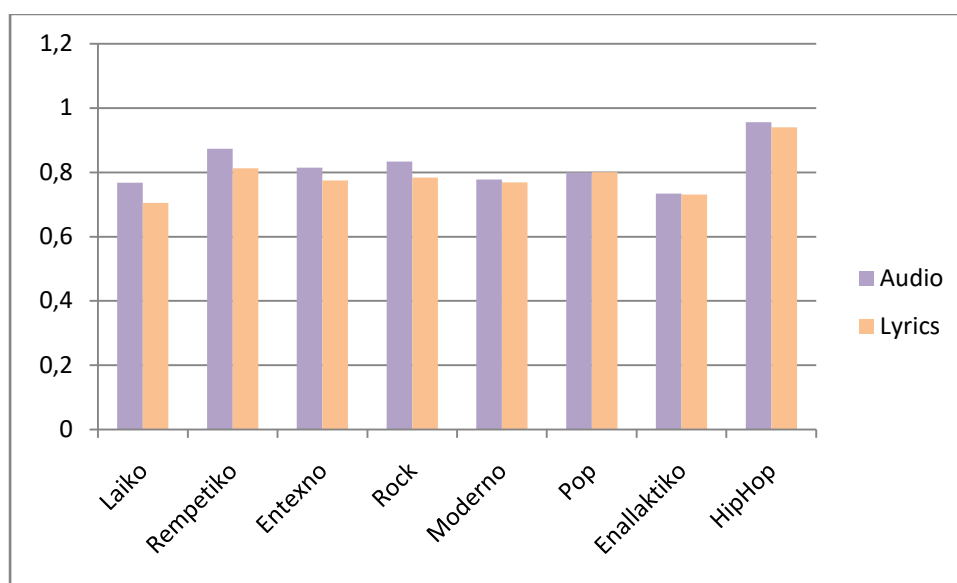
Γράφημα 15. Καμπύλη AUC-ROC Audio και Lyrics με cross-validation



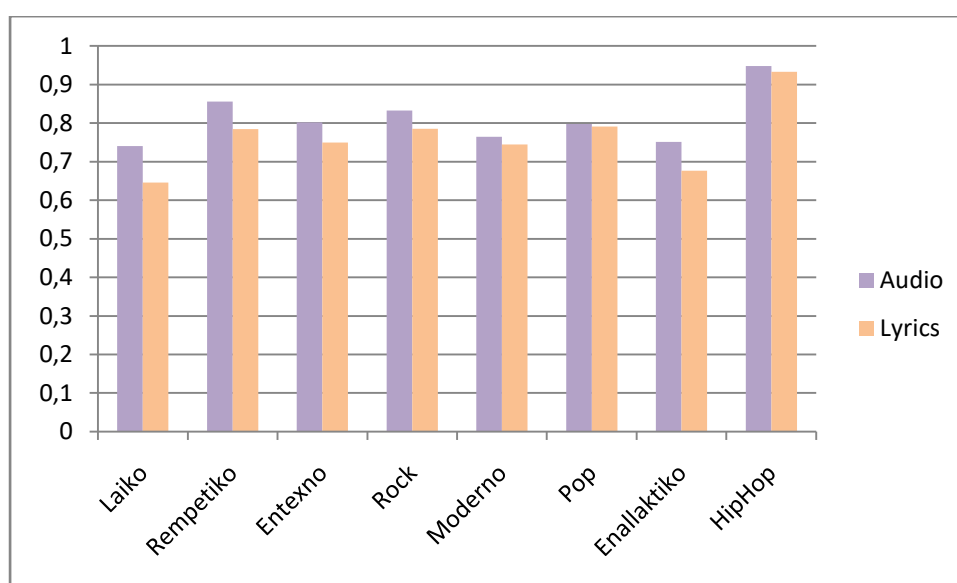
Γράφημα 16. Καμπύλη AUC-ROC Audio και Lyrics με train-test split

Τέλος, οι τιμές της ακρίβειας στα audio δεδομένα με cross-validation, είναι μεγαλύτερες στα είδη «Λαϊκό», «Ρεμπέτικο», «Έντεχνο», «Ροκ» και «Hip/Hop» κατά 0,05 μονάδες μέσο όρο, ενώ στα είδη «Μοντέρνο», «Ποπ» και «Εναλλακτικό» οι τιμές είναι όμοιες (με διαφορά χιλιοστών) με τα δεδομένα στίχων.

Με τη μέθοδο train-test split, οι τιμές της ακρίβειας στα audio δεδομένα είναι μεγαλύτερες από τα δεδομένα στίχων, σε όλα τα είδη κατά 0,05 μονάδες μέσο όρο, εκτός από το είδος «Ποπ», όπου οι τιμές είναι ίδιες με διαφορά χιλιοστών.



Γράφημα 17. Ακρίβεια Audio και Lyrics με cross-validation



Γράφημα 18. Ακρίβεια Audio και Lyrics με test-train split

6 Συμπεράσματα και μελλοντικές προεκτάσεις

Η πρώτη σύγκριση που πραγματοποιήθηκε ήταν μεταξύ των αρχικών δεδομένων και των δεδομένων μετά την επεξεργασία τους με την τεχνική SMOTE. Τα αποτελέσματα δεν είχαν μεγάλες αποκλίσεις, όμως σε γενικές γραμμές καλύτερα αποτελέσματα απέδωσαν τα επεξεργασμένα δεδομένα. Ως εκ τούτου για τη διεξαγωγή των υπολοίπων συγκρίσεων χρησιμοποιήθηκαν τα επεξεργασμένα δεδομένα.

Στη συνέχεια έγινε σύγκριση μεταξύ των βασικότερων χαρακτηριστικών του ήχου: τον ρυθμό, το τονικό ύψος και το ηχόχρωμα. Σύμφωνα με τη βιβλιογραφία τα χαρακτηριστικά τα οποία δε βασίζονται στο ηχόχρωμα όπως το τονικό ύψος και ο ρυθμός αποδίδουν χειρότερα από τα χαρακτηριστικά ηχοχρώματος στις περισσότερες περιπτώσεις. Πράγματι, καλύτερη απόδοση και με τις δύο μεθόδους αξιολόγησης είχαν τα χαρακτηριστικά ηχοχρώματος, σε όλες τις μετρικές αξιολόγησης, ενώ χειρότερα αποτελέσματα είχαν τα ρυθμικά χαρακτηριστικά.

Η τελευταία σύγκριση που έγινε ήταν μεταξύ δεδομένων ήχου και δεδομένων στίχων. Προηγούμενες έρευνες έχουν δείξει πως οι λέξεις δεν είναι τόσο αποτελεσματικές όσο ο ήχος στην ταξινόμηση της μουσικής. Τα αποτελέσματα των πειραμάτων ήταν καλύτερα για τα δεδομένα ήχου σε όλες τις μετρικές αξιολόγησης και με τις δύο μεθόδους, αλλά χωρίς μεγάλες διαφορές. Για παράδειγμα στη μετρική αξιολόγησης F1 οι διαφορές δεν ήταν μεγαλύτερες από 0.1 μονάδες. Συνεπώς οι στίχοι περιέχουν χρήσιμες πληροφορίες για την ταξινόμηση της μουσικής σε είδη, αλλά όχι στον ίδιο βαθμό με τον ήχο.

Ο αλγόριθμος Deep learning απέδωσε αρκετά ικανοποιητικά αποτελέσματα, παρεμφερή με προηγούμενων ερευνών. Λίγο καλύτερα ήταν τα αποτελέσματα που προέκυψαν από τη μέθοδο train-test split. Σε αυτό ίσως οφείλεται ότι σε αυτή τη μέθοδο χρησιμοποιείται το μεγαλύτερο μέρος του δείγματος για εκπαίδευση, και το υπόλοιπο για αξιολόγηση, ενώ με τη μέθοδο cross-validation το δείγμα χωρίζεται σε μέρη, και όλα τα μέρη χρησιμοποιούνται για εκπαίδευση και για αξιολόγηση σε κύκλους, άρα το δείγμα αξιολόγησης είναι περισσότερο αντιπροσωπευτικό.

Ως μελλοντική έρευνα θα μπορούσαν να διεξαχθούν περισσότερα πειράματα με ελληνική μουσική και μεγαλύτερα σύνολα δεδομένων με αλγορίθμους βαθιάς μάθησης. Επιπλέον θα μπορούσε να δοθεί έμφαση στα δεδομένα των στίχων, καθώς και αυτά περιέχουν χρήσιμες πληροφορίες για την ταξινόμηση. Ακόμα θα μπορούσε να δημιουργηθούν συμβολικές

αναπαραστάσεις, οι οποίες είναι σπάνια διαθέσιμες, για υπάρχοντα σύνολα δεδομένων , αλλά και νέα, καθώς από τα συμβολικά δεδομένα μπορούν να εξαχθούν υψηλού επιπέδου μουσικές έννοιες, που δύσκολα μπορούν να αποδοθούν από τους υπόλοιπους τύπους δεδομένων.

Βιβλιογραφία

1. Lidy, Thomas and Andreas Rauber. "Evaluation of Feature Extractors and Psycho-Acoustic Transformations for Music Genre Classification." ISMIR (2005).
2. Καρύδης Ιωάννης. «Εισαγωγή στην ανάκτηση και εξόρυξη μουσικής πληροφορίας.» Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών(2005).
3. Li, Tao and Mitsunori Ogihara. "Music genre classification with taxonomy." Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. 5 (2005): v/197-v/200 Vol. 5.
4. G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," in IEEE Transactions on Speech and Audio Processing, vol. 10, no. 5, pp. 293-302, July 2002.
5. Hu, Xiao, J. Stephen Downie and Andreas F. Ehmman. "Lyric Text Mining in Music Mood Classification." ISMIR (2009).
6. Hu, Xiao. "Music and Mood: Where Theory and Reality Meet." (2010). [music and mood final]
7. Besson, M., F. Fai ta, I. Peretz, A.-M. Bonnel, and J. Requin. "Singing in the Brain: Independence of Lyrics and Tunes." Psychological Science 9, no. 6 (November 1998): 494–98.
8. Li, Tao & Ogihara, Mitsunori & Peng, Wei & Shao, Bo & Zhu, Shenghuo. (2009). Music Clustering With Features From Different Information Sources. IEEE Transactions on Multimedia. 11. 477-485. 10.1109/TMM.2009.2012942
9. Thayer, R.: The biopsychology of mood and arousal. Oxford University Press, USA (1989)
10. Bergstra. Algorithms for classifying recorded music by genre. Master's thesis, Universite de ´ Montreal, 2006.
11. Aucouturier, Jean-Julien and Fran ois Pachet. "Representing Musical Genre: A State of the Art." Journal of New Music Research 32 (2003): 83 - 93.

12. Li, Tao & Tzanetakis, George. (2003). Factors in automatic musical genre classification of audio signals. 143 - 146. 10.1109/ASPAA.2003.1285840.
13. François Pachet and Daniel Cazaly. 2000. A taxonomy of musical genres. In Content-Based Multimedia Information Access - Volume 2 (RIAO '00). LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE, Paris, FRA, 1238–1245
14. N. Scaringella, G. Zoia and D. Mlynek, "Automatic genre classification of music content: a survey," in IEEE Signal Processing Magazine, vol. 23, no. 2, pp. 133-141, March 2006.
15. Nam, J., Choi, K., Lee, J., Chou, S., & Yang, Y. (2019). Deep Learning for Audio-Based Music Classification and Tagging: Teaching Computers to Distinguish Rock from Bach. IEEE Signal Processing Magazine, 36, 41-51.
16. Ι. Βλαχάβας, Π. Κεφαλάς, Ν. Βασιλειάδης, Φ. Κόκκορας, Η. Σακελλαρίου.«Τεχνητή νοημοσύνη» Γ' Έκδοση, Εκδόσεις Γκιούρδας,2006.
17. Tom M. Mitchell." The Discipline of Machine Learning". Pittsburgh(2006)
18. Ian H. Witten, Eibe Frank .Data mining: Practical machine Learning tools and Techniques(3rd. ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.(2011)
19. Li, Tao & Ogihara, Mitsunori & Li, Qi. (2003). A Comparative Study on Content-Based Music Genre Classification. 282-289. 10.1145/860435.860487.
20. Bergstra, James & Casagrande, Norman & Erhan, Dumitru & Eck, Douglas & Kégl, Balázs. (2006). Aggregate features and ADABOOST for music classification. Machine Learning. 65. 473-484. 10.1007/s10994-006-9019-7.
21. Y. Panagakis and C. Kotropoulos, "Music genre classification via Topology Preserving Non-Negative Tensor Factorization and sparse representations," 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, 2010, pp. 249-252.
22. Yang, D., Lee, W.S.: Disambiguating music emotion using software agents. In: Proc. of International Conference on Music Information Retrieval (2004)
23. Lee, Honglak & Pham, Peter & Largman, Yan & Ng, Andrew. (2009). Unsupervised feature learning for audio classification using convolutional deep belief networks. NIPS. 1096-1104.

24. Hamel, Philippe and Douglas Eck. "Learning Features from Music Audio with Deep Belief Networks." ISMIR (2010).
25. Yandre M.G. Costa, Luiz S. Oliveira, and Carlos N. Silla. 2017. An evaluation of Convolutional Neural Networks for music classification using spectrograms. Appl. Soft Comput. 52, C (March 2017), 28–38. DOI:<https://doi.org/10.1016/j.asoc.2016.12.024>
26. Choi, Keunwoo & Fazekas, George & Sandler, Mark. (2016). Automatic tagging using deep convolutional neural networks.
27. Zhang, Chiyuan & Evangelopoulos, Georgios & Voinea, Stephen & Rosasco, Lorenzo & Poggio, Tomaso. (2014). A Deep Representation for Invariance And Music Classification. Proceedings - ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing. 10.1109/ICASSP.2014.6854954.
28. Oramas, S., Barbieri, F., Nieto, O., & Serra, X. (2018). Multimodal Deep Learning for Music Genre Classification.
29. Bengani, S., Vadivel, S., & Jothi, J.A. (2019). Efficient Music Auto-Tagging with Convolutional Neural Networks. Journal of Computer Science, 15, 1203-1208. [8apo deep]
30. Ben Krose, Patrick van der Smagt. "An introduction to Neural Networks". The University of Amsterdam(1996).
31. Alvertos Benroumpi. "AffWild Net and Aff-Wild Dtabase."(2019). Διαθέσιμο στο διαδικτυακό τόπο: "<https://www.groundai.com/project/affwild-net-and-aff-wild-database/1>
32. Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, W. Philip Kegelmeyer. "SMOTE: Synthetic Minority Over-sampling Technique". Journal of Artificial Intelligence Research 16 (2002) 321–357
33. Eibe Frank, Mark A. Hall, and Ian H. Witten. DATA MINING: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Series in Data Management Systems. 4th edition
34. Grigorios Tsoumakas, Ioannis Katakis. Multi-Label Classification: An Overview. International Journal of Data Warehousing & Mining,3(3), 1-13, July–September 2007.
35. Jesse Read, Bernhard Pfahringer, Geoff Holmes, Eibe Frank. Classifier Chains for Multi-label Classification. Machine Learning Journal. Springer. Vol. 85(3), (2011).

36. Sarang Narkhede. Understanding AUC - ROC Curve.Towards Data Science. Διαθέσιμο στο διαδικτυακό τόπο : <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>
37. Renuka Joshi, Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures: How to evaluate the performance of a model in Azure ML and understanding “Confusion Metrics”,Related Posts. Διαθέσιμο στο διαδικτυακό τόπο: <https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/>
38. Berrar Daniel. Cross-Validation. 10.1016/B978-0-12-809633-8.20349-X.(2018)
39. Scheirer, Eric D. and Malcolm Slaney. “Construction and evaluation of a robust multifeature speech/music discriminator.” 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing 2 (1997): 1331-1334 vol.2.
40. Tao Li and M. Ogihara, "Toward intelligent music information retrieval," in IEEE Transactions on Multimedia, vol. 8, no. 3, pp. 564-574, June 2006.
41. Cataltepe, Z., Yaslan, Y. & Sonmez, A. Music Genre Classification Using MIDI and Audio Features. EURASIP J. Adv. Signal Process. **2007**, 036409 (2007).
42. Skowronek, Janto, Martin F. McKinney and Steven van de Par. “Ground truth for automatic music mood classification.” ISMIR (2006).
43. Laurier, Cyril, Jens Grivolla and Perfecto Herrera. “Multimodal Music Mood Classification Using Audio and Lyrics.” 2008 Seventh International Conference on Machine Learning and Applications (2008): 688-693.
44. Mayer, Rudolf & Neumayer, Robert & Rauber, Andreas. (2008). Rhyme and Style Features for Musical Genre Classification by Song Lyrics.. ISMIR 2008 - 9th International Conference on Music Information Retrieval. 337-342.
45. McKay, Cory & Fujinaga, Ichiro. (2008). Combining Features Extracted from Audio, Symbolic and Cultural Sources.. 597-602.
- 46.Mcvicar, Matt & Freeman, Tim & Bie, Tijl. (2011). Mining the Correlation between Lyrical and Audio Features and the Emergence of Mood.. Proceedings of the 12th International Society for Music Information Retrieval Conference, ISMIR 2011. 783-788.

47. Μακρής Δημοσθένης. Introducing the Greek Audio Dataset. Master thesis, Ionian University, 2014.
48. Bob L. Sturm. 2012. An analysis of the GTZAN music genre dataset. In Proceedings of the second international ACM workshop on Music information retrieval with user-centered and multimodal strategies (MIRUM '12). Association for Computing Machinery, New York, NY, USA, 7–12.
49. Ayush Kumar, Bojja Krishna, Sai Siva, G. Sumin Reddy, M.R. Rashmi, "Music genre classification", International Research Journal of Engineering and Technology (IRJET), Apr 2017.
50. Makris, D., Kermanidis, K.L., Karydis, I.: The Greek audio dataset. In: Iliadis, L., Maglogiannis, I., Papadopoulos, H., Sioutas, S., Makris, C. (eds.) Artificial Intelligence Applications and Innovations. IFIP AICT, vol. 437, pp. 165–173. Springer, Heidelberg (2014)
51. Kalpesh Patil, Rishabh Raj, Chandrakanth BKC. Music Genre Classification. Automatic speech recognition. Master thesis. Indian Institute of Technology Bombay.2017.
52. Humphrey, Eric J., Juan Pablo Bello and Yann LeCun. "Moving Beyond Feature Design: Deep Architectures and Automatic Feature Learning in Music Informatics." ISMIR (2012).
53. Humphrey, E.J., Bello, J.P., & LeCun, Y. (2013). Feature learning and deep architectures: new directions for music informatics. Journal of Intelligent Information Systems, 41, 461-481
54. J. Schmidhuber.(2015). "Deep Learning in Neural Networks: An Overview".Neural Networks. 85-117.
55. Ludovic Arnold, Sébastien Rebecchi, Sylvain Chevallier, Hélène Paugam-Moisy. An Introduction to Deep Learning. European Symposium on Artificial Neural Networks (ESANN), Apr 2011, Bruges, Belgium.
56. Ian Goodfellow,Yoshua Bengio,Aaron Courville. "Deep Learning". MIT Press(2016).
57. Michael A. Nielsen." Neural Networks and Deep Learning". Determination Press (2018).
58. S. Lang, F. Bravo-Marquez, C. Beckham, M. Hall, and E. Frank [WekaDeeplearning4j: a Deep Learning Package for Weka based on DeepLearning4j](#), In *Knowledge-Based Systems*, Volume 178, 15 August 2019, Pages 48-50. DOI: 10.1016/j.knosys.2019.04.013 ([author version](#))

59. Y. Bengio, A. Courville and P. Vincent, "Representation Learning: A Review and New Perspectives," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1798-1828, Aug. 2013.
60. Kok, J. N., E. J. Boers, W. A. Kusters, P. van der Putten and M. Poel. "Artificial Intelligence: Definition, Trends, Techniques, and Cases". Artificial intelligence(2009).