



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

ΕΡΓΑΣΤΗΡΙΟ ΕΥΦΥΩΝ ΣΥΣΤΗΜΑΤΩΝ, ΠΕΡΙΕΧΟΜΕΝΟΥ ΚΑΙ
ΑΛΛΗΛΕΠΙΔΡΑΣΗΣ

Ανάπτυξη εφαρμογής κινητού για αναγνώριση και αντικατάσταση συγχορδιών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

Χριστίνας Μπακιρτζή

Επιβλέπων : Γ. Στάμου
Αν. Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2021



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Ανάπτυξη εφαρμογής κινητού για αναγνώριση και αντικατάσταση συγχορδιών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

Χριστίνας Μπακιρτζή

Επιβλέπων : Γ. Στάμου
Αν. Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 27^η Οκτωβρίου 2021.

(Υπογραφή)

.....
Γ. Στάμου
Αν. Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....
Α.Γ. Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....
Σ. Κόλλιας
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2021

(Υπογραφή)

.....

Μπακιρτζή Χριστίνα

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

Copyright © Μπακιρτζή Χριστίνα, 2021

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Περίληψη

Ο σκοπός της παρούσας διπλωματικής εργασίας ήταν η ανάπτυξη μιας εφαρμογής κινητού Android για την αναγνώριση και την αντικατάσταση των συγχορδιών ενός μουσικού κομματιού. Οι ακολουθίες συγχορδιών είναι τα δομικά στοιχεία από τα οποία κατασκευάζεται η τονική μουσική και η ανάλυση και επεξεργασία τους αποτελεί ένα πολύ ενδιαφέρον ζήτημα που έχει απασχολήσει έντονα τον τομέα της ανάκτησης μουσικής πληροφορίας.

Υιοθετήθηκε ένα μοντέλο βαθιάς μάθησης για την αναγνώριση των συγχορδιών που αντιστοιχούν σε ένα μουσικό απόσπασμα, το οποίο επιλέχθηκε να λειτουργεί σε εξωτερικό διακομιστή και να επικοινωνεί με την εφαρμογή μέσω αιτημάτων διαδικτύου. Ταυτόχρονα, οι έξοδοι του μοντέλου υπεβλήθησαν σε επιπρόσθετη επεξεργασία, για την εξαγωγή δύο προτάσεων αντικατάστασης για κάθε μία από τις συγχορδίες. Οι προτάσεις βασίστηκαν σε στατιστικά στοιχεία που συλλέχθηκαν από ένα σύνολο δεδομένων προόδων συγχορδιών, τα οποία εξέταζαν την συχνότητα εμφάνισης διαδοχικών μουσικών διαστημάτων μεταξύ των συγχορδιών κάθε ακολουθίας. Εξετάστηκαν διάφοροι τρόποι συγκέντρωσης αλλά και χρήσης των στοιχείων αυτών, με στόχο τη μεγιστοποίηση της απόδοσης του συνολικού συστήματος στην παροχή εύηχων αλλά και ενδιαφερουσών προτάσεων αντικατάστασης.

Οι μουσικές λειτουργίες συνδυάστηκαν σε μια εφαρμογή κινητού φιλική προς το χρήστη, συμβατή με λειτουργικό σύστημα Android, η οποία παρέχει τη δυνατότητα ηχογράφησης, αποθήκευσης και αναπαραγωγής ήχων μέσω του μικροφώνου και των ηχείων της συσκευής. Η εφαρμογή επικοινωνεί με απομακρυσμένο διακομιστή ώστε ο χρήστης να λαμβάνει πληροφορίες για τις συγχορδίες που συνοδεύουν οποιοδήποτε μουσικό απόσπασμα επιλέξει να ηχογραφήσει, καθώς και για προτάσεις αντικατάστασης κάθε μίας από αυτές. Το τελικό προϊόν αξιολογήθηκε ως προς τα επιμέρους αλλά και τα συνολικά του αποτελέσματα, ενώ εξετάστηκαν και μελλοντικές επεκτάσεις του.

Λέξεις Κλειδιά: Αναγνώριση Συγχορδιών, Αντικατάσταση Συγχορδιών, Εφαρμογή Κινητού, Βαθιά Μάθηση, Νευρωνικά Δίκτυα, Ανάκτηση Μουσικής Πληροφορίας, Συγχορδία, Νότα, Ματζόρε, Μινόρε, Ηχητικό Σήμα

Abstract

The scope of this thesis was the development of an Android mobile application for identifying and substituting the chords of a musical piece. Chord sequences are the building blocks of tonal music and their analysis and processing is a very interesting issue in the field of music information retrieval.

A deep learning model was adopted to recognize the chords corresponding to a music track, which was decided to run on an external server and communicate with the mobile application through internet requests. Moreover, the model's outputs were further processed to extract two substitution proposals for each chord. The suggestions were based on statistic numbers collected from a set of chord progressions, which examined the frequency of occurrence of consecutive musical intervals between the chords of each sequence. Various ways of collecting and using these data were examined, to maximize the efficiency of the overall system in providing compatible and interesting replacement proposals.

These music features were combined in a user-friendly mobile application compatible with the Android operating system, which provides the ability to record, save and play sounds through the device's microphone and speakers. The application communicates with a remote server so that the user receives information about the chords that correspond to the music track he chooses to record, as well as suggestions for substituting each of them. The final product was evaluated in terms of its individual and overall results, while its future extensions were also examined.

Keywords: Chord Recognition, Chord Substitution, Mobile Application, Deep Learning, Neural Networks, Music Information Retrieval, Chord, Note, Major, Minor, Audio Signal

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες στους Σπύρο Κανταρέλη και Εντμουντ Ντερβάκο για την πολύτιμη βοήθειά τους και την εξαιρετική συνεργασία μας καθ' όλη τη διάρκεια εκπόνησης της παρούσας διπλωματικής εργασίας. Επίσης, ευχαριστώ πολύ τον φίλο μου Κυριάκο Πετρόλια για την ανεκτίμητη συμβολή του στην ανάπτυξη της εφαρμογής με τις ιδέες και τις παρεμβάσεις του, όπως επίσης και τον Vivien Seguy για την ευγενική παραχώρηση μέρους της δουλειάς του για το μοντέλο μηχανικής μάθησης. Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου για την υπομονή, την αγάπη και την συνεχή στήριξή τους.

Πίνακας περιεχομένων

1	Εισαγωγή.....	1
1.1	Αντικείμενο διπλωματικής.....	1
1.2	Οργάνωση κειμένου.....	2
2	Θεωρητικό υπόβαθρο	3
2.1	Μουσική Θεωρία.....	3
2.1.1	Μουσική Νότα.....	3
2.1.2	Κλίμακες και Τρόποι	4
2.1.3	Βαθμίδες και Διαστήματα.....	4
2.1.4	Μουσική Συγχορδία.....	6
2.2	Στοιχεία επεξεργασίας Σήματος	7
2.2.1	Βραχυπρόθεσμος μετασχηματισμός <i>Fourier</i>	7
2.2.2	<i>Mel-Spectrogram</i>	7
2.3	Μηχανική & Βαθιά Μάθηση	8
2.3.1	Μηχανική Μάθηση.....	8
2.3.2	Βαθιά Μάθηση.....	9
2.3.3	Βαθιά Μάθηση σε κινητές συσκευές	10
3	Αναγνώριση & Αντικατάσταση Συγχορδιών.....	11
3.1	Αναγνώριση Συγχορδιών	11
3.1.1	Μέθοδοι αναγνώρισης συγχορδιών	12
3.2	Αντικατάσταση Συγχορδιών.....	13
3.2.1	Μέθοδοι Αντικατάστασης με χρήση μουσικής θεωρίας.....	13
3.2.2	Σχετικές Έρευνες	15
4	Ανάλυση Συστήματος.....	17
4.1	Αρχιτεκτονική Συστήματος.....	17
4.2	Περιγραφή Λειτουργιών	18
4.2.1	Διακομιστής.....	18
4.2.2	Αναγνώριση Συγχορδιών.....	19
4.2.3	Αντικατάσταση Συγχορδιών	24
4.2.4	<i>Android Application</i>	25

5	Υλοποίηση.....	29
5.1	Διακομιστής / Server.....	29
5.2	Μοντέλο Αναγνώρισης Συγχορδιών	31
5.2.1	Αρχιτεκτονική μοντέλου	31
5.2.2	Εκπαίδευση μοντέλου.....	33
5.3	Σύστημα Αντικατάστασης Συγχορδιών	34
5.3.1	Απλοποίηση Συγχορδιών – N-grams.....	35
5.3.2	Αλλαγή κλειδιού - Dataset	35
5.3.3	Μετατροπή ακολουθιών σε διαστήματα.....	38
5.3.4	Επιλογή N-grams.....	39
5.3.5	Εξαγωγή αποτελεσμάτων.....	40
5.4	Android Application.....	41
6	Αξιολόγηση Συστήματος.....	44
6.1	Μοντέλο Αναγνώρισης Συγχορδιών	44
6.2	Αξιολόγηση Συστήματος Αντικατάστασης	46
6.2.1	Βάρος N-gram.....	46
6.2.2	Πίνακας Bigrams.....	47
6.2.3	Πίνακας Bigrams Μείζονα – Ελάσσονα.....	49
6.2.4	Πίνακας Trigrams.....	51
6.2.5	Πίνακας Trigrams Μείζονα – Ελάσσονα.....	52
6.2.6	Συνδυαστικός πίνακας	53
6.2.7	Τελική επιλογή - Συμπεράσματα.....	54
6.3	Έλεγχος της Εφαρμογής	56
7	Επίλογος.....	57
7.1	Σύνοψη και συμπεράσματα	57
7.2	Μελλοντικές επεκτάσεις.....	58
8	Βιβλιογραφία.....	60

Κατάλογος εικόνων

Εικόνα 2.2.1 Παράδειγμα φασματογράμματος[3]	8
Εικόνα 4.1.1 Διάγραμμα χρήσης εφαρμογής “ReChord”	18
Εικόνα 4.2.1 Διαδικασία ανάπτυξης εφαρμογής Μηχανικής Μάθησης	19
Εικόνα 4.2.2 Παράδειγμα JSON απάντησης του διακομιστή.....	24
Εικόνα 4.2.3 Αρχική διεπαφή χρήστη.....	25
Εικόνα 4.2.4 Κουμπά 'Stop'	25
Εικόνα 4.2.5 Αναπαραγωγή ήχου.....	25
Εικόνα 4.2.6 Αναμονή αποτελεσμάτων.....	26
Εικόνα 4.2.7 Δεύτερη πρόταση αντικατάστασης	27
Εικόνα 4.2.8 Πρώτη πρόταση αντικατάστασης	27
Εικόνα 4.2.9 Αναγνώριση συγχορδιών	27
Εικόνα 4.2.10 Επιβεβαίωση κλείσιμου εφαρμογής	28
Εικόνα 5.2.1 Παράδειγμα One – Hot Encoding [37]	33
Εικόνα 6.3.1 Σφάλμα Διακομιστή.....	56
Εικόνα 6.3.2 Σφάλμα σύνδεσης στο διαδίκτυο.....	56

Κατάλογος Πινάκων

Πίνακας 2.1.3.1 Βαθμίδες κλιμάκων	5
Πίνακας 2.1.3.2 Διαστήματα φθόγγων.....	5
Πίνακας 4.2.2.1 Σύγκριση μεθόδων εξαγωγής συμπερασμάτων ML σε εφαρμογές Android.	20
Πίνακας 5.3.2.1 Διαχωρισμός δεδομένων ανά μουσικό είδος	36
Πίνακας 5.3.3.1 Αντιστοίχιση ρωμαϊκών αριθμών σε αραβικούς	38
Πίνακας 5.3.3.2 Αντιστοίχιση διαστημάτων σε αριθμό ημιτόνιων.....	38
Πίνακας 5.3.5.1 Αντιστοίχιση συγχορδιών σε αραβικούς αριθμούς.....	40
Πίνακας 6.2.2.1 Αποτελέσματα πίνακα Bigrams	48
Πίνακας 6.2.3.1 Αποτελέσματα πίνακα Bigrams min/maj.....	49
Πίνακας 6.2.4.1 Αποτελέσματα πίνακα Trigrams	51
Πίνακας 6.2.5.1 Αποτελέσματα πίνακα Trigrams min/maj.....	52
Πίνακας 6.2.6.1 Αποτελέσματα πίνακα Bigrams και Trigrams min/maj.....	53
Πίνακας 6.2.7.1 Σύνοψη αποτελεσμάτων αξιολόγησης.....	54

1

Εισαγωγή

Τα τελευταία χρόνια, έχει παρατηρηθεί μια εκρηκτική αύξηση στη χρήση τεχνολογικών μέσων για την σύνθεση και την αναπαραγωγή μουσικής. Ερασιτέχνες αλλά και επαγγελματίες μουσικοί καταφεύγουν συχνά στην αναζήτηση των συγχορδιών που συνοδεύουν τα αγαπημένα τους κομμάτια, ενώ συνθέτες προσπαθούν να εμπλουτίσουν τα έργα τους υιοθετώντας εναλλακτικές προτάσεις συστημάτων αντικατάστασης και παραγωγής μουσικής. Με τη συνεχή χρήση κινητών συσκευών και έξυπνων εφαρμογών που χαρακτηρίζει την σημερινή εποχή, υπάρχει μια αυξανόμενη ζήτηση για εφαρμογές κινητού που ενσωματώνουν μεθόδους μηχανικής και βαθιάς μάθησης και προσφέρουν μουσικές υπηρεσίες στο ευρύ κοινό. Η αποδοτική ενσωμάτωση μοντέλων μηχανικής μάθησης σε κινητές συσκευές είναι ένα θέμα που απασχολεί όλο και περισσότερους ειδικούς στο χώρο της τεχνολογίας, ενώ δοκιμάζονται συνεχώς νέες προσεγγίσεις για προβλήματα μουσικής ταξινόμησης, παραγωγής και παροχής προτάσεων, τόσο ακαδημαϊκά όσο και βιομηχανικά.

1.1 Αντικείμενο διπλωματικής

Στην παρούσα εργασία αναπτύχθηκε μια εφαρμογή κινητού η οποία ενσωματώνει μεθόδους βαθιάς μάθησης για την αναγνώριση των συγχορδιών ενός μουσικού κομματιού, ενώ ταυτόχρονα εξετάστηκε η ανάπτυξη ενός συστήματος αντικατάστασης συγχορδιών συλλέγοντας στατιστικά δεδομένα, το οποίο ενσωματώθηκε επίσης στην εφαρμογή. Η αναγνώριση συγχορδιών ως τομέας έρευνας έχει ποικίλλες επιστημονικές χρήσεις, όπως την σημασιολογική ανάλυση ενός μουσικού κομματιού ή την κατάτμηση ενός τραγουδιού σε χαρακτηριστικά τμήματα για περαιτέρω ανάλυση, αλλά έχει και καθημερινές χρήσεις, για παράδειγμα τον προσδιορισμό των σωστών συγχορδιών για την αναπαραγωγή τραγουδιών από κάποιο μουσικό όργανο. Επιπλέον, η εξαγωγή προτάσεων αντικατάστασης συγχορδιών επιτρέπει τη μελωδική και αρμονική βελτίωση ενός μουσικού αποσπάσματος, εφαρμόζοντας

οικείες κλίμακες και συγχορδίες με νέους και διαφορετικούς τρόπους. Η τέχνη της αντικατάστασης συγχορδιών είναι ιδιαίτερα χρήσιμη για τη σύνθεση και την εναλλαγή τραγουδιών με πρωτότυπα και δημιουργικά στοιχεία. Μεγάλο μέρος του ευρωπαϊκού κλασικού ρεπερτορίου και η συντριπτική πλειοψηφία των τραγουδιών blues, jazz και rock βασίζονται σε μοτίβα προόδων συγχορδιών. Νέοι συνθέτες συχνά δυσκολεύονται να χρησιμοποιήσουν πιο περίτεχνα συγχορδίες, επειδή είναι λιγότερο σίγουροι για τη μέθοδο ομαλής ενσωμάτωσής τους στη μουσική με τρόπο πρωτότυπο και συγκροτημένο. Ο συνδυασμός των παραπάνω λειτουργιών σε μία εφαρμογή κινητού με φιλικό περιβάλλον προς το χρήστη, θα καθιστούσε τη δημιουργία και την αναπαραγωγή μουσικής πολύ πιο εύκολη και γρήγορη για κάθε έμπειρο αλλά και άπειρο χρήστη.

1.2 Οργάνωση κειμένου

Η παρούσα διπλωματική εργασία οργανώνεται σε οκτώ κεφάλαια.

Στο πρώτο κεφάλαιο παρουσιάζεται το αντικείμενο της διπλωματικής και η οργάνωση του κειμένου.

Στο δεύτερο κεφάλαιο γίνεται μια σύντομη αναφορά σε βασικές γνώσεις θεωρίας που χρειάζεται να έχει ο χρήστης για την βέλτιστη κατανόηση της εργασίας.

Στο τρίτο κεφάλαιο αναλύονται η αναγνώριση και η αντικατάσταση συγχορδιών σαν τομείς έρευνας, όπως επίσης και προσεγγίσεις που έχουν προηγηθεί στον χώρο της τεχνολογίας. Το τέταρτο κεφάλαιο αναφέρεται στην αρχιτεκτονική του συστήματος που αναπτύξαμε αλλά και στις λειτουργίες που επιτελεί κάθε κομμάτι του συστήματος.

Στο πέμπτο κεφάλαιο περιγράφεται η υλοποίηση του συστήματος.

Στο έκτο κεφάλαιο αξιολογείται η απόδοση και η ακρίβεια του συστήματος.

Στο έβδομο κεφάλαιο καταλήγουμε σε συμπεράσματα και προτείνουμε μελλοντικές επεκτάσεις πάνω στο αντικείμενο της εργασίας μας.

Το όγδοο κεφάλαιο αποτελεί την παράθεση της βιβλιογραφίας που χρησιμοποιήθηκε.

2

Θεωρητικό υπόβαθρο

Στο κεφάλαιο αυτό αναλύονται οι βασικές έννοιες που χρειάζεται να γνωρίζει ο αναγνώστης για να κατανοήσει πλήρως το αντικείμενο της εργασίας.

2.1 Μουσική Θεωρία

2.1.1 Μουσική Νότα

Μουσική είναι η τέχνη και η επιστήμη που ασχολείται με τους ήχους και σκοπός της είναι να εκφράζει, με κατάλληλους συνδυασμούς, ήχους, ανθρώπινες ιδέες και συναισθήματα [1]. Όπως κάθε γλώσσα έχει τα γράμματα και τους κανόνες της, έτσι και η μουσική διέπεται από τα δικά της γράμματα και τους δικούς της κανόνες, με τη διαφορά ότι στη θέση των γραμμάτων χρησιμοποιεί μουσικούς ήχους. Τους ήχους αυτούς τους ονομάζουμε μουσικούς φθόγγους ή μουσικές νότες. Κάθε νότα ορίζεται ως ένα ηχητικό κύμα με σταθερή συχνότητα ανάμεσα στα 32 με 8000 Hz. Το διάστημα που παρεμβάλλεται μεταξύ δύο μουσικών νοτών που η αναλογία συχνοτήτων τους είναι 1:2, ονομάζεται οκτάβα, και οι νότες αυτές έχουν το ίδιο όνομα καθώς έχουν πανομοιότυπα. Στη σύγχρονη δυτική μουσική, έχουν καθοριστεί 12 θεμελιώδεις συχνοότητες, οι 12 νότες, οι οποίες επαναλαμβάνονται ανά οκτάβα. Αυτό φαίνεται εύκολα στα πλήκτρα ενός πιάνο.



Τα γράμματα στο κάτω μέρος των λευκών πλήκτρων είναι ονομασίες νοτών, ενώ τα μαύρα πλήκτρα είναι επίσης νότες. Το διάστημα μεταξύ δύο διαφορετικών διαδοχικών φθόγγων της ίδιας οκτάβας το ονομάζουμε ημιτόνιο και αποτελεί το μικρότερο μουσικό διάστημα στο σύστημα που μελετάμε, ενώ το διάστημα που ισούται με δύο ημιτόνια ονομάζεται τόνος. Στην ελληνική γραφή, έχουμε τις νότες ΝΤΟ ΡΕ ΜΙ ΦΑ ΣΟΛ ΛΑ ΣΙ, ενώ στη λατινική γραφή έχουμε κατ' αντιστοιχία: C D E F G A B, οι οποίες αντιπροσωπεύουν τα λευκά πλήκτρα. Για

να μπορούμε να αναφερόμαστε σε όλο το 12-φθογγο σύστημα όμως, χρειαζόμαστε και δύο σημεία αλλοίωσης τα οποία υποδηλώνουν αν μια νότα είναι αυξημένη ή ελαττωμένη κατά ένα ημιτόνιο. Τα σημεία αυτά ονομάζονται αντίστοιχα δίεση (#) και ύφεση (b). Έτσι, για παράδειγμα η νότα Ντο δίεση (C#) αντιπροσωπεύει τη νότα που βρίσκεται ένα ημιτόνιο πάνω από τη φυσική Ντο και ταυτόχρονα ένα ημιτόνιο κάτω από τη φυσική Ρε – είναι το μαύρο πλήκτρο ανάμεσά τους. Η Ντο δίεση παρουσιάζει ηχητική ταύτιση με τη Ρε ύφεση (Db), έχουν δηλαδή την ίδια συχνότητα και την ίδια ακριβώς θέση στα πλήκτρα. Αυτοί οι φθόγγοι ονομάζονται εναρμόνιοι. Παρατηρούμε πως κάθε 12 πλήκτρα, οι ίδιες νότες επαναλαμβάνονται.

2.1.2 Κλίμακες και Τρόποι

Ορίζουμε μια νότα ως τονική. Είδαμε πως αν καλύψουμε ένα διάστημα οκτάβας, φτάνουμε ξανά σε μια νότα με το ίδιο όνομα και τόνο με την τονική αυτή. Με τον όρο κλίμακα ή σκάλα ορίζουμε τη διαδοχή συνεχόμενων φθόγγων, στην περίπτωση μας μεταξύ των επτά νοτών που αναφέρθηκαν παραπάνω, για να καλύψουμε το διάστημα αυτό προς τα πάνω αλλά και προς τα κάτω, χωρίς τη χρήση εναρμόνιων φθόγγων. Χρωματική κλίμακα ονομάζεται εκείνη που δημιουργείται με τη χρήση και των 12 μουσικών φθόγγων. Έχει σημασία ο τρόπος σχηματισμού της διαδοχής κάθε κλίμακας, ως προς τη σχέση απόστασης μεταξύ των νοτών που την απαρτίζουν. Όταν σε μια κλίμακα στην ανιούσα διαδοχή των φθόγγων σχηματίζονται στη σειρά 2 φορές τόνος, μία φορά ημιτόνιο, 3 φορές τόνος και μια φορά ημιτόνιο (T-T-H-T-T-H), ο τρόπος αυτός σχηματισμού της ονομάζεται μείζονας, και είναι ο τρόπος που ακολουθεί η κλίμακα που χρησιμοποιεί μόνο φυσικές νότες και ξεκινάει από τη νότα Ντο. Ο φυσικός ελάσσων τρόπος, που με νότες χωρίς αλλοιώσεις ξεκινάει από τη νότα Λα, αποτελείται από τη διαδοχή 1 τόνου, 1 ημιτόνιου, 2 τόνων, 1 ημιτόνιου και 2 τόνων (T-H-T-T-H-T). Χρησιμοποιώντας μόνο φυσικές νότες και με τονική κάθε φυσική νότα υπάρχουν επτά διαφορετικοί τρόποι δημιουργίας κλιμάκων, ο Μείζον/Ιωνικός (τρόπος του Ντο / major scale), ο Δώριος τρόπος (τρόπος του Ρε / dorian scale), ο Φρύγιος τρόπος (τρόπος του Μι / phrygian scale), ο Λύδιος τρόπος (τρόπος του Φα / lydian scale), ο Μιξολύδιος τρόπος (τρόπος του Σολ / mixolydian scale), ο Ελάσσων/Αιολικός τρόπος (τρόπος του Λα / minor scale) και ο Λόκριος τρόπος (τρόπος του Σι / locrian scale). Μπορούμε να κατασκευάσουμε κλίμακες οποιουδήποτε τρόπου ξεκινώντας από όποια νότα επιθυμούμε, με την προϋπόθεση ότι τηρούμε τους κανόνες διαδοχής τόνων και ημιτόνιων.

2.1.3 Βαθμίδες και Διαστήματα

Στις οκτάφθογγες κλίμακες, που δομούνται σύμφωνα με το Μείζονα ή τον Ελάσσονα τρόπο, έχει επικρατήσει η θέση που κατέχει κάθε φθόγγος στις κλίμακες αυτές να χαρακτηρίζεται ως

"Βαθμίδα", με κάθε βαθμίδα να έχει τη δική της ονομασία [2]. Υπάρχουν επτά βαθμίδες, καθώς η όγδοη νότα στην κορυφή της κλίμακας αποτελεί επανάληψη της πρώτης και θεωρητικά αντιπροσωπεύει την αρχή της επόμενης κλίμακας. Οι βαθμίδες, προσδιορίζονται συνήθως με ρωμαϊκούς αριθμούς ως εξής:

Πίνακας 2.1.3.1 Βαθμίδες κλιμάκων

Βαθμίδα	Ονομασία
I	Τονική ή 1 ^η
II	Επιτονική ή 2 ^η
III	Τρίτη ή Μέση ή 3 ^η
IV	Υποδεσπόζουσα ή 4 ^η
V	Δεσπόζουσα ή 5 ^η
VI	Επιδεσπόζουσα ή 6 ^η
VII	Προσαγωγέας ή 7 ^η

Ως διάστημα στη μουσική ορίζουμε τη διαφορά στη συχνότητα μεταξύ δύο ήχων. Ο παρακάτω πίνακας παρουσιάζει τα πιο ευρέως χρησιμοποιούμενα ονόματα για τα διαστήματα μεταξύ νοτών μιας κλίμακας 12 φθόγγων που απέχουν μεταξύ τους ένα ημιτόνιο.

Πίνακας 2.1.3.2 Διαστήματα φθόγγων

Διάστημα	Αριθμός Ημιτονίων
Τέλεια ομοφωνία	0
Δευτέρα μικρή	1
Δευτέρα μεγάλη	2
Τρίτη μικρή	3
Τρίτη μεγάλη	4
Τετάρτη καθαρή	5
Τετάρτη αυξημένη	6
Πέμπτη ελαττωμένη	6
Πέμπτη καθαρή	7
Έκτη μικρή	8
Έκτη μεγάλη	9
Εβδόμη μικρή	10
Εβδόμη μεγάλη	11

Το είδος ενός διαστήματος εξαρτάται από τον αριθμό των ημιτονίων που περιέχει. Τα είδη είναι επτά: μικρό, μεγάλο, καθαρό, ελαττωμένο, αυξημένο, δις ελαττωμένο και δις αυξημένο, με τα δύο τελευταία να χρησιμοποιούνται σπάνια. Τα διαστήματα που χαρακτηρίζονται ως μικρά ή μεγάλα δεν μπορούν να χαρακτηριστούν καθαρά και αντίστροφα τα καθαρά διαστήματα δεν μπορούν να χαρακτηριστούν μικρά ή μεγάλα. Όταν μεταβάλλεται η οξύτητα των φθόγγων που το σχηματίζουν, το διάστημα αλλοιώνεται. Το αλλοιωμένο διάστημα διατηρεί το ίδιο όνομα, αλλά αλλάζει το είδος του.

2.1.4 Μουσική Συγχορδία

Το Oxford University ορίζει τη μουσική συγχορδία ως: «Οποιοσδήποτε ταυτόχρονος συνδυασμός από νότες, αλλά συνήθως όχι λιγότερες από 3. Η χρήση συγχορδιών είναι το βασικό θεμέλιο της αρμονίας.» Αυτός ο ορισμός αποδέχεται ως συγχορδία οποιονδήποτε πολλαπλό αριθμό νοτών που παίζονται ταυτόχρονα και έτσι ένας συνδυασμός πλήκτρων στο πιάνο θεωρείται μια συγχορδία. Η συγχορδία κατέχει έναν καθοριστικό ρόλο στον τομέα της μουσικής, αφού η ταυτόχρονη ήχηση διαφορετικών τόνων, δίνει διαφορετικά ηχητικά αποτελέσματα και επιτρέπει στο συνθέτη να δώσει στο έργο του το ύφος που επιθυμεί. Είναι προφανές ότι υπάρχει δυνατότητα για πάρα πολλούς συνδυασμούς νοτών και εξ' αυτού πάρα πολλές συγχορδίες.

Ονομάζουμε θεμέλιο νότα την πρώτη νότα κάθε συγχορδίας, η οποία αποτελεί τη βάση της και είθιστα το όνομά της να χρησιμοποιείται για την αναπαράσταση της συγχορδίας. Στο τονικό μουσικό σύστημα μια συγχορδία σχηματίζεται με επάλληλους φθόγγους που απέχουν μεταξύ τους διάστημα τρίτης, δηλαδή με την προσθήκη της τρίτης, της πέμπτης, της έβδομης κ.ο.κ. νότας πάνω στο θεμέλιο φθόγγο. Οι υπόλοιποι φθόγγοι ονομάζονται τρίτη, πέμπτη, έβδομη, ένατη, κ.ο.κ. Η βασικότερη και απλούστερη μορφή συγχορδίας είναι η τρίφωνη συγχορδία σε ευθεία θέση, αυτή δηλαδή που αποτελείται από τρεις (3) νότες. Οι τρίφωνες συγχορδίες, διακρίνονται ανάλογα με τα διαστήματα που περιλαμβάνουν, σε μείζονες, ελάσσονες, αυξημένες και ελαττωμένες. Ακόμα μια διάκριση των συγχορδιών είναι σε σύμφωνες, όταν οι ήχοι τους ικανοποιούν την ακοή μας, και διάφωνες, όταν οι ήχοι τους απαιτούν μια λύση που πρέπει να καταλήξει στο άκουσμα μιας άλλης σύμφωνης. Σύμφωνες συγχορδίες είναι οι μείζονες και οι ελάσσονες, οι οποίες είναι και οι κύριες συγχορδίες που θα μας απασχολήσουν στην παρούσα εργασία.

Οι μείζονες (major - maj) συγχορδίες αποτελούνται από την πρώτη θεμέλιο νότα, η οποία δίνει και το όνομα στη συγχορδία, τη δεύτερη νότα σε απόσταση τρίτης μεγάλης από την πρώτη και την τρίτη νότα σε απόσταση πέμπτης καθαρής από την θεμέλιο νότα.

Οι ελάσσονες (minor – min) συγχορδίες αποτελούνται από τη θεμέλιο νότα η οποία δίνει το όνομα στη συγχορδία, τη δεύτερη νότα σε απόσταση τρίτης μικρής από τη βάση και την τρίτη νότα σε απόσταση πέμπτης καθαρής από την πρώτη νότα.

Όπως οι φθόγγοι μιας κλίμακας στο τονικό σύστημα, έτσι και οι συγχορδίες που σχηματίζονται πάνω σε αυτούς, χαρακτηρίζονται από τη θέση τους μέσα στην κλίμακα. Απαριθμούνται με ρωμαϊκούς αριθμούς και ονομάζονται όπως και οι βαθμίδες που έχουν για θεμέλιο.

2.2 Στοιχεία επεξεργασίας Σήματος

Στην προσπάθειά μας να εξάγουμε πληροφορίες από αρχεία ήχου και μουσικής, χρειάζεται να επεξεργαστούμε ηχητικά σήματα. Ένα ηχητικό σήμα είναι η ηλεκτρονική αναπαράσταση των μεταβολών πίεσης που μεταδίδονται στον αέρα.

2.2.1 Βραχυπρόθεσμος μετασχηματισμός Fourier

Το πιο χρήσιμο ίσως εργαλείο για την επεξεργασία ηχητικού σήματος είναι ο μετασχηματισμός Fourier. Πρόκειται για μια μαθηματική φόρμουλα που μεταφέρει το ηχητικό σήμα από το πεδίο του χρόνου στο πεδίο της συχνότητας, δημιουργώντας έτσι το φάσμα του σήματος. Ο βραχυπρόθεσμος μετασχηματισμός Fourier (Short-Time Fourier Transform – STFT) είναι ένας μετασχηματισμός που σχετίζεται με το Fourier και χρησιμοποιείται για τον προσδιορισμό της ημιτονοειδούς συχνότητας και του περιεχομένου φάσης των τοπικών τμημάτων ενός σήματος, καθώς αυτό αλλάζει με την πάροδο του χρόνου. Ουσιαστικά, όταν το σήμα δεν έχει σταθερή συχνότητα, δηλαδή δεν είναι περιοδικό, ο αλγόριθμος SFTF υπολογίζει τον γρήγορο μετασχηματισμό Fourier (Fast Fourier Transform – FFT) σε μικρά τμήματα ίσου μήκους του σήματος, τα επονομαζόμενα παράθυρα $\omega(t)$ (window segments), τα οποία επικαλύπτονται μεταξύ τους. Έχουμε τον παρακάτω τύπο:

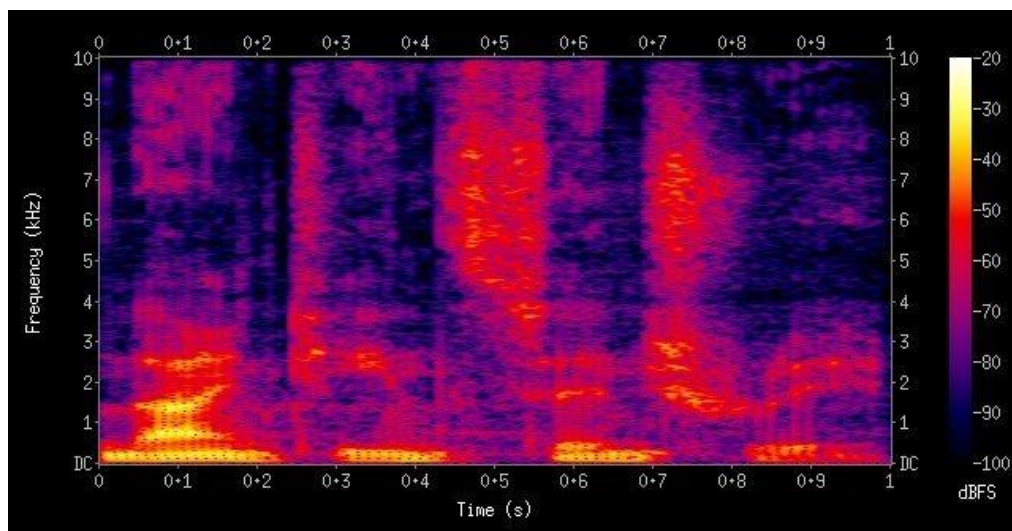
$$SFTF\{x(t)\}(\tau, \omega) \equiv X(\tau, \omega) = \int_{-\infty}^{\infty} x(t)\omega(t - \tau)e^{-i\omega t} dt$$

Το αποτέλεσμα που παράγεται είναι η οπτικοποίηση της συχνότητας του σήματος ως συνάρτηση του χρόνου, το λεγόμενο φασματογράφημα του σήματος.

2.2.2 Mel-Spectrogram

Ένα φασματογράφημα (spectrogram) σήματος αποτελεί έναν τρόπο απεικόνισης ήχων που παρέχει πληροφορίες σχετικά με τις συχνότητες που συνθέτουν τον ήχο. Πρόκειται για ένα διάγραμμα του οποίου ο οριζόντιος άξονας αντιπροσωπεύει το χρόνο ενώ ο κάθετος άξονας τη συχνότητα του σήματος. Παράλληλα δίνεται πληροφορία για το πλάτος και την ένταση του

σήματος. Τα φασματογραφήματα χρησιμοποιούνται πολύ συχνά στον χώρο της μουσικής επεξεργασίας.



Εικόνα 2.2.1 Παράδειγμα φασματογράμματος[3]

Αφού υπολογιστεί το spectrogram ενός σήματος μπορούμε να υπολογίσουμε και το mel-spectrogram. Κλίμακα mel, ονομασία που προέρχεται από τη λέξη μελωδία, ονομάζεται μία αντιληπτική κλίμακα τόνων που οι ακροατές αντιλαμβάνονται ότι ισαπέχουν μεταξύ τους. Το σημείο αναφοράς μεταξύ αυτής της κλίμακας και της μέτρησης της κανονικής συχνότητας ορίζεται με την εκχώρηση ενός αντιληπτικού τόνου 1000 mels σε έναν τόνο 1000 Hz, 40 dB πάνω από το όριο του ακροατή. Πάνω από περίπου 500 Hz, όλο και μεγαλύτερα διαστήματα κρίνεται από τους ακροατές ότι παράγουν ίδια αύξηση τόνων[4]. Ένα mel-spectrogram είναι ένα φασματογράφημα όπου οι συχνότητες μετατρέπονται στην κλίμακα mel. Προσεγγίζει τον τρόπο λειτουργίας του ανθρώπινου ακουστικού συστήματος και μπορεί να φανεί ότι το φασματογράφημα εξομαλύνεται στον τομέα συχνοτήτων, με υψηλή ακρίβεια στις χαμηλές συχνότητες και χαμηλή ακρίβεια στις υψηλές συχνότητες.

2.3 Μηχανική & Βαθιά Μάθηση

2.3.1 Μηχανική Μάθηση

Για να μπορέσουμε να ορίσουμε πλήρως τον όρο μηχανική μάθηση, χρειάζεται να γνωρίζουμε αρχικά τι είναι η μάθηση. Σύμφωνα με τον Simon (1983), "η μάθηση σηματοδοτεί προσαρμοστικές αλλαγές σε ένα σύστημα, με την έννοια ότι αυτές του επιτρέπουν να κάνει την ίδια εργασία ή εργασίες της ίδιας κατηγορίας, πιο αποδοτικά και αποτελεσματικά την επόμενη φορά" [5]. Η μηχανική μάθηση (Machine Learning - ML) είναι ένα πεδίο της επιστήμης της Τεχνητής Νοημοσύνης (Artificial Intelligence - AI) που στοχεύει στην εκμάθηση υπολογιστικών συστημάτων. Όπως ορίστηκε από τον Tom M. Mitchel (1997), «ένα πρόγραμμα

υπολογιστή λέγεται ότι μαθαίνει από μια εμπειρία E ως προς μια κλάση εργασιών T και ένα μέτρο επίδοσης P , αν η επίδοσή του στις εργασίες της κλάσης T , όπως αποτιμάται από το μέτρο P , βελτιώνεται με την εμπειρία E » [6]. Οι αλγόριθμοι μηχανικής μάθησης κατασκευάζουν ένα μαθηματικό μοντέλο με τη χρήση κάποιων δεδομένων εκπαίδευσης, το οποίο χρησιμοποιούν κατάλληλα για να καταφέρουν να προβλέψουν και να αποφασίσουν για κάποια άλλα, άγνωστα για αυτούς δεδομένα, με όσο το δυνατόν μεγαλύτερη ακρίβεια, γενικεύοντας το μοντέλο αυτό. Οι αλγόριθμοι μηχανικής μάθησης ταξινομούνται σε τρεις κύριες κατηγορίες βάσει του τρόπου εκμάθησης: την επιβλεπόμενη μάθηση (Supervised learning), τη μη επιβλεπόμενη μάθηση (Unsupervised learning) και την ενισχυμένη μάθηση (Reinforcement learning). Επίσης, ανάλογα με το επιθυμητό αποτέλεσμα, τα προβλήματα που επιλύει η μηχανική μάθηση χωρίζονται σε προβλήματα ταξινόμησης, παλινδρόμησης, συσταδοποίησης, εκτίμησης πιθανότητας και μείωσης διαστατικότητας.

2.3.2 Βαθιά Μάθηση

Ένα τεχνητό νευρωνικό δίκτυο (Artificial Neural Network) είναι ένα δίκτυο από απλούς υπολογιστικούς κόμβους, τους νευρώνες, διασυνδεδεμένους μεταξύ τους. Κάθε νευρώνας δέχεται ένα σύνολο εισόδων από διαφορετικές πηγές και παράγει μια έξοδο. Η κατανομή των νευρώνων γίνεται σε επίπεδα, από τα οποία κάποια είναι κρυφά (hidden), και καθένα κατέχει διαφορετικό ρόλο, επιτελώντας ξεχωριστές λειτουργίες.

Η Βαθιά Μάθηση (Deep Learning) αποτελεί ένα εξαιρετικά ενδιαφέρον κομμάτι της μηχανικής μάθησης και μπορεί να χαρακτηριστεί ως μια εξέλιξη των απλών νευρωνικών δικτύων. Βασίζεται στη χρήση πολύ περισσότερων κρυφών επιπέδων από αυτά των απλών νευρωνικών δικτύων, με στόχο την επίτευξη μεγαλύτερης ακρίβειας αποτελεσμάτων και την εξαγωγή χαρακτηριστικών υψηλού επιπέδου. Η βαθιά μάθηση έχει φέρει επανάσταση στον τρόπο με τον οποίο ο κόσμος επεξεργάζεται, μοντελοποιεί και ερμηνεύει δεδομένα.

2.3.2.1 Συνελκτικά Νευρωνικά Δίκτυα

Προβλήματα με πολυδιάστατα δεδομένα εισόδου δημιούργησαν την ανάγκη χρήσης πολλών επιπέδων νευρώνων με πολυάριθμες συνδέσεις μεταξύ τους, τα οποία αντιμετωπίστηκαν μέσω τεχνικών βαθιάς μάθησης και κυρίως μέσω Συνελκτικών Νευρωνικών Δικτύων (Convolutional Neural Networks – CNN). Ο όρος συνελκτικό νευρωνικό δίκτυο χρησιμοποιήθηκε για να περιγράψει την αρχιτεκτονική των νευρωνικών δικτύων τα οποία εφαρμόζονται σε δισδιάστατους πίνακες και βασίζονται σε χωρικά οριοθετημένη νευρωνική είσοδο. Η συγκεκριμένη αρχιτεκτονική έχει αναφερθεί και ως η τεχνική των μοιρασμένων βαρών ή των τοπικών πεδίων υποδοχής. Η ιδιαιτερότητά τους, οφείλεται στην ικανότητά τους

να αναγνωρίζουν συγκεκριμένα (πολλά και διαφορετικά) στοιχεία της εισόδου πετυχαίνοντας με αυτόν τον τρόπο μεγάλες ακρίβειες στην κατηγοριοποίηση.

2.3.3 Βαθιά Μάθηση σε κινητές συσκευές

Εμπνευσμένοι από την εξαιρετική απόδοση της βαθιάς μάθησης, οι άνθρωποι προσπαθούν σήμερα να ωθήσουν όλο και περισσότερο τη βαθιά μάθηση σε κινητές συσκευές για να παρέχουν έξυπνες υπηρεσίες υψηλής ποιότητας. Πιστεύεται ότι η βαθιά μάθηση θα διαδραματίσει εξαιρετικά σημαντικό ρόλο στην εξέλιξη των κινητών εφαρμογών [7]. Οι εφαρμογές που βασίζονται στη βαθιά μάθηση μπορούν να διασπαστούν σε δύο κύριες εργασίες, την εκπαίδευση και την εξαγωγή συμπερασμάτων. Στη φάση της εκπαίδευσης, ένα μεγάλο σύνολο δεδομένων εκπαίδευσης χρησιμοποιείται για την αυτόματη προσαρμογή των εκπαιδευσιμων παραμέτρων στο βαθύ νευρωνικό δίκτυο (Deep Neural Network – DNN). Έχουμε λοιπόν ένα μοντέλο DNN εκπαιδευμένο για μια συγκεκριμένη εφαρμογή το οποίο στη συνέχεια θα χρησιμοποιήσουμε για το κομμάτι εξαγωγής συμπερασμάτων, όπως για την αναγνώριση μιας εικόνας που το μοντέλο δεν έχει ξαναδεί ποτέ. Λόγω της περιορισμένης χωρητικότητας υπολογισμού και μπαταρίας, και οι δύο αυτές εργασίες επηρεάζουν σημαντικά τις κινητές συσκευές, καθώς οι διαθέσιμοι υπολογιστικοί πόροι είναι συχνά ανεπαρκείς. Αναζητούνται, λοιπόν, συχνά εναλλακτικές λύσεις και ιδέες ως προς την βελτίωση της επίδοσης των εφαρμογών και τη μείωση της επιβάρυνσης της κινητής συσκευής.

3

Αναγνώριση & Αντικατάσταση

Συγχορδιών

Στην εργασία αυτή μας απασχόλησαν δύο βασικά θέματα που αφορούν τις μουσικές συγχορδίες, η αναγνώριση των συγχορδιών ενός μουσικού κομματιού και η αντικατάστασή τους.

3.1 Αναγνώριση Συγχορδιών

Η αναγνώριση συγχορδιών είναι η εργασία "εξαγωγής ή μεταγραφής μιας ακολουθίας συγχορδιών από μια ηχογράφηση μουσικού ήχου" όπως ορίζεται από το MIREX (Music Information Retrieval Evaluation eXchange [8]), τον μεγαλύτερο διαγωνισμό για την αξιολόγηση αλγορίθμων ανάκτησης μουσικής πληροφορίας. Έχει επιστημονικές χρήσεις, όπως τη σημασιολογική ανάλυση ενός κομματιού ή την κατάτμηση ενός τραγουδιού σε χαρακτηριστικά τμήματα για περαιτέρω ανάλυση, αλλά και πιο καθημερινές, όπως όταν ένας μουσικός προσπαθεί να συνοδεύσει με την κιθάρα του ένα τραγούδι.

Η αυτόματη εκτίμηση συγχορδιών είναι πολύ χρήσιμη για την αρμονική ανάλυση της μουσικής, καθώς μόλις γίνει γνωστό το αρμονικό περιεχόμενο ενός κομματιού, μια ακολουθία συγχορδιών μπορεί να χρησιμοποιηθεί για επιπρόσθετη δομική ανάλυση υψηλότερου επιπέδου όπου μπορούν να οριστούν φράσεις ή μορφές. Οι ακολουθίες συγχορδιών είναι επίσης μια καλή αναπαράσταση μεσαίου επιπέδου μουσικών σημάτων για εφαρμογές όπως η αναζήτηση, η τμηματοποίηση και η αναγνώριση ομοιότητας μουσικής και κατηγοριοποίηση ήχου. Οι ακολουθίες αυτές αποτελούν τα δομικά στοιχεία από τα οποία κατασκευάζεται η μουσική. Επιπλέον, για προσωπική χρήση ερασιτεχνών αλλά και επαγγελματιών μουσικών, εφαρμογές αναγνώρισης συγχορδιών έχουν αποκτήσει μεγάλη δημοτικότητα στις μέρες μας, καθώς διευκολύνουν την εκμάθηση και την αναπαραγωγή μουσικών κομματιών με μουσικά όργανα, χωρίς την προϋπόθεση εύρεσης κάποιας παρτιτούρας. Η ανάπτυξη αποτελεσματικών μεθόδων αυτόματης αναγνώρισης συγχορδιών έχει απασχολήσει έντονα τον επιστημονικό και το μουσικό κλάδο τα τελευταία χρόνια, ενώ όλο και περισσότεροι ερευνητές προσελκύνονται στον τομέα της Ανάκτησης Μουσικής Πληροφορίας.

3.1.1 Μέθοδοι αναγνώρισης συγχορδιών

Παρακάτω αναλύονται οι βασικές μέθοδοι που χρησιμοποιούνται σήμερα για την αναγνώριση συγχορδιών ενός μουσικού κομματιού.

3.1.1.1 Άνθρωπος

Έμπειροι, εκπαιδευμένοι μουσικοί και άνθρωποι με έφεση στον τομέα, συχνά έχουν τη δυνατότητα να αναγνωρίζουν τις συγχορδίες που συνοδεύουν ένα μουσικό κομμάτι χωρίς τη χρήση κάποιας εφαρμογής ή κάποιου αναγνωριστικού συστήματος. Το σύνολο αυτό των μουσικών που μπορούν να αναγνωρίσουν την τονική νότα μιας συγχορδίας απλώς με το άκουσμά της, αποτελεί μειονότητα βέβαια ανάμεσα στο συνολικό αριθμό των ατόμων που ασχολούνται με τη μουσική. Γνωρίζοντας, ωστόσο, τη θεμέλιο νότα μιας συγχορδίας, ακόμα και πιο άπειροι μουσικοί μπορούν συνήθως να αναγνωρίσουν τους χρωματισμούς της. Ο συνδυασμός μουσικών γνώσεων θεωρίας με τον εντοπισμό των συναισθημάτων που προκαλεί το κάθε μουσικό κομμάτι, μπορούν να φανούν σημαντικοί αρωγοί σε αυτήν την προσπάθεια αναγνώρισης. Παρόλα αυτά, οι περισσότεροι μουσικοί καταφεύγουν τελικά στην εύρεση κάποιας παρτιτούρας ή εφαρμογής για την επιβεβαίωση της επιλογής τους.

3.1.1.2 Μηχανή

Η αυτόματη αναγνώριση συγχορδιών ήταν πάντα ένα πολύ ενδιαφέρον ερευνητικό θέμα στον τομέα της ανάκτησης πληροφοριών μουσικής (Music Information Retrieval - MIR) με μεγάλες προκλήσεις. Βασίζεται αρχικά στην εξαγωγή πληροφορίας για τις νότες που ακούγονται κάθε χρονική στιγμή, και έπειτα στην αξιοποίηση της πληροφορίας αυτής για τον καθορισμό της συγχορδίας. Δεδομένου ότι οι μουσικές συγχορδίες και οι ακολουθίες συγχορδιών είναι εξαιρετικά αφηρημένες και περιγραφικές αναπαραστάσεις της πολυφωνικής μουσικής, συνήθως παίζουν σημαντικό ρόλο και σε άλλες εργασίες του τομέα MIR, όπως στην αναγνώριση διασκευασμένων τραγουδιών [9], στην ταξινόμηση μουσικών συναισθημάτων [10], στην ταξινόμηση μουσικού είδους [11], στην ανάλυση μουσικής δομής [12] και στην μεταγραφή μελωδίας [13]. Η σωστή μοντελοποίηση της διαδικασίας μεταγραφής συγχορδιών έχει χαρακτηριστεί ως ιδιαίτερα δύσκολη, λόγω της πολύπλοκης δομής του μικτού ακουστικού σήματος και της μη προφανούς σχέσης με τα διακριτά σύμβολα συγχορδιών. Επομένως, οι περισσότερες προηγούμενες έρευνες σχετικά με την αυτόματη αναγνώριση συγχορδιών επικεντρώνονται σε μεθόδους που βασίζονται σε δεδομένα και σε μοντέλα μηχανικής μάθησης γενικής χρήσης, εφόσον μάλιστα οι παραδοσιακοί μέθοδοι επεξεργασίας σήματος φτάνουν σε τέλος. Τα τελευταία χρόνια, η ενίσχυση των υπολογιστικών δυνατοτήτων και των τεχνικών εκπαίδευσης έχουν κάνει τη βαθιά μάθηση μεγάλη επιτυχία σε διάφορους ερευνητικούς τομείς, ειδικά στην αναγνώριση ομιλίας και στην όραση υπολογιστή. Όσον αφορά την αναγνώριση

συγχορδιών, τα παραδοσιακά μοντέλα μηχανικής μάθησης κινούνται σταδιακά προς τα δίκτυα βαθιάς μάθησης, χωρίς τα τελευταία να έχουν επικρατήσει πλήρως ακόμα. Το πλεονέκτημά τους έγκειται στο γεγονός ότι τα βαθιά νευρωνικά δίκτυα μπορούν να εκπαιδευτούν σε χαρακτηριστικά χαμηλότερου επιπέδου σε σύγκριση με παλαιότερες τεχνολογίες, μειώνοντας την απαιτούμενη προεπεξεργασία των δεδομένων και αυξάνοντας την απόδοση του συνολικού δικτύου. Ερευνητικές ομάδες έχουν στρέψει την προσοχή τους στον τομέα δοκιμάζοντας διάφορες τεχνικές εξαγωγής χαρακτηριστικών [14], [15], ενώ ταυτόχρονα έχουν αναπτυχθεί και εφαρμογές που ενσωματώνουν αντίστοιχα μοντέλα [16]. Στην παρούσα εργασία εφαρμόστηκαν μέθοδοι βαθιάς μάθησης για την αναγνώριση μουσικών συγχορδιών.

3.2 Αντικατάσταση Συγχορδιών

Στη θεωρία της μουσικής, η αντικατάσταση συγχορδιών είναι η τεχνική της χρήσης μιας συγχορδίας στη θέση μιας άλλης σε μια ακολουθία συγχορδιών. Θεωρητικά, κάθε συγχορδία μπορεί να αντικαταστήσει οποιαδήποτε άλλη, αρκεί η νέα συγχορδία να υποστηρίζει τη μελωδία. Στην πράξη, όμως, μόνο μερικές επιλογές ακούγονται μουσικά και στιλιστικά κατάλληλες για μια δεδομένη μελωδία.

3.2.1 Μέθοδοι Αντικατάστασης με χρήση μουσικής θεωρίας

Ακολουθώντας τη μουσική θεωρία, υπάρχουν αρκετοί τρόποι διαφοροποίησης και αντικατάστασης μιας συγχορδίας σε μια ακολουθία συγχορδιών. Δύο είναι οι βασικοί τρόποι: η εναλλαγή της συγχορδίας χωρίς να ξεφεύγει σημαντικά από την πρωτότυπη, για παράδειγμα προσθέτοντας ή αφαιρώντας μία νότα, και η αντικατάσταση της αρχικής συγχορδίας με μία διαφορετική, που όμως συχνά οδηγεί φυσικά σε αυτή, αντικαθιστώντας μόνο ένα μέρος της διάρκειάς της. Παρακάτω αναλύονται περιεκτικά οι συνηθέστεροι μέθοδοι αντικατάστασης με βάση τη θεωρία.

3.2.1.1 Επανάληψη

Αυτός ο κανόνας επιτρέπει την επανάληψη οποιασδήποτε συγχορδίας, εφόσον η επανάληψη διαρκεί το ίδιο χρονικό διάστημα με την αρχική συγχορδία, δηλαδή κάθε επαναλαμβανόμενη συγχορδία διαρκεί το μισό χρόνο.

3.2.1.2 Εμπλουτισμός συγχορδιών

Η τζαζ μουσική τείνει να χρησιμοποιεί πιο πολύπλοκες συγχορδίες από την κλασική μουσική. Είναι, επομένως, σύνηθες να αντικαθίστανται οι απλές συγχορδίες από πιο περίπλοκες, που δημιουργούνται με την προσθήκη επιπλέον νοτών σε βασικές συγχορδίες. Για παράδειγμα, μια

συγχορδία C7 (Ντο εβδόμη) θα αντικατασταθεί συχνά από μια πιο σύνθετη C7 9 11. Ομοίως, μια συγχορδία Cm (Ντο μινόρε), η οποία περιέχει μόνο τρεις νότες, αντικαθίσταται συχνά από μια Cm7 9, με πέντε νότες. Δεδομένου ότι υπάρχουν πολλοί πιθανοί τρόποι εμπλουτισμού μιας συγχορδίας, δεν είναι πρακτικό να γραφούν όλοι ως κανόνες. Ένα παράδειγμα θα μπορούσε να είναι η αντικατάσταση μιας C7 σε C9.

3.2.1.3 Σχετική ελάσσονα

Αυτός ο κανόνας προέρχεται από την κλασική αρμονία και αντικατοπτρίζει την ισοδυναμία μεταξύ μείζονων και σχετικών ελάσσονων συγχορδιών, οι οποίες απαρτίζονται από σχεδόν το ίδιο σύνολο νοτών. Ένα παράδειγμα είναι η αντικατάσταση της συγχορδίας C (που αποτελείται από τις νότες C-E-G) από την Am (με νότες A-C-E).

3.2.1.4 Αντικατάσταση Τρίτονης

Πρόκειται ίσως για τον πιο χαρακτηριστικό κανόνα της τζαζ μουσικής, που είναι το βασικό είδος μουσικής στο οποίο χρησιμοποιούνται συνεχώς αντικαταστάσεις συγχορδιών. Δηλώνει μια ισοδυναμία μεταξύ των έβδομων συγχορδιών και του τριτόνου τους (ένα μουσικό διάστημα αποτελούμενο από τρεις παρακείμενους ολόκληρους τόνους). Ο κανόνας μπορεί να εξηγηθεί με όρους κλασικής αρμονίας. Παρότι αυτές οι δύο συγχορδίες αντιτίθενται στον κύκλο των πέμπτων, μοιράζονται την ίδια τρίτη και έβδομη. Για παράδειγμα, αντικαθιστούμε την C7 με την F#7.

3.2.1.5 Κανόνας Προετοιμασίας

Οι κανόνες προετοιμασίας επιτρέπουν την αντικατάσταση μιας συγχορδίας με δύο ή περισσότερες συγχορδίες. Για παράδειγμα, η C θα μπορούσε να αντικατασταθεί από G7 / C. Ουσιαστικά ο κανόνας επιτρέπει σε κάθε συγχορδία να «προετοιμαστεί» από την έβδομη συγχορδία της. Κάτι τέτοιο αυξάνει την αίσθηση της προόδου, χωρίς να δημιουργεί νέα αρμονία. Ένα άλλο αλλά κάπως ισοδύναμο είδος προετοιμασίας είναι η αντικατάσταση με μια μικρή έβδομη συγχορδία, αντικαθιστώντας την G7 με Dm7 / G7.

3.2.1.6 Μετάβαση στην τέταρτη συγχορδία

Αυτός ο κανόνας εισάγει τέταρτες συγχορδίες σε ακολουθίες. Οι τέταρτες συγχορδίες είναι σταθερές συγχορδίες που τονίζουν την τονικότητα της συγχορδίας που αντικαταστάθηκε. Ο κανόνας μπορεί να εκφραστεί αντικαθιστώντας την C7 με C7 / F.

3.2.1.7 Πίσω διάδοση της εβδομής

Αυτός είναι ένας πιο πολύπλοκος κανόνας που ασχολείται με την αναδρομική διάδοση των εβδομών συγχορδιών. Αυτό το φαινόμενο έχει επισημανθεί από τον Steedman (1984) [17] και χαρακτηρίζεται απαραίτητο για τη δημιουργία μιας πλήρους γραμματικής ακολουθιών συγχορδιών τζαζ. Ο κανόνας δεν τροποποιεί την ακολουθία αυτή καθαυτή, αλλά μόνο τη χρονική δομή της. Δηλώνει ότι μια εβδομη συγχορδία μπορεί με κάποιο τρόπο να μετακινηθεί προς τα πίσω στο χρόνο, τονίζοντας έτσι τον ρόλο της προετοιμασίας της, προβλέποντας την εμφάνισή της. Σε μια ακολουθία X X C7 Y η αντικατάσταση θα ήταν X C7 Y Y.

3.2.1.8 Αριστερή Διαγραφή

Τέλος, μερικές συγχορδίες μπορεί να διαγράφονται περιστασιακά, για άλλη μια φορά χωρίς να αλλάζει το αρμονικό περιεχόμενο. Αυτό συμβαίνει συνήθως μετά την εφαρμογή του προηγούμενου κανόνα.

Το σύνολο κανόνων που περιγράφεται παραπάνω δεν είναι σε καμία περίπτωση εξαντλητικό (είναι πράγματι ενδιαφέρον ερευνητικό ζήτημα η παρουσίαση ενός ελάχιστου και πλήρους συνόλου κανόνων που θα επέτρεπαν την αναδημιουργία όλων των ακολουθιών). Αποτελεί μια προσπάθεια σύνοψης των σημαντικότερων κανόνων αντικατάστασης συγχορδιών που βασίζονται στη μουσική θεωρία και χρησιμοποιούνται έντονα, ιδίως στην τζαζ και ίσως και στην pop μουσική [18]. Θέλοντας να δοκιμάσουμε μια πιο επαναστατική προσέγγιση η οποία δεν βασίζεται απλώς στη μουσική θεωρία, αλλά περισσότερο στην παρατήρηση των μοτίβων ανάμεσα σε μουσικά κομμάτια διαφόρων ειδών και προτείνει συγχορδίες που διαφέρουν περισσότερο από την αρχική, στην τρέχουσα εργασία η αντικατάσταση συγχορδιών δεν βασίστηκε στους παραπάνω κανόνες αλλά σε στατιστικά δεδομένα.

3.2.2 Σχετικές Έρευνες

Μέχρι σήμερα έχει αναπτυχθεί μια σειρά από εργαλεία που βοηθούν τους αρχάριους συνθέτες στη σύνθεση ακολουθιών συγχορδιών, προτείνοντας συγχορδίες που θα μπορούσαν να συμπληρώσουν μία ακολουθία ή να αντικαταστήσουν κομμάτια αυτής για την διαμόρφωση ενός διαφορετικού αποτελέσματος. Τα περισσότερα συστήματα έχουν σχεδιαστεί για να αντιμετωπίζουν το μουσικό πρόβλημα της εναρμόνισης μιας δεδομένης μελωδίας [19], [20], [21]. Ενώ όλα αυτά τα συστήματα χρησιμοποιούν μια προσέγγιση με γνώμονα τα δεδομένα, κάθε σύστημα έχει έναν διαφορετικό στόχο για τις προτάσεις συγχορδιών. Για παράδειγμα, έχουν αναπτυχθεί μοντέλα που επιτρέπουν στους χρήστες να προσαρμόσουν τη συνολική ποιότητα "διάθεσης" του κομματιού, μέσω ετικετών όπως το "happy" (χαρούμενο) και έτσι προσαρμόζεται το βάρος μεταξύ των μεταβατικών πινάκων ενός κρυμμένου μοντέλου Markov

για το μείζον και το ελάχισον κλειδί [21]. Οι χρήστες μπορούν επίσης να τροποποιήσουν τις συγχορδίες μεμονωμένα, αλλά οι προτεινόμενες συγχορδίες περιορίζονται μόνο στις πιο τυπικές, δεδομένου του τοπικού μουσικού πλαισίου. Αντίθετα, άλλα συστήματα επιτρέπουν στους χρήστες να εξερευνήσουν πιο αποτελεσματικά ένα ευρύ φάσμα παραλλαγών σε ακολουθίες συνοδείας. Η προσέγγιση που περιγράφεται στο [19] υιοθετεί μια υβριδική τεχνική που χρησιμοποιεί τη θεωρία της μουσικής για να καθοδηγήσει την επιλογή των συγχορδιών για συνοδεία. Διαφορετικά, το [22] παρουσιάζει στον χρήστη σύνολα διαφορετικών συγχορδιών που είναι κατάλληλα αλλά πιο σπάνια χρησιμοποιούμενα σε ένα συγκεκριμένο μουσικό πλαίσιο. Η υποστήριξη για εξερεύνηση καινοτόμων και ποικίλων εμπνευσμένων παραδειγμάτων είναι κρίσιμη, καθώς τέτοια συστήματα βοηθούν τους χρήστες να δημιουργήσουν περισσότερο καινοτόμα και ποικίλα τεχνουργήματα [23], [24].

Ταυτόχρονα, προσεγγίσεις βασισμένες στη βαθιά μάθηση έδειξαν πρόσφατα μεγάλες βελτιώσεις σε εργασίες μηχανικής μάθησης μεγάλων συνόλων δεδομένων. Ειδικά για χρονικές ακολουθίες, τα επαναλαμβανόμενα νευρωνικά δίκτυα (Recurrent Neural Networks - RNN) και τα δίκτυα μακροπρόθεσμης μνήμης (Long Short Term Memory - LSTM) έχουν αποδειχθεί ότι είναι πολύ ισχυρά μοντέλα στον τομέα της αναγνώρισης χειρόγραφων, ομιλίας και συναισθημάτων. Σήμερα, ακόμη και έρευνες παραγωγής και αντικατάστασης μουσικής έχουν προσαρμόσει όλο και περισσότερο μοντέλα RNN/LSTM σε δύο κύριες ροές - το ένα στοχεύει στη δημιουργία ολοκληρωμένων μουσικών ακολουθιών [25], [26] και το άλλο επικεντρώνεται στη δημιουργία μουσικών συμπληρωμάτων και προτάσεων, όπως είναι η ακολουθία μελωδίας, συγχορδιών και ντραμς ή οι αντικαταστάσεις μέρους αυτών [27], [14]. Ακόμα, έχουν γίνει προσπάθειες δημιουργίας μοντέλων βασισμένων σε στατιστική ανάλυση μοτίβων, σε συνδυασμό με μεθόδους μηχανικής μάθησης, τα οποία είχαν μάλιστα αρκετά ικανοποιητικά αποτελέσματα απόδοσης με προοπτικές βελτίωσης [28], [29].

4

Ανάλυση Συστήματος

Στο κεφάλαιο αυτό θα ακολουθήσει η περιγραφή της αρχιτεκτονικής του συστήματος και θα γίνει η ανάλυση απαιτήσεων των λειτουργιών του.

4.1 Αρχιτεκτονική Συστήματος

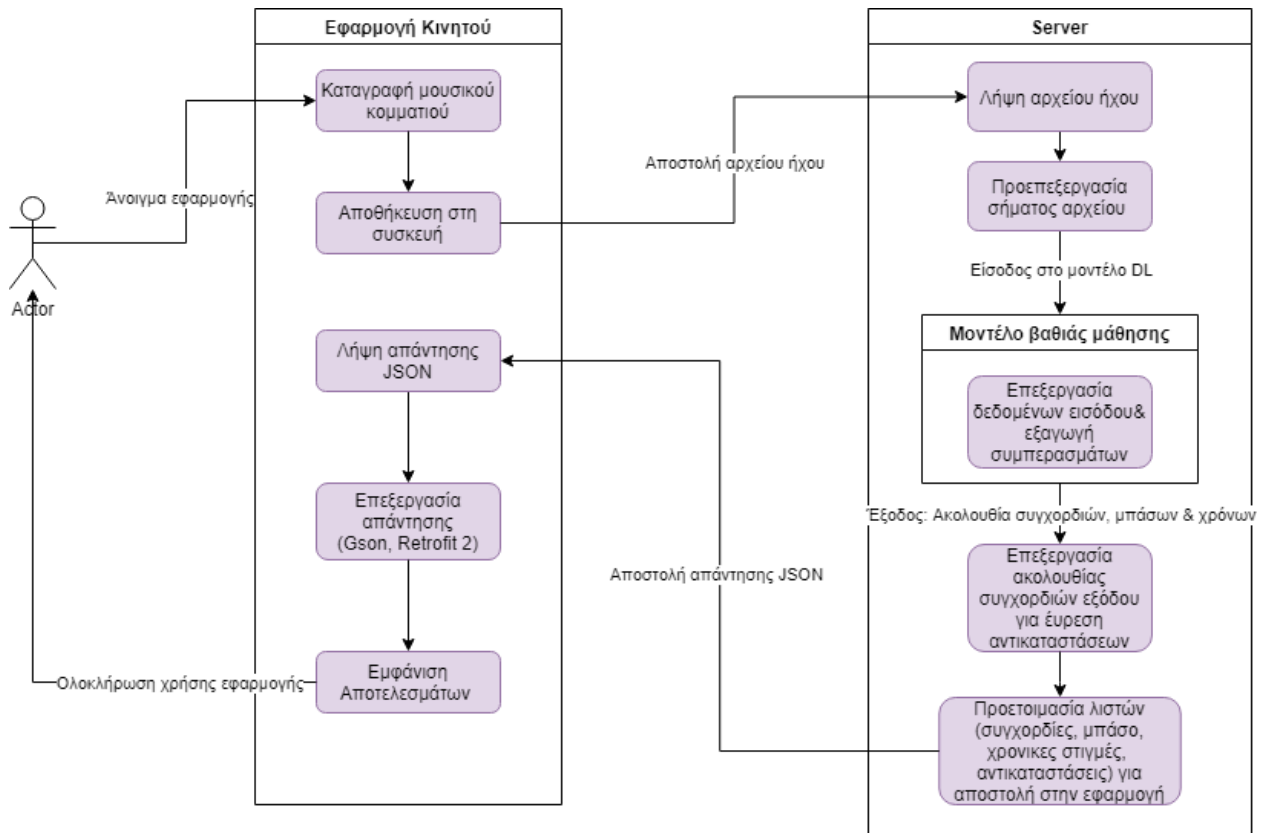
Βασικό προϊόν της εργασίας μας αποτελεί μία εφαρμογή κινητών συσκευών Android με την ονομασία “ReChord”, η οποία δίνει τη δυνατότητα στον χρήστη να ηχογραφήσει ένα μουσικό κομμάτι, να το αποθηκεύσει και να το αναπαράγει μέσα από τη συσκευή του, να λάβει πληροφορία για τις συγχορδίες του τραγουδιού αλλά και δύο εναλλακτικές προτάσεις αντικαταστάσεων για κάθε μία από τις συγχορδίες αυτές. Για την επίτευξη των παραπάνω λειτουργιών αναπτύχθηκε μια εφαρμογή, συμβατή με λογισμικό Android, μέσω της οποίας γίνεται η καταγραφή ήχου από το μικρόφωνο της συσκευής. Το αρχείο που καταγράφεται αποθηκεύεται στη συσκευή σε μορφή .wav και μεταφέρεται μέσω αιτήματος HTTP σε εξωτερικό διακομιστή (server), στον οποίο έχει δημιουργηθεί web εφαρμογή ειδικά σχεδιασμένη για να δέχεται τέτοιου είδους αιτήματα. Μετά τη λήψη του αιτήματος από τον διακομιστή, ξεκινά η προεπεξεργασία ηχητικού σήματος του αρχείου, ώστε να διαμορφωθεί κατάλληλα και να λειτουργήσει ως είσοδος σε ένα μοντέλο βαθιάς μάθησης για την αναγνώριση συγχορδιών. Το μοντέλο επεξεργάζεται τα δεδομένα εισόδου και εξάγει τρεις λίστες συμβολοσειρών – μία ακολουθία συγχορδιών, μια ακολουθία για το αντίστοιχο μπάσο κάθε συγχορδίας και μία ακολουθία αντίστοιχων χρονικών στιγμών εκκίνησης. Στη συνέχεια, η ακολουθία συγχορδιών αναλύεται από ένα python script και συγκρίνεται με συγκεκριμένους πίνακες συχνοτήτων εμφάνισης διαστημάτων μεταξύ συγχορδιών για την εύρεση προτάσεων αντικαταστάσεων. Με αυτόν τον τρόπο δημιουργούνται δύο νέες λίστες συμβολοσειρών, που περιλαμβάνουν από μία εναλλακτική πρόταση αντικατάστασης για την κάθε συγχορδία της αρχικής ακολουθίας.

Έχουμε, λοιπόν, πέντε τελικές λίστες συμβολοσειρών που αναπαριστούν:

- την ακολουθία αναγνωρισμένων συγχορδιών του μουσικού κομματιού
- το αντίστοιχο μπάσο για κάθε συγχορδία
- το χρόνο εκκίνησης κάθε συγχορδίας με βάση το χρονοδιάγραμμα του κομματιού
- την πρώτη εναλλακτική πρόταση αντικατάστασης κάθε συγχορδίας

- τη δεύτερη εναλλακτική πρόταση αντικατάστασης κάθε συγχορδίας.

Οι τελικές αυτές λίστες συγκεντρώνονται και αποστέλλονται σε μορφή JSON πίσω στην εφαρμογή, η οποία λαμβάνει τα αποτελέσματα, τα επεξεργάζεται και τα τροποποιεί κατάλληλα ώστε να τα εμφανίσει στον χρήστη. Έτσι ολοκληρώνεται ένα βασικό σενάριο χρήσης της εφαρμογής, όπως φαίνεται και στο παρακάτω διάγραμμα:



Εικόνα 4.1.1 Διάγραμμα χρήσης εφαρμογής “ReChord”

4.2 Περιγραφή Λειτουργιών

Παρακάτω αναλύονται οι λειτουργίες που επιτελεί η τελική εφαρμογή “ReChord”.

4.2.1 Διακομιστής

Στον ρόλο του απομακρυσμένου διακομιστή χρησιμοποιήθηκε μία εικονική μηχανή στην οποία με τη βοήθεια ενός microframework της python, του Flask, δημιουργήθηκε μια εφαρμογή ιστότοπου (web app). Η εφαρμογή αυτή έχει διαμορφωθεί έτσι ώστε να δέχεται αιτήματα HTTP με την IP διεύθυνσή της, και να ανταποκρίνεται είτε με κωδικό επιτυχίας HTTP 200 και κάποιο αποτέλεσμα, είτε με τον αντίστοιχο κωδικό αποτυχίας που προκάλεσε ένα τυχόν σφάλμα. Με αυτόν τον τρόπο ο διακομιστής λαμβάνει το αίτημα της εφαρμογής και χρησιμοποιεί τον κώδικα αναγνώρισης και αντικατάστασης συγχορδιών για να παράγει αποτελέσματα, τα οποία στέλνει ως απάντηση του αιτήματος που έλαβε από την εφαρμογή.

πίσω στην πηγή προέλευσης. Για την λειτουργία αυτή απαιτείται σύνδεση στο διαδίκτυο από την κινητή συσκευή, ώστε να μπορούν τα σταλούν τα αιτήματα HTTP μεταξύ εφαρμογής και διακομιστή.

4.2.2 Αναγνώριση Συγχορδίων

Χρησιμοποιώντας μεθόδους βαθιάς μάθησης, η εφαρμογή που αναπτύχθηκε παρέχει τη δυνατότητα αναγνώρισης ακολουθιών συγχορδίων, μελετώντας αρχεία ήχου που ο χρήστης καταγράφει με τη συσκευή του, μέσω μικροφώνου. Για να επιτευχθεί αυτό, δημιουργήθηκε η ανάγκη αναζήτησης τρόπων ενσωμάτωσης μοντέλων μηχανικής μάθησης σε εφαρμογές συμβατές με λογισμικό Android. Κατά την ανάπτυξη τέτοιου είδους εφαρμογών, οι κύριες εργασίες που χρειάζεται να υλοποιηθούν είναι ο σχεδιασμός προϊόντων, η δημιουργία και η εκπαίδευση του μοντέλου μηχανικής μάθησης, η εφαρμογή αλγορίθμων εξαγωγής συμπερασμάτων και η ανάπτυξη υφιστάμενων μοντέλων μηχανικής μάθησης. Ανάλογα με την περίπτωση, μπορούν να επανεκπαιδευτούν υπάρχοντα μοντέλα ή να δημιουργηθούν και να εκπαιδευτούν νέα.

Σχεδίαση	Δημιουργία και Εκπαίδευση	Εξαγωγή Συμπερασμάτων	Ανάπτυξη
προσδιορισμός των στοιχείων του προϊόντος και χρήση μοτίβων σχεδίασης ML για την επίτευξή τους	Δημιουργία μοντέλου μηχανικής μάθησης ή επιλογή προεκπαιδευμένων μοντέλων	Ανάλυση δεδομένων χρησιμοποιώντας εκπαιδευμένα μοντέλα που εκτελούνται σε συσκευή Android ή στο cloud.	Εγκατάσταση και ενημέρωση μοντέλων ML για την εφαρμογή.

Εικόνα 4.2.1 Διαδικασία ανάπτυξης εφαρμογής Μηχανικής Μάθησης

Η εξαγωγή συμπερασμάτων (inference) είναι η διαδικασία χρήσης ενός μοντέλου μηχανικής μάθησης που έχει ήδη εκπαιδευτεί για την εκτέλεση μιας συγκεκριμένης εργασίας. Μια βασική απόφαση που κληθήκαμε να λάβουμε ως προγραμματιστές λογισμικού Android στη συγκεκριμένη εργασία ήταν αν η εξαγωγή συμπερασμάτων θα εκτελούνταν στην ίδια τη συσκευή ή θα χρησιμοποιούσε κάποια υπηρεσία cloud ή έναν απομακρυσμένο διακομιστή (server) στον οποίο θα μπορούσαμε να έχουμε πρόσβαση από απόσταση. Πρόκειται για ένα ζήτημα που απασχολεί όλους τους προγραμματιστές εφαρμογών κινητού στις οποίες χρησιμοποιείται η μηχανική μάθηση, καθώς και οι δύο επιλογές έχουν προτερήματα και μειονεκτήματα ανάλογα με τις περιπτώσεις χρήσης και τα χαρακτηριστικά της εκάστοτε εφαρμογής.

Ακολουθούν ορισμένα από τα ζητήματα που χρειάστηκε να λάβουμε υπόψιν κατά τη λήψη αυτής της απόφασης:

Πίνακας 4.2.2.1 Σύγκριση μεθόδων εξαγωγής συμπερασμάτων ML σε εφαρμογές Android

Ζήτημα	Εξαγωγή Συμπερασμάτων απευθείας στη συσκευή	Εξαγωγή συμπερασμάτων σε υπηρεσία cloud / server
Καθυστέρηση	Η χαμηλότερη καθυστέρηση ενισχύει την εμπειρία σε πραγματικό χρόνο	Η ασύγχρονη επικοινωνία και το διαθέσιμο εύρος ζώνης μπορούν να επηρεάσουν την καθυστέρηση
Υπολογιστικοί Πόροι	Οι πόροι της συγκεκριμένης συσκευής, όπως η επεξεργαστική ισχύς και ο αποθηκευτικός χώρος, μπορούν να περιορίσουν την απόδοση	Οι πόροι που βασίζονται στο cloud είναι πιο ισχυροί και ο αποθηκευτικός χώρος είναι περισσότερος
Offline/Online	Η δυνατότητα λειτουργίας εκτός σύνδεσης είναι ένα πλεονέκτημα για λειτουργία με κακή ή μη υπάρχουσα υποδομή δικτύου	Απαιτείται σύνδεση δικτύου
Κόστος	Χρήση μπαταρίας, χρόνος λήψης μοντέλου για τους τελικούς χρήστες, δέσμευση υπολογιστικών πόρων	Εύρος ζώνης για μεταφορά δεδομένων για τους τελικούς χρήστες, υπολογιστικές χρεώσεις για προγραμματιστές
Ασφάλεια - Ιδιωτικότητα	Τα δεδομένα του χρήστη δεν φεύγουν ποτέ από τη συσκευή, ωστόσο το μοντέλο ML είναι πιο ευάλωτο στους ανταγωνιστές με αντίστροφη μηχανική	Τα δεδομένα ενδέχεται να φύγουν από τη συσκευή, μπορεί να απαιτηθούν πρόσθετες προφυλάξεις. Μοντέλο προστατευμένο από ανταγωνιστές/hackers.
Ενημερώσεις μοντέλου	Ενδέχεται να χρειάζεται περισσότερος χρόνος και υπολογιστικοί πόροι για την ενημέρωση του μοντέλου	Άμεση ενημέρωση μοντέλου στο cloud/server που ενσωματώνεται σε όλες τις κινητές συσκευές αυτόματα

4.2.2.1 Εξαγωγή Συμπερασμάτων στη συσκευή

Για την εξαγωγή συμπερασμάτων ενός μοντέλου μηχανικής μάθησης απευθείας στην κινητή συσκευή, γίνεται φόρτωση των τροποποιημένων παραμέτρων του μοντέλου στην εφαρμογή. Για να κάνει μια πρόβλεψη, η εφαρμογή εκτελεί όλους τους υπολογισμούς συμπερασμάτων

τοπικά στη συσκευή, με τη δική της CPU ή GPU χωρίς να χρειάζεται να επικοινωνήσει με κάποιο διακομιστή. Ένας αριθμός frameworks βαθιάς μάθησης, όπως το Caffe2 [30] και το TensorFlow Lite¹, υποστηρίζουν την ενσωμάτωση μοντέλων μηχανικής & βαθιάς μάθησης απευθείας σε φορητές συσκευές. Αυτά τα frameworks υπολογίζουν τα συμπεράσματα χρησιμοποιώντας εξαγόμενα μοντέλα που έχουν εκπαιδευτεί σε ισχυρούς διακομιστές.

Η ανεξαρτησία από τη σύνδεση δικτύου είναι το βασικό προτέρημα αυτής της επιλογής. Δεν απαιτείται σύνδεση στο διαδίκτυο, καθώς δεν χρειάζεται να σταλεί κάποιο αίτημα μέσω διαδικτύου ούτε και η αναμονή απάντησης - αντίθετα, το αποτέλεσμα λαμβάνεται (σχεδόν) αμέσως, ανάλογα βέβαια και με τα χαρακτηριστικά της συσκευής. Ταυτόχρονα, τα δεδομένα που εισάγονται στο μοντέλο παραμένουν στη συσκευή καθ' όλη τη διάρκεια της διαδικασίας και δεν εξέρχονται από αυτήν, επομένως δεν υπάρχει ο ίδιος κίνδυνος υποκλοπής τους με την μεταφορά στο δίκτυο. Ωστόσο, ακόμη και με τις βελτιστοποιήσεις που πραγματοποιούνται με τις βιβλιοθήκες λογισμικού που προαναφέρθηκαν, η εξαγωγή συμπερασμάτων απευθείας στη συσκευή σε ορισμένες περιπτώσεις μπορεί να είναι πιο αργή από την εκτέλεση της ίδιας διαδικασίας σε ισχυρούς διακομιστές. Αυτά τα κενά απόδοσης οφείλονται κυρίως σε περιορισμούς στο υλικό παλαιότερων κινητών συσκευών, π.χ. σχετικά μικρή μνήμη και έλλειψη GPU. Η αναποτελεσματικότητα επιδεινώνεται όταν μια εφαρμογή χρειάζεται ένα πιο ακριβές μοντέλο ή χρειάζεται να φορτώσει πολλά μοντέλα, π.χ., συνδέοντας την εκτέλεση ενός μοντέλου οπτικής αναγνώρισης χαρακτήρων και ενός μοντέλου μετάφρασης κειμένου. Για να διασφαλιστεί ότι η εξαγωγή συμπερασμάτων ολοκληρώνεται σε εύλογο χρονικό διάστημα, συγκεκριμένα μοντέλα για κινητά [31] θυσιάζουν συχνά την ακρίβεια συμπερασμάτων και έτσι μπορούν να αποκλείσουν πολύπλοκα σενάρια εφαρμογών, ή επιλέγουν να αποκλείσουν χρήστες με λιγότερο προηγμένες συσκευές από την επιλογή εγκατάστασης και χρήσης της εφαρμογής.

Λόγω της φύσης της μηχανικής μάθησης, τα μοντέλα που εκπαιδεύονται μπορεί να γίνουν αρκετά μεγάλα. Αν και αυτό δεν αποτελεί πρόβλημα όταν εκτελούνται σε απομακρυσμένο διακομιστή, ίσως κάτι τέτοιο να είναι αρκετά περιοριστικό όταν εκτελούνται σε έναν υπολογιστή - πελάτη. Οι κινητές συσκευές είναι πιο περιορισμένες όσον αφορά την αποθήκευση, τη μνήμη, τους υπολογιστικούς πόρους και τους περιορισμούς κατανάλωσης ενέργειας. Ως αποτέλεσμα, τα μοντέλα στη συσκευή πρέπει να είναι πολύ μικρότερα από τα αντίστοιχα που εκτελούνται σε κάποια απομακρυσμένη υπηρεσία cloud. Τυπικά, αυτό σημαίνει ότι είναι λιγότερο ισχυρά. Παρόλο που η πλήρης ενσωμάτωση του μοντέλου στη συσκευή είναι μια εύλογη επιλογή για απλές εργασίες και νεότερες κινητές συσκευές, είναι λιγότερο κατάλληλη για πολύπλοκες εργασίες και παλαιότερες κινητές συσκευές [32].

¹ <https://www.tensorflow.org/lite>

4.2.2.2 Εξαγωγή συμπερασμάτων σε υπηρεσία cloud ή απομακρυσμένο server

Εναλλακτικά, οι εφαρμογές κινητού μπορούν επίσης να στείλουν τα αιτήματά τους για εξαγωγή συμπερασμάτων σε συγκροτημένα μοντέλα σε απομακρυσμένες υπηρεσίες cloud και διακομιστές. Ένας αριθμός συστημάτων εξυπηρέτησης μοντέλων [33] έχει προταθεί για τη διαχείριση διαφορετικών εκδόσεων μοντέλων που διατίθενται για την διαδικασία αυτή. Αυτά τα συστήματα συχνά επικεντρώνονται στη μεγιστοποίηση της απόδοσης εξαγωγής συμπερασμάτων με την ομαδοποίηση εισερχόμενων αιτημάτων. Κάτι τέτοιο, βέβαια, μπορεί να αυξήσει τον χρόνο αναμονής ορισμένων αιτημάτων και κατά συνέπεια να έχει αρνητικές επιπτώσεις σε αιτήματα από άκρο σε άκρο. Για τη χρήση τέτοιων συστημάτων, οι προγραμματιστές πρέπει να καθορίσουν με μη αυτόματο τρόπο το ακριβές μοντέλο DNN (DotNetDuke - σύστημα διαχείρισης περιεχομένου ιστού) που θα χρησιμοποιούν μέσω εκτεθειμένων endpoints API. Για προγραμματιστές κινητών συσκευών, αυτή η χειροκίνητη επιλογή μοντέλου δεν λαμβάνει υπόψιν το αντίκτυπο των δυναμικών συνθηκών δικτύου κινητής τηλεφωνίας, οι οποίες μπορούν να καταλάβουν σημαντικό μέρος του χρόνου εξαγωγής συμπερασμάτων [34]. Τέτοιες στατικές προσπάθειες χρόνου ανάπτυξης μπορεί να οδηγήσουν στην επιλογή ενός ταχύτερου αλλά λιγότερο ακριβούς μοντέλου DNN ή να διακινδυνεύσουν παραβιάσεις SLA (Service-Level Agreement) επιλέγοντας ένα πιο εξελιγμένο μοντέλο DNN [35].

Η εξαγωγή συμπερασμάτων που βασίζεται σε εφαρμογές cloud παρέχει τη δυνατότητα να υποστηρίζεται μια πληθώρα σεναρίων εφαρμογών, απλών και σύνθετων, σε διάφορες κινητές συσκευές, παλιές και νέες. Ένα από τα μεγάλα οφέλη αυτής της λύσης είναι η δυνατότητα ενσωμάτωσης στην εφαρμογή βελτιωμένων μοντέλων άμεσα – χρειάζεται μόνο η ενημέρωση του μοντέλου στον διακομιστή. Αν το μοντέλο βρίσκεται αυτούσιο στη συσκευή, κάτι τέτοιο δεν είναι τόσο απλό, αφού πρέπει να προωθηθεί με κάποιο τρόπο το βελτιωμένο μοντέλο σε όλες τις υπάρχουσες εγκατεστημένες εκδόσεις εφαρμογών. Επιπλέον, όταν το μοντέλο δεν ενσωματώνεται στην εφαρμογή δεν υπάρχει ο φόβος υποκλοπής του αλγόριθμου με αντίστροφη μηχανική από ανταγωνιστές, καθώς βρίσκεται προστατευμένο σε συγκεκριμένο διακομιστή. Ωστόσο πρέπει να δοθεί προσοχή στα πρωτόκολλα ασφαλείας που χρησιμοποιεί η εφαρμογή, ιδιαίτερα όταν μεταφέρει προσωπικά ή ευαίσθητα δεδομένα στο δίκτυο.

4.2.2.3 Εξαγωγή συμπερασμάτων στην εφαρμογή “ReChord”

Είναι προφανές ότι και οι δύο επιλογές έχουν βάσιμα προτερήματα και μειονεκτήματα, ενώ δεν υπάρχει σαφής απάντηση για το ποιά μέθοδος είναι καλύτερη. Οι περιπτώσεις χρήσης και τα χαρακτηριστικά της εκάστοτε εφαρμογής ορίζουν εάν η βέλτιστη λύση είναι η ενσωμάτωση των μοντέλων μηχανικής και βαθιάς μάθησης στη συσκευή ή αν είναι προτιμότερη η χρήση κάποιας υπηρεσίας cloud ή κάποιου διακομιστή. Λαμβάνοντας υπόψιν ότι η πλήρης

ενσωμάτωση του μοντέλου στην εφαρμογή έχει αρκετούς περιορισμούς, επιλέξαμε να ακολουθήσουμε την επιλογή της εξαγωγής συμπερασμάτων σε απομακρυσμένο διακομιστή, δημιουργώντας μια εικονική μηχανή με την υπηρεσία του Okeanos². Η επιλογή αυτή έγινε μετά από αρκετές δοκιμές και των δύο ενδεχομένων ενσωμάτωσης τεχνητής νοημοσύνης σε Android εφαρμογή. Συγκεκριμένα, δοκιμάσαμε την πλήρη ενσωμάτωση του μοντέλου στην συσκευή με τη βοήθεια βιβλιοθηκών όπως οι onnxruntime³, pytorch⁴, DNNLibrary⁵ και Tflite⁶, με διάφορες μετατροπές στον κώδικά μας, όμως πάντα υπήρχαν προβλήματα με το μέγεθος του μοντέλου και τις λειτουργίες που επιθυμούσαμε να εκτελεί αυτό. Αντίθετα, επικοινωνώντας με απομακρυσμένο διακομιστή, θυσιάζουμε ίσως λίγο την άμεση διαθεσιμότητα, όμως τα αποτελέσματα είναι ακριβή και μπορούμε να ενημερώνουμε το μοντέλο βαθιάς μάθησης οποιαδήποτε στιγμή, χωρίς να επηρεάζεται η εφαρμογή, ανεξάρτητα από τη συσκευή στην οποία λειτουργεί, και χωρίς να περιορίζουμε το κοινό που μπορεί να χρησιμοποιήσει την εφαρμογή απαιτώντας συγκεκριμένους υπολογιστικούς πόρους και δυνατότητες.

Όπως προαναφέρθηκε, η εφαρμογή μεταφέρει μέσω αιτήματος HTTP POST ένα αρχείο ήχου, το οποίο περιέχει τη μουσική πληροφορία που έχει καταγράψει ο χρήστης με το μικρόφωνο της συσκευής του. Στη συνέχεια, αφού το αρχείο φτάσει στον διακομιστή, γίνεται προεπεξεργασία του ηχητικού σήματος ώστε να διαμορφωθεί κατάλληλα και να εισαχθεί στο μοντέλο βαθιάς μάθησης. Το μοντέλο έχει εκπαιδευτεί ώστε να μπορεί να αναγνωρίσει την ακολουθία συγχορδιών που συνοδεύει ένα μουσικό κομμάτι και παράγει ως έξοδο τρεις λίστες συμβολοσειρών, αναγνωρίζοντας μόνο μείζονες και ελάσσονες συγχορδίες. Η πρώτη λίστα περιλαμβάνει τις συγχορδίες, η δεύτερη το αντίστοιχο μπάσο για κάθε συγχορδία και η τρίτη λίστα το χρόνο εκκίνησης της κάθε συγχορδίας, με βάση το χρονοδιάγραμμα που ορίζει το μουσικό αρχείο, κατ' απόλυτη αντιστοιχία δεικτών με την πρώτη λίστα. Οι λίστες αυτές εισάγονται στο μοντέλο αντικατάστασης συγχορδιών και μετά την απαραίτητη επεξεργασία στέλνονται πίσω στο χρήστη, δίνοντάς του τις επιθυμητές πληροφορίες για τις συγχορδίες του μουσικού κομματιού που κατέγραψε.

² <https://okeanos.grnet.gr/home/>

³ <https://onnxruntime.ai/>

⁴ <https://pytorch.org/>

⁵ <https://github.com/JDAI-CV/DNNLibrary/>

⁶ <https://www.tensorflow.org/lite/>

4.2.3 Αντικατάσταση Συγχορδιών

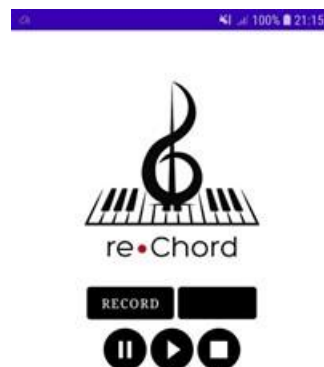
Μια βασική λειτουργία της εφαρμογής που πραγματεύεται η συγκεκριμένη εργασία είναι η παροχή ενός συστήματος προτάσεων αντικαταστάσεων συγχορδιών για κάθε μουσικό κομμάτι που έχει ηχογραφήσει ο χρήστης. Με την ηχογράφηση ενός μουσικού κομματιού και την αποστολή του στο μοντέλο μηχανικής μάθησης, ο χρήστης λαμβάνει πληροφορία για τις συγχορδίες που συνοδεύουν το μουσικό αρχείο που δημιουργήθηκε. Με δεδομένη πλέον μια ακολουθία συγχορδιών, η εφαρμογή προτείνει δύο εναλλακτικές επιλογές για κάθε μία από τις συγχορδίες αυτές. Ουσιαστικά, δημιουργήθηκαν πίνακες συχνοτήτων με βάση ένα σύνολο δεδομένων ακολουθιών συγχορδιών αποτελούμενο από διάφορα μουσικά είδη. Οι πίνακες αυτοί καταγράφουν τη συχνότητα εμφάνισης συνδυασμών συγχορδιών ως προς τα διαστήματα που αυτές απέχουν μεταξύ τους και χρησιμοποιήθηκαν για την εξαγωγή συμπερασμάτων στην εφαρμογή. Κάθε συγχορδία της ακολουθίας που εξάγεται από το μοντέλο βαθιάς μάθησης εξετάζεται ως προς τα διαστήματα που απέχει από τις γειτονικές της συγχορδίες στην ακολουθία και επιλέγεται για αυτήν, ως προτεινόμενη αντικαταστάτρια, η συγχορδία που έχει τη μεγαλύτερη συχνότητα εμφάνισης σε αντίστοιχη θέση με βάση τους πίνακες. Μάλιστα προτείνεται και μία δεύτερη συγχορδία για να υπάρχει ποικιλία στις επιλογές του χρήστη. Με αυτόν τον τρόπο, δημιουργούνται δύο νέες λίστες συγχορδιών, με μήκος ίδιο με την αρχική ακολουθία, οι οποίες περιλαμβάνουν για κάθε συγχορδία της πρωτότυπης ακολουθίας τις αντίστοιχες προτεινόμενες συγχορδίες αντικατάστασης. Η μία λίστα περιλαμβάνει τις πρώτες προτάσεις αντικατάστασης ενώ η δεύτερη τις δεύτερες προτάσεις αντικατάστασης κάθε συγχορδίας. Οι λίστες αυτές αποστέλλονται πίσω στην εφαρμογή σε μορφή JSON και παρουσιάζονται κατάλληλα στον χρήστη μέσα από τη διεπαφή χρήστη. Ένα παράδειγμα της απάντησης JSON που στέλνει ο διακομιστής στην εφαρμογή παρουσιάζεται στην παρακάτω εικόνα:

```
{"Chords": ["G", "C#", "G", "Em", "C"],  
  "Bass": ["D", "D", "G", "E", "C"],  
  "Time": ["0.22", "0.59", "0.96", "3.19", "5.05"],  
  "AlternativeChords": ["Fm", "Am", "C", "Cm", "D"],  
  "AlternativeChords2": ["F", "B", "B", "E", "C#m"]}
```

Εικόνα 4.2.2 Παράδειγμα JSON απάντησης του διακομιστή

4.2.4 Android Application

Η εφαρμογή Android που συγκεντρώνει όλες τις παραπάνω λειτουργίες ονομάστηκε “ReChord” και έχει ως στόχο την διευκόλυνση μουσικών στην αναγνώριση των συγχορδίων που συνοδεύουν τα αγαπημένα τους μουσικά κομμάτια και στην πρόταση αντικαταστάσεων για κάθε μία από αυτές. Η αρχική διεπαφή χρήστη της εφαρμογής αποτελείται από το λογότυπο της εφαρμογής και τα κουμπιά Record και Stop, όπως επίσης και από τα κουμπιά συμβόλων παύσης, έναρξης και τερματισμού αναπαραγωγής ενός μουσικού κομματιού. Ανάλογα με το σενάριο χρήσης, εμφανίζονται περισσότερες δυνατότητες και επιλογές στον χρήστη.



Εικόνα 4.2.3 Αρχική διεπαφή χρήστη

- **Record:** Πατώντας το κουμπί ‘Record’, ο χρήστης ξεκινάει μια διαδικασία ηχογράφησης μέσω του μικροφώνου της συσκευής του. Απαραίτητη προϋπόθεση αποτελεί η συναίνεσή του στην απόκτηση δικαιώματος χρήσης του μικροφώνου από την εφαρμογή. Κατά τη διάρκεια της ηχογράφησης, τα κουμπιά αναπαραγωγής ήχου είναι απενεργοποιημένα και μόνο το κουμπί ‘Stop’ ενεργοποιείται.



Εικόνα 4.2.4 Κουμπί ‘Stop’

- **Stop:** Το κουμπί ‘Stop’ ενεργοποιείται μόνο όταν ο χρήστης ξεκινήσει κάποια ηχογράφηση, πατώντας το κουμπί ‘Record’. Το ‘Stop’ σταματάει την ηχογράφηση και αποθηκεύει σε ένα αρχείο .wav με τυχαίο κωδικό ονόματος που ακολουθείται από την ετικέτα “_recording.wav” στον εσωτερικό χώρο αποθήκευσης της συσκευής. Ο χρήστης έχει ερωτηθεί για την παραχώρηση στην εφαρμογή άδειας ανάγνωσης και γραφής (read και write) στον εσωτερικό χώρο της συσκευής του. Κατά την ολοκλήρωση της ηχογράφησης ενεργοποιείται πλέον το



κουμπί ‘Play’, ενώ εμφανίζεται και ένα ακόμα κουμπί ‘Find Chords’.

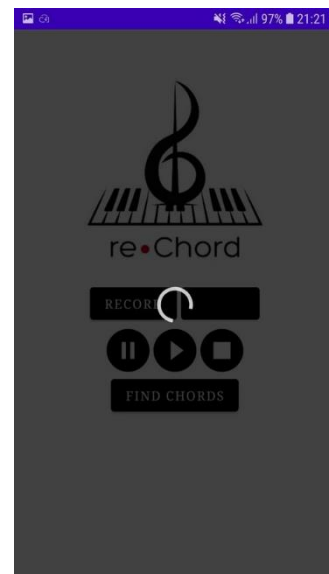
- **Play Recording:** Το μαύρο κυκλικό κουμπί με το λευκό τρίγωνο που βρίσκεται στη μέση της διεπαφής συμβολίζει, όπως είθισται, το κουμπί αναπαραγωγής μουσικής. Αφού ο χρήστης έχει ολοκληρώσει μια



Εικόνα 4.2.5 Αναπαραγωγή ήχου

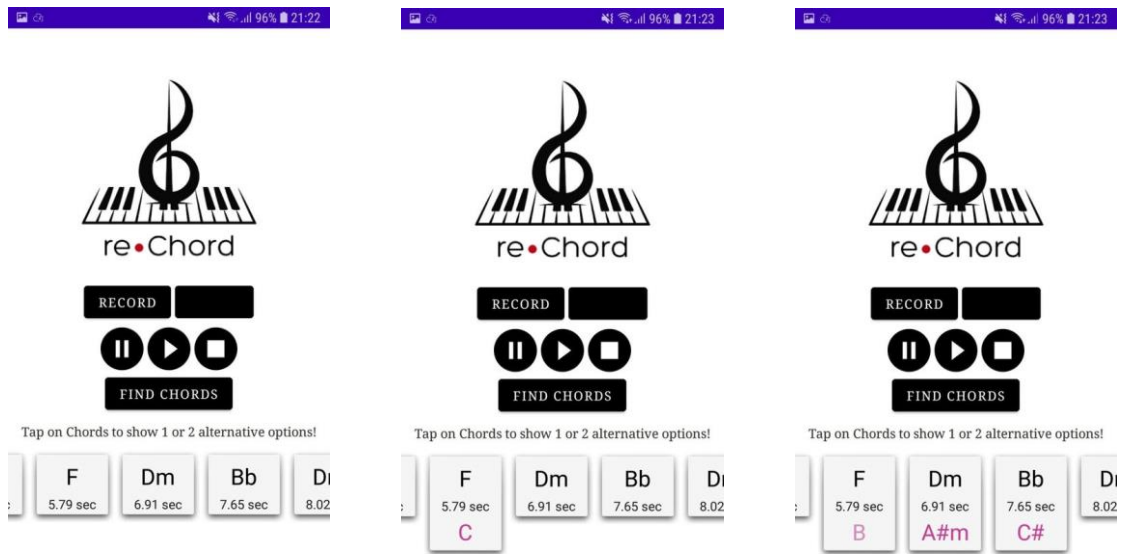
εγγραφή ήχου, πατώντας το συγκεκριμένο κουμπί μπορεί να αναπαράγει όσες φορές θέλει τον ήχο που αποθήκευσε τελευταίο στην συσκευή του. Μάλιστα, κατά την αναπαραγωγή εμφανίζεται στην οθόνη ένας οπτικοποιητής ήχου, όπως φαίνεται στην εικόνα, ο οποίος ρυθμίζει την κίνησή του ανάλογα με τη συχνότητα και την ένταση του μουσικού κομματιού την κάθε χρονική στιγμή.

- **Pause Playing Recording:** Το αριστερό κυκλικό κουμπί με τις δύο κάθετες λευκές γραμμές συμβολίζει την παύση αναπαραγωγής ήχου. Είναι ενεργοποιημένο μόνο όταν βρίσκεται σε εξέλιξη κάποια αναπαραγωγή μουσικού αρχείου. Δίνει τη δυνατότητα στο χρήστη να σταματήσει προσωρινά την αναπαραγωγή ήχου, όμως πατώντας ξανά το κουμπί 'Play', ο χρήστης να μπορεί να συνεχίσει να ακούει το μουσικό κομμάτι από το σημείο στο οποίο σταμάτησε.
- **Stop Playing Recording:** Το τελευταίο κουμπί που αφορά την αναπαραγωγή ήχου είναι αυτό που συμβολίζεται με ένα λευκό τετράγωνο και χρησιμοποιείται για τον τερματισμό αναπαραγωγής ήχου. Πατώντας το κουμπί αυτό, ο χρήστης τερματίζει την αναπαραγωγή, ενώ πατώντας ξανά το κουμπί 'Play', η αναπαραγωγή ξεκινά πάλι από την αρχή του αρχείου και όχι από το σημείο που είχε σταματήσει. Το κουμπί αυτό μπορεί να ενεργοποιηθεί ακόμα κι όταν μια αναπαραγωγή βρίσκεται σε παύση, και να επιφέρει τα ίδια αποτελέσματα.
- **Find Chords:** Το κουμπί 'Find Chords' εμφανίζεται στην οθόνη του χρήστη αφού εκείνος ολοκληρώσει την πρώτη του ηχογράφιση κατά την τρέχουσα χρήση της εφαρμογής. Με το πάτημα του κουμπιού, το αρχείο ήχου αποστέλλεται μέσω HTTP POST request στον διακομιστή, για να λάβει τις επιθυμητές πληροφορίες για τις συγχορδίες που συνοδεύουν το μουσικό κομμάτι που ηχογράφησε. Η αναμονή διαρκεί μερικά δευτερόλεπτα, ανάλογα με το μέγεθος του εκάστοτε αρχείου ήχου, μέχρι να ληφθεί η απάντηση του αιτήματος. Κατά τη διάρκεια της αναμονής εμφανίζεται ένας κυκλικός φορτωτής που υπερκαλύπτει την οθόνη και δεν επιτρέπει στον χρήστη να εκτελέσει κάποια άλλη ενέργεια. Σε περίπτωση αποτυχίας του αιτήματος εμφανίζεται μήνυμα αποτυχίας στον χρήστη, ενώ σε περίπτωση επιτυχίας εμφανίζονται τα αποτελέσματα της απάντησης.
- **Chord Visualization:** Μετά από ένα επιτυχές αίτημα HTTP POST που στάλθηκε στον διακομιστή, η απάντηση που λαμβάνεται εμφανίζεται στον χρήστη ως εξής:



Εικόνα 4.2.6 Αναμονή αποτελεσμάτων

Η ακολουθία συγχορδιών που συνοδεύει το κομμάτι εμφανίζεται στην οθόνη, με κάθε συγχορδία να βρίσκεται σε ένα ξεχωριστό μπλοκ μιας οριζόντια κυλιόμενης προβολής. Κάτω από την ετικέτα κάθε συγχορδίας εμφανίζεται η χρονική στιγμή σε δευτερόλεπτα στην οποία ξεκινά η αναπαραγωγή της συγκεκριμένης συγχορδίας, με βάση το χρονοδιάγραμμα του ηχογραφημένου αρχείου. Κάθε συγχορδία συνοδεύεται από δύο εναλλακτικές προτάσεις αντικατάστασης, οι οποίες εμφανίζονται ακριβώς από κάτω από το μπλοκ που την περιέχει, με διαφορετικό χρώμα, αν ο χρήστης πατήσει πάνω στο όνομά της. Πατώντας μία φορά πάνω στη συγχορδία εμφανίζεται η πρώτη εναλλακτική πρόταση αντικατάστασης της συγχορδίας, με έντονο χρώμα. Πατώντας ξανά την συγχορδία εμφανίζεται η δεύτερη πρόταση αντικατάστασης, με λιγότερο έντονο χρώμα. Στην οθόνη του χρήστη υπάρχει οδηγία που εξηγεί την λειτουργία αυτή καθώς δεν είναι προφανής.



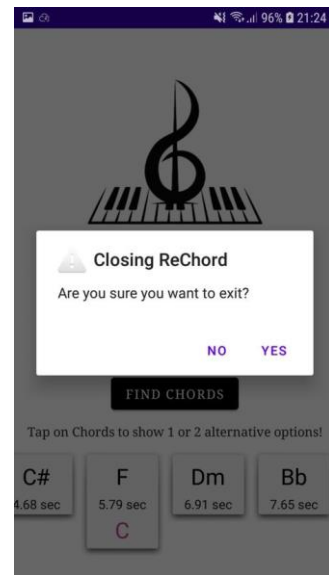
Εικόνα 4.2.9 Αναγνώριση συγχορδιών

Εικόνα 4.2.8 Πρώτη πρόταση αντικατάστασης

Εικόνα 4.2.7 Δεύτερη πρόταση αντικατάστασης

-
-
-
-
-
-
-
-

- **Έξοδος:** Η έξοδος από την εφαρμογή πραγματοποιείται πατώντας το κουμπί «πίσω» της συσκευής. Η εφαρμογή ζητά επιβεβαίωση από τον χρήστη για το κλείσιμό της, σε περίπτωση που το κουμπί πατήθηκε κατά λάθος:



Εικόνα 4.2.10 Επιβεβαίωση κλείσιμου εφαρμογής

5

Υλοποίηση

Σε αυτό το κεφάλαιο αναλύεται ο τρόπος και τα μέσα υλοποίησης κάθε μέρους του συστήματος.

5.1 Διακομιστής / Server

Επιλέξαμε να ενσωματώσουμε το μοντέλο μηχανικής μάθησης στην εφαρμογή μας, επικοινωνώντας με απομακρυσμένο διακομιστή στον οποίο υπάρχει και λειτουργεί το μοντέλο. Η υλοποίηση του διακομιστή επετεύχθη με τη βοήθεια της υπηρεσίας Okeanos[36]. Πρόκειται για μια υπηρεσία cloud που δημιουργήθηκε για την ελληνική έρευνα και την ακαδημαϊκή κοινότητα, η οποία επιτρέπει τη δημιουργία εικονικών μηχανών, δικτύων και χώρων αποθήκευσης. Με άδεια από τη σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου, δημιουργήθηκε μια εικονική μηχανή (Virtual Machine – VM) με λειτουργικό σύστημα Ubuntu 16.04.3 LST, 4 CPU, 4GB μέγεθος μνήμης και 5GB μέγεθος δίσκου, με σύνδεση δικτύου Internet (public IPv4) και IP ‘83.212.112.103’. Η πρόσβασή μας στο VM γίνεται μέσω SSH, το οποίο είναι ένα ασφαλές δικτυακό πρωτόκολλο που επιτρέπει τη μεταφορά δεδομένων μεταξύ δύο υπολογιστών. Τυπικά έχουμε έναν Server που τρέχει μόνιμα την υπηρεσία SSHD (SSH Daemon) η οποία ακούει από προεπιλογή στην πόρτα 22, και έναν απομακρυσμένο υπολογιστή που θέλει να συνδεθεί στο Server με έναν SSH client. Για τη σύνδεση αυτή χρησιμοποιήθηκε η εφαρμογή “Putty”⁷, η οποία σχεδιάστηκε για να επιτρέπει τη σύνδεση και τη διαχείριση απομακρυσμένων λειτουργικών που υποστηρίζουν το πρωτόκολλο SSH.

Ο διακομιστής χρησιμοποιήθηκε για την εξαγωγή συμπερασμάτων του μοντέλου βαθιάς μάθησης που είναι υπεύθυνο για την αναγνώριση συγχορδίων και στη συνέχεια για την εύρεση προτάσεων αντικατάστασης. Για την ανάπτυξη του κώδικα, χρησιμοποιήθηκε η γλώσσα Python, καθώς πρόκειται για μια γλώσσα προγραμματισμού γενικού σκοπού, υψηλού επιπέδου, που εστιάζει στην αναγνωσιμότητα του κώδικα, και αποτελεί ιδανική επιλογή για ανάπτυξη ιστοσελίδων, διαθέτοντας μεγάλες τυπικές βιβλιοθήκες που καθιστούν τον κώδικα ανάπτυξης ιστού απλό και σύντομο. Η Python πρωταρχικά σχεδιάστηκε για διακομιστές ιστού

⁷ <https://www.putty.org/>

και ειδικότερα για την αντιμετώπιση της εισερχόμενης κίνησης στον διακομιστή. Αναπτύξαμε, λοιπόν, μια εφαρμογή web server με τη βοήθεια του Flask, ενός microframework της Python το οποίο παρέχει τη βασική λειτουργικότητα του πλαισίου ιστού και επιτρέπει την προσθήκη περισσότερων χαρακτηριστικών ανάλογα με τις ανάγκες του προγραμματιστή. Το Flask ονομάζεται microframework επειδή καθιστά τη βασική λειτουργικότητα απλή αλλά επεκτάσιμη από άποψη ανάπτυξης, ενώ ταυτόχρονα μπορεί να χρησιμοποιηθεί για την εξοικονόμηση χρόνου κατά την κατασκευή διαδικτυακών εφαρμογών. Μέσω του Flask, δημιουργήσαμε μία Διεπαφή Προγραμματισμού Εφαρμογών (Application Programming Interface – API), η οποία έχει σχεδιαστεί για να επιστρέφει αποτελέσματα σε συγκεκριμένη διεύθυνση URL, “http://83.212.112.103:5000/upload/”. Χρησιμοποιούμε τη μέθοδο HTTP POST για να στείλουμε αίτημα στην παραπάνω διεύθυνση, με παράμετρο ένα αρχείο ήχου ‘.wav’. POST είναι η μέθοδος HTTP που έχει σχεδιαστεί για την αποστολή φορτίων δεδομένων σε διακομιστή από έναν καθορισμένο πόρο. Οι περισσότερες κοινές φόρμες HTML στον Ιστό λειτουργούν χρησιμοποιώντας αυτήν τη μέθοδο αιτήματος, η οποία συνήθως μεταδίδει σχετικά μικρά φορτία δεδομένων σε έναν δέκτη. Η μέθοδος αυτή επιτρέπει την αποστολή δεδομένων ως πακέτο σε ξεχωριστή επικοινωνία με το σενάριο επεξεργασίας, κάτι που σημαίνει ότι τα δεδομένα που αποστέλλονται μέσω της μεθόδου POST δεν είναι ορατά στη διεύθυνση URL. Αν το αίτημα είναι επιτυχημένο επιστρέφεται μια απάντηση από τον διακομιστή στον αποστολέα με τον κωδικό επιτυχούς αιτήματος (HTTP 200) και στην δική μας περίπτωση, τα αποτελέσματα της κλήσης σε μορφή JSON. Τα αποτελέσματα απαρτίζονται ουσιαστικά από πέντε λίστες συμβολοσειρών: την ακολουθία αναγνωρισμένων συγχορδίων, το μπάσο, τους αντίστοιχους χρόνους κάθε μιας από τις συγχορδίες και τις προτεινόμενες αντικαταστάσεις συγχορδίων, και αποτελούν το προϊόν που παράχθηκε από την επεξεργασία του ηχητικού αρχείου μέσω του μοντέλου βαθιάς μάθησης και του συστήματος αντικατάστασης συγχορδίων.

Για να αντιμετωπίσουμε τα ζητήματα ασφάλειας που προκύπτουν με τη χρήση απομακρυσμένου διακομιστή για την επεξεργασία δεδομένων της εφαρμογής, ενεργοποιήσαμε το πρωτόκολλο HTTPS μέσω του προγράμματος Certbot ACME Client⁸ για χρήση ασφαλούς πρωτόκολλου HTTP, εκδίδοντας δωρεάν πιστοποιητικό με το “Let’s Encrypt”⁹ (μια δωρεάν αρχή πιστοποίησης ανοιχτού κώδικα που εκδίδει βραχύβια πιστοποιητικά 90 ημερών). Με αυτόν τον τρόπο γίνεται η διασφάλιση κρυπτογραφημένης και πιστοποιημένης σύνδεσης με τον διακομιστή.

⁸ <https://certbot.eff.org/>

⁹ <https://letsencrypt.org/>

Από προεπιλογή, το Flask τρέχει στην τοπική IP διεύθυνση του υπολογιστή 127.0.0.1, στο όνομα localhost, που χρησιμοποιείται για να δηλώσει την ίδια τη συσκευή. Κάθε συσκευή χρησιμοποιεί αυτήν την διεύθυνση για τη δημιουργία σύνδεσης IP στο ίδιο το μηχάνημα ή τον υπολογιστή που χρησιμοποιείται από τον τελικό χρήστη, οπότε ήταν αδύνατο να έχουμε πρόσβαση στον διακομιστή από απομακρυσμένες συσκευές με αυτήν την διεύθυνση. Για την αντιμετώπιση αυτού του προβλήματος, ορίστηκε ως host διεύθυνση της εφαρμογής Flask η διεύθυνση 0.0.0.0. Στο πλαίσιο των διακομιστών, η διεύθυνση 0.0.0.0 αντιπροσωπεύει όλες τις διευθύνσεις IPv4 στον τοπικό υπολογιστή. Έτσι έχουμε πρόσβαση στον διακομιστή μέσω της IP διεύθυνσής του, 83.212.112.103.

5.2 Μοντέλο Αναγνώρισης Συγχορδιών

Το Flask χρησιμοποιήθηκε για να πλαισιώσει τον κώδικα Python με τον οποίο τρέχει το μοντέλο βαθιάς μάθησης που χρησιμοποιείται για την αναγνώριση συγχορδιών. Στον τομέα ταξινόμησης ήχου και μουσικής συνήθως χρησιμοποιούνται παρεμφερή μοντέλα με αυτό που χρησιμοποιήθηκε στην εργασία μας, με ξεχωριστά χαρακτηριστικά που ποικίλλουν ανάλογα με τις ανάγκες της εκάστοτε εργασίας. Το συγκεκριμένο μοντέλο πρόκειται για μια ελαφρώς απλοποιημένη έκδοση του μοντέλου βαθιάς μάθησης που χρησιμοποιείται για την αναγνώριση συγχορδιών στην εφαρμογή ChordAI¹⁰, το οποίο όμως πλησιάζει πολύ το γνήσιο, με τη βασική διαφορά ότι αναγνωρίζει μόνο μείζονες και ελάσσονες τρίφωνες συγχορδίες.

5.2.1 Αρχιτεκτονική μοντέλου

Ο κώδικας της web εφαρμογής που δημιουργήθηκε δέχεται ως είσοδο μουσικά αρχεία .wav, τα οποία φτάνουν σε αυτή μέσω των αιτημάτων που στέλνονται στον διακομιστή. Αυτά τα αρχεία είναι είτε μονοφωνικά είτε στερεοφωνικά, ανάλογα με το μικρόφωνο της εκάστοτε συσκευής. Η διαφορά μεταξύ μονοφωνικού και στερεοφωνικού ήχου έγκειται στον αριθμό των καναλιών που χρησιμοποιούνται για την εγγραφή και την αναπαραγωγή του. Τα μονοφωνικά σήματα εγγράφονται και αναπαράγονται χρησιμοποιώντας ένα μόνο κανάλι ήχου, ενώ οι στερεοφωνικοί ήχοι εγγράφονται και αναπαράγονται χρησιμοποιώντας δύο κανάλια ήχου. Η διαδικασία που ακολουθείται μετά την λήψη των αρχείων ξεκινά με τον υπολογισμό του βραχυπρόθεσμου μετασχηματισμού Fourier (Short-time Fourier transform – STFT). Αυτός ο μετασχηματισμός σχετίζεται με το Fourier και χρησιμοποιείται για τον προσδιορισμό της ημιτονοειδούς συχνότητας και του περιεχομένου φάσης των τοπικών τμημάτων ενός σήματος καθώς αυτό αλλάζει με την πάροδο του χρόνου. Ουσιαστικά υπολογίζεται ο γρήγορος

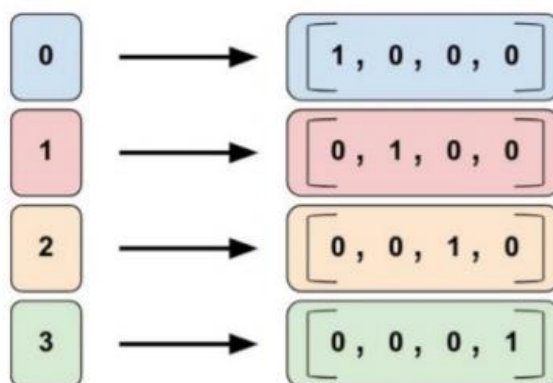
¹⁰ <https://play.google.com/store/apps/details?id=com.chordai&hl=el&gl=US>

μετασχηματισμός Fourier (Fast Fourier Transform – FFT) κατά μήκος της εισόδου, χρησιμοποιώντας το πρωτόκολλο συρόμενου παραθύρου (sliding window protocol). Στη συνέχεια υπολογίζουμε το μέγεθος και τη δύναμη κάθε FFT, για να μπορέσουμε να εξάγουμε τα mel-spectrograms που τους αντιστοιχούν, καθένα από τα οποία έχει 256 συντελεστές. Κατασκευάζουμε, λοιπόν, ομάδες των 16 διαδοχικών mel-spectrograms, ομαδοποιώντας και κανονικοποιώντας τα χαρακτηριστικά πλαισίων με βήμα 8 ώστε να υπάρχει επικάλυψη. Οι ομάδες αυτές μπορούν ξεχωριστά να θεωρηθούν ως εικόνες 2-D 256 x 16 και χρησιμοποιούνται ως είσοδοι στο μοντέλο βαθιάς μάθησης. Οι διαστάσεις της εισόδου του μοντέλου αντιστοιχούν σε (batch size, sequence size, channels, height, width) και καταλήγουν στις τιμές (1, ?, 1, 16, 128), όπου ‘?’ είναι η τιμή που παίρνει κάθε φορά η μεταβλητή sequence size.

Το ίδιο το μοντέλο αποτελείται από ένα συνελκτικό νευρωνικό δίκτυο (Convolutional Neural Network – CNN) με αρχιτεκτονική που ταιριάζει στο πρότυπο VGG, ένα καινοτόμο μοντέλο αναγνώρισης αντικειμένων που υποστηρίζει έως και 19 επίπεδα. Ο όρος συνελκτικό νευρωνικό δίκτυο χρησιμοποιήθηκε για να περιγράψει την αρχιτεκτονική των νευρωνικών δικτύων τα οποία εφαρμόζονται σε διδιάστατους πίνακες και βασίζονται σε χωρικά οριοθετημένη νευρωνική είσοδο. Η ιδιαιτερότητά τους, οφείλεται στην ικανότητά τους να αναγνωρίζουν συγκεκριμένα (πολλά και διαφορετικά) στοιχεία εντός της εισόδου, πετυχαίνοντας με αυτό τον τρόπο μεγάλες ακρίβειες στην κατηγοριοποίηση. Στην συγκεκριμένη περίπτωση το CNN δέχεται ουσιαστικά σαν είσοδο εικόνες 2 διαστάσεων, δηλαδή τις παραπάνω ομάδες mel-spectrograms. Το CNN επεξεργάζεται κάθε μία από τις δυοδιάστατες αυτές εικόνες και στη συνέχεια εξάγει μικρότερες εικόνες περισσότερων στρωμάτων. Στη συνέχεια τα στρώματα αυτά συγχωνεύονται σε ένα ενιαίο, με διαδοχική τοποθέτηση σε σειρά, δημιουργώντας εκ νέου μια εικόνα με μικρό ύψος και αρκετά μεγαλύτερο πλάτος. Οι έξοδοι του VGG τροφοδοτούνται σε ένα μακροπρόθεσμο βραχυχρόνο δίκτυο μνήμης (LSTM - Long Short Term Memory) 2 επιπέδων. Το LSTM είναι ένας τύπος επαναλαμβανόμενου νευρωνικού δικτύου (RNN - Recurrent neural network), που υπερέρχει στη μάθηση, την επεξεργασία και την ταξινόμηση διαδοχικών δεδομένων. Ως είσοδοι κάθε LSTM χρησιμοποιούνται οι έξοδοι του προηγούμενου και του επόμενου LSTM, όπως επίσης και η έξοδος του VGG. Με αυτόν τον τρόπο λαμβάνουμε πληροφορία και για τις γειτονικές περιοχές κάθε εισόδου ώστε να υπάρχει συνοχή, γι’ αυτό και επιλέχθηκε RNN διπλής κατεύθυνσης. Οι έξοδοι του LSTM τροφοδοτούνται με τη σειρά τους σε δύο πλήρως συνδεδεμένα νευρωνικά δίκτυα 3 στρωμάτων. Το πρώτο νευρωνικό δίκτυο χρησιμοποιεί 12 λογότυπα και ταξινομεί το μπάσο (η χρωματική κλίμακα έχει 12 τιμές) και το δεύτερο έχει 24 λογότυπα, για όλες τις μείζονες και ελάσσονες συγχορδίες (12 x 2). Σε αυτό το σημείο να υπενθυμίσουμε ότι το μοντέλο βαθιάς μάθησης που περιγράφεται εδώ είναι ρυθμισμένο έτσι ώστε να εντοπίζει μόνο μείζονες και ελάσσονες τρίφωνες συγχορδίες, χωρίς επεκτάσεις και παραπάνω πληροφορίες.

5.2.2 Εκπαίδευση μοντέλου

Σε προβλήματα μηχανικής και βαθιάς μάθησης, χρησιμοποιούνται συναρτήσεις απώλειας/κόστους (loss functions) για τη βελτιστοποίηση του εκάστοτε μοντέλου κατά τη διάρκεια της εκπαίδευσης. Ο στόχος είναι σχεδόν πάντα να ελαχιστοποιηθεί η συνάρτηση απώλειας. Όσο μικρότερη είναι η απώλεια, τόσο καλύτερο είναι το μοντέλο. Έστω $X = \{x_1, \dots, x_N\} \in X^N$ τα δεδομένα / οι εικόνες εκπαίδευσης και $Y = \{y_1, \dots, y_N\} \in Y^N$ οι συσχετιζόμενες ετικέτες τους με one-hot κωδικοποιημένες κλάσεις ($y_i \in \{0, 1\}^{1 \times c}$, c είναι ο αριθμός των κλάσεων). One-hot κωδικοποίηση σημαίνει ότι οι ετικέτες είναι κωδικοποιημένες έτσι ώστε η τιμή εξόδου που αντιστοιχεί σε κάθε δείγμα να είναι ένα διάνυσμα μεγέθους c , όπου όλα τα στοιχεία είναι ίσα με 0, εκτός από το στοιχείο που δείχνει τη θέση της κλάσης στην οποία ανήκει το παράδειγμα, το οποίο έχει τιμή ίση με 1.



Εικόνα 5.2.1 Παράδειγμα One – Hot Encoding [37]

Έστω $f: X \rightarrow Y$ το μοντέλο νευρωνικού δικτύου με παραμέτρους θ , το οποίο αντιστοιχεί τα χαρακτηριστικά εισόδου σε πιθανότητες υπό όρους κλάσης $f(x)_j = \hat{p}(y = j|x)$. Η εκάστοτε συνάρτηση κόστους $L(f(x), y)$ μετρά τη διαφορά (το σφάλμα) μεταξύ της πραγματικής ετικέτας y και της προβλεπόμενης κατανομής της ετικέτας $f(x)$ από το νευρωνικό δίκτυο. Στην επιβλεπόμενη μάθηση, υπολογίζονται οι παράμετροι θ της f , ελαχιστοποιώντας τον εμπειρικό κίνδυνο στα δεδομένα εκπαίδευσης μέσω της εξίσωσης:

$$\min_f \frac{1}{N} \sum_{i=1}^N L(f(x_i), y_i) \quad (1)$$

Για το παρόν μοντέλο χρησιμοποιήθηκε η διασταυρούμενη εντροπία (cross - entropy) ως συνάρτηση κόστους, η οποία είναι μια από τις πιο σημαντικές συναρτήσεις κόστους και χρησιμοποιείται κυρίως για τη βελτιστοποίηση μοντέλων ταξινόμησης. Ορίζεται ως:

$$L(f(x_i), y_i) = \sum_{j=1}^c -y_{ij} \log(f(x_i)_j) \quad (2)$$

και έτσι η εξίσωση (1) μπορεί να εκφραστεί εκ νέου ως:

$$\min_f \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^c -y_{ij} \log(f(x_i)_j) \quad (3)$$

Το μοντέλο νευρωνικού δικτύου f εκτιμάται ελαχιστοποιώντας τον παραπάνω όρο σε σχέση με τις παραμέτρους του, μέσω διαδικασιών στοχαστικής βελτιστοποίησης. Η απώλεια δια-εντροπίας μετρά την επίδοση ενός μοντέλου ταξινόμησης του οποίου η απόδοση είναι μια τιμή πιθανότητας μεταξύ 0 και 1 και αυξάνεται καθώς η προβλεπόμενη πιθανότητα αποκλίνει από την πραγματική ετικέτα. Ένα τέλειο μοντέλο θα είχε μηδενική απώλεια cross entropy [38].

Το μοντέλο βαθιάς μάθησης εξάγει αποτελέσματα για τη συγχορδία και το μπάσο που αντιστοιχεί σε κάθε κομμάτι των δεδομένων. Με βάση τη συνάρτηση κόστους και την one-hot κωδικοποίηση των αποτελεσμάτων, διαμορφώνεται το τελικό αποτέλεσμα. Τα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου προέρχονται από διάφορες πηγές δεδομένων, όπως τα RWC Music Database [39], Jazz Audio-Aligned Harmony (JAAM) Dataset [40]. Ωστόσο, επειδή το μοντέλο προοριζόταν και για εμπορική χρήση, υπήρχε πρόβλημα με τις άδειες χρήσης των παραπάνω δεδομένων και σε συνδυασμό με την έλλειψη απόλυτης ακρίβειας των δεδομένων αυτών, αποφασίστηκε η δημιουργία νέων δεδομένων εκπαίδευσης χειροκίνητα, από μουσικά κομμάτια διαφόρων ειδών. Έτσι, το μοντέλο απέκτησε μεγαλύτερη ακρίβεια αποτελεσμάτων, εκπαιδευμένο σε ένα μοναδικό σύνολο δεδομένων.

5.3 Σύστημα Αντικατάστασης Συγχορδιών

Μετά την αναγνώριση της ακολουθίας συγχορδιών που συνοδεύει το αρχείο ήχου, σειρά έχει η εύρεση αντικαταστάσεων για κάθε μία από τις συγχορδίες αυτές. Για την επίτευξη αυτής της λειτουργίας, έγινε συλλογή στατιστικών στοιχείων ώστε να συνοψιστούν δεδομένα από ακολουθίες συγχορδιών διαφόρων μουσικών κομματιών και να συγκριθούν μεταξύ τους. Εξετάζοντας μια πληθώρα ακολουθιών, το απλούστερο στατιστικό στοιχείο που θα μπορούσε να χρησιμοποιηθεί για ένα σύστημα προτάσεων αντικαταστάσεων είναι οι μετρήσεις συχνότητας εμφάνισης των συγχορδιών στις ακολουθίες αυτές, δηλαδή το πόσο συχνά χρησιμοποιείται μια συγκεκριμένη συγχορδία ή στοιχεία αυτής σε μία σύνθεση. Η χρήση στατιστικών έχει το μεγάλο πλεονέκτημα ότι με μία γρήγορη σάρωση των δεδομένων μπορούμε να υπολογίσουμε τη συχνότητα εμφάνισης κάθε συγχορδίας σε μια σύνθεση, οπότε η διαδικασία είναι εύκολη και δεν χρειάζεται πολύ χρόνο. Ωστόσο, τα απλά στατιστικά στοιχεία δεν επαρκούν για το σκοπό αυτής της εργασίας, καθώς η συχνότητα εμφάνισης συγχορδιών δεν παρέχει πληροφορίες σχετικά με τη σειρά με την οποία εμφανίζονται οι συγχορδίες.

Το παραπάνω πρόβλημα μπορεί να αντιμετωπιστεί χρησιμοποιώντας N-grams - μοτίβα που αποτελούνται από N διαδοχικές συγχορδίες που εμφανίζονται σε μία ακολουθία συγχορδιών. Το N-gram είναι ένα τυπικό εργαλείο για την κατανόηση της φυσικής γλώσσας και χρησιμοποιείται στο χώρο της ανάκτησης μουσικής πληροφορίας και ιδίως στην ανάλυση μελωδικού περιγράμματος [41], [42]. Για την ανάλυσή μας, θεωρούμε ότι κάθε συγχορδία που ανήκει στη σύνθεση αποτελεί μια τριάδα των βασικών νοτών, ώστε να επιτραπεί η θεμελιώδης κατανόηση της δομής της συγχορδίας. Κάθε τέτοια τρίφωνη συγχορδία, χαρακτηρίζεται από μια ετικέτα με το όνομα της τονικής της νότας, την πληροφορία για το αν είναι μείζονα ή ελάσσονα και οποιαδήποτε παραπάνω πληροφορία χρειάζεται για την πλήρη αναπαράστασή της.

5.3.1 Απλοποίηση Συγχορδιών – N-grams

Υπάρχουν δύο σημαντικά ζητήματα που πρέπει να ληφθούν υπόψιν κατά τον ορισμό των N-grams συγχορδιών. Το πρώτο είναι εάν πρέπει να επιτρέπονται διαδοχικές επαναλήψεις της ίδιας συγχορδίας. Το δεύτερο είναι πώς να συνυπολογίσουμε τη διάρκεια της κάθε συγχορδίας. Τα δύο θέματα αυτά σχετίζονται μεταξύ τους. Οι συγχορδίες στα δεδομένα μας έχουν απλοποιηθεί, οπότε δεν τίθεται θέμα διαδοχικής επανάληψης της ίδιας συγχορδίας αφού οι διπλότυπες εμφανίσεις σε σειρά έχουν εξαλειφθεί. Μετά την απλοποίηση, όλες οι διαδοχικές καταχωρήσεις των οποίων το τμήμα συγχορδίας είναι πανομοιότυπο έχουν συγχωνευτεί σε μία μόνο καταχώριση. Το ίδιο έχει συμβεί και για τις σύμφωνες συγχορδίες, αυτές δηλαδή που έχουν διαφορετική ονομασία αλλά έχουν ακριβώς τον ίδιο ήχο, καθώς αποτελούνται από τις ίδιες νότες. Έτσι, κάθε γειτονικό ζεύγος συγχορδιών είναι διαφορετικό και κάθε μία από τις ακολουθίες είναι ένα κατάλληλο N-gram. Η πληροφορία που μας ενδιαφέρει στην συγκεκριμένη εργασία είναι μόνο το όνομα της κάθε συγχορδίας και ο μείζον ή ελάσσον τρόπος της, έχοντας συνολικά 24 διαφορετικές συγχορδίες (12 ματζόρε – 12 μινόρε). Επιπλέον, δεν λήφθηκε υπόψιν η διάρκεια της κάθε συγχορδίας στην συλλογή στατιστικών στοιχείων, εφόσον η μελέτη επικεντρώθηκε κυρίως στην σειρά με την οποία εμφανίζονται οι συγχορδίες και όχι στο χρονικό μήκος αυτών. Οι παραδοχές και απλοποιήσεις αυτές σίγουρα ελλοχεύουν τον κίνδυνο υπεραπλοποίησης του μουσικού κομματιού και παράλειψης συγκεκριμένων λεπτομερειών και μοτίβων, γι' αυτό και είναι απαραίτητο να αναφερθούν.

5.3.2 Αλλαγή κλειδιού - Dataset

Γνωρίζοντας ότι όλα τα μουσικά κομμάτια μπορούν να μεταφερθούν σε διαφορετικά κλειδιά με την προϋπόθεση ότι διατηρείται η σχέση διαστημάτων μεταξύ των συγχορδιών, έχει ενδιαφέρον η μελέτη των αλλαγών συγχορδιών συγκριτικά με την πρώτη συγχορδία του κομματιού. Έτσι μπορεί κανείς να μεταφέρει κάθε κομμάτι τοπικά, με τέτοιο τρόπο ώστε

καθένα να ξεκινά με έναν κωδικό που έχει την ίδια συγχορδία ως ρίζα, έστω A. Δεδομένου ότι το A είναι απλώς ονομαστικό, μπορούν να χρησιμοποιηθούν οι ρωμαϊκοί αριθμοί I, II, III, IV, V, VI, VII για να αναπαραστήσουν τις επτά νότες, μαζί με τα σημεία αλλοίωσης φυσικά και οποιαδήποτε άλλη πληροφορία παρέχεται για τη συγχορδία.

Οι ακολουθίες που χρησιμοποιήθηκαν ως σύνολο δεδομένων για την παρούσα ανάλυση προέρχονται από την ιστοσελίδα Hooktheory¹¹, η οποία δίνει τη δυνατότητα στους εγγεγραμμένους της χρήστες να μεταγράψουν οποιοδήποτε υπάρχον μουσικό κομμάτι με τις προβλεπόμενες μουσικές πληροφορίες και να το ανεβάσουν στην ιστοσελίδα ώστε να είναι προσβάσιμο από όλους τους υπόλοιπους εγγεγραμμένους και μη χρήστες. Συνολικά τα δεδομένα αποτελούνταν από 8716 διαφορετικά μουσικά κομμάτια, τα οποία ανήκαν στα παρακάτω είδη (ένα κομμάτι μπορεί να ανήκει σε παραπάνω από ένα είδος μουσικής):

Πίνακας 5.3.2.1 Διαχωρισμός δεδομένων ανά μουσικό είδος

Είδος Μουσικής	Αριθμός μουσικών κομματιών
Soundtrack	1422
Pop	1325
Video-game	1249
Rock	1237
Electronical	1070
Dance	679
J-pop	547
Alternative	496
R&B	472
Hip-hop-rap	441
House	432
Indie	387
Classical	331
Singer-songwriter	311
Jazz	272
Soul	244
Country	239

¹¹ www.hooktheory.com

Folk	238
Blues	231
K-pop	211
Experimental	161
Disney	148
Children's	142
Metal	139
Punk	136
Alt-country	130
Reggae	86
Latin	81
Vocal	79
Techno	72
Worship	56
Holiday	41
World	37

Τελικά, αγνοώντας τις περιπτώσεις παύσης ή επανάληψης της ίδιας συγχορδίας και τις ακολουθίες με λιγότερες από τρεις συγχορδίες, το σύνολο δεδομένων αποτελείται από 8515 μουσικές ακολουθίες συγχορδιών. Ο αριθμός των συνολικών συγχορδιών είναι 119876, ενώ υπάρχουν 732 διαφορετικές συγχορδίες οι οποίες εμφανίζονται τουλάχιστον μια φορά. Στην παρούσα εργασία τα δεδομένα μας δόθηκαν έτοιμα, σε μορφή λιστών, σε γλώσσα Python. Η διαδικασία άντλησης και προεπεξεργασίας των δεδομένων περιγράφεται στην διπλωματική εργασία του Σ. Κανταρέλη [43].

Οι ακολουθίες συγχορδιών που απαρτίζουν το παραπάνω σύνολο δεδομένων αναπαρίστανται με ρωμαϊκούς αριθμούς, χωρίς να παραλείπονται βέβαια και τα συμπληρωματικά χαρακτηριστικά κάθε μίας από αυτές, όπως ο οπλισμός και ο μουσικός τρόπος. Για την αναπαράσταση μείζονων συγχορδιών χρησιμοποιούνται κεφαλαία γράμματα στους ρωμαϊκούς αριθμούς ενώ για τις ελάσσονες συγχορδίες χρησιμοποιούνται πεζά. Με τη χρήση ρωμαϊκών αριθμών αντιμετωπίστηκε άμεσα το πρόβλημα μεταφοράς κλειδιού, αφού με αυτήν την αναπαράσταση μπορούμε να έχουμε πληροφορία για τις βαθμίδες και τα διαστήματα, και όχι για συγκεκριμένα ονόματα συγχορδιών. Για παράδειγμα, η ακολουθία [I, III, IV, #VI, I] με ρίζα τη Ντο αντιστοιχεί στην ακολουθία συγχορδιών [C, E, F, A#, C], με ρίζα τη Ρε στην [D,

F#, G, B#, D], με ρίζα τη Φα στην [F, A, Bb, D#, F], κ.ο.κ.. Χρησιμοποιώντας python scripts, μετατρέψαμε τις ακολουθίες βαθμίδων σε αντίστοιχες ακολουθίες διαστημάτων.

5.3.3 Μετατροπή ακολουθιών σε διαστήματα

Το σύνολο δεδομένων είναι ανεξάρτητο από το κλειδί, δηλαδή όλες οι μουσικές ακολουθίες μπορούν να αντιμετωπιστούν σαν να ξεκινούν από το ίδιο κλειδί, "το κλειδί του 0". Έτσι, κάθε συγχορδία i/I έχει ρίζα το 0. Επειδή λαμβάνονται υπόψη μόνο οι ρίζες των συγχορδιών, η ποιότητα της συγχορδίας (π.χ. μείζονα/ελάσσονα) δεν έχει σημασία. Επομένως, τα i/I, ii/II, κ.λπ. δεν χρειάζεται να διαφοροποιηθούν. Αρχικά αντιστοιχίστηκε κάθε ρωμαϊκός αριθμός σε έναν αντίστοιχο του αραβικού συστήματος αρίθμησης, σύμφωνα με τον παρακάτω πίνακα:

Πίνακας 5.3.3.1 Αντιστοίχιση ρωμαϊκών αριθμών σε αραβικούς

I/VI#	IIb/I#	II	IIIb/III#	III/IVb	IV/III#	Vb/IV#	V	VIb/V#	VI	VIIb/VII#	VII/IIb
0	1	2	3	4	5	6	7	8	9	10	11

Σημειώνεται ότι δεν μας ενδιαφέρει σε αυτό το σημείο αν ο ρωμαϊκός αριθμός ήταν γραμμένος με κεφαλαία ή με πεζά γράμματα. Οι αριθμοί που προέκυψαν από την αντιστοίχιση αποθηκεύτηκαν σε μια καινούργια λίστα. Στη συνέχεια, μετατρέψαμε τους αριθμούς αυτούς σε διαστήματα, όπου το διάστημα εκφράζεται ανάλογα με τον αριθμό των ημιτόνιων που εμπλέκονται σε αυτό. Δεδομένου ότι οι ρωμαϊκοί αριθμοί (δηλαδή οι ρίζες των συγχορδιών) αγνοούν τις διαφορές στις οκτάβες, χρειαζόμαστε μόνο δώδεκα διαστήματα (0 - 11 ημιτόνια). Η μετατροπή των αριθμών σε διαστήματα έγινε ακολουθώντας τους παρακάτω κανόνες:

Πίνακας 5.3.3.2 Αντιστοίχιση διαστημάτων σε αριθμό ημιτόνιων

Διάστημα	Αριθμός
Τέλεια ομοφωνία	0
Δευτέρα μικρή	1, -11
Δευτέρα μεγάλη	2, -10
Τρίτη μικρή	3, -9
Τρίτη μεγάλη	4, -8
Τετάρτη καθαρή	5, -7
Τετάρτη αυξημένη / Πέμπτη ελαττωμένη	6, -6
Πέμπτη καθαρή	7, -5
Έκτη μικρή	8, -4

Έκτη μεγάλη	9, -3
Εβδόμη μικρή	10, -2
Εβδόμη μεγάλη	11, -1

Δεδομένων αυτών των κωδικοποιήσεων, η απόσταση μεταξύ των δύο συγχορδιών X και Y είναι $(X - Y) \bmod 12$ (το modulo χρησιμοποιείται για να έχουμε μόνο θετικούς αριθμούς από το 0 έως το 11 και να αποφύγουμε τους αρνητικούς). Για μια ακολουθία συγχορδιών $[X_0, X_1, \dots, X_n]$ το σύνολο των μεσοδιαστημάτων που της αντιστοιχεί είναι η ακολουθία που συμπληρώνεται με βάση των παρακάτω κανόνα:

$$Intervals[i] = \begin{cases} 0, & i = 0 \\ (X_i - X_{i-1}) \bmod 12, & i \in [1, n] \end{cases}$$

Για παράδειγμα, για μια ακολουθία $[V, i, bIII, bVI, V]$, με μετατροπή στο αραβικό αριθμητικό σύστημα έχουμε την λίστα $[7, 0, 3, 8, 7]$. Χρησιμοποιώντας τους παραπάνω κανόνες, η αντίστοιχη ακολουθία διαστημάτων είναι : $[0, (0 - 7) \bmod 12, (3 - 0) \bmod 12, (8 - 3) \bmod 12, (7 - 8) \bmod 12] = [0, 5, 3, 5, 11]$. Η διαδικασία ολοκληρώθηκε προγραμματιστικά με τη βοήθεια python scripts αυτόματα για ολόκληρο το σύνολο δεδομένων. Αποθηκεύσαμε σε ξεχωριστές λίστες τα λατινικά γράμματα, τα διαστήματα σε αριθμούς, τον τρόπο της κάθε συγχορδίας (min, maj) και τις εκτεταμένες πληροφορίες (add9, sus4, 7, κ.ο.κ.).

5.3.4 Επιλογή N-grams

Για τη συλλογή στατιστικών στοιχείων χρησιμοποιήσαμε τις παραπάνω λίστες διαστημάτων, σε συνδυασμό με τις λίστες που περιγράφουν τον τρόπο κάθε συγχορδίας. Αποθηκεύσαμε όλες τις λίστες με τους ρωμαϊκούς αριθμούς, τον τρόπο κάθε συγχορδίας, τα διαστήματα και τις εκτεταμένες πληροφορίες σε ένα python dictionary, χρησιμοποιώντας ως αναγνωριστικό κλειδί για κάθε ακολουθία το ξεχωριστό track_id που συνοδεύει κάθε κομμάτι στο σύνολο δεδομένων. Έχοντας λοιπόν όλη την πληροφορία για κάθε συγχορδία οποιασδήποτε ακολουθίας, δοκιμάσαμε να δημιουργήσουμε πίνακες που περιέχουν την συχνότητα εμφάνισης N-grams. Καθώς το μοντέλο βαθιάς μάθησης που χρησιμοποιήθηκε για την τρέχουσα εργασία αναγνωρίζει μόνο μείζονες και ελάσσονες συγχορδίες, θεωρήσαμε λογικό να εργαστούμε στο ίδιο πλαίσιο και για το σύστημα αντικαταστάσεων συγχορδιών. Επιλέξαμε να συλλέξουμε στατιστικά δεδομένα για την συχνότητα εμφάνισης bigrams (ακολουθίες δύο στοιχείων $[X, Y]$) και trigrams (ακολουθίες τριών στοιχείων $[X, Y, Z]$) διαστημάτων, ως πιο αντιπροσωπευτικά, καθώς εξετάζουν τις αποστάσεις κάθε συγχορδίας από την προηγούμενη και την επόμενη της σε διαστήματα. Αξίζει να υπενθυμίσουμε ότι τα N-grams αποτελούνται από διαστήματα (δηλαδή αριθμούς από το 0 έως το 11 και όχι ονόματα συγχορδιών), οι οποίοι αναπαριστούν

σχέσεις μεταξύ συγχορδιών, ανεξαρτήτων κλειδιού. Θεωρήσαμε ενδιαφέρουσα την αξιολόγηση συστημάτων αντικατάστασης που είτε χρησιμοποιούσαν N-grams που περιελάμβαναν πληροφορία μόνο για την ρίζα της συγχορδίας (λαμβάνοντας δεδομένα από τις λίστες διαστημάτων), είτε N-grams αποτελούμενα από την ρίζα αλλά και τον τρόπο (μείζονα – ελάσσονα) της συγχορδίας (λαμβάνοντας δεδομένα από τις λίστες διαστημάτων αλλά και τρόπων). Ουσιαστικά διατρέχοντας τις λίστες διαστημάτων ή και τρόπων του αρχικού dictionary που περιλαμβάνει την πληροφορία για όλες τις ακολουθίες μουσικών κομματιών, και δημιουργώντας δυάδες ή τριάδες από διαδοχικές τιμές διαστημάτων (συνοδευόμενες ή όχι από τον τρόπο της συγχορδίας), μετρήσαμε τις εμφανίσεις καθενός από αυτά τα N-grams στο σύνολο δεδομένων μας. Όλα τα αποτελέσματα αποθηκεύτηκαν σε ξεχωριστούς πίνακες N-grams με τη μορφή pythion dictionaries. Ως κλειδί χρησιμοποιήθηκε το ίδιο το N-gram και ως τιμή του κάθε κλειδιού το σύνολο των καταμετρημένων εμφανίσεών του στις ακολουθίες συγχορδιών. Μετά από μελέτη των αποτελεσμάτων και αξιολόγηση των επιλογών μας, επιλέξαμε να χρησιμοποιήσουμε στην τελική εφαρμογή ένα σύστημα αντικατάστασης που βασίζεται σε συνδυασμό πληροφοριών από bigrams και trigrams που περιέχουν και πληροφορία για τον τρόπο κάθε συγχορδίας, ματζόρε ή μινόρε.

5.3.5 Εξαγωγή αποτελεσμάτων

Το μοντέλο μηχανικής μάθησης, αφού επεξεργαστεί το αρχείο ήχου που έστειλε στο διακομιστή ο χρήστης, εξάγει ως αποτέλεσμα μια ακολουθία συγχορδιών, στην οποία κάθε συγχορδία αναπαρίσταται από ένα γράμμα της χρωματικής κλίμακας (A, B, C, D, E, F, G), ένα ή κανένα σύμβολο οπλισμού (# δίεση, b ύφεση), και τον τρόπο που τη χαρακτηρίζει (m για τις ελάσσονες συγχορδίες, κενό για τις μείζονες). Στόχος είναι η εύρεση μιας πρότασης αντικατάστασης για κάθε συγχορδία της ακολουθίας, βασισμένη σε στοιχεία για την συχνότητα εμφάνισης συγκεκριμένων συνδυασμών διαστημάτων μεταξύ συγχορδιών σε μουσικές ακολουθίες. Αρχικά, η ακολουθία συγχορδιών μετατρέπεται σε μια λίστα αριθμών από το 0 έως το 11, σύμφωνα με τον παρακάτω πίνακα:

Πίνακας 5.3.5.1 Αντιστοίχιση συγχορδιών σε αραβικούς αριθμούς

C/B#	D ^b /C#	D	E ^b /D#	E/F ^b	F/E#	G ^b /F#	G	A ^b /G#	A	B ^b /A#	B/C ^b
0	1	2	3	4	5	6	7	8	9	10	11

Ο τρόπος κάθε συγχορδίας αποθηκεύεται σε μία ξεχωριστή ακολουθία, με αντίστοιχους δείκτες (indexes), ενώ χρησιμοποιούμε μια μεταβλητή 'key' με τον αριθμό που αντιστοιχεί στην πρώτη συγχορδία της ακολουθίας για μπορούμε να μετατρέψουμε ξανά τα διαστήματα σε συγχορδίες της σωστής κλίμακας.

Χρησιμοποιώντας την λίστα των αριθμών που παρήχθη, δημιουργούμε μια καινούργια λίστα διαστημάτων με την βοήθεια της συνάρτησης *Intervals(i)*. Διατρέχοντας την ακολουθία αυτή, συλλέγουμε διαδοχικές δυάδες (bigrams) και τριάδες (trigrams) διαστημάτων. Για κάθε τιμή X_i που θέλουμε να αντικαταστήσουμε στη θέση i , αναζητούμε στον πίνακα με τις συχνότητες bigrams τα ζευγάρια $[X_{i-1}, Y]$ και $[Y, X_{i+1}]$ και στον πίνακα με τις συχνότητες trigrams τις τριάδες $[X_{i-1}, Y, X_{i+1}]$, όπου X_{i-1} η τιμή της ακολουθίας στη θέση $i-1$ και X_{i+1} η τιμή της ακολουθίας στη θέση $i+1$. Επιλέγουμε ως βέλτιστη πρόταση αντικατάστασης την τιμή $Y \neq X_i$ που δίνει το μέγιστο άθροισμα $\text{Bigrams}[X_{i-1}, Y] * 0.7 + \text{Bigrams}[Y, X_{i+1}] * 0.3 + \text{Trigrams}[X_{i-1}, Y, X_{i+1}]$. Τα βάρη πολλαπλασιασμού κάθε bigram καθορίστηκαν μετά από αξιολόγηση όλων των πιθανών συνδυασμών για τη βέλτιστη απόδοση του συστήματος, στην οποία καταλήξαμε στο συμπέρασμα ότι χρειάζεται να δίνεται μεγαλύτερη βαρύτητα στο bigram που περιλαμβάνει την τιμή X_{i-1} . Αν το X_i είναι το πρώτο ή το τελευταίο στοιχείο της λίστας, η απόφασή μας στηρίζεται αποκλειστικά στη συχνότητα εμφάνισης του bigram $[X_{i-1}, Y]$ και του trigram $[X_{i-2}, X_{i-1}, Y]$ ή του bigram $[Y, X_{i+1}]$ και του trigram $[Y, X_{i+1}, X_{i+2}]$ αντίστοιχα. Το στοιχείο Y αντικαθιστά το X_i στην ακολουθία διαστημάτων, και ακριβώς αντίστροφα η λίστα διαστημάτων μετατρέπεται σε αριθμούς του αραβικού συστήματος αρίθμησης (χρησιμοποιώντας το κλειδί key της ακολουθίας για την μεταφορά διαστημάτων στην κατάλληλη κλίμακα) και έπειτα στις αντίστοιχες συγχορδίες χρωματικής κλίμακας, παρέχοντας στον χρήστη την εναλλακτική λύση που αναζητούσε. Η τελική πρόταση συγχορδίας στη θέση i αποθηκεύεται σε μία λίστα που περιλαμβάνει τις προτάσεις αντικατάστασης, κατ' αντιστοιχία θέσεων με την αρχική ακολουθία συγχορδιών. Η ίδια διαδικασία ακολουθείται για όλες τις συγχορδίες της λίστας, με αποτέλεσμα να δημιουργείται μια καινούργια λίστα που περιλαμβάνει όλες τις εναλλακτικές επιλογές αντικατάστασης συγχορδιών. Μάλιστα για κάθε συγχορδία παρέχεται και μία δεύτερη επιλογή αντικατάστασης, επιλέγοντας το Y με τη δεύτερη μεγαλύτερη συχνότητα εμφάνισης, δημιουργώντας άλλη μια λίστα με τις δεύτερες επιλογές αντικατάστασης συγχορδιών. Οι πέντε λίστες που παράχθηκαν από το συνολικό σύστημα επιστρέφονται ως απάντηση στον χρήστη.

5.4 Android Application

Η τελική εφαρμογή Android αναπτύχθηκε στο Android Studio σε λειτουργικό σύστημα Windows 10, και οι δοκιμές για την εφαρμογή έγιναν σε φυσική συσκευή, μέσω σύνδεσης καλωδίου με τον υπολογιστή. Αξίζει να σημειωθεί ότι το Android Studio παρέχει την επιλογή testing σε προσομοιωτή (Nexus 6 API 22), λύση η οποία απορρίφθηκε εξαιτίας της εξαιρετικής επιβράδυνσης που επέφερε στον υπολογιστή αλλά και της ανάγκης χρήσης μικροφώνου για τον έλεγχο των λειτουργιών της εφαρμογής, δυνατότητα που δεν υποστηρίζεται από τον

προσομοιωτή. Η ανάπτυξη του κώδικα έγινε αποκλειστικά σε γλώσσες Java και Kotlin, ενώ για την διεπαφή χρήστη χρησιμοποιήθηκαν αρχεία xml.

Για την ηχογράφηση μουσικού κομματιού μέσω του μικροφώνου της συσκευής χρησιμοποιήθηκαν οι έτοιμες κλάσεις AudioRecord¹² και MediaRecorder¹³, ενώ έγινε και περαιτέρω επεξεργασία σήματος με την κλάση ExtAudioRecorder για την βέλτιστη ποιότητα καταγραφής του ήχου. Το ηχογραφημένο αρχείο αποθηκεύεται σε μορφή .wav, καθώς αυτή είναι η μορφή εισόδου που δέχεται ο κώδικας του μοντέλου βαθιάς μάθησης, ενώ η αναπαραγωγή ήχου επιτυγχάνεται μέσω της κλάσης MediaPlayer¹⁴, που αποτελεί το πρωταρχικό API αναπαραγωγής ήχου και βίντεο για Android.

Η εφαρμογή επικοινωνεί μέσω HTTP POST requests με τον διακομιστή, με τη βοήθεια του Retrofit 2. Το Retrofit 2 είναι ένας εξαιρετικά χρήσιμος πελάτης HTTP για Android που επιτρέπει στις εφαρμογές να συνδέονται με ένα API Ιστού με ασφάλεια και πολύ λιγότερο κώδικα. Στην εφαρμογή που πραγματεύεται η τρέχουσα εργασία, το εργαλείο αυτό βοηθάει και στην διαχείριση των JSON responses που στέλνει ο διακομιστής στη συσκευή, ώστε να μετατραπούν στην επιθυμητή μορφή για την παρουσίαση στο χρήστη. Η Javascript Object Notation (JSON) είναι μια τυπική μορφή κειμένου για την αναπαράσταση δομημένων δεδομένων με βάση τη σύνταξη αντικειμένου Javascript. Συνήθως χρησιμοποιείται για τη μετάδοση δεδομένων σε διαδικτυακές εφαρμογές (π.χ. αποστολή ορισμένων δεδομένων από τον διακομιστή στον υπολογιστή -πελάτη, ώστε να μπορεί να εμφανίζεται σε μια ιστοσελίδα ή αντίστροφα) [44].

Η απάντηση που λαμβάνει το Retrofit 2 ακολουθεί τη μορφή της κλάσης "chord_model" που αποτελείται από πέντε λίστες συμβολοσειρών, με ονόματα "Chords", "Bass", "Times", "AlternativeChords" και "AlternativeChords2". Κάθε στοιχείο της λίστας "Chords" αναπαριστά μια συγχορδία, χρησιμοποιώντας το όνομα της τονικής της νότας, συμπεριλαμβανομένης της πληροφορίας αν είναι ματζόρε ή μινόρε. Κάθε συγχορδία καταλαμβάνει ένα chordView μαζί με τα στοιχεία που βρίσκονται στην αντίστοιχη θέση στις λίστες "Times", "AlternativeChords" και "AlternativeChords2". Η "AlternativeChords" περιλαμβάνει τις πρώτες και η "AlternativeChords2" τις δεύτερες εναλλακτικές προτάσεις αντικατάστασης συγχορδιών, και βρίσκονται σε απόλυτη αντιστοιχία με την λίστα "Chords", όπως και η λίστα "Times" που περιλαμβάνει τους χρόνους εκκίνησης κάθε συγχορδίας του μουσικού κομματιού, χρησιμοποιώντας ως χρονοδιάγραμμα το ηχογραφημένο αρχείο. Έτσι η απάντηση που έστειλε ο διακομιστής παρουσιάζεται στον χρήστη σε ένα UI με μία μόνο

¹² <https://developer.android.com/reference/android/media/AudioRecord>

¹³ <https://developer.android.com/reference/android/media/MediaRecorder>

¹⁴ <https://developer.android.com/guide/topics/media/mediaplayer>

δραστηριότητα, την κλάση MainActivity.java, και διάφορα views για τη βέλτιστη εμπειρία χρηστών.

6

Αξιολόγηση Συστήματος

Σε αυτό το κεφάλαιο παρουσιάζεται η αξιολόγηση του συνολικού συστήματος.

6.1 Μοντέλο Αναγνώρισης Συγχορδιών

Απαραίτητη προϋπόθεση για τη χρήση του μοντέλου βαθιάς μάθησης που περιεγράφηκε παραπάνω στην τελική εφαρμογή αποτέλεσε ο έλεγχος και η αξιολόγηση των αποτελεσμάτων που παράγει, για να υπολογιστεί η ακρίβεια των απαντήσεων που λαμβάνει ο χρήστης. Οι ακολουθίες συγχορδιών που χρησιμοποιήθηκαν για την αξιολόγηση αυτή προέρχονται από την ιστοσελίδα *isophonics*¹⁵. Πρόκειται για αρχεία κειμένου *.lab*, αποτελούμενα από τρεις στήλες που αντιστοιχούν σε χρόνο έναρξης, χρόνο μετατόπισης και ονομασία της συγχορδίας αντίστοιχα. Τα αρχεία ακολουθούν το συντακτικό σύστημα που πρότεινε ο Christopher Harte το 2010 και περιγράφεται στη διδακτορική του διατριβή [45] (η οποία υπερισχύει του εγγράφου πρότασης ISMIR του 2005 [46]). Επιλέχθηκε να χρησιμοποιηθεί ως σύνολο δεδομένων η δισκογραφία των Beatles, αποτελούμενη από 180 μουσικά κομμάτια, καθώς ήταν η πιο ασφαλής και ακριβής επιλογή εκ των διαθέσιμων datasets, ελεγμένη σε βάθος από τον Christopher Harte και την κοινότητα MIR (Music Information Retrieval) [47]. Χρησιμοποιήθηκε, επιπλέον, το αντίστοιχο σύνολο μουσικών κομματιών σε μορφή *.flac*, το οποίο μετατράπηκε εξ ολοκλήρου σε αρχεία *.wav*, για να μπορέσει να λειτουργήσει ως είσοδος του μοντέλου βαθιάς μάθησης. Να σημειώσουμε ότι εξαιτίας των δικαιωμάτων χρήσης μουσικών κομματιών, είναι εξαιρετικά δύσκολο να βρεθούν μουσικά σύνολα δεδομένων διαθέσιμα για αξιολογήσεις ερευνών.

Αποστέλλοντας αιτήματα HTTP POST στον διακομιστή με τη βοήθεια της εφαρμογής Postman¹⁶ και αποθηκεύοντας τα αποτελέσματα του μοντέλου βαθιάς μάθησης σε αρχεία κειμένου *.txt* που χρησιμοποιούν το συντακτικό σύστημα που περιεγράφηκε παραπάνω, καταφέραμε να συγκεντρώσουμε τα δεδομένα μας και με python scripts να τα συγκρίνουμε με το σύνολο δεδομένων αξιολόγησης. Αξίζει να αναφερθεί ότι το μοντέλο βαθιάς μάθησης που χρησιμοποιήθηκε έχει εκπαιδευτεί για να αναγνωρίζει μόνο μείζονες και ελάσσονες τρίφωνες

¹⁵ <http://www.isophonics.net/>

¹⁶ <https://www.postman.com/>

συγχορδίες, χωρίς να εντοπίζει κάποια παραπάνω πληροφορία για αυτές. Επομένως η σύγκριση των αποτελεσμάτων έγινε μόνο ως προς αυτήν την έκταση της πληροφορίας. Για να αξιολογηθεί η ποιότητα των αποτελεσμάτων, ακολουθήθηκε το πρότυπο αξιολόγησης MIREX (Music Information Retrieval Evaluation eXchange) [8]. Το MIREX χρησιμοποιεί συνήθως ανάκληση συμβόλου συγχορδιών (chord symbol recall - CSR) για να εκτιμήσει πόσο ταιριάζουν οι προβλεπόμενες συγχορδίες με τις πραγματικές, ακολουθώντας τον παρακάτω τύπο:

$$CSR = \frac{\text{totaldurationofsegmentswhereannotationequatestimation}}{\text{totaldurationofannotatedsegments}}$$

όπου 'totaldurationofsegmentswhereannotationequatestimation' είναι η συνολική διάρκεια των τμημάτων στα οποία η πραγματική συγχορδία του συνόλου ελέγχου ταυτίζεται με την εκτιμώμενη και 'totaldurationofannotatedsegments' είναι η συνολική διάρκεια τμημάτων που χαρακτηρίζονται από κάποια συγχορδία. Ακολουθώντας τον Christopher Harte (2010, παράγραφος 8.1.2, [45]), μπορούμε να αναπαραστήσουμε τους πραγματικούς και τους εκτιμώμενους σχολιασμούς συγχορδιών ως συνεχείς τμηματοποιήσεις του ήχου και να υπολογίσουμε το CSR λαμβάνοντας υπόψη το μήκος των σωστά επικαλυπτόμενων τμημάτων. Αυτός ο τρόπος υπολογισμού του CSR είναι ο πιο ακριβής, καθώς η ακρίβεια της μεθόδου βάσει πλαισίου, όπως υπολογιζόταν παλαιότερα, περιορίζεται από το εκάστοτε καθορισμένο μήκος πλαισίου. Ταυτόχρονα, η μετρική που χρησιμοποιήθηκε είναι και υπολογιστικά πιο αποτελεσματική, καθώς μειώνει τον αριθμό των συγκρίσεων τμημάτων. Επειδή όμως τα μουσικά κομμάτια ποικίλλουν σε διάρκεια, σταθμίζουμε το CSR με το μήκος του τραγουδιού κατά τον υπολογισμό του μέσου όρου για ένα δεδομένο κομμάτι. Ο τελικός αριθμός που παράγεται αναφέρεται ως σταθμισμένη ανάκληση συμβόλου συγχορδίας (weighted chord symbol recall - WCSR). Ωστόσο, μια δυσκολία που χρειάζεται να ληφθεί υπόψιν στο τελικό συμπέρασμα είναι το γεγονός ότι ο χρόνος στα αρχεία αξιολόγησης και στα αρχεία των εκτιμώμενων αποτελεσμάτων ίσως δεν είναι ακριβώς ισοσταθμισμένος, καθώς υπάρχει ένα περιθώριο χρόνου αντιστάθμισης στα αρχεία ήχου που χρησιμοποιήθηκαν για την εξαγωγή των αρχείων κειμένου του συνόλου ελέγχου και του παραγόμενου συνόλου. Ο μέσος όρος WCSR ολόκληρου του συνόλου δεδομένων που αντιστοιχεί στα αποτελέσματά μας, όπως προέκυψε εκτελώντας το αρχείο 'compare.py' που σύγκρινε μεταξύ τους όλα τα αντίστοιχα αρχεία, ισούται με 0.79. Η τιμή αυτή υποδεικνύει μια πολύ ικανοποιητική ακρίβεια του μοντέλου που χρησιμοποιήθηκε στην εφαρμογή για την πρόβλεψη συγχορδιών που αντιστοιχούν σε μουσικά αρχεία ήχου. Σίγουρα υπάρχουν περιθώρια βελτίωσης, όμως η τρέχουσα μορφή του μοντέλου αποτελεί μια εξαιρετικά αποδοτική επιλογή.

6.2 Αξιολόγηση Συστήματος Αντικατάστασης

Στα πλαίσια ανάπτυξης της εφαρμογής αναπτύχθηκε επίσης σύστημα προτάσεων αντικατάστασης για συγχορδίες που ανήκουν σε μουσικές ακολουθίες. Για την αξιολόγηση του συστήματος αυτού, διαιρέσαμε το σύνολο δεδομένων των 8515 ακολουθιών συγχορδιών που είχαμε στη διάθεσή μας, όπως περιεγράφηκε στο Κεφάλαιο 5, σε δύο ομάδες, αποδίδοντας κάθε φορά με τυχαία επιλογή το 80% του συνόλου στα δεδομένα εκπαίδευσης (6812 ακολουθίες συγχορδιών) και το υπόλοιπο 20% στα δεδομένα ελέγχου (1703 ακολουθίες συγχορδιών). Τα δεδομένα εκπαίδευσης χρησιμοποιήθηκαν για να εξαχθούν εκ νέου πίνακες με bigrams και trigrams. Βάσει αυτών των πινάκων, έγινε προσπάθεια πρόβλεψης κάθε συγχορδίας του συνόλου ελέγχου. Η απόδοση του συστήματος κρίθηκε από το ποσοστό ορθών προβλέψεων που εξήγαγε το σύστημα παρέχοντας είτε μία είτε δύο προτάσεις αντικατάστασης, συγκρίνοντας με τα δεδομένα ελέγχου. Ωστόσο, υπενθυμίζεται ότι το παρόν σύστημα δεν πρόκειται για σύστημα πρόβλεψης αλλά αντικατάστασης συγχορδιών, επομένως δεν αποσκοπούμε σε μεγάλη ακρίβεια πρόβλεψης των δεδομένων. Επιπλέον, τα δεδομένα που χρησιμοποιήθηκαν ανήκουν σε διάφορα μουσικά είδη, καθώς δεν υπήρχαν αρκετά δείγματα ώστε να γίνει διαχωρισμός ανά είδος. Στόχος ενός συστήματος αντικαταστάσεων συγχορδιών είναι συνήθως η παροχή προτάσεων που απέχουν από τις συνήθεις επιλογές, πάντα με την προϋπόθεση ότι το τελικό αποτέλεσμα είναι εύηχο. Για αυτόν το λόγο, για κάθε πρόταση αντικατάστασης συγχορδίας ελέγξαμε αν η πρόταση αυτή υπήρχε σε κάποια άλλη θέση της αρχικής ακολουθίας συγχορδιών στην οποία ανήκει, διότι σε αυτήν την περίπτωση είμαστε σίγουροι ότι η πρόταση αντικατάστασης είναι συμβατή με την κλίμακα και τον τρόπο του συγκεκριμένου κομματιού.

6.2.1 Βάρος N-gram

Αρχικά, για να δημιουργήσουμε τους πίνακες N-grams, κληθήκαμε να επιλέξουμε τον τρόπο υπολογισμού του βάρους κάθε N-gram. Όπως αναφέρθηκε προηγουμένως, κάθε ακολουθία συγχορδιών στα δεδομένα μας είναι μια σειρά βαθμίδων αναπαριστώμενη από ρωμαϊκούς αριθμούς. Για ένα N-gram συγχορδιών, το βάρος του αντιπροσωπεύει τη συχνότητα εμφάνισής του στο σύνολο δεδομένων μας. Δοκιμάστηκαν δύο διαφορετικοί τρόποι υπολογισμού του βάρους κάθε N-gram συγχορδιών. Έστω ότι το Π είναι μια ακολουθία διαστημάτων που δημιουργείται από μια δεδομένη ακολουθία εισόδου μετά από την καθορισμένη απλοποίηση. Έστω g_1, \dots, g_k η απαρίθμηση όλων των μοναδικών N-grams που εμφανίζονται στο Π και για κάθε i , $1 \leq i \leq k$, έστω m_i οι φορές που το N-gram g_i εμφανίζεται στο Π . Για κάθε i , $1 \leq i \leq k$, ορίζουμε ως βάρος c_i που έχει ανατεθεί στο g_i τον αριθμό $m_i / (m_1 + \dots + m_k)$. Εναλλακτικά, ορίσαμε ως βάρος c_i του g_i τον ίδιο τον αριθμό των εμφανίσεών του στο Π ,

δηλαδή $c_i = m_i$. Παρατηρήσαμε ότι τα αποτελέσματα απόδοσης των δύο αυτών τρόπων υπολογισμού του βάρους δεν παρουσιάζουν διαφορές, οπότε επιλέξαμε τον δεύτερο τρόπο για την τελική μορφή υλοποίησης του συστήματος, ως απλούστερου υπολογιστικά.

6.2.2 Πίνακας Bigrams

Στη συνέχεια έγινε έλεγχος της απόδοσης του συστήματος χρησιμοποιώντας πίνακα με bigrams που περιελάμβανε πληροφορία μόνο για το όνομα της τονικής νότας κάθε συγχορδίας, χωρίς κανένα παραπάνω χαρακτηριστικό, καθώς και για τη συχνότητα εμφάνισης που αντιστοιχεί σε κάθε bigram σύμφωνα με τα δεδομένα εκπαίδευσης. Υπενθυμίζεται ότι όλα τα N-grams αποτελούνται από διαστήματα και όχι ονόματα συγχορδιών, όπως αναλύθηκε στο Κεφάλαιο 5, και ότι οι πίνακες N-grams που αναφέρονται έχουν την μορφή rython dictionaries στην υλοποίηση. Με κατάλληλη επεξεργασία των δεδομένων, αφαιρέθηκαν όλες οι διαδοχικές επαναλήψεις συγχορδιών με την ίδια ονομασία, ακόμα κι αν αυτές διέφεραν στο αν είναι μείζονες ή ελάσσονες, καθώς αυτή η πληροφορία δεν λήφθηκε υπόψιν στην τρέχουσα δοκιμή. Συνολικά, ο πίνακας περιελάμβανε τελικά κατά μέσο όρο 79876 καταχωρήσεις bigrams (ανάλογα με τον τυχαίο διαχωρισμό των δεδομένων σε δεδομένα ελέγχου και εκπαίδευσης).

Για να βρεθεί η πρόταση αντικατάστασης της τιμής X_i που αντιστοιχεί στην θέση i της ακολουθίας, εξετάστηκαν ταυτόχρονα όλα τα ζευγάρια $[X_{i-1}, X_i]$ του πίνακα, όπως επίσης και τα ζευγάρια $[X_i, X_{i+1}]$, όπου X_{i-1} και X_{i+1} οι τιμές που βρίσκονται πριν και μετά τη X_i στην ακολουθία του συνόλου δεδομένων ελέγχου αντίστοιχα. Η τελική πρόταση αντικατάστασης προέκυψε ως εκείνη η επιλογή εκ των 12 διαθέσιμων τιμών για τη X_i που συγκέντρωσε συνολικά τη μεγαλύτερη συχνότητα εμφάνισης με βάση τον πίνακα bigrams, όπως αυτός διαμορφώθηκε από το σύνολο εκπαίδευσης. Μάλιστα, προσπαθώντας να βελτιστοποιήσουμε το σύστημα αντικατάστασης, μελετήσαμε το αντίκτυπο χρήσης διαφορετικού βάρους στο bigram της προηγούμενης και της επόμενης τιμής από την X_i στην ακολουθία, για τη διαμόρφωση του τελικού αποτελέσματος. Δοκιμάστηκε, δηλαδή, ο πολλαπλασιασμός της συχνότητας εμφάνισης καθενός από τα ζευγάρια $[X_{i-1}, X_i]$ και $[X_i, X_{i+1}]$ με διαφορετικό συντελεστή α_i και β_i αντίστοιχα, όπου $\alpha_i + \beta_i = 1$, για την εύρεση της συνολικής συχνότητας εμφάνισης του εκάστοτε X_i και την επιλογή της μέγιστης τιμής $\alpha_i * [X_{i-1}, X_i] + \beta_i * [X_i, X_{i+1}]$. Σημειώνεται ότι για το πρώτο και το τελευταίο στοιχείο κάθε ακολουθίας, το αποτέλεσμα διαμορφώθηκε χρησιμοποιώντας μόνο το bigram $[X_i, X_{i+1}]$ και $[X_{i-1}, X_i]$ αντίστοιχα.

Τα αποτελέσματά μας συνοψίζονται στον παρακάτω πίνακα, όπου βρέθηκε ο μέσος όρος ποσοστού επιτυχίας της πρότασης X_i , ως προς την πρόβλεψη της πραγματικής τιμής του συνόλου δεδομένων ελέγχου στην θέση i , σε 20 τυχαίες δοκιμές διαίρεσης του συνόλου δεδομένων. Η προτελευταία στήλη του πίνακα αντιπροσωπεύει τον αριθμό των προτάσεων

αντικατάστασης X_i (εκφρασμένο σε ποσοστό κατά μέσο όρο) οι οποίες ακόμα κι αν δεν ταυτίζονταν με την πραγματική τιμή των δεδομένων ελέγχου, εμφανίζονταν ξανά σε κάποια θέση της αρχικής ακολουθίας. Κάτι τέτοιο επιβεβαιώνει την συμβατότητά τους με την κλίμακα του κομματικού και τη σχετικότητα με το δεδομένο μουσικό απόσπασμα. Στην τελευταία στήλη παρουσιάζεται το αντίστοιχο συνολικό ποσοστό επιτυχούς πρόβλεψης, αν το αποτέλεσμα αποτελείται από δύο εναλλακτικές προτάσεις αντικατάστασης για την τιμή της X_i .

Πίνακας 6.2.2.1 Αποτελέσματα πίνακα Bigrams

Συντελεστής βαρύτητας α_i	Συντελεστής βαρύτητας β_i	Ποσοστό Επιτυχίας για 1 πρόταση αντικατάστασης	Ποσοστό σχετικότητας για την 1 ^η πρόταση αντικατάστασης	Ποσοστό επιτυχίας για 2 προτάσεις αντικατάστασης
1.0	0.0	24%	68%	41%
0.9	0.1	25%	68%	41%
0.8	0.2	25%	67%	42%
0.7	0.3	26%	66%	42%
0.6	0.4	28%	65%	43%
0.5	0.5	27%	64%	43%
0.4	0.6	28%	62%	43%
0.3	0.7	27%	60%	42%
0.2	0.8	26%	60%	42%
0.1	0.9	24%	58%	42%
0.0	1.0	24%	56%	40%

Κατά μέσο όρο αναζητήθηκαν αντικαταστάσεις για 21658 τιμών X_i με βάση το σύνολο δεδομένων ελέγχου και την προεπεξεργασία για ίδιες συγχορδίες σε σειρά. Παρατηρούμε ότι το ποσοστό επιτυχίας για μία πρόταση αντικατάστασης δεν ξεπερνάει το 28%, το οποίο μάλιστα επιτυγχάνεται με την επιβολή συντελεστή βάρους 0.6 είτε στα bigrams που περιλαμβάνουν την προηγούμενη είτε σε αυτά που περιλαμβάνουν την επόμενη τιμή διαστήματος από τη X_i . Το ποσοστό αυτό δεν φαίνεται να είναι αρκετά υψηλό. Ωστόσο, οι προτάσεις συγχορδιών φαίνεται να είναι συμβατές με την κλίμακα του κομματιού, ειδικά στις περιπτώσεις που η επιλογή μας βασίζεται περισσότερο στο bigram $[X_{i-1}, X_i]$. Συμπεραίνουμε, δηλαδή, ότι οι προτάσεις μας εντάσσονται στο μουσικό πλαίσιο του κομματιού και δεν είναι υπερβολικά τολμηρές τις περισσότερες φορές. Μάλιστα, αν το σύστημα διαμορφωθεί έτσι

ώστε να παρέχει δύο εναλλακτικές προτάσεις αντικατάστασης, το ποσοστό επιτυχούς πρόβλεψης φτάνει το 43% για $\alpha_i = 0.6$ και $\beta_i = 0.4$, το οποίο αποτελεί μια εξαιρετικά ικανοποιητική τιμή.

Όπως προαναφέρθηκε, χρειάζεται να λάβουμε υπόψιν ότι τα δεδομένα μας προέρχονται από μουσικά κομμάτια διαφόρων ειδών μουσικής, που ακολουθούν διαφορετικά μοτίβα ήχου, όπως επίσης και το γεγονός ότι είχαμε στη διάθεσή μας μόνο 6812 ακολουθίες συγχορδιών για την συλλογή των στατιστικών στοιχείων ως δεδομένα εκπαίδευσης. Ωστόσο, το συγκεκριμένο μοντέλο προτείνει ονόματα συγχορδιών αντικατάστασης χωρίς να παρέχει κάποια πληροφορία για τον τρόπο των συγχορδιών αυτών και για αυτόν τον λόγο δεν εξυπηρετεί τους χρήστες της εφαρμογής. Το αποτέλεσμα είναι αρκετά ελλιπές, επομένως η χρήση ενός τέτοιου μοντέλου για την εξαγωγή συμπερασμάτων στην τελική εφαρμογή Android δεν θα ήταν ιδιαίτερα αποδοτική.

6.2.3 Πίνακας Bigrams Μείζονα – Ελάσσονα

Συνεχίσαμε ακολουθώντας την ίδια διαδικασία, αυτή τη φορά λαμβάνοντας υπόψιν αν η κάθε συγχορδία είναι μείζονα ή ελάσσονα. Υπενθυμίζεται ότι αυτή είναι και όλη η πληροφορία που μας παρέχει το μοντέλο αναγνώρισης συγχορδιών, καθώς έχει εκπαιδευτεί για να αναγνωρίζει μόνο μείζονες και ελάσσονες συγχορδίες. Στο συγκεκριμένο μοντέλο, λοιπόν, υπάρχουν 24 δυνατές επιλογές για κάθε τιμή X_{i,x_i} , όπου $x_i = 'min'$ για ελάσσονες (minor) συγχορδίες και $x_i = 'maj'$ για μείζονες (major) συγχορδίες, έναντι των 12 του προηγούμενου μοντέλου.

Δημιουργήθηκε νέος αντίστοιχος πίνακας bigrams – συχνότητας εμφάνισης, που σε αυτήν την περίπτωση περιελάμβανε κατά μέσο όρο 80872 bigrams. Ο αριθμός αυτός είναι μεγαλύτερος από τις καταχωρήσεις της προηγούμενης μεθόδου καθώς, για παράδειγμα, το bigram δύο συνεχόμενων συγχορδιών μιας ακολουθίας που διαφέρουν μόνο στο αν είναι μείζονες ή ελάσσονες ενώ χαρακτηρίζονται από την ίδια τονική νότα, δεν προστίθεται στον προηγούμενο πίνακα bigrams, αφού οι συγχορδίες θεωρούνται ίδιες (έχουν ίδια ετικέτα ονόματος). Ωστόσο, το bigram αυτό προστίθεται στον τρέχοντα πίνακα bigrams μείζονων – ελάσσονων συγχορδιών.

Τα αποτελέσματα που λάβαμε αναζητώντας προτάσεις X_{i,x_i} συνοψίζονται παρακάτω:

Πίνακας 6.2.3.1 Αποτελέσματα πίνακα Bigrams min/maj

Συντελεστής βαρύτητας α_i	Συντελεστής βαρύτητας β_i	Ποσοστό Επιτυχίας για 1 πρόταση αντικατάστασης	Ποσοστό σχετικότητας για την 1 ^η πρόταση αντικατάστασης	Ποσοστό επιτυχίας για 2 προτάσεις αντικατάστασης
----------------------------------	---------------------------------	--	--	--

1.0	0.0	20%	51%	36%
0.9	0.1	22%	51%	38%
0.8	0.2	25%	53%	40%
0.7	0.3	26%	53%	40%
0.6	0.4	26%	52%	39%
0.5	0.5	25%	51%	39%
0.4	0.6	25%	50%	39%
0.3	0.7	24%	49%	37%
0.2	0.8	23%	47%	37%
0.1	0.9	22%	45%	36%
0.0	1.0	20%	39%	34%

Σημειώνεται ότι η προτελευταία στήλη του πίνακα αντιπροσωπεύει τον αριθμό των προτάσεων X_{i,x_i} (εκφρασμένο σε ποσοστό κατά μέσο όρο) οι οποίες ακόμα κι αν δεν ταυτίζονταν με την πραγματική τιμή των δεδομένων ελέγχου στην θέση i , εμφανίζονταν στην αρχική ακολουθία τουλάχιστον μία φορά. Η τελευταία στήλη περιέχει και πάλι το μέσο όρο ποσοστού επιτυχίας όταν το σύστημα εξάγει δύο προτάσεις αντικατάστασης για κάθε θέση i .

Παρατηρώντας τα παραπάνω αποτελέσματα καταλαβαίνουμε ότι τα ποσοστά επιτυχίας δεν διαφέρουν σημαντικά από τα προηγούμενα. Παρουσιάζουν κατά γενική εικόνα μία μικρή μείωση, επειδή σε αυτήν την περίπτωση προσφέρεται περισσότερη πληροφορία στο χρήστη και το σύστημα επιλέγει ανάμεσα σε 24 και όχι 12 τιμές. Επομένως συνολικά δεν μειώνεται ιδιαίτερα η απόδοση του συστήματος αντικατάστασης σε σχέση με πριν και ταυτόχρονα βελτιώνεται η ακρίβεια και η ειδικότητα των αποτελεσμάτων, προσθέτοντας σημαντικές λεπτομέρειες για τον τρόπο κάθε συγχορδίας. Βέλτιστο ποσοστό επετεύχθη με επιβολή μεγαλύτερου βάρους στο bigram $[X_{i-1,x_{i-1}}, X_{i,x_i}]$, για τιμές συντελεστή βάρους $\alpha_i = 0.6$ και $\alpha_i = 0.7$. Το αποτέλεσμα είναι απόλυτα συμβατό με τη μουσική θεωρία όπου είθισται η κάθε συγχορδία να εξαρτάται περισσότερο από εκείνη που προηγείται αυτής σε μία μουσική ακολουθία. Χαμηλότερο ποσοστό επιτυχίας λαμβάνουμε όταν η πρόταση αντικατάστασης βασίζεται αποκλειστικά στην προηγούμενη ή την επόμενη συγχορδία με συντελεστή 1.0, επιβεβαιώνοντας την επιλογή μας να αποφασίζουμε για την τιμή του X_{i,x_i} εξετάζοντας τις αποστάσεις του και από τις δύο συγχορδίες που την πλαισιώνουν.

Οι προτεινόμενες συγχορδίες φαίνεται να συνάδουν με την κλίμακα του κομματιού, αφού όσο $\alpha_i \geq \beta_i$ το 50% των συγχορδιών αυτών εμφανίζεται ξανά στην αρχική ακολουθία, ενώ το ποσοστό μειώνεται έως το 39% όσο η επιλογή βασίζεται περισσότερο στο bigram

$[X_{i,x_i}, X_{i+1,x_{i+1}}]$. Οι τιμές αυτές των ποσοστών είναι χαμηλότερες από πριν, όπως είναι λογικό, εφόσον πλέον έχουμε τις διπλάσιες διαθέσιμες επιλογές για την πρόταση X_{i,x_i} . Ωστόσο θα μπορούσαμε να τις θεωρήσουμε αρκετά ικανοποιητικές για ένα σύστημα αντικαταστάσεων δεδομένων των παραμέτρων της έρευνάς μας. Επιπλέον, αν επιλέξουμε να παρέχουμε στο χρήστη δύο εναλλακτικές προτάσεις συγχορδιών, τα ποσοστά ορθής πρόβλεψης της συγχορδίας που υπάρχει στην θέση i της αρχικής ακολουθίας φτάνουν το 40%.

Σημαντικό ρόλο στη διαμόρφωση των αποτελεσμάτων διαδραμάτισε το γεγονός ότι αυτή τη φορά είχαμε περισσότερες καταχωρήσεις του συνόλου εκπαίδευσης στον πίνακα bigrams, εξαιτίας των λιγότερων διπλότυπων ζευγαριών σε σειρά. Κατά μέσο όρο αναζητήθηκαν αντικαταστάσεις για 21911 προτάσεις X_{i,x_i} με βάση το σύνολο δεδομένων ελέγχου και την προεπεξεργασία για συνεχόμενες επαναλήψεις συγχορδιών στις ακολουθίες.

6.2.4 Πίνακας Trigrams

Εξετάστηκε, επίσης, η χρήση trigrams για την εξαγωγή βέλτιστων αποτελεσμάτων. Θεωρήσαμε ότι έχει ενδιαφέρον η αναζήτηση του ποσοστού επιτυχούς πρόβλεψης της συγχορδίας στη θέση i , εξετάζοντας trigrams της μορφής $[X_{i-1}, X_i, X_{i+1}]$ στα οποία η X_i βρίσκεται στη δεύτερη (μεσαία) θέση. Η πρόταση αντικατάστασης για την τιμή X_i καθορίζεται τώρα από το trigram $[X_{i-1}, X_i, X_{i+1}]$ που αντιστοιχεί στη μέγιστη συχνότητα εμφάνισης στα δεδομένα εκπαίδευσης. Για την εύρεση της πρώτης και της τελευταίας συγχορδίας κάθε ακολουθίας χρησιμοποιήθηκαν τα trigrams $[X_0, X_1, X_2]$ και $[X_{i-2}, X_{i-1}, X_i]$ αντίστοιχα. Υπενθυμίζεται ότι τα δεδομένα μας αποτελούνται από ακολουθίες τουλάχιστον τριών συγχορδιών.

Ο πίνακας trigrams περιελάμβανε κατά μέσο όρο 73060 καταχωρήσεις από το σύνολο εκπαίδευσης, έχοντας αφαιρέσει τις διπλές συγχορδίες που βρίσκονταν σε σειρά. Όπως είναι προφανές, σε αυτήν την περίπτωση δεν χρησιμοποιήθηκαν βάρη, καθώς για κάθε πιθανή τιμή του X_i δόθηκε μόνο μια επιλογή trigram για μέτρηση συχνότητας εμφάνισης. Χρησιμοποιώντας τον πίνακα που δεν περιείχε πληροφορίες εκτός του ονόματος που χαρακτηρίζει κάθε συγχορδία και της συχνότητας εμφάνισης κάθε trigram, έχοντας 12 διαθέσιμες επιλογές για την εκάστοτε τιμή πρότασης X_i , για 20 τυχαίες επαναλήψεις έχουμε:

Πίνακας 6.2.4.1 Αποτελέσματα πίνακα Trigrams

Ποσοστό Επιτυχίας για 1 πρόταση αντικατάστασης	Ποσοστό σχετικότητας για την 1 ^η πρόταση αντικατάστασης	Ποσοστό επιτυχίας για 2 προτάσεις αντικατάστασης
34%	69%	53%

Ο μέσος όρος ποσοστού επιτυχίας για μία πρόταση αντικατάστασης είναι 34%, ποσοστό που ξεπερνάει όλα τα προηγούμενα αποτελέσματα. Ωστόσο, η πληροφορία που παρέχεται είναι και πάλι ελλιπής και δεν είναι ιδανική για χρήση στην εφαρμογή.

Το ποσοστό των προτάσεων συγχορδιών που εμφανίζονται ξανά στην αρχική ακολουθία είναι 69%. Προτείνοντας δύο εναλλακτικές συγχορδίες, το ποσοστό επιτυχίας φτάνει το 53%. Οι υψηλές αυτές τιμές υποδεικνύουν ίσως μια μορφή υπερεξειδίκευσης του μοντέλου αντικαταστάσεων, καθώς προτείνονται πολύ συχνά συγχορδίες που υπάρχουν ήδη στην ακολουθία, γεγονός που ίσως οφείλεται στο μικρό αριθμό διαθέσιμων δεδομένων εκπαίδευσης. Ένα ιδεατό σύστημα προτάσεων αντικατάστασης θα θέλαμε να προτείνει συγχορδίες που ταιριάζουν στο μουσικό πλαίσιο του κομματιού αλλά ταυτόχρονα να μην είναι εντελώς ίδιες με τις ήδη υπάρχουσες συγχορδίες της ακολουθίας. Συνολικά αναζητήθηκαν κατά μέσο όρο 21576 συγχορδίες X_i για αντικατάσταση, ενώ ο πίνακας trigrams περιελάμβανε περίπου 72977 trigrams.

6.2.5 Πίνακας Trigrams Μείζονα – Ελάσσονα

Τέλος εκτελέσαμε την ίδια διαδικασία για τον πίνακα Trigrams μείζονων και ελάσσονων συγχορδιών με τις αντίστοιχες μεθόδους εξαγωγής αποτελεσμάτων, έχοντας 24 διαθέσιμες επιλογές για την εκάστοτε τιμή της X_{i,x_i} .

Πίνακας 6.2.5.1 Αποτελέσματα πίνακα Trigrams min/maj

Ποσοστό Επιτυχίας για 1 πρόταση αντικατάστασης	Ποσοστό σχετικότητας για την 1 ^η πρόταση αντικατάστασης	Ποσοστό επιτυχίας για 2 προτάσεις αντικατάστασης
38%	63%	59%

Το ποσοστό επιτυχίας για μία πρόταση συγχορδίας έφτασε το 38%, με 22561 αναζητήσεις κατά μέσο όρο, σε έναν πίνακα με 76945 καταχωρήσεις. Αυτό το αποτέλεσμα φαίνεται να ξεπερνάει σημαντικά όλα τα προηγούμενα ποσοστά επιτυχίας για μία πρόταση αντικατάστασης και να πετυχαίνει μια πολύ ικανοποιητική τιμή, δεδομένου και του περιορισμένου συνόλου δεδομένων. Το 63% των προτάσεων συγχορδιών αυτού του μοντέλου εμφανίζονταν ξανά στις αρχικές ακολουθίες, ενώ για δύο προτάσεις αντικατάστασης, το ποσοστό επιτυχούς πρόβλεψης της συγχορδίας X_{i,x_i} φτάνει το 59%.

Παρότι οι τιμές αυτές είναι οι πιο υψηλές σε σχέση με τα προηγούμενα μοντέλα, με τη χρήση trigrams παρατηρείται μια υπερεξειδίκευση του μοντέλου, καθώς πολλά trigrams του πίνακα έχουν μοναδιαία συχνότητα εμφάνισης και το αποτέλεσμα που παρέχει το σύστημα ίσως να μην αποκλίνει αρκετά από τη «συνηθέστερη» επιλογή. Παρέχονται, δηλαδή, πιο συμβατικές

επιλογές, ενώ κάποιοι συνδυασμοί διαστημάτων αποκλείονται εντελώς. Ένα σύστημα προτάσεων αντικατάστασης θα θέλαμε ιδανικά να έχει τη δυνατότητα να προσφέρει εναλλακτικές προτάσεις συγχορδιών στον χρήστη που ίσως ο ίδιος να μην σκεφτόταν. Επιπλέον, όπως είδαμε και στις δοκιμές bigrams, η πρόταση συγχορδίας ίσως θα έπρεπε να βασίζεται λίγο περισσότερο σε εκείνη που προηγείται αυτής στην αρχική ακολουθία. Χρησιμοποιώντας trigrams κάτι τέτοιο δεν είναι εφικτό.

6.2.6 Συνδυαστικός πίνακας

Προσπαθώντας να καταλήξουμε σε μια ενδιάμεση βέλτιστη λύση για την εφαρμογή μας, εκτελέσαμε τις ίδιες διαδικασίες ελέγχου που περιεγράφηκαν παραπάνω, συνδυάζοντας στοιχεία από bigrams και trigrams, πάντα συμπεριλαμβανομένης της πληροφορίας μείζονος ή ελάσσονος συγχορδίας. Για την πρώτη και την τελευταία συγχορδία κάθε ακολουθίας, το αποτέλεσμα λήφθηκε υπολογίζοντας το άθροισμα της συχνότητας εμφάνισης του bigram $[X_{i,x_i}, X_{i+1,x_{i+1}}]$ και του trigram $[X_{i,x_i}, X_{i+1,x_{i+1}}, X_{i+2,x_{i+2}}]$ ή του bigram $[X_{i-1,x_{i-1}}, X_{i,x_i}]$ και του trigram $[X_{i-2,x_{i-2}}, X_{i-1,x_{i-1}}, X_{i,x_i}]$ αντίστοιχα. Για τις υπόλοιπες συγχορδίες υπολογίζονταν κάθε φορά το άθροισμα συχνότητας εμφάνισης bigrams και trigram $[X_{i,x_i}, X_{i+1,x_{i+1}}] * \alpha_i + [X_{i-1,x_{i-1}}, X_{i,x_i}] * \beta_i + [X_{i,x_i}, X_{i+1,x_{i+1}}, X_{i+2,x_{i+2}}]$, όπου α_i και β_i οι συντελεστές βαρύτητας bigram της προηγούμενης και της επόμενης συγχορδίας αντίστοιχα. Τα αποτελέσματα που λάβαμε συνοψίζονται στον παρακάτω πίνακα:

Πίνακας 6.2.6.1 Αποτελέσματα πίνακα Bigrams και Trigrams min/maj

Συντελεστής βαρύτητας α_i	Συντελεστής βαρύτητας β_i	Ποσοστό Επιτυχίας για 1 πρόταση αντικατάστασης	Ποσοστό σχετικότητας για την 1 ^η πρόταση αντικατάστασης	Ποσοστό επιτυχίας για 2 προτάσεις αντικατάστασης
1.0	0.0	27%	56%	41%
0.9	0.1	29%	57%	42%
0.8	0.2	29%	58%	43%
0.7	0.3	30%	59%	44%
0.6	0.4	29%	57%	43%
0.5	0.5	29%	55%	43%
0.4	0.6	28%	54%	42%
0.3	0.7	28%	53%	42%
0.2	0.8	27%	51%	41%

0.1	0.9	27%	49%	40%
0.0	1.0	26%	47%	38%

Κατά μέσο όρο αναζητήθηκαν αντικαταστάσεις για 21908 τιμές X_{i,x_i} με βάση το σύνολο δεδομένων ελέγχου και την προεπεξεργασία για διπλές τιμές.

Παρατηρούμε ότι με τον συνδυασμό της πληροφορίας από τα trigrams και τα bigrams επιτυγχάνουμε ένα μέγιστο ποσοστό επιτυχίας 30%, για συντελεστές βαρύτητας $\alpha_i = 0.7$ και $\beta_i = 0.3$. Για αυτές τις τιμές συντελεστών βαρύτητας, φαίνεται να έχουμε και το μέγιστο ποσοστό προτάσεων συγχορδιών που ανήκουν σίγουρα στην κλίμακα της ακολουθίας, που αγγίζει το 59%. Επιλέγοντας να εξάγουμε 2 εναλλακτικές προτάσεις για κάθε συγχορδία, το ποσοστό επιτυχίας του βέλτιστου συνδυασμού συντελεστών φτάνει έως και το 44%. Τα ποσοστά αυτά είναι πολύ ικανοποιητικά και για την απόδοση του συστήματος σε σχέση με τα προηγούμενα μοντέλα, και ως προς την επιτυχή πρόβλεψη της συγχορδίας στη θέση i αλλά και ως προς τη συνάφεια της πρότασης με την κλίμακα του μουσικού κομματιού. Αποφεύγεται, κατά το δυνατόν, ο κίνδυνος υπερεξειδίκευσης των αποτελεσμάτων, εφόσον ακόμα και με δύο επιλογές το ποσοστό επιτυχίας δεν υπερβαίνει το 50%, ενώ ταυτόχρονα αυξάνεται η χαμηλή ακρίβεια αποτελεσμάτων των μεθόδων που χρησιμοποιούν μόνο bigrams και δίνεται περισσότερη σημασία στην συγχορδία που προηγείται της ζητούμενης σε σχέση με την επόμενη.

6.2.7 Τελική επιλογή - Συμπεράσματα

Μετά από αυτήν την διαδικασία αξιολόγησης του συστήματος προτάσεων αντικαταστάσεων συγχορδιών, διαμορφώθηκε ο παρακάτω πίνακας που συνοψίζει τα βέλτιστα αποτελέσματα κάθε μεθόδου:

Πίνακας 6.2.7.1 Σύνοψη αποτελεσμάτων αξιολόγησης

	Ποσοστό Επιτυχίας 1 ^{ης} επιλογής	Ποσοστό Σχετικότητας 1 ^{ης} επιλογής	Ποσοστό Επιτυχίας 1 ^{ης} και 2 ^{ης} εναλλακτικής επιλογής
Bigrams	28%	65%	42%
Bigrams min/maj	26%	53%	40%
Trigrams	34%	69%	53%
Trigrams min/maj	38%	63%	59%

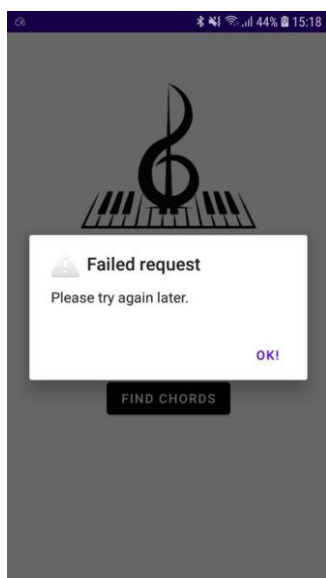
Bigrams + Trigrams min/maj	30%	59%	44%
---------------------------------------	-----	-----	-----

Επιλέξαμε να χρησιμοποιήσουμε στην εφαρμογή μας τη συνδυαστική μέθοδο χρήσης Bigrams και Trigrams Μείζονων και Ελάσσονων συγχορδιών με συντελεστές $\alpha_i = 0.7$ και $\beta_i = 0.3$, λαμβάνοντας πληροφορία για τους πίνακες συχνοτήτων εμφάνισης N-grams και από τις 8515 ακολουθίες συγχορδιών που μας δόθηκαν, για ακόμα καλύτερα αποτελέσματα. Αυτό το μοντέλο παρουσιάζει τον καλύτερο συνδυασμό ακρίβειας αλλά και εναλλακτικότητας αποτελεσμάτων, γι' αυτό και αποτελεί την πιο συμφέρουσα επιλογή. Για βέλτιστα αποτελέσματα και εξάλειψη τυχόν εσφαλμένων προτάσεων, αποφασίσαμε να παρέχουμε στον χρήστη δύο εναλλακτικές προτάσεις συγχορδιών.

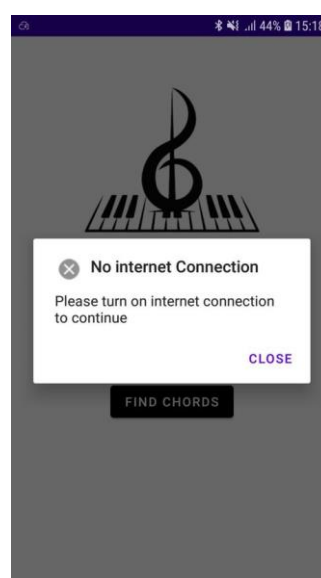
Υπάρχουν αρκετές προοπτικές βελτίωσης του συστήματος, με χρήση μεγαλύτερου συνόλου δεδομένων και διαχωρισμό ανάλογα με το μουσικό είδος κάθε κομματιού, όμως το τρέχον σύστημα φαίνεται να είναι αρκετό για προτάσεις εναλλακτικών συγχορδιών, που διαφέρουν από την πρώτη και τη συνήθη επιλογή του μέσου χρήστη. Εξ' άλλου στόχος του συστήματος αντικαταστάσεων είναι η πρόταση μιας εναλλακτικής συγχορδίας που θα προσθέτει κυριολεκτικά και μεταφορικά μια διαφορετική νότα στο μουσικό κομμάτι. Η συγχορδία που επιλέγεται κάθε φορά με βάση τις αποστάσεις της από την προηγούμενη και την επόμενη συγχορδία, αποτελεί μια επιλογή που ως προς τα διαστήματα, εμφανίζεται σίγουρα σε αρκετές ακολουθίες του συνόλου δεδομένων και μάλιστα η αξιολόγηση του μοντέλου απέδειξε ότι περίπου το 60% των προτάσεων ανήκουν σίγουρα στην κλίμακα της αρχικής ακολουθίας. Συνεπώς μπορούμε με ασφάλεια να ισχυριστούμε ότι κάθε πρόταση αντικατάστασης δεν πρόκειται για μία συγχορδία που δεν σχετίζεται με τον τόνο, την κλίμακα και τη ροή του μουσικού κομματιού.

6.3 Έλεγχος της Εφαρμογής

Για τον έλεγχο της εφαρμογής, εκτελέστηκαν πολλά και διαφορετικά σενάρια χρήσης, προσομοιώνοντας την εφαρμογή σε 5 διαφορετικές φυσικές συσκευές Android. Με τη βοήθεια breakpoints που διακόπτουν την εκτέλεση της εφαρμογής σε μια καθορισμένη γραμμή κώδικα, μπορέσαμε να εξετάσουμε μεταβλητές, να αξιολογήσουμε εκφράσεις και να ακολουθήσουμε την εκτέλεση των λειτουργιών της εφαρμογής γραμμή προς γραμμή για να προσδιοριστούν οι αιτίες σφαλμάτων χρόνου εκτέλεσης. Μάλιστα προστέθηκαν κατάλληλα μηνύματα σφαλμάτων σε κάθε περίπτωση αποτυχίας χρήσης της εφαρμογής, ώστε ο χρήστης να ενημερώνεται για την αιτία του εκάστοτε σφάλματος (π.χ. αποτυχία σύνδεσης στο διαδίκτυο, σφάλματα διακομιστή).



Εικόνα 6.3.1 Σφάλμα Διακομιστή



Εικόνα 6.3.2 Σφάλμα σύνδεσης στο διαδίκτυο

Εξετάστηκε επίσης η απόδοση και ο χρόνος απόκρισης του συνολικού συστήματος σε αιτήματα αναγνώρισης και αντικατάστασης συγχορδιών μέσα από την κινητή συσκευή. Ως αποτέλεσμα, διαπιστώθηκε ότι ανεξάρτητα από τους υπολογιστικούς πόρους και την αρχιτεκτονική κάθε συσκευής, ο χρόνος αναμονής της απάντησης επηρεάζεται μόνο από το μέγεθος και την ποιότητα του μουσικού αρχείου που επιθυμούμε να αναλυθεί. Όσο πιο μεγάλο είναι το αρχείο ήχου, τόσο περισσότερος χρόνος απαιτείται για την αποστολή και την επεξεργασία του, όπως είναι λογικό. Επιπλέον, οποιαδήποτε ενημέρωση χρειάστηκε να εφαρμοστεί στο σύστημα αναγνώρισης και αντικατάστασης συγχορδιών δεν επηρέασε την εγκατεστημένη εφαρμογή στις κινητές συσκευές. Οι ενημερώσεις εφαρμόστηκαν κατευθείαν στον διακομιστή και ήταν διαθέσιμες για χρήση άμεσα σε οποιαδήποτε συσκευή, χωρίς να χρειάζεται ενημέρωση της έκδοσης εφαρμογής σε κάθε κινητό. Καταφέραμε, λοιπόν, να αξιοποιήσουμε επιτυχώς και τα πλεονεκτήματα χρήσης απομακρυσμένου διακομιστή για την επικοινωνία με τα μοντέλα.

7

Επίλογος

7.1 Σύνοψη και συμπεράσματα

Στην παρούσα εργασία αναπτύχθηκε μια εφαρμογή κινητών συσκευών Android που ενσωματώνει μεθόδους βαθιάς μάθησης για την αναγνώριση συγχορδιών και ταυτόχρονα παρέχει προτάσεις αντικατάστασης συγχορδιών, αντλώντας πληροφορίες από ένα σύστημα βασισμένο σε στατιστικά στοιχεία συχνότητας εμφάνισης μουσικών διαστημάτων. Πρόκειται για έναν πρωτοποριακό συνδυασμό δύο πολύ χρήσιμων μουσικών λειτουργιών σε μια εφαρμογή, η οποία θα μπορούσε να επιφέρει θετικές εξελίξεις στον τομέα της μουσικής τεχνολογίας.

Μετά από αξιολόγηση των διαθέσιμων επιλογών και αρκετές δοκιμές κάθε μίας από αυτές, επιλέχθηκε η εξαγωγή συμπερασμάτων του μοντέλου βαθιάς μάθησης να πραγματοποιείται σε απομακρυσμένο διακομιστή και όχι στην ίδια τη συσκευή, καθώς φαίνεται ότι τα πλεονεκτήματα αυτής της μεθόδου υπερέχουν των μειονεκτημάτων. Με την υλοποίηση και την εφαρμογή αυτής της επιλογής, διασφαλίστηκε η ταχύτητα και η ακρίβεια των αποτελεσμάτων του μοντέλου ανεξάρτητα από τα χαρακτηριστικά και τους υπολογιστικούς πόρους της εκάστοτε συσκευής στην οποία λειτουργεί η εφαρμογή.

Εξετάστηκε, επίσης, η βέλτιστη μέθοδος συλλογής και χρήσης στατιστικών στοιχείων για τη συχνότητα εμφάνισης συνδυασμών διαστημάτων μεταξύ συγχορδιών σε μουσικά κομμάτια για ένα σύστημα αντικαταστάσεων. Διαπιστώθηκε ότι, αποσκοπώντας στην πρόταση συγχορδιών που συμφωνούν μουσικά με το πλαίσιο ενός κομματιού αλλά ταυτόχρονα παρεκκλίνουν λίγο από το βασικό μοτίβο του, χρειάζεται να χρησιμοποιήσουμε ένα συνδυασμό στοιχείων, εξετάζοντας δυάδες αλλά και τριάδες διαδοχικών διαστημάτων μεταξύ συγχορδιών που εμφανίζονται σε μουσικές ακολουθίες. Τα αποτελέσματα που λάβαμε μετά από μια τέτοια ανάλυση ήταν αρκετά ικανοποιητικά για ένα σύστημα αντικαταστάσεων εφαρμογής κινητού.

Οι λειτουργίες αυτές συγκεντρώθηκαν σε μία εφαρμογή που απαιτεί σύνδεση δικτύου για επικοινωνία με απομακρυσμένο διακομιστή, και παρουσιάστηκαν σε μία φιλική προς το χρήστη διεπαφή, εμπλουτισμένη με λογότυπα, εικονίδια και σαφείς οδηγίες χρήσης.

7.2 Μελλοντικές επεκτάσεις

Μελλοντικές επεκτάσεις που αφορούν την παρούσα εργασία θα μπορούσαν να εφαρμοστούν σε διάφορα τμήματα της παραπάνω εφαρμογής, καθώς αποτελείται από αρκετά επιμέρους κομμάτια που δέχονται βελτιώσεις. Αρχικά έντονο ενδιαφέρον δείχνει να έχει η βελτίωση του μοντέλου βαθιάς μάθησης για την αναγνώριση συγχορδιών, ίσως ενσωματώνοντας την δυνατότητα αναγνώρισης περισσότερων χαρακτηριστικών μιας συγχορδίας και όχι μόνο το όνομά της και τον μείζονα ή ελάσσονα τρόπο. Απαιτείται η επανεκπαίδευση του μοντέλου, ίσως με περισσότερα και διαφορετικότερα δεδομένα, σε μουσικά κομμάτια που χρησιμοποιούν πιο εξειδικευμένες συγχορδίες (όπως συμβαίνει στην τζαζ μουσική, όπου συναντώνται συχνά συγχορδίες της μορφής 'C13b9#11'). Η αναγνώριση συγχορδιών θα μπορούσε να γίνεται σε πραγματικό χρόνο ηχογράφησης, ενώ θα ήταν πολύ ενδιαφέρουσα και η παροχή πληροφορίας για τη θέση στην οποία θα ήταν βέλτιστη η αναπαραγωγή της, ίσως ανάλογα και με το μουσικό όργανο που επιλέγει ο χρήστης. Επιπλέον, ένας εναλλακτικός τρόπος εξοικονόμησης πόρων δικτύου ίσως θα ήταν η συμπίεση του αρχείου ήχου wav που εισάγεται στο μοντέλο, μετατρέποντάς το για παράδειγμα σε μορφή mp3. Έτσι θα μειώναμε και τον χρόνο απόκρισης των αιτημάτων HTTP της εφαρμογής.

Ως προς την αντικατάσταση συγχορδιών, μια επέκταση του παραπάνω συστήματος θα αποτελούσε η ανάπτυξη ενός ακριβέστερου αλγορίθμου για την παραγωγή αρμονικής εναλλαγής, συνδυάζοντας τη θεωρία και τους κανόνες αντικατάστασης συγχορδιών, δίνοντας τη δυνατότητα στον χρήστη να επιλέξει ίσως και το μουσικό πλαίσιο στο οποίο επιθυμεί να εντάσσεται η σύνθεσή του. Η αντικατάσταση συγχορδιών με τον τρόπο που προσεγγίστηκε στην παρούσα εργασία είναι ιδιαίτερα χρήσιμη, ωστόσο δεν μπορεί να παρέχει μια λύση που να ταιριάζει σε όλους. Υπάρχουν διάφοροι παράγοντες που επηρεάζουν το πόσο μεγάλες αλλαγές μπορεί να δεχτεί ένας συνθέτης για να διαφοροποιήσει ένα μέρος της ακολουθίας συγχορδιών του. Οι μουσικές και στιλιστικές επιδιώξεις του επηρεάζουν την δεκτικότητά του σε σπάνιες συγχορδίες, και αν η πρόθεσή του αποσκοπεί στην δημιουργία ενός πιο οικείου μουσικού κομματιού για τον ακροατή, οι συστάσεις αντικατάστασης οφείλουν να ακολουθούν πιο συνηθισμένα μοτίβα. Εάν, αντίθετα, θέλει να ξεφύγει από τα συνηθισμένα μουσικά πρότυπα, οι προτάσεις θα μπορούσαν να αφορούν πιο σπάνιους συνδυασμούς. Μια ακόμη ιδέα, λοιπόν, θα ήταν η ανάπτυξη ενός μηχανισμού που επιτρέπει ρητά στους συνθέτες να προσαρμόζουν το πόσο τυπικές ή περιπετειώδεις θέλουν να είναι οι προτάσεις αντικατάστασης. Το σύστημα θα μπορούσε επίσης να επιτρέπει στους συνθέτες να ορίζουν το μουσικό είδος στο οποίο επιθυμούν να ανήκει η σύνθεσή τους, και έχοντας διεξάγει μια πιο λεπτομερή ανάλυση ανάλογα με την κατηγορία του κομματιού, το σύστημα να καταφεύγει σε κατάλληλα προσαρμοσμένους πίνακες ανάλογα με αυτήν την επιλογή. Η χρήση τεχνητής νοημοσύνης ή

κάποιας μουσικής οντολογίας για την βελτίωση του συστήματος συστάσεων συγχορδιών, χρησιμοποιώντας ένα μεγαλύτερο σύνολο δεδομένων διαχωρισμένο ανάλογα με το μουσικό είδος κάθε κομματιού, θα μπορούσε να σημειώσει εξαιρετική βελτίωση της απόδοσης και των δυνατοτήτων της εφαρμογής. Αδιαμφισβήτητα πρόκειται για έναν τομέα που όσο αναπτύσσεται θα επιφέρει εξαιρετικά θετικές επιδράσεις, τόσο στην εξέλιξη της μουσικής, όσο και στην διεύρυνση των ιδεών ερασιτεχνών αλλά και επαγγελματιών μουσικών.

8

Βιβλιογραφία

- [1] I. Δ. Χριστοφίλου, “*Θεωρία της Μουσικής.*” Παγκράτι - Αθήνα: Εκδόσεις “MUSIC LOVERS,” 1985.
- [2] E. Prout, *Harmony, Its Theory and Practice (30th edition, revised and largely rewritten ed.)*. London: Augener; Boston: Boston Music Co., 1903.
- [3] J.-P. Briot, G. Hadjeres, and F.-D. Pachet, “Deep Learning Techniques for Music Generation -- A Survey,” no. September, 2017, [Online]. Available: <http://arxiv.org/abs/1709.01620>.
- [4] S. S. Stevens, J. Volkman, and E. B. Newman, “A Scale for the Measurement of the Psychological Magnitude Pitch,” *J. Acoust. Soc. Am.*, vol. 8, no. 3, pp. 185–190, 1937, doi: 10.1121/1.1915893.
- [5] J. R. Hayes and H. A. Simon, *Understanding written problem instructions*. In L. W. Gregg (Ed.), *Knowledge and cognition*. Lawrence Erlbaum, 1974.
- [6] T. Mitchel, *Machine learning*, English. London: McGraw-Hill Pub. Co. (ISE Editions), 1997.
- [7] N. D. Lane, C. Forlivesi, and F. Kawsar, “Squeezing Deep Learning into Mobile and Embedded Devices,” vol. 16, no. 3, pp. 82–88, 2017, doi: 10.1109/ISCA.2016.11.
- [8] J. S. Downie, “The music information retrieval evaluation exchange (2005-2007): A window into music information retrieval research,” *Acoust. Sci. Technol.*, vol. 29, no. 4, pp. 247–255, 2008, doi: 10.1250/ast.29.247.
- [9] J. P. Bello, “Audio-based cover song retrieval using approximate chord sequences: testing shifts, gaps, swaps and beats.”
- [10] H.-T. Cheng, Y.-H. Yang, Y.-C. Lin, I.-B. Liao, and H. H. Chen, “Automatic chord recognition for music classification and retrieval.”
- [11] R. B. Dannenberg and N. Hu, “Polyphonic Audio Matching for Score Following and Intelligent Audio Editors,” *Proc 2003 ICMC/ICMA*, no. May, pp. 27–33, 2003, [Online]. Available: <https://www.cs.cmu.edu/~rbd/papers/icmc03matchingaudio.pdf>.
- [12] J. Pauwels, F. Kaiser, and G. Peeters, “Combining Harmony-Based and Novelty-Based Approaches for Structural Segmentation.”

- [13] A. Laaksonen, "Automatic melody transcription based on chord transcription," *Proc. 15th Int. Soc. Music Inf. Retr. Conf. ISMIR 2014*, no. Ismir, pp. 119–124, 2014.
- [14] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, "Audio Chord Recognition with Recurrent Neural Networks."
- [15] F. Korzeniowski and G. Widmer, "A fully convolutional deep auditory model for musical chord recognition," *IEEE Int. Work. Mach. Learn. Signal Process. MLSP*, vol. 2016-November, Nov. 2016, doi: 10.1109/MLSP.2016.7738895.
- [16] Nomad AI and Bellec Research, "Chord AI - Mobile Application." [Online]. Available: <https://www.chordai.net/>.
- [17] M. Chemillier, "Steedman's grammar for jazz chord sequences," *Soft Comput.*, no. special issue on Formal Systems and Music.
- [18] F. Pachet, "Surprising Harmonies," Paris, France, Mar. 1999. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-03112196>.
- [19] C.-H. Chuan and E. Chew, "A Hybrid System for Automatic Generation of Style-Specific Accompaniment," 2007.
- [20] D. Morris, I. Simon, and S. Basu, "Exposing Parameters of a Trained Dynamic Model for Interactive Music Creation," 2008, [Online]. Available: www.aaai.org.
- [21] I. Simon, D. Morris, and S. Basu, "MySong: Automatic accompaniment generation for vocal melodies," *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 725–734, 2008, doi: 10.1145/1357054.1357169.
- [22] C. A. Huang, C.-Z. Anna, D. Duvenaud, and K. Z. Gajos, "ChordRipple: Recommending Chords to Help Novice Composers Go Beyond the Ordinary," pp. 241–250, 2016, doi: 10.1145/2856767.2856792.
- [23] B. A. Nijstad, W. Stroebe, and H. F. M. Lodewijkx, "Cognitive stimulation and interference in groups: Exposure effects in an idea generation task," *J. Exp. Soc. Psychol.*, vol. 38, no. 6, pp. 535–544, Nov. 2002, doi: 10.1016/S0022-1031(02)00500-0.
- [24] P. Siangliulue, K. C. Arnold, K. Z. Gajos, and S. P. Dow, "Toward Collaborative Ideation at Scale-Leveraging Ideas from Others to Generate More Creative and Diverse Ideas," *Proc. 18th ACM Conf. Comput. Support. Coop. Work Soc. Comput.*, doi: 10.1145/2675133.
- [25] G. Schaefer, B. Krawczyk, M. E. Celebi, and H. Iyatomi, "An ensemble classification approach for melanoma diagnosis," *Memetic Comput.*, vol. 6, no. 4, pp. 233–240, Dec. 2014, doi: 10.1007/s12293-014-0144-8.
- [26] E. Nasr-Esfahani *et al.*, "Melanoma detection by analysis of clinical images using

- convolutional neural network,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, Oct. 2016, vol. 2016-
Octob, pp. 1373–1376, doi: 10.1109/EMBC.2016.7590963.
- [27] A. E. Coca, D. C. Correa, and L. Zhao, “Computer-aided music composition with LSTM neural network and chaotic inspiration,” *Proc. Int. Jt. Conf. Neural Networks*, 2013, doi: 10.1109/IJCNN.2013.6706747.
- [28] G. Tzanetakis, N. Hu, and R. B. Dannenberg, “Toward an intelligent editor for jazz music,” no. May 2014, 2003, doi: 10.1142/9789812704337.
- [29] U. S. Cunha and G. Ramalho, “An intelligent hybrid model for chord prediction An Intelligent Hybrid Model for Chord Prediction,” no. December, 2013, doi: 10.1017/S1355771899002071.
- [30] J. Y. Vajda Peter, “Delivering real-time AI in the palm of your hand - Facebook Engineering,” 2016. <https://engineering.fb.com/2016/11/08/android/delivering-real-time-ai-in-the-palm-of-your-hand/>.
- [31] A. G. Howard *et al.*, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” 2017.
- [32] S. S. Ogden and T. Guo, “Characterizing the Deep Neural Networks Inference Performance of Mobile Applications,” 2019.
- [33] W. X. Crankshaw Daniel and Z. G. J. Franklin Michael, “Clipper: A Low-Latency Online Prediction Serving System,” Boston, MA, USA, 2017.
- [34] M. Satyanarayanan, P. Bahl, R. Cáceres, and N. Davies, “The case for VM-based cloudlets in mobile computing,” *IEEE Pervasive Comput.*, vol. 8, no. 4, pp. 14–23, Oct. 2009, doi: 10.1109/MPRV.2009.82.
- [35] S. S. Ogden and T. Guo, “Mobile Deep Inference Made Efficient by Edge Computing.” 2018.
- [36] V. Koukis, C. Venetsanopoulos, and N. Koziris, “~Okeanos: Building a cloud, cluster by cluster,” *IEEE Internet Comput.*, vol. 17, no. 3, pp. 67–71, 2013, doi: 10.1109/MIC.2013.43.
- [37] V. Kataria, “Feature Column in TensorFlow,” Aug. 20, 2019. <https://medium.com/@vipin.kataria2209/feature-column-in-tensorflow-4e107ce7cb4a>.
- [38] B. B. Damodaran, V. Seguy, R. Flamary, and N. Courty, “An Entropic Optimal Transport Loss for Learning Deep Neural Networks under Label Noise in Remote Sensing Images.”
- [39] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, “RWC Music Database: Popular, Classical, and Jazz Music Databases.”

- [40] V. Eremenko, E. Demirel, B. Bozkurt, and X. Serra, “JAAH: Audio-aligned jazz harmony dataset,” Jun. 2018, doi: 10.5281/ZENODO.1290737.
- [41] J. S. Downie, “Evaluating a simple approach to music information retrieval : conceiving melodic n-grams as text,” *undefined*, 1999.
- [42] S. Doraisamy and S. Rüger, “Robust polyphonic music retrieval with N-grams,” *J. Intell. Inf. Syst.*, vol. 21, no. 1, Jul. 2003, doi: 10.1023/A:1023553801115.
- [43] Σ. Β. Κανταρέλης, “Ανάπτυξη Οντολογίας και Αλγορίθμου για την Αρμονική Ανάλυση Ακολουθιών Μουσικών Συγχορδιών,” Αθήνα, 2019.
- [44] F. Pezoa, J. L. Reutter, F. Suarez, M. Ugarte, and D. Vrgoč, “Foundations of JSON Schema,” doi: 10.1145/2872427.2883029.
- [45] C. Harte, “Towards automatic extraction of harmony information from music signals,” no. August, p. 283, 2010.
- [46] C. Harte, M. Sandler, S. Abdallah, and E. Gómez, “Symbolic representation of musical chords: A proposed syntax for text annotations,” *ISMIR 2005 - 6th Int. Conf. Music Inf. Retr.*, pp. 66–71, 2005.
- [47] J. Futrelle and J. S. Downie, “Interdisciplinary Communities and Research Issues in Music Information Retrieval,” 2002.