ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΠΛΗΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Αυτόματη ταξινόμηση και ανάλυση Βυζαντινής Μουσικής με τεχνολογίες Τεχνητής Νοημοσύνης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Νεόφυτος Χ. Παπασάββας

Επιβλέπων : Γεώργιος Στάμου
Καθηγητής Ε.Μ.Π

Αθήνα, Νοέμβριος, 2020

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΠΛΗΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Αυτόματη ταξινόμηση και ανάλυση Βυζαντινής Μουσικής με τεχνολογίες Τεχνητής Νοημοσύνης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Νεόφυτος Χ. Παπασάββας

Επιβλέπων : Γεώργιος Στάμου
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 18η Νοεμβρίου 2020.

.....	.....	.....
Γεώργιος Στάμου	Στέφανος Κόλιας	Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π	Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π

Αθήνα, Νοέμβριος, 2020

.....

Νεόφυτος Χ. Παπασάββας

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

Περίληψη

Η παρούσα διπλωματική εργασία ασχολείται με την ταξινόμηση ψαλμωδιών στους οκτώ ήχους της Βυζαντινής Μουσικής. Θα γίνει χρήση τεχνολογιών Τεχνητής Νοημοσύνης και Βαθιάς Μηχανικής Μάθησης. Το παρόν θέμα ανήκει στον κλάδο της Ανάκτησης Μουσικής Πληροφορίας (Music Information Retrieval, MIR) και πρόκειται για κάτι καινοτόμο, αφού παρόλο που η ταξινόμηση μουσικής ανά είδος είναι τομέας στον οποίο που έχουν γίνει πολλές έρευνες με πολύ καλά αποτελέσματα, δεν έχει γίνει κάτι αντίστοιχο για Βυζαντινή Μουσική, η οποία έχει αξιοσημείωτους ιδιωματισμούς σε σχέση με την ευρωπαϊκή-δυτική μουσική.

Η εποχή μας χαρακτηρίζεται από την εξάπλωση της τεχνολογίας και των αυτοματισμών σε όλους τους τομείς της καθημερινότητάς σε πολύ μεγάλο βαθμό. Αυτό έχει άμεση επίδραση και στον τομέα της μουσικής, με σημαντικά αποτελέσματα και πολύ καλές προσπάθειες για δημιουργία νέας μουσικής σύνθεσης. Η ταξινόμηση τραγουδιών ανά κατηγορία έχει χρησιμοποιηθεί σε πολλούς τομείς της καθημερινότητας όπως τα συστήματα προτάσεων και μπορεί να χρησιμοποιηθεί σαν μέρος άλλων πιο σύνθετων εργασιών.

Η προσέγγιση του θέματος θα γίνει χρησιμοποιώντας επιβλεπόμενη μάθηση σε ηχογραφήσεις ψαλμωδιών στους οκτώ ήχους της βυζαντινής μουσικής. Οι ηχογραφήσεις θα είναι υψηλής ποιότητας για να είναι όσο το δυνατό πιο ακριβή τα χαρακτηριστικά που θα εξαχθούν. Τα μουσικά αρχεία θα επεξεργαστούν και από αυτά θα εξαχθούν χαρακτηριστικά, τα οποία θα περαστούν σε νευρωνικό δίκτυο για εκπαίδευση αφού έρθουν στην κατάλληλη μορφή. Το δίκτυο θα αποτελείται από 2 ή περισσότερα κρυφά επίπεδα. Θα δοκιμαστούν διάφορες τεχνικές προεπεξεργασίας δεδομένων και βελτιστοποίηση του νευρωνικού δικτύου, για βελτίωση αποτελεσμάτων.

Λέξεις Κλειδιά

Ψηφιακή Επεξεργασία Σήματος, Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Βαθιά Μηχανική Μάθηση, Βυζαντινή Μουσική, Ταξινόμηση Μουσικού Είδους, Ανάκτηση Μουσικής Πληροφορίας.

Abstract

The present dissertation deals with the classification of psalms into the eight sounds of Byzantine Music. Artificial Intelligence and Deep Machine Learning technologies will be used. This topic belongs to the field of Music Information Retrieval (MIR) and it is something innovative, since although the classification of music by genre is an area that a lot of research has been done with very good results, nothing similar has been done for Byzantine Music, which has remarkable idioms in relation to European-Western music.

Our age is characterized by the spread of technology and automation in all areas of everyday life to a great extent. This has a direct impact in the field of music, with significant results and very good efforts to create a new music composition. Classification of songs by category has been used in many areas of everyday life such as recommender systems and can be used as part of other more complex tasks.

The subject will be approached by using supervised learning in recordings of psalms in the eight sounds of Byzantine music. The recordings will be of high quality so that the extracted features will be as accurate as possible. The music files will be processed and features will be extracted from them, which will be passed on to a neural network for training once they are in the appropriate format. The network will consist of 2 or more hidden levels. Various data preprocessing techniques and neural network optimization will be tested to improve results.

Keywords

Digital Signal Processing, Artificial Intelligence, Machine Learning, Deep Learning, Byzantine Music, Music Genre Classification, Music Information Retrieval.

Ευχαριστίες

Ευχαριστώ πρωτίστως τον καθηγητή κύριο Γεώργιο Στάμου που μου έδωσε την ευκαιρία να πραγματοποιήσω τη συγκεκριμένη διπλωματική, προσαρμοσμένη στις δικές μου προτιμήσεις.
Ευχαριστώ επίσης την υποψήφια διδάκτορα κυρία Ναταλία Κωτσάνη, η οποία με καθοδήγησε σε όλη τη διάρκεια της εκπόνησης της παρούσας διπλωματικής εργασίας, παρά τις υποχρεώσεις, τις δυσκολίες και την απόσταση.
Τέλος, θα ήταν μεγάλη παράληψή μου εάν δεν ευχαριστούσα τους γονείς και την οικογένειά μου, που με στήριξαν σε όλα τα επίπεδα καθόλη τη διάρκεια της πενταετούς φοίτησής μου στο Εθνικό Μετσόβιο Πολυτεχνείο, καθώς επίσης και τους συμφοιτητές και φίλους μου Σάββα, Μηνά, Σταύρο, Λένο, Ειρήνη και Μαρία, που ομόρφυναν την καθημερινότητά μου και έκαναν τα τελευταία πέντε χρόνια της ζωής μου αξέχαστα.

Νεόφυτος Παπασάββας

Πίνακας Περιεχομένων

1 Εισαγωγή – Θεωρητικό Υπόβαθρο 1

1.1 Τεχνητή Νοημοσύνη 1

1.1.1 Μηχανική Μάθηση 1

1.1.1.1 Επιβλεπόμενη Μάθηση 2

1.1.1.2 Μη Επιβλεπόμενη Μάθηση 2

1.1.1.3 Ενισχυτική Μάθηση 3

1.1.2 Πλήρως Συνδεδεμένα Νευρωνικά Δίκτυα (Fully Connected Neural Networks) 3

1.1.3 Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks, CNN) 6

1.1.4 Επαναλαμβανόμενα Νευρωνικά Δίκτυα (Recurent Neural Networks, RNN) .. 9

1.1.5 Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (Long Short-Term Memory, LSTM) 10

1.1.6 Μετρικές Αξιολόγησης 10

1.2 Ανάκτηση Μουσικής Πληροφορίας (Music Information Retrieval, MIR) 13

1.3 Βυζαντινή Μουσική 15

1.4 Προηγούμενη Έρευνα 19

1.4.1 Έρευνες στον τομέα της Βυζαντινής Μουσικής 19

1.4.2 Έρευνες στον τομέα της ταξινόμησης του είδους της μουσικής 21

2 Επεξεργασία Δεδομένων και Εξαγωγή Χαρακτηριστικών 23

2.1 Δημιουργία Συνόλου Δεδομένων 23

2.2 Χαρακτηριστικά Ηχητικών Σημάτων 25

2.2.1 Mel-frequency cepstral coefficients (MFCC) 26

2.2.2 Chroma Feature ή Chromagram 27

2.2.3 Φασματικό Κέντρο (Spectral Centroid) 28

2.2.4 Φασματική Διάθεση (Spectral Roll-Off) 28

2.2.5 Ρίζα Μέσης Τετραγωνικής Ενέργειας (Root Mean Square Energy RMSE) ... 28
2.2.6 Φασματικό Εύρος Ζώνης (Spectral Bandwidth) ... 28
2.2.7 Ρυθμός Αλλαγής Πρόσημου (Zero Crossing Rate) ... 28
2.3 Εξαγωγή Χαρακτηριστικών και Προεπεξεργασία Δεδομένων ... 29
2.3.1 Αύξησης Δεδομένων (Data Augmentation) ... 29
2.3.2 Διαίρεση κομματιού ... 30
2.3.3 Κανονικοποίηση δεδομένων (Normalization) ... 30
2.3.4 Παραγέμισμα (Padding) ... 30
2.3.5 Προσθήκη πληροφορίας από το τέλος του κομματιού ... 31
3 Μοντέλα που θα δοκιμαστούν και Εξαγωγή Αποτελεσμάτων ... 32
3.1 Προτεινόμενα Μοντέλα ... 32
3.2 Δοκιμή Μοντέλων Ταξινόμησης Βαθιάς Μάθησης και Βελτιστοποίηση Υπερπαραμέτρων ... 35
3.2.1 Βελτιστοποίηση Batch Size ... 42
3.2.2 Βελτιστοποίηση Αριθμού εποχών εκπαίδευσης μοντέλου ... 44
3.2.3 Βελτιστοποίηση Βελτιστοποιητή και Ρυθμού Μάθησης ... 51
3.2.4 Βελτιστοποίηση Αριθμού Κρυφών Επιπέδων και Αριθμού Νευρώνων κάθε Κρυφού Επιπέδου ... 53
3.2.5 Χρήση τεχνικής Cross Validation ... 65
4 Συμπεράσματα και Μελλοντικές Επεκτάσεις ... 67
4.1 Συμπεράσματα ... 67
4.2 Μελλοντικές Επεκτάσεις και Ανοικτά Πεδία ... 69
Κατάλογος Σχημάτων ... 71
Κατάλογος Πινάκων ... 73
Βιβλιογραφία ... 75

1

Εισαγωγή – Θεωρητικό Υπόβαθρο

1.1 Τεχνητή Νοημοσύνη

Η Τεχνητή Νοημοσύνη αποτελεί τον κλάδο της πληροφορικής που ασχολείται με τη σχεδίαση συστημάτων που πλησιάζουν την ανθρώπινη συμπεριφορά, αφού προσομοιάζουν κάποια στοιχειώδη ευφυΐα-μάθηση. Η ανάπτυξη τέτοιων συστημάτων αποσκοπεί στο να δώσει σε μηχανές τη δυνατότητα της προσαρμοστικότητας στα διάφορα προβλήματα, της εξαγωγής συμπερασμάτων και της επίλυση προβλημάτων. Για την ανάπτυξή αυτού του κλάδου, περιπλέκονται πολλές επιστήμες, όπως αυτή της πληροφορικής, της ψυχολογίας, της φιλοσοφίας, της νευρολογίας, της επιστήμης μηχανών και τον τομέα που θα ασχοληθεί η κάθε εφαρμογή που θα αναπτυχθεί. Ο κλάδος της τεχνητής νοημοσύνης επεκτείνεται και χρησιμοποιείται σε πολλά άλλα πεδία όπως η μηχανική όραση, η επεξεργασία φυσικής γλώσσας η ρομποτική και πολλά άλλα.

1.1.1 Μηχανική Μάθηση

Η Μηχανική Μάθηση ανήκει στον κλάδο της τεχνητής νοημοσύνης. Ο τομέας αυτός ασχολείται με την κατασκευή αλγορίθμων οι οποίοι αποκτούν γνώση από ορισμένα δεδομένα και λαμβάνουν αποφάσεις με βάση αυτές τις γνώσεις. Ο βασικός στόχος ενός μοντέλου μηχανικής μάθησης είναι να γενικεύει την εμπειρία που αποκτά από τα δεδομένα που δέχεται [1] [2], ούτως ώστε να αντεπεξέρχεται σε πρωτόγνωρες εργασίες με μεγάλη ακρίβεια. Η μηχανική μάθηση βρίσκει εφαρμογή στα φίλτρα spam (spam filtering), στην οπτική αναγνώριση χαρακτήρων (Optical Character Recognition, OCR), στις μηχανές αναζήτησης και στην υπολογιστική όραση. Η διαδικασία εκμάθησης που γίνεται κατά τη μηχανική μάθηση, χωρίζεται σε 3 κατηγορίες, ανάλογα με τον τρόπο που γίνεται, την επιβλεπόμενη μάθηση, τη μη επιβλεπόμενη μάθηση και την ενισχυτική μάθηση.

1.1.1.1 Επιβλεπόμενη Μάθηση

Ο όρος Επιβλεπόμενη Μάθηση, αναφέρεται στην κατηγορία της μηχανικής όπου τα δεδομένα εκπαίδευσης (σύνολο δεδομένων) που παρέχονται στο σύστημα για εκπαίδευση και αξιολόγηση βρίσκονται στη μορφή δεδομένο-επιθυμητή έξοδος. Η επιθυμητή έξοδος ονομάζεται και ετικέτα. Το σύνολο δεδομένων χωρίζεται σε δεδομένα εκπαίδευσης και δεδομένα δοκιμής. Τα δεδομένα εκπαίδευσης χρησιμοποιούνται από το νευρωνικό μοντέλο για να εκπαιδευτεί και να διαμορφώσει μια συνάρτηση η οποία με κάποια άγνωστη είσοδο, να μπορεί να αναθέσει την επιθυμητή ετικέτα στην έξοδο. Τα δεδομένα δοκιμής χρησιμοποιούνται μετά το στάδιο της εκπαίδευσης για να αξιολογηθεί η ακρίβεια του μοντέλου. Η χρήση της Επιβλεπόμενης Μάθησης, μπορεί να λύσει δύο ειδών προβλήματα με βάση το είδος των ετικετών, προβλήματα Κατηγοριοποίησης (Classification) και προβλήματα παλινδρόμησης (Recursion). Στα προβλήματα κατηγοριοποίησης οι ετικέτες είναι διακριτές και αντιπροσωπεύουν η κάθε μια κάποια κλάση. Σκοπός του νευρωνικού μοντέλου είναι μετά την εκπαίδευσή του να ταξινομήσει τυχαία είσοδο σε κάποια από τις υπάρχουσες κλάσεις του εκάστοτε προβλήματος. Μερικά παραδείγματα κατηγοριοποίησης που χρησιμοποιούν επιβλεπόμενη μάθηση είναι η αυτόματη ταξινόμηση εικόνων σε κατηγορίες [3] ή η αυτόματη ταξινόμηση τραγουδιών σε διάφορα είδη [4]. Σε αυτή την κατηγορία ανήκει και η τρέχουσα εργασία. Στα προβλήματα παλινδρόμησης οι ετικέτες παίρνουν τιμές που εκφράζουν κάποια ποσότητα. Σε αυτή την κατηγορία προβλημάτων, το νευρωνικό μοντέλο έχει σκοπό να προβλέψει μια τιμή με βάση τα δεδομένα εισόδου. Παραδείγματα αυτών των προβλημάτων έχουμε στις προβλέψεις θερμοκρασίας.

1.1.1.2 Μη Επιβλεπόμενη Μάθηση

Η διαφορά της Μη Επιβλεπόμενης Μάθησης από την Επιβλεπόμενη Μάθηση είναι ότι τα δεδομένα εκπαίδευσης που παρέχονται στο σύστημα για εκπαίδευση και αξιολόγηση δεν συμπεριλαμβάνουν ετικέτα μαζί τους. Ως αποτέλεσμα, σκοπός ενός μοντέλου που εκπαιδεύεται με Μη Επιβλεπόμενη Μάθηση είναι η Συσταδοποίηση (Clustering), ο διαχωρισμός δηλαδή των δεδομένων σε συστάδες-ομάδες βάση κοινών χαρακτηριστικών που αναγνωρίζει. Στη Μη Επιβλεπόμενη Μάθηση, ο αριθμός των συστάδων μπορεί να είναι είτε γνωστός, όταν για παράδειγμα το πρόβλημα που επιλύει είναι ο διαχωρισμός μηνυμάτων αλληλογραφίας σε κακόβουλα και μη, όπου έχουμε 2 συστάδες, είτε άγνωστος, όταν έχουμε να διαχωρίσουμε ένα τυχαίο σύνολο μουσικών κομματιών στο είδος τους. Προφανώς, αφού δεν υπάρχει αυστηρός χαρακτηρισμός-ετικέτα στα δεδομένα, το αποτέλεσμα ενός μοντέλου που θα εκπαιδευτεί με τη χρήση Μη Επιβλεπόμενης Μάθησης δε μπορεί να αξιολογηθεί.

1.1.1.3 Ενισχυτική Μάθηση

Η Ενισχυτική Μάθηση διαφέρει κατά πολύ από τα προηγούμενα δύο είδη μάθησης. Συγκεκριμένα, δεν υπάρχει η έννοια της εκπαίδευσης και της αξιολόγησης, αλλά η έννοια της ανταμοιβής και της τιμωρίας. Στη γενική περίπτωση, έχουμε ένα πράκτορα ο οποίος πραγματοποιεί κάποιες κινήσεις προσπαθώντας να μεγιστοποιήσει μια αθροιστική μεταβλητή-στόχο και ανταμείβεται ή τιμωρείται για τις κινήσεις που επιλέγει. Ο πράκτορας χρησιμοποιεί την ήδη υπάρχουσα γνώση του για να πάρει αποφάσεις και ανάλογα από την τιμωρία ή την ανταμοιβή που θα λάβει προσαρμόζει ανάλογα τις επόμενες αποφάσεις. Τα συστήματα που χρησιμοποιούν Ενισχυτική Μάθηση είναι πάρα πολλά λόγω της γενικής φύσης και του ευρέως φάσματος τους των προβλημάτων που επιλύουν. Μερικά παραδείγματα που χρησιμοποιείται είναι μοντέλα που αναπτύσσονται για να παίζουν παιχνίδια όπως σκάκι, ντάμα [5] και μοντέλα που αναπτύσσονται σε αυτοοδηγούμενα οχήματα.

1.1.2 Πλήρως Συνδεδεμένα Νευρωνικά Δίκτυα (Fully Connected Neural Networks)

Τα Πλήρως Συνδεδεμένα Νευρωνικά Δίκτυα αποτελούν την πιο διαδεδομένη μορφή νευρωνικών δικτύων. Έχουν σαν δομική τους μονάδα το μοντέλο του τεχνητού νευρώνα, ο οποίος απαρτίζεται από τα εξής μέρη:

- Έχει μια σταθερή τιμή στην είσοδό του $x_0 = 1$, η οποία καλείται πόλωση.
- Δέχεται n εισόδους πραγματικών τιμών $x_1, x_2, \ldots, x_n$.
- Έχει n +1 τιμές $w_0, w_1, \ldots, w_n$, οι οποίες αντιστοιχίζονται με τις ανάλογες εισόδους και η τιμή τους ρυθμίζεται κατά την εκπαίδευση του συστήματος. Αυτά τα βάρη πολλαπλασιάζονται με τις τιμές της εκάστοτε εισόδου.
- Οι τιμές της εισόδου πολλαπλασιάζονται με τα βάρη και περνούν από μια συνάρτηση ενεργοποίησης. Αυτή είναι η τελική έξοδος του νευρώνα.

Οι πιο συνηθισμένες συναρτήσεις ενεργοποίησης είναι:

- Σιγμοειδής Συνάρτηση

  Η Σιγμοειδής Συνάρτηση χαρακτηρίζεται από τη σχέση

$$f(x) = \frac{1}{1 + e^{-x}}$$

  Έχει πεδίο τιμών (0, 1).

- Υπερβολική Εφαπτομένη

  Η Υπερβολική Εφαπτομένη χαρακτηρίζεται από τη σχέση

$$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1$$

  Έχει πεδίο τιμών (-1, 1).

- Rectified Linear Unit (ReLU)

  Η Rectified Linear Unit (ReLU) χαρακτηρίζεται από τη σχέση

$$f(x) = \max(0, x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

- Softmax

  Η Softmax χαρακτηρίζεται από τη σχέση

$$f(x) = \frac{e^{x_i}}{\sum_{j=0}^{n} e^{x_j}}$$

Τα Πλήρως Συνδεδεμένα Νευρωνικά Δίκτυα αποτελούνται από επίπεδα τέτοιων νευρώνων, όπου κάθε νευρώνας του κάθε επιπέδου δέχεται σαν είσοδο την έξοδο όλων των νευρώνων του προηγούμενου επιπέδου.

Κάθε τέτοιος νευρώνας μπορεί να μπορεί να προσεγγίσει μόνο γραμμικές συναρτήσεις. Για να προσεγγιστούν μη γραμμικές συναρτήσεις χρησιμοποιούνται τα πολυεπίπεδα νευρωνικά δίκτυα. Τα πολυεπίπεδα νευρωνικά δίκτυα έχουν εκτός από το επίπεδο εισόδου, το οποίο αποτελείται από τις εισόδους του συστήματος, 1 ή περισσότερα επίπεδα νευρώνων. Το τελευταίο επίπεδο νευρώνων ονομάζεται επίπεδο εξόδου. Τα υπόλοιπα επίπεδα νευρώνων ονομάζονται κρυφά επίπεδα. Μπορεί να είναι 1 ή περισσότερα τα κρυφά επίπεδα και αν είναι περισσότερα από 1 τότε αναφερόμαστε την κατηγορία του Βαθιού Νευρωνικού Δικτύου. Λόγω της αρχιτεκτονικής τους, η χρήση των βαθιών νευρωνικών επιπέδων γίνεται σε εφαρμογές βαθιάς μάθησης όπου εξάγονται χαρακτηριστικά και αναλύονται, όπως για παράδειγμα στην κατηγοριοποίηση εικόνων. Κάθε κρυφό επίπεδο, ανάλογα με τη διάταξή του, αναγνωρίζει διαφορετικά χαρακτηριστικά από το επόμενο, πιο γενικά ή πιο σύνθετα. Στο τέλος όλα τα χαρακτηριστικά λαμβάνονται υπόψιν για την τελική πρόβλεψη.

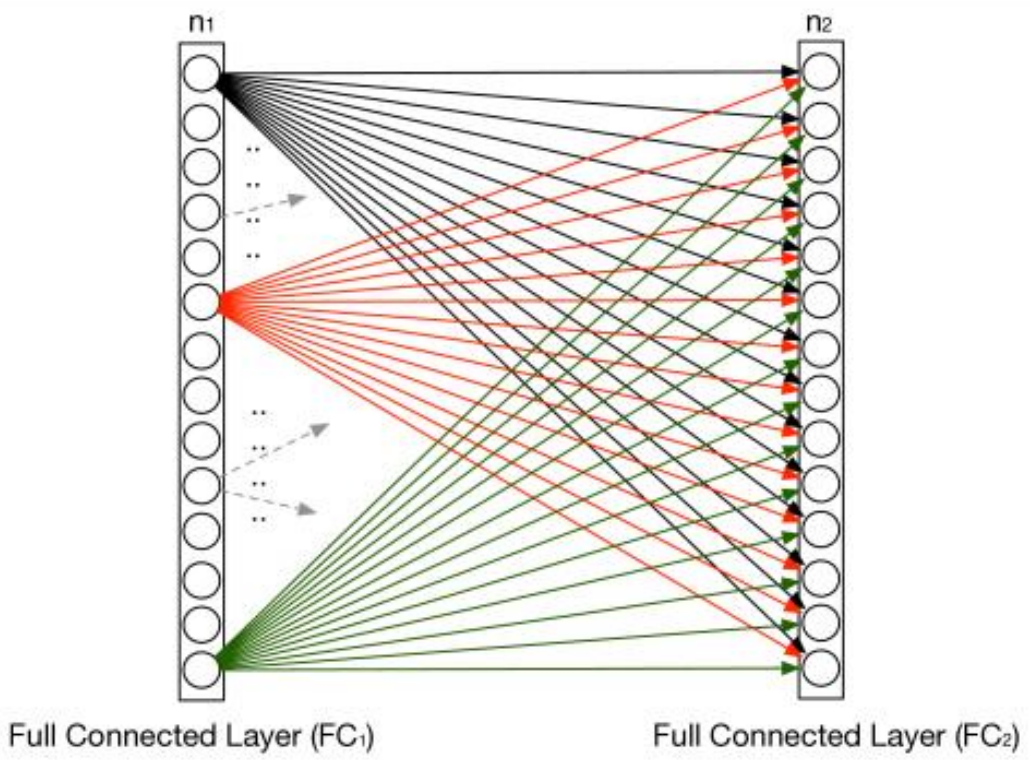
Για να μπορεί να προβλέψει σωστά ένα τέτοιο δίκτυο, πρέπει να περάσει πρώτα από μια διαδικασία εκπαίδευσης, ούτως ώστε να αποκομίσει τα σωστά χαρακτηριστικά από την κάθε διαφορετική κλάση κατηγοριοποίησης και να είναι σε θέση να τις διαχωρίζει με μεγάλη ακρίβεια. Ουσιαστικά στη διαδικασία εκπαίδευσης, το δίκτυο προσαρμόζει τις τιμές των βαρών του κάθε νευρώνα, ούτως ώστε να έχει την επιθυμητή πρόβλεψη στην έξοδο. Η

διαδικασία εκπαίδευσης του δικτύου γίνεται με τη χρήση ενός συνόλου δεδομένων, όπου υπάρχουν παραδείγματα εισόδου μαζί με την κατηγορία στην οποία ανήκουν και πρέπει να ταξινομηθούν. Για την εκπαίδευση ενός νευρωνικού δικτύου, ο πιο γνωστός αλγόριθμος είναι ο αλγόριθμος ανάστροφης μετάδοσης λάθους (back propagation). Αυτός λειτουργεί σε δύο φάσεις, τη φάση του πρόσθιου περάσματος, και τη φάση του ανάστροφου περάσματος. Κατά τη φάση του πρόσθιου περάσματος, η είσοδος περνάει στο δίκτυο και η έξοδος του κάθε επιπέδου περνά σαν είσοδος στο επόμενο επίπεδο, μέχρι το επίπεδο εξόδου, όπου παράγεται το διάνυσμα εξόδου, η πρόβλεψη. Εκεί ξεκινά η φάση του ανάστροφου περάσματος. Συγκρίνονται το διάνυσμα εξόδου/πρόβλεψη του δικτύου με την πραγματική κλάση του τρέχοντος παραδείγματος και υπολογίζεται το τετραγωνικό σφάλμα μεταξύ των δύο διανυσμάτων από τον τύπο

$$E_k = \sum_{i=1}^{m} (t_{kj} - y_{kj})^2$$

ο οποίος λαμβάνει υπόψιν του τις τιμές της εισόδου, της εξόδου και των βαρών του κάθε επιπέδου του δικτύου. Αφού υπολογιστούν οι μερικές παράγωγοι του του παραπάνω σφάλματος ως προς κάθε βάτος του δικτύου, εφαρμόζεται η μέθοδος βελτιστοποίησης επικλινούς καθόδου (gradient descent optimization) και γίνονται προσαρμογές στις τιμές των βαρών.

Ο παραπάνω αλγόριθμος συνεχίζεται μέχρι το τετραγωνικό σφάλμα να φτάσει κάτω από ένα συγκεκριμένο όριο, ή συνήθως μέχρι να περαστεί συγκεκριμένος αριθμός εποχών, επαναλήψεων δηλαδή των δεδομένων εκπαίδευσης.



1.1.2.1 Πλήρως Συνδεδεμένο Επίπεδο

1.1.3 Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks, CNN)

Τα Συνελικτικά Νευρωνικά Δίκτυα ανήκουν στην κλάση των βαθιών νευρωνικών δικτύων και χρησιμοποιούνται κυρίως στην ανάπτυξη εφαρμογών αναγνώρισης εικόνων και βίντεο, επεξεργασίας φυσικής γλώσσας [6] και συστήματα προτάσεων [7]. Η ονομασία τους προέρχεται από τη μαθηματική πράξη της συνέλιξης, την οποία και χρησιμοποιούν κατά βάση.

Τα Συνελικτικά Νευρωνικά Δίκτυα εξειδικεύονται στην ανάλυση εικόνων και βίντεο. Με τη λήψη μιας εικόνας, επικεντρώνονται στα σημεία τα οποία είναι πιο χαρακτηριστικά ανάλογα και με το πρόβλημα το οποίο επιλύουν, με πολύ μεγάλη ανοχή στις παραμορφώσεις των εικόνων. Ο τρόπος με τον οποίο λειτουργούν είναι ο εξής: σε πρώτο στάδιο γίνεται η εξαγωγή χαρακτηριστικών από τον κάθε νευρώνα, σε τοπικό επίπεδο, χρησιμοποιώντας την είσοδο που έχει. Έπειτα, αφού κρατηθεί η πληροφορία για το κάθε χαρακτηριστικό που έχει εξαχθεί, μειώνεται η σημαντικότητα της θέσης της, για να δωθεί προτεραιότητα και σε άλλα χαρακτηριστικά να αναγνωριστούν. Σε επόμενο στάδιο γίνεται η αντιστοίχιση χαρακτηριστικών, η οποία περιλαμβάνει την αποθήκευση των χαρακτηριστικών σε χάρτες χαρακτηριστικών σε κάθε επίπεδο του Συνελικτικού Νευρωνικού Δικτύου. Ο κάθε ένας χάρτης χαρακτηριστικών προέρχεται από τη συνέλιξη της εισόδου του επιπέδου μαζί με ένα φίλτρο. Το επόμενο βήμα είναι με μια διαδικασία ενεργοποίησης. Ακολούθως, γίνεται υποδειγματολειψία μετά από κάθε συνελικτικό επίπεδο, ούτως ώστε να μειωθούν οι διαστάσεις στους χάρτες χαρακτηριστικών. Τελευταίο στάδιο του τρόπου λειτουργίας των Συνελικτικών Νευρωνικών Δικτύων αποτελεί η αντιστοίχηση προβλέψεων, η οποία γίνεται στα τελευταία επίπεδα των Δικτύου τα οποία είναι πλήρως συνδεδεμένα επίπεδα. Ο λόγος είναι για να μετατραπεί η πληροφορία στην επιθυμητή έξοδο και να γίνει η πρόβλεψη.

Το κάθε Συνελικτικό Νευρωνικό Δίκτυο απαρτίζεται από συγκεκριμένα επίπεδα επεξεργασίας, τα οποία αθροιστικά οδηγούν στο αποτέλεσμα. Κάθε τέτοιο επίπεδο αποτελείται από συγκεκριμένους νευρώνες. Παρακάτω επεξηγούνται ορισμένα επίπεδα επεξεργασίας.

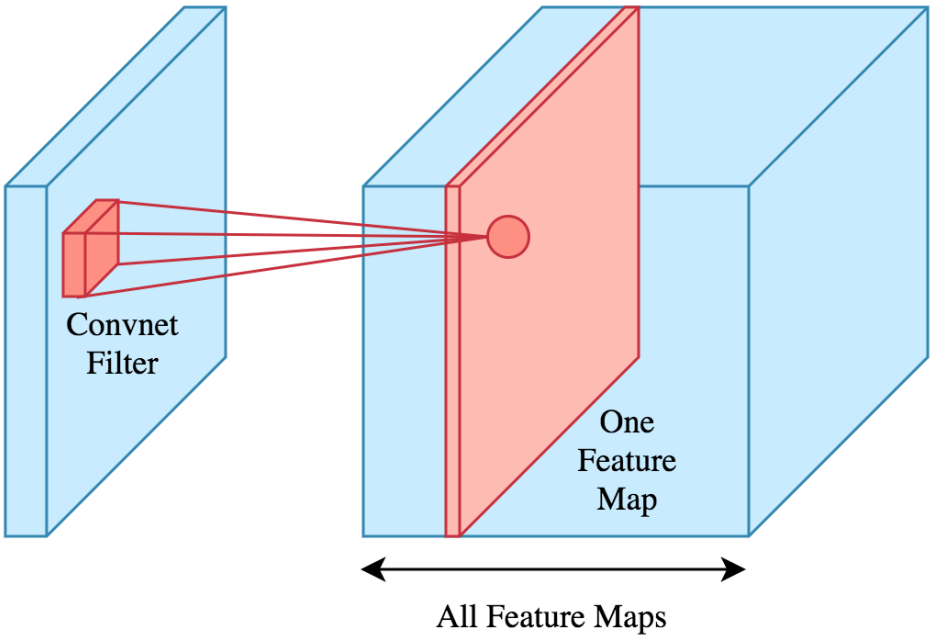
Επίπεδο Εισόδου

Το επίπεδο εισόδου είναι το πρώτο επίπεδο επεξεργασίας που συναντάται σε μια διάταξη Συνελικτικού Νευρωνικού Δικτύου. Είναι υπεύθυνο για την τροφοδότηση των δεδομένων εισόδου στο δίκτυο. Οι διαστάσεις του καθορίζονται από τα δεδομένα εισόδου.

Συνελικτικό Επίπεδο

Το συνελικτικό επίπεδο αποτελεί την ουσία του δικτύου. Σε αυτό το επίπεδο γίνεται η επεξεργασία των χαρακτηριστικών και η εξαγωγή των χαρτών των χαρακτηριστικών. Πιο συγκεκριμένα, σε αυτό το επίπεδο γίνεται η πράξη της συνέλιξης μεταξύ των δεδομένων εισόδου και ενός φίλτρου. Τα στοιχεία εισόδου θα πολλαπλασιαστούν με τα στοιχεία του φίλτρου. Επειδή το φίλτρο έχει πολύ μικρότερες διαστάσεις από τα δεδομένα εισόδου,

χρησιμοποιείται η τεχνική του παραθύρου που ολισθαίνει με τέτοιο τρόπο ούτως ώστε να περάσει από όλες τις τιμές. Αποτέλεσμα της πράξης της συνέλιξης των δύο διανυσμάτων είναι ο χάρτης χαρακτηριστικών ο οποίος αποθηκεύεται, αφού πρώτα περάσει από μια συνάρτηση ενεργοποίησης. Ανάλογα με τον αριθμό και τις διαστάσεις των φίλτρων που επιλέγονται, δημιουργούνται αντίστοιχοι πίνακες χαρακτηριστικών, οι οποίοι στοιβάζονται και δημιουργούν στην έξοδο αυτού του επιπέδου ένα διάνυσμα 1 διάστασης μεγαλύτερης από τις διαστάσεις των δεδομένων εισόδου, με ύψος όσος είναι ο αριθμός των φίλτρων. Το μέγεθος των πινάκων χαρακτηριστικών καθορίζεται από το μέγεθος του φίλτρου, τον αριθμό τους, το βήμα με το οποίο θα ολισθαίνει το φίλτρο πάνω στα δεδομένα εισόδου και το παραγέμισμα του περιθωρίου. Το μέγεθος του φίλτρου καθορίζει τον όγκο της πληροφορίας που περιέχει το κάθε στοιχείο της εξόδου, το βήμα ολίσθησης καθορίζει τις διαστάσεις του κάθε ενός πίνακα χαρακτηριστικών και το παραγέμισμα χρησιμοποιείται σε περίπτωση που θέλουμε να κρατήσουμε σταθερές τις διαστάσεις εισόδου-εξόδου και να μην έχουμε μείωση τους.

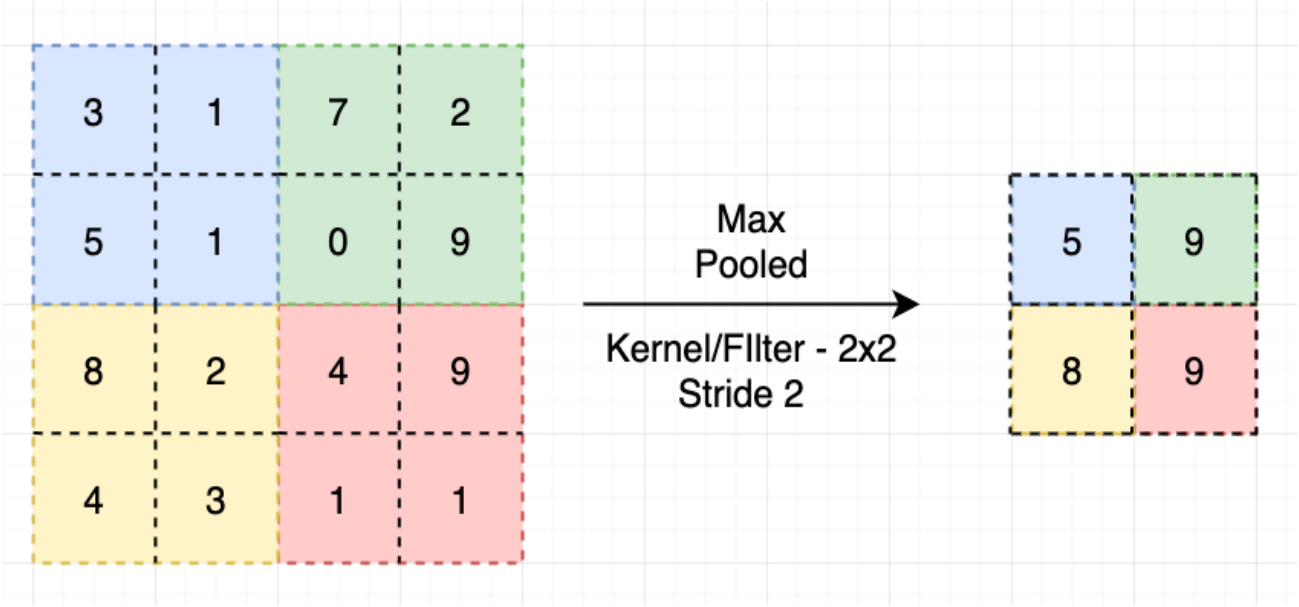


1.1.3.1 Πίνακας Χαρακτηριστικών

Συγκεντρωτικό Επίπεδο

Μετά το συνελικτικό επίπεδο, συχνά υπάρχει ένα συγκεντρωτικό επίπεδο, το οποίο πραγματοποιεί μια υποδειγματοληψία στους πίνακες χαρακτηριστικών που εξάγει το συνελικτικό επίπεδο. Σκοπός της δειγματοληψίας αυτής είναι η μείωση των διαστάσεων των δεδομένων μειώνοντας μόνο το μήκος και το πλάτος τους και όχι τον αριθμό των

φίλτρων. Η διαδικασία αυτή επιτυγχάνει την πιο γρήγορη εκπαίδευση του συστήματος, καθώς επίσης αντιμετωπίζει το πρόβλημα της υπερπροσαρμογής. Η υποδειγματοληψία γίνεται με την τεχνική του παραθύρου που ολισθαίνει πάνω από όλες τις τιμές εισόδου του επιπέδου και σε κάθε βήμα κρατά μόνο μια τιμή από τα στοιχεία που περνά. Ο τρόπος επιλογής της τιμής καθορίζεται από το χρήστη. Πιθανοί τρόποι είναι η επιλογή της μέγιστης τιμής των στοιχείων του παραθύρου, επιλογή του μέσου όρου των στοιχείων του παραθύρου και το άθροισμα των στοιχείων του παραθύρου. Το μέγεθος παραθύρου όπως επίσης και το βήμα ολίσθησης ρυθμίζονται από το χρήστη.



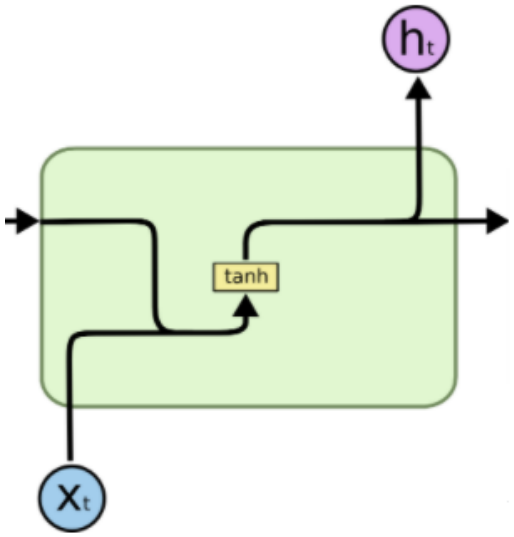
1.1.3.2 Συγκεντρωτικό Επίπεδο

Πλήρως Συνδεδεμένο Επίπεδο

Έπειτα από μια σειρά συνελικτικών και συγκεντρωτικών επιπέδων, το δίκτυο ολοκληρώνεται με ορισμένες ομάδες πλήρως συνδεδεμένων επιπέδων. Ένα πλήρως συνδεδεμένο επίπεδο αποτελείται από ένα αριθμό νευρώνων που επιλέγεται από το χρήστη, όπου κάθε νευρώνας είναι συνδεδεμένος με ολους τους νευρώνες του προηγούμενου επιπέδου. Η είσοδος ενός τέτοιου επιπέδου πρέπει να είναι διάνυσμα μιας διάστασης. Επειδή σε μια διάταξη Συνελικτικού Νευρωνικού Δικτύου η είσοδος του πλήρως συνδεδεμένου επιπέδου θα είναι πολυδιάστατο διάνυσμα, πριν το πρώτο πλήρως συνδεδεμένο επίπεδο τα δεδομένα ισοπεδώνονται και έρχονται σε μορφή διανύσματος μιας διάστασης, ούτως ώστε να μην χαθεί πληροφορία. Τα πλήρως συνδεδεμένα επίπεδα σκοπό έχουν την πρόβλεψη της επιθυμητής κλάσης της εισόδου με μεγάλη ακρίβεια

1.1.4 Επαναλαμβανόμενα Νευρωνικά Δίκτυα (Recurent Neural Networks, RNN)

Η αδυναμία των προαναφερθέντων Νευρωνικών Δικτύων να αξιοποιήσουν δεδομένα τα οποία είναι ακολουθιακά, όπως για παράδειγμα τις λέξεις ενός κειμένου, οδήγησε τους ερευνητές του κλάδου της Τεχνητής Νοημοσύνης στη δημιουργία των Επαναλαμβανόμενων Νευρωνικών Δικτύων. Τα δίκτυα αυτά, μπορούν να μιμηθούν την ανθρώπινη μνήμη, αφού θυμούνται εισόδους που χρησιμοποιήθηκαν προηγουμένως για την παραγωγή εξόδων. Έτσι, κάθε έξοδος του δικτύου δεν εξαρτάται μόνο από δεδομένα της εισόδου εκείνη την χρονική στιγμή και τους παραμέτρους του δικτύου αλλά, και από μια κρυφή κατάσταση (hidden state) που χρησιμοποιεί το δίκτυο για να 'θυμάται' τις προηγούμενες εισόδους που είχε. Επομένως, ένα RNN δεν έχει σταθερή έξοδο για δύο περιπτώσεις με την ίδια είσοδο, αφού τα παρελθοντικά δεδομένα εισόδου μπορεί να διαφέρουν σε κάθε περίπτωση.



1.1.4.1 RNN νευρώνας

Ο υπολογισμός της επόμενης κρυφής κατάστασης εξαρτάται από την είσοδο που περνάει στο σύστημα, καθώς επίσης και από την προηγούμενη κρυφή κατάσταση. Ο υπολογισμός της κρυφής κατάστασης γίνεται με την εξίσωση

$$h_t = \sigma_h(W_h x_t + U_h h_{t-1} + b_h)$$

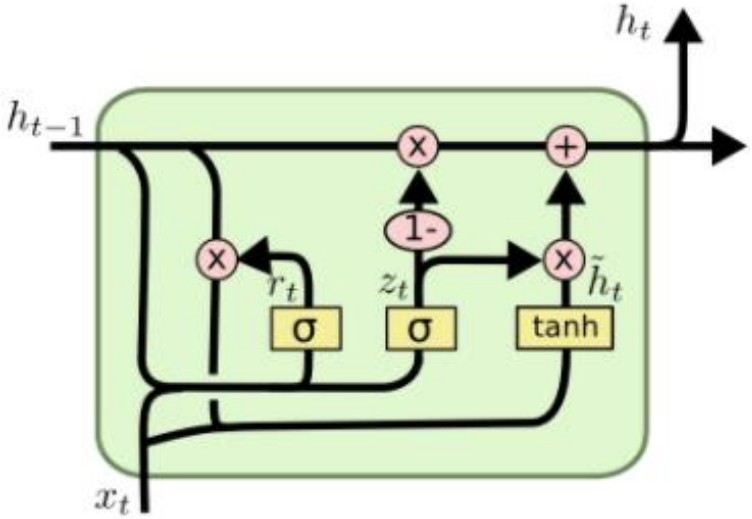
Ο υπολογισμός της εξόδου γίνεται με την εξίσωση

$$y_t = \sigma_y (W_y h_t + b_y)$$

Οι μεταβλητές W_h, U_h, W_y, b_h, b_y είναι διανύσματα παραμέτρων και οι διαστάσεις τους εξαρτώνται από το input size και το hidden size. Τέλος, τα σ_h, σ_y είναι συναρτήσεις ενεργοποίησης.

1.1.5 Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (Long Short-Term Memory, LSTM)

Τα Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (Long Short-Term Memory, LSTM) είναι μια τροποποιημένη εκδοχή των RNN, που μπορούν να διατηρούν πληροφορία για μεγάλα χρονικά διαστήματα. Η βασική διαφορά που έχουν τα LSTM σε σχέση με τα RNN είναι ότι εκτός από την κρυφή κατάσταση (hidden state) έχουν και την κατάσταση κυττάρου (cell state). Επιπλέον, τα κύτταρα LSTM έχουν μια επιπλέον πύλη που ονομάζεται πύλη άγνοιας (forget gate). Παρακάτω παρουσιάζεται η εσωτερική δομή ενός νευρώνα ενός δικτύου LSTM.



1.1.5.1 LSTM νευρώνας

1.1.6 Μετρικές Αξιολόγησης

Για την αξιολόγηση της επίδοσης και της αξιοπιστίας ενός συστήματος που έχει αναπτυχθεί με τεχνολογίες Τεχνητής Νοημοσύνης χρειάζονται κάποιες μετρικές. Οι μετρικές αυτές θα πρέπει να είναι καθολικές ανάλογα με τη φύση του κάθε προβλήματος, ούτως ώστε να μπορεί να γίνεται σύγκριση μεταξύ διαφόρων συστημάτων που επιλύουν το ίδιο πρόβλημα, ή απλά να φαίνεται η αξιοπιστία του συστήματος και εάν αυτό μπορεί να χρησιμοποιηθεί για τον εκάστοτε σκοπό. Οι προβλέψεις που κάνει το κάθε σύστημα χαρακτηρίζονται με τον εξής τρόπο: **True Positive (TP)** ονομάζεται η πρόβλεψη που έχει γίνει από το σύστημα αν είναι σωστή και η προβλεπόμενη κλάση είναι θετική. **True Negative (TN)** ονομάζεται η πρόβλεψη που έχει γίνει από το σύστημα αν είναι σωστή και η προβλεπόμενη κλάση είναι αρνητική. **False Positive (FP)** ονομάζεται η πρόβλεψη που έχει γίνει από το σύστημα αν είναι λανθασμένη και η προβλεπόμενη κλάση είναι θετική. **False Negative (FN)** ονομάζεται η πρόβλεψη που έχει γίνει από το σύστημα αν είναι λανθασμένη και η προβλεπόμενη κλάση είναι αρνητική.

Με βάση τους παραπάνω χαρκτηρισμούς των προβλέψεων που κάνει το κάθε σύστημα, χρησιμοποιούνται οι παρακάτω μετρικές αξιολόγησης των συστημάτων, οι οποίες είναι μερικές από τις πιο σημαντικές:

- **Accuracy**

Η μετρική αξιολόγησης Accuracy ή ακρίβεια αποτελεί τον λόγο των σωστά ταξινομημένων δειγμάτων προς το σύνολο όλων των δειγμάτων.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Η μετρική Accuracy είναι από τις πιο σημαντικές μετρικές και χρησιμοποιείται πολύ συχνά, αφού πρόκειται για την πιο άμεση αξιολόγηση ενός μοντέλου.

- **Precision**

Η μετρική αξιολόγησης Precision αποτελεί τον λόγο των σωστά ταξινομημένων θετικών προβλέψεων προς το σύνολο των προβλέψεων που έχουν ταξινομηθεί ως θετικές.

$$Precision = \frac{TP}{TP + FP}$$

Η μετρική Precision χρησιμοποιείται συχνά όταν η εγκυρότητα της πρόβλεψης που θα κάνει το σύστημα είναι μεγάλης σημασίας.

- **Recall**

Η μετρική αξιολόγησης Recall αποτελεί το λόγο των σωστά ταξινομημένων θετικών προβλέψεων προς το σύνολο των θετικών προβλέψεων.

$$Recall = \frac{TP}{TP + FN}$$

Η μετρική Recall χρησιμοποιείται όταν σκοπός του προβλήματος είναι η μεγιστοποίηση των θετικών προβλέψεων.

- **F1 Score**

Η μετρική αξιολόγησης F1 Score αποτελεί τον αρμονικό μέσο όρο των μετρικών Precision και Recall.

$$F1\ Score = 2\ \times \frac{Precision\ \times Recall}{Precision + Recall}$$

Η μετρική αξιολόγησης F1 Score χρησιμοποιείται όταν στο πρόβλημα που επιλύεται απαιτείται καλή επίδοση Precision και Recall. Χρησιμοποιείται όταν το σύνολο δεδομένων δεν έχει ισοκατανεμημένα δεδομένα εισόδου, δηλαδό όταν υπάρχουν πολλά δείγματα κάποιας κλάσης, ενώ πολύ λίγα στις υπόλοιπες κλάσεις.

1.2 Ανάκτηση Μουσικής Πληροφορίας (Music Information Retrieval, MIR)

Τα τελευταία χρόνια, η χρήση της τεχνητής νοημοσύνης στον τομέα της μουσικής έχει τραβήξει την προσοχή σε πολλούς ερευνητές. Πολλές είναι οι έρευνες που έχουν γίνει, και μάλιστα με πάρα πολύ καλά αποτελέσματα. Ο κλάδος της ανάκτησης μουσικής πληροφορίας έχει υπόβαθρο εκτός από την τεχνητή νοημοσύνη, στη μουσικολογία, τη ψυχολογία, την ανάλυση σήματος Οι τομείς που χρησιμοποιείται η ανάκτηση μουσικής πληροφορίας είναι τα συστήματα προτάσεων, διαχωρισμός κομματιών και αναγνώριση οργάνων, αυτόματη μεταγραφή μουσικής, κατηγοριοποίηση μουσικής και τέλος δημιουργία μουσικής.

Συστήματα προτάσεων (Recommender systems) [8] [9]

Συστήματα προτάσεων για μουσική ήδη υπάρχουν, τα οποία χρησιμοποιούν ετικέτες για να προσδιορίσουν ομοιότητα μεταξύ τραγουδιών. Πρόσφατα όμως ξεκίνησαν προσπάθειες που χρησιμοποιούν τεχνικές ανάκτησης μουσικής πληροφορίας για εύρεση ομοιότητας μεταξύ τραγουδιών χρησιμοποιώντας στοιχεία των τραγουδιών που εξάγονται από αυτά και όχι ετικέτες που τοποθετήθηκαν από ανθρώπους. Τα συστήματα προτάσεων χρησιμοποιούνται από διάφορες υπηρεσίες και εφαρμογές οι οποίες παρέχουν στο χρήστη τραγούδια και του προτείνουν άλλα τραγούδια με βάση αυτά που έχει ήδη ψάξει.

Διαχωρισμός πηγών και αναγνώριση οργάνων (Source separation and instrument recognition) [10] [11][12][13]

Σε αυτό τον τομέα αναγνωρίζονται τα διάφορα όργανα που απαρτίζουν ένα τραγούδι. Επίσης διαχωρίζεται το τραγούδι στα επιμέρους όργανα που απαρτίζεται και εξάγεται η μελωδία που το καθένα από αυτά ερμηνεύει. Ο τομέας αυτός αποτελεί πρόκληση αυτή την εποχή και πολλές έρευνες ασχολούνται με αυτό, προκειμένου να πετύχουν υψηλά ποσοστά ακρίβειας.

Αυτόματη μεταγραφή μουσικής (Automatic music transcription) [14] [15]

Με τον όρο μεταγραφή μουσικής εννοούμε τη μετατροπή ενός ηχητικού αρχείου σε συμβολική σημειογραφία. Για να επιτευχθεί αυτό, χρειάζεται επεξεργασία του ηχητικού αρχείου σε πολλούς τομείς και εξαγωγή χαρακτηριστικών του ως προς την τονικότητα, το ρυθμό, και εξαγωγή αρμονικών, γεγονός που καθιστά τον τομέα αυτόν ιδιαίτερα απαιτητικο, με τη δυσκολία να ανεβαίνει αισθητά όσο αυξάνεται ο αριθμός των οργάνων.

Αυτόματη κατηγοριοποίηση μουσικής (Automatic categorization) [16] [17] [18]

Μια κοινή εργασία για τον κλάδο της ανάκτησης μουσικής πληροφορίας είναι η κατηγοριοποίηση της μουσικής στο είδος της. Στη μηχανική μάθηση χρησιμοποιούνται Μηχανές Διανυσμάτων Υποστήριξης (SVM) για την υλοποίηση αυτής της εργασίας, έχοντας σημειώσει μάλιστα αρκετά μεγάλη ακρίβεια στα αποτελέσματα. Η αυτόματη κατηγοριοποίηση μουσικής χρησιμοποιείται σε συστήματα προτάσεων για να προτείνει τραγούδια που ανήκουν στην ίδια κατηγορία και κατ' επέκταση είναι όμοια μεταξύ τους, καθώς επίσης και στην αυτόματη δημιουργία λιστών αναπαραγωγής τραγουδιών τα οποία να έχουν συνοχή.

Δημιουργία/Σύνθεση μουσικής (Music Generation Composition) [19]

Ο τομέας αυτός είναι σχετικά πρώιμος και δύσκολος. Αν και ήδη έχουνε γίνει πολλές προσπάθειες αξιόλογες, τα αποτελέσματα δεν είναι επαρκώς καλά.

Η εργασία αυτή αποτελεί μια έρευνα σε αυτό τον τομέα της τεχνητής νοημοσύνης και πιο συγκεκριμένα στην κατηγοριοποίηση μουσικής, αφού θέμα της είναι η κατηγοριοποίηση ψαλμωδιών σε 8 κατηγορίες-ήχους που θα εξηγηθούν παρακάτω. Θα ληφθούν υπόψη οι υπερσύγχρονες μέθοδοι που χρησιμοποιούνται για αυτή την εργασία και θα δοκιμαστούν επίσης καινούριες με σκοπό την εύρεση του καλύτερου δυνατού αποτελέσματος. Κατά την έρευνα που έγινε προτού ξεκινήσει η υλοποίηση του ταξινομητή, δε βρέθηκαν άλλες παρεμφερείς έρευνες ταξινόμησης ήχων στον τομέα της Βυζαντινής Μουσικής.

1.3 Βυζαντινή Μουσική

Η βυζαντινή μουσική (ψαλτική τέχνη) αποτελεί τη μουσική έκφραση της Ορθόδοξης υμνολογίας. Ξεκίνησε να αναπτύσσεται ως ξεχωριστό μουσικό είδος ήδη από τον 3^{ο} αιώνα μ.Χ. και χρησιμοποίησε την αρχαιοελληνική θεωρία περί μουσικής, δηλαδή ενσωμάτωσε τους τρόπους των αρχαίων Ελλήνων μουσικών και το θεωρητικό τους υπόβαθρο. [20] Η γλώσσα που χρησιμοποιήθηκε στην υμνογραφία είναι η ελληνική. Χρησιμοποιείται κυρίως σαν λατρευτική μουσική. Πρόκειται για μονοφωνική μουσική με συνοδεία ισοκρατήματος.

Στη βυζαντινή μουσική χρησιμοποιούνται διαφορετικοί φθόγγοι και τόνοι από ότι στην ευρωπαϊκή μουσική. Συγκεκριμένα, χρησιμοποιούνται οι φθόγγοι πΑ, Βου, Γα, Δι, κΕ, Ζω και νΗ, με τα κεφαλαία γράμματα να είναι τα πρώτα 7 γράμματα του ελληνικού αλφαβήτου. Ενώ στην ευρωπαϊκή μουσική το σύστημα είναι συγκερασμένο, δηλαδή μόνο μείζονες τόνοι και ημιτόνια, στη βυζαντινή μουσική υπάρχουν μείζονες, ελάσσονες και ελάχιστοι τόνοι. [21]

Υπάρχουν 3 είδη μελοποιίας, το Ειρμολογικό, το Στιχηραρικὸ και το Παπαδικό. Με τον όρο ‘είδος μελοποιίας’ αναφερόμαστε στην ανάπτυξη της μουσικής φράσης σε σχέση με τις συλλαβές του κειμένου, δηλαδή το πόσους χρόνους θα διαρκεί η κάθε συλλαβή. Οι τρείς όροι που προαναφέρθηκαν αναφέρονται σε σύντομα, αργά και πολύ αργά μέλη. Αυτό αφορά την μουσική ανάπτυξη της φράσης και όχι τη χρονική αγωγή.

Τα είδη Ειρμολογικό και Στιχηραρικό χωρίζονται σε Σύντομο Ειρμολογικό, Αργό Ειρμολογικό, Σύντομο Στιχηραρικό και Αργό Στιχηραρικό, αντίστοιχα. Στο Σύντομο Ειρμολογικό είδος υπάρχει η μελοποιία στην οποία κατά βάση ένας φθόγγος αναλογεί σε μια συλλαβή. Στο Αργό Ειρμολογικό είδος υπάρχει η μελοποιία στην οποία κατά βάση δύο φθόγγοι αναλογούν σε μια συλλαβή. Στο Σύντομο Στιχηραρικό είδος έχουμε τη μελοποιία στην οποία κατά βάση δύο φθόγγοι αναλογούν σε μια συλλαβή. Στο Αργό Στιχηραρικό είδος έχουμε τη μελοποιία στην οποία κατά βάση τέσσερις φθόγγοι αναλογούν σε μια συλλαβή. Τέλος, στο Παπαδικό είδος, έχουμε τη μελοποιία στην οποία κατά βάση οκτώ φθόγγοι και πάνω αναλογούν σε μια συλλαβή. [22]

Υπάρχουν 8 ήχοι, ο Πρώτος ήχος, ο Δεύτερος, ο Τρίτος, ο Τέταρτος, ο Πλάγιος του Πρώτου, ο Πλάγιος του Δευτέρου, ο Βαρύς και ο Πλάγιος του Τετάρτου. Ο κάθε ήχος έχει τα δικά του γνωρίσματα, τα οποία είναι το γένος στο οποίο ανήκει, το απήχημα, η κλίμακα που χρησιμοποιεί, οι δεσπόζοντες φθόγγοι και οι καταλήξεις. Σε κάθε ήχο, κάποια από αυτά τα χαρακτηριστικά διαφοροποιούνται για μέλη που βρίσκονται σε διαφορετικό είδος μελοποιίας. [23]

Το γένος είναι ομάδα ήχων που χρησιμοποιεί ίδια ή συγγενή τετράχορδα. Υπάρχουν τρία γένη, το Διατονικό, το Χρωματικό και το Εναρμόνιο.

Στο Διατονικό γένος διακρίνονται τα εξής διαστήματα:

Νη-Πα: 12 μόρια
Πα-Βου: 10 μόρια
Βου-Γα: 8 μόρια
Γα-Δι: 12 μόρια
Δι-Κε: 12 μόρια
Κε-Ζω: 10 μόρια
Ζω-Νη΄: 8 μόρια

τα οποία επαναλαμβάνονται σε όλες τις οκτάβες. [24]

Στο Χρωματικό γένος διακρίνουμε 2 κλίμακες, τη Μαλακή Χρωματική και τη Σκληρή Χρωματική. Στη Μαλακή Χρωματική κλίμακα υπάρχουν τα εξής διαστήματα:

Δι-Κε: 8 μόρια
Κε-Ζω: 14 μόρια
Ζω-Νη΄: 8 μόρια
Νη'-Πα΄: 12 μόρια

τα οποία επαναλαμβάνονται ανά 5 φθόγγους. [24]

Στη Σκληρή Χρωματική κλίμακα υπάρχουν τα εξής διαστήματα:

Πα-Βου: 6 μόρια
Βου-Γα: 20 μόρια
Γα-Δι: 4 μόρια
Δι-Κε: 12 μόρια

τα οποία επαναλαμβάνονται ανά 5 φθόγγους. [24]

Στο Εναρμόνιο γένος υπάρχουν τα εξής διαστήματα:

Γα-Δι: 12 μόρια
Δι-Κε: 12 μόρια
Κε-Ζω: 6 μόρια

τα οποία επαναλαμβάνονται ανά 4 φθόγγους. [24]

Στην εργασία αυτή θα επικεντρωθούμε στην ταξινόμηση ύμνων Σύντομου Ειρμολογικού και Σύντομου Στιχηραρικού είδους μελοποιίας, αφού το σύνολο δεδομένων απαρτίζεται από ύμνους αυτού του είδους. Για αυτό το λόγο θα αναφερθούν παρακάτω τα βασικά χαρακτηριστικά των οκτώ ήχων για αυτά τα είδη μελοποιίας.

Πρώτος ήχος Σύντομος Ειρμολογικός: Ανήκει στο Διατονικό γένος και χρησιμοποιεί τη διατονική κλίμακα. Έχει ως βάση το Πα και δεσπόζοντες φθόγγοι είναι οι Πα και Δι.

Πρώτος ήχος Σύντομος Στιχηραρικός: Ανήκει στο Διατονικό γένος και χρησιμοποιεί τη διατονική κλίμακα. Έχει ως βάση το Πα και δεσπόζοντες φθόγγοι είναι οι Πα και Γα.

Δεύτερος ήχος Σύντομος Ειρμολογικός: Ανήκει στο Χρωματικό γένος και χρησιμοποιεί τη σκληρή χρωματική κλίμακα. Έχει ως βάση το Πα και δεσπόζοντες φθόγγοι είναι οι Πα και Δι.

Δεύτερος ήχος Σύντομος Στιχηραρικός: Ανήκει στο Χρωματικό γένος και χρησιμοποιεί τη μαλακή χρωματική κλίμακα. Έχει ως βάση το Δι και δεσπόζοντες φθόγγοι είναι οι Βου και Δι.

Τρίτος ήχος Σύντομος Ειρμολογικός: Ανήκει στο Εναρμόνιο γένος και χρησιμοποιεί τη εναρμόνιο κλίμακα. Έχει ως βάση το Γα και δεσπόζοντες φθόγγοι είναι οι Πα, Γα και Κε.

Τρίτος ήχος Σύντομος Στιχηραρικός: Ανήκει στο Εναρμόνιο γένος και χρησιμοποιεί τη εναρμ όνιο κλίμακα. Έχει ως βάση το Γα και δεσπόζοντες φθόγγοι είναι οι Πα, Γα, και Κε.

Τέταρτος ήχος Σύντομος Ειρμολογικός (ή αλλιώς 'Λέγετος'): Ανήκει στο διατονικό γένος και χρησιμοποιεί τη διατονική κλίμακα. Έχει ως βάση το Βου και δεσπόζοντες φθόγγοι είναι οι Πα, Βου, και Δι.

Τέταρτος ήχος Σύντομος Στιχηραρικός: Ανήκει στο διατονικό γένος και χρησιμοποιεί τη διατονική κλίμακα. Έχει ως βάση το Δι και δεσπόζοντες φθόγγοι είναι οι Πα, και Δι.

Ήχος Πλάγιος του Πρώτου Σύντομος Ειρμολογικός: Ανήκει στο διατονικό γένος και χρησιμοποιεί τη διατονική κλίμακα. Έχει ως βάση το Κε και δεσπόζοντες φθόγγοι είναι οι Κε και Νη΄.

Ήχος Πλάγιος του Πρώτου Σύντομος Στιχηραρικός: Ανήκει στο διατονικό γένος και χρησιμοποιεί τη διατονική κλίμακα. Έχει ως βάση το Πα και δεσπόζοντες φθόγγοι είναι οι Πα, Δι και Κε.

Ήχος Πλάγιος του Δευτέρου Σύντομος Ειρμολογικός: Ανήκει στο Χρωματικό γένος και χρησιμοποιεί τη μαλακή χρωματική κλίμακα. Έχει ως βάση το Δι και δεσπόζοντες φθόγγοι είναι οι Βου και Δι.

Ήχος Πλάγιος του Δευτέρου Σύντομος Στιχηραρικός: Ανήκει στο Χρωματικό γένος και χρησιμοποιεί τη σκληρή Χρωματική κλίμακα. Έχει ως βάση το Πα και δεσπόζοντες φθόγγοι είναι οι Πα και Δι.

Ήχος Βαρύς Σύντομος Ειρμολογικός: Ανήκει στο Εναρμόνιο γένος και χρησιμοποιεί την εναρμόνιο κλίμακα. Έχει ως βάση το Γα και δεσπόζοντες φθόγγοι είναι οι Πα, Γα και Δι.

Ήχος Βαρύς Σύντομος Στιχηραρικός: Ανήκει στο διατονικό γένος και χρησιμοποιεί τη διατονική κλίμακα. Έχει ως βάση το Γα και δεσπόζοντες φθόγγοι είναι οι Πα, Γα και Δι.

Ήχος Πλάγιος του Τετάρτου Σύντομος Ειρμολογικός: Ανήκει στο διατονικό γένος και χρησιμοποιεί τη διατονική κλίμακα. Έχει ως βάση το Νη και δεσπόζοντες φθόγγοι είναι οι Νη, Βου και Δι.

Ήχος Πλάγιος του Τετάρτου Σύντομος Στιχηραρικός: Ανήκει στο διατονικό γένος και χρησιμοποιεί τη διατονική κλίμακα. Έχει ως βάση το Νη και δεσπόζοντες φθόγγοι είναι οι Νη, Βου και Δι. [25]

1.4 Προηγούμενη Έρευνα

Τα τελευταία χρόνια, η ταξινόμηση των τραγουδιών με βάση το είδος τους έχει αναπτυχθεί αρκετά και πολλές έρευνες έχουν γίνει σε αυτό. Στην εργασία αυτή έχουμε χρησιμοποιήσει ιδέες από ορισμένους ταξινομητές [16][18], καθώς επίσης και παράμετροι που χρησιμοποιούνται γενικά για την κατηγοριοποίηση των ήχων στη Βυζαντινή Μουσική. Επίσης η Βυζαντινή Μουσική και η Τούρκικη Μουσική έχουν πολλά κοινά, τόσο στους ήχους, οι οποίοι ονομάζονται μακάμια (makams), όσο και στα μικροδιαστήματα που χρησιμοποιούνται. Μερικές έρευνες που έχουν γίνει στους τομείς της Βυζαντινής Μουσικής και της ταξινόμησης είδους της μουσικής παρατίθενται παρακάτω.

1.4.1 Έρευνες στον τομέα της Βυζαντινής Μουσικής

'An optical music recognition system for the notation of the Orthodox Hellenic Byzantine Music' [26]
Στο παραπάνω άρθρο οι συγγραφείς αναγνωρίζουν τα σύμβολα της βυζαντινής μουσικής, αφού πρώτα περιγράφουν τη δομή του συστήματος και προτείνουν αλγόριθμους αναγνώρισης των συμβόλων. Χρησιμοποιώντας αλγόριθμο KNN (K Nearest Neighbours), έχει επιτευχθεί ακρίβεια ταξινόμησης 99.3% .

'Optical character recognition of the Orthodox Hellenic Byzantine Music notation'[27]
Στο επιστημονικό άρθρο αυτό, οι συγγραφείς ασχολήθηκαν με την αναγνώριση των συμβόλων της βυζαντινής μουσικής, έχοντας είσοδο μια εικόνα. Έχουν χρησιμοποιήσει τον ταξινομητή KNN (K Nearest Neighbours), μαζί με κάποια επεξεργασία δεδομένων πριν και μετά την είσοδο τους στον ταξινομητή. Η ακρίβεια που έχει επιτευχθεί στην έρευνα αυτή ήταν 99.4% για εκτυπωμένο μουσικό κείμενο, ενώ για χειρόγραφο μουσικό κείμενο ήταν 73.2%.

'Optical recognition of psaltic Byzantine chant notation' [28]
Σαν συνέχεια του παραπάνω επιστημονικού άρθρου, έχει γραφτεί αυτό, στο οποίο οι συγγραφείς ασχολήθηκαν με την αναγνώριση των συμβόλων της βυζαντινής μουσικής, έχοντας είσοδο μια εικόνα, καθώς επίσης και το διαχωρισμό σελίδων και την αφαίρεση των στίχων. Πρόκειται για μια μεταγενέστερη προσέγγιση του θέματος (6 χρόνια μετά), λαμβάνοντας υπόψη και θέματα ορθογραφίας της γραφής και διαφόρων μουσικών φράσεων που υπάρχουν, για βελτιωμένα αποτελέσματα. Δοκιμάστηκε σε διάφορα βιβλία η αναγνώριση συμβόλων, τα οποία έχουν διαφορετική γραμματοσειρά συμβόλων. Τελικά, η ακρίβεια που έχει επιτευχθεί σε αυτή την έρευνα είναι μεγαλύτερη από 99% σε όλα τα βιβλία.

‘Format Tuning In Byzantine Chant’ [29]
Στο παραπάνω άρθρο, οι συγγραφείς ασχολήθηκαν με συντονισμό φόρμας (formant tuning) σε Βυζαντινή εκκλησιαστική ψαλμωδία. Οι ηχογραφήσεις που χρησιμοποιήθηκαν ήταν από 10 διαφορετικούς ψάλτες. Η μέθοδος ανάλυσης περιελάμβανε μια ημι-αυτόματη τμηματοποίηση του ηχητικού υλικού, εξαγωγή μετρήσεων σε PRAAT και την τελική μετα-επεξεργασία σε MATLAB. Σύμφωνα με τα αποτελέσματα, η τεχνική χρησιμοποιείται από τους μοντέρνους Βυζαντινούς εκτελεστές.

‘An empirical comparison of machine learning techniques for chant classification’ [30]
Στο παραπάνω άρθρο οι συγγραφείς δημιουργούν ομάδες ψαλμών από ηχητική αναγνώριση της φωνής. Εξήγαγαν χαρακτηριστικά από τα ηχητικά αποσπάσματα παραστάσεων και εκπαίδευσαν ένα σύστημα χρησιμοποιώντας Hidden Markov Models(HMM). Επίσης έχουν χρησιμοποιήσει την τεχνική cross-validation για την αξιολόγηση του μοντέλου.

Οι παραπάνω μελέτες που έχουν γίνει δε θα αναφερθούν παρακάτω, αλλά μπορούν να χρησιμοποιηθούν μαζί με την παρούσα εργασία για περεταίρω έρευνα στον τομέα της Βυζαντινής Μουσικής.

1.4.2 Έρευνες στον τομέα της ταξινόμησης του είδους της μουσικής

'Automatic Makam recognition using chroma features' [16]
Στο επιστημονικό άρθρο αυτό, οι συγγραφείς ασχολήθηκαν με την ταξινόμηση τραγουδιών σε makams. Χρησιμοποιήθηκαν τα chroma features των τραγουδιών για την ταξινόμησή τους, και με κατάλληλη προεπεξεργασία των τραγουδιών, κατέληξαν σε ταξινόμηση μόνο σε 9 makam και ακρίβεια 89%. Χρησιμοποίησαν επιβλεπόμενη μάθηση και SVM (Support Vector Machine) για την ταξινόμηση. Επίσης κανονικοποίησαν τα δεδομένα προτού τα περάσουν στο μοντέλο.

'Features For Analysis Of Makam Music' [31]
Ο συγγραφέας αυτού του άρθρου αυτού απαριθμεί χαρακτηριστικά που μπορούν να χρησιμοποιηθούν για την ανάλυση του κάθε μακάμ, καθώς επίσης και το διαχωρισμό μεταξύ των διαφόρων μακάμ. Πιο συγκεκριμένα αναφέρει τα εξής 6 χαρακτηριστικά: Η κλίμακα που χρησιμοποιεί ο κάθε ήχος και τα διαστήματα μεταξύ των φθόγγων, το μελωδικό εύρος των νοτών, το συνολικό διάστημα δηλαδή στο οποίο κινείται ο κάθε ήχος, τη γενικό τρόπο με τον οποίο αναπτύσσεται η μελωδία, δηλαδή αν ανεβαίνει και μετά κατεβαίνει τονικά, τυπικές μελωδικές φράσεις που χρησιμοποιούνται συχνά, τυπικές μεταβάσεις από τον ήχο του κομματιού σε ένα άλλο για μια φράση και μετά επιστροφή στον ήχο του κομματιού, και τέλος τα πεντάχορδα ή τετράχορδα που απαρτίζουν την κλίμακα καθώς επίσης και τις νότες που είναι τονικές, δεσπόζουσες και προσαγωγείς.

'Music Genre Classification using MFCC, SVM and BPNN' [18]
Σε αυτό το άρθρο προτείνονται 2 ταξινομητές του είδους της μουσικής. Ένας ταξινομητής Support Vector Machine (SVM) και ένας ταξινομητής Back-Propagation Neural Network (BPNN). Και στους 2 ταξινομητές σαν δεδομένα χρησιμοποιούνται τα Mel-Frequency Cepstral Coefficients (MFCC) τα οποία εξάγονται από τα μουσικά αρχεία. Τελικά η ακρίβεια που έχει επιτευχθεί είναι 83% με τη χρήση SVM και 95% με τη χρήση BPNN.

'Classification of Classic Turkish Music Makams' [32]
Στο άρθρο αυτό ταξινομείται κλασσική τούρκικη μουσική σε 6 μακάμς. Χρησιμοποιήθηκε ως ταξινομητής ένα Πιθανοτικό Νευρωνικό Δίκτυο (Probabilistic Neural Network PNN), το οποίο είναι ένα νευρωνικό δίκτυο 3 επιπέδων εμπρόσθιας τροφοδότησης. Ως δεδομένα εισόδου στο νευρωνικό δίκτυο χρησιμοποιούνται τα Mel-Frequency Cepstral Coefficients (MFCC) τα οποία εξάγονται από τα μουσικά αρχεία. Τέλος, αξιολογήθηκε μεταξύ άλλων και η επίδραση που έχει το μήκος του κάθε κομματιού στην απόδοση του συστήματος. Η ακρίβεια που επιτεύχθηκε είναι 89.4% με μήκος κομματιού 6 δευτερόλεπτα.

‘Musical Genre Classification of Audio Signals’ [33]
Σε αυτή την έρευνα, χρησιμοποιείται ταξινομητής μουσικού είδους. Τα χαρακτηριστικά που χρησιμοποιούν είναι η ποιότητα τόνου ή η χροιά της μουσικής, χαρακτηριστικά του ρυθμού της μουσικής και χαρακτηριστικά της τονικότητας της μουσικής. Σαν ταξινομητές χρησιμοποιήθηκαν οι Simple Gaussian, Gaussian Mixture Model (GMM) και K Nearest Neighbours (KNN). Το αποτέλεσμα που επιτεύχθηκε ήταν ακρίβεια 61%, το οποίο ήταν πολύ καλό για τη χρονολογία 2002, αφού συγκρίθηκε και με αποτελέσματα ταξινόμησης που έχουν δώσει άνθρωποι.

Η ερευνά που έχει γίνει για τους σκοπούς αυτής της διπλωματικής επεκτάθηκε σε πιο ευρύ φάσμα δημοσιεύσεων, εν τούτοις μελετήθηκαν εκτεταμένα οι παραπάνω έρευνες.

2

Επεξεργασία Δεδομένων και Εξαγωγή Χαρακτηριστικών

2.1 Δημιουργία Συνόλου Δεδομένων

Για να εκπαιδευτεί το νευρωνικό σύστημα, θα πρέπει να επεξεργαστεί μια μεγάλη ποσότητα πληροφοριών για να αποκομίσει τα χαρακτηριστικά της κάθε κλάσης, ούτως ώστε να μπορεί να έχει επαρκές πληροφορίες για ταξινόμηση μιας νέας εισόδου. Τα χαρακτηριστικά των αρχείων του συνόλου δεδομένων θα πρέπει να είναι ευδιάκριτα και με όσο το δυνατό λιγότερο θόρυβο, ούτως ώστε οι διαφοροποιήσεις μεταξύ των κλάσεων να γίνονται πιο εύκολα και πιο αποδοτικά.

Οι περισσότερες από τις ηχογραφήσεις Βυζαντινής Μουσικής προέρχονται από ζωντανές εκκλησιαστικές λειτουργίες, χρησιμοποιώντας μάλιστα ερασιτεχνικό εξοπλισμό, με αποτέλεσμα η ποιότητα του ηχητικού αρχείου να είναι πολύ χαμηλή για να μπορεί να εξάγει τις σωστές πληροφορίες ένα σύστημα. Επίσης πολλές από τις ηχογραφήσεις ψαλμωδιών με χρήση επαγγελματικού εξοπλισμού έχουν γίνει με χορωδίες αντί μονοφωνίες, γεγονός που καθιστά δυσκολότερη την εξαγωγή πληροφοριών για το ψαλτικό κομμάτι λόγω ζητημάτων συγχρονισμού της χορωδίας και ομοιογένειας. Για τους παραπάνω λόγους, η εύρεση συνόλου δεδομένων που να πληροί όλες τις απαιτήσεις γίνεται δύσκολη.

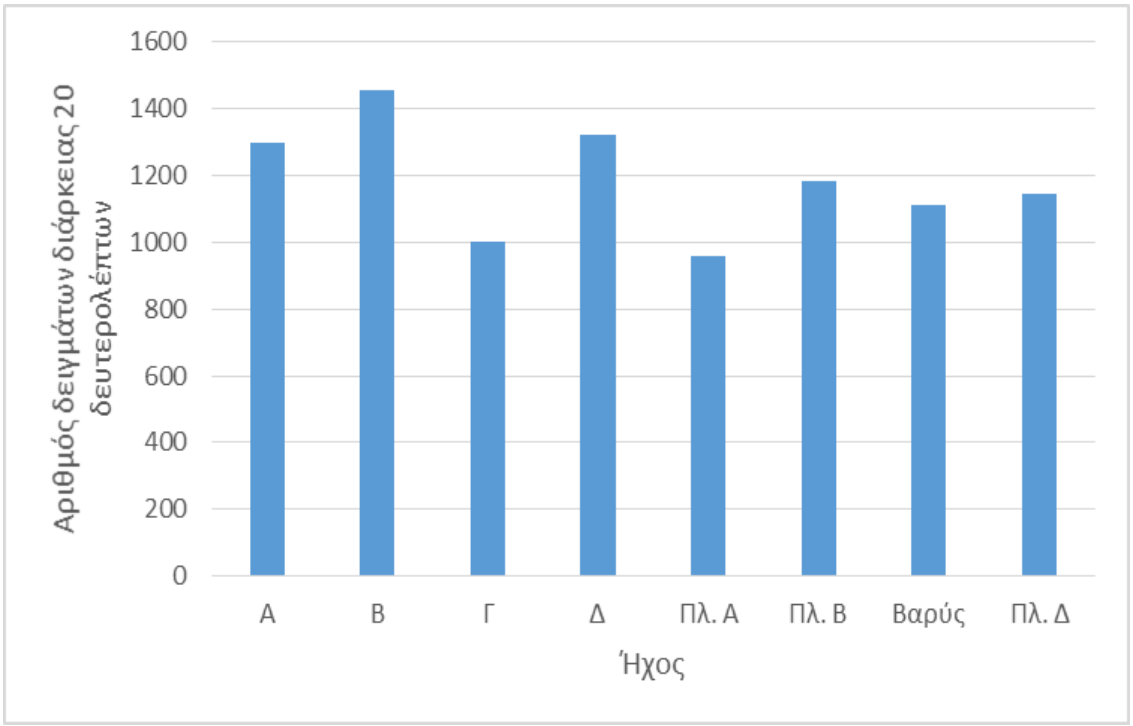
Το σύνολο δεδομένων που χρησιμοποιήθηκε είναι σε μεγάλο βαθμό ηχογραφήσεις που έχει κάνει ο Άρχων Πρωτοψάλτης Ιεράς Αρχιεπισκοπής Αμερικής κύριος Φώτιος Κετσετζής με χρήση επαγγελματικού εξοπλισμού και συνοδεία ισοκρατήματος. Πρόκειται για κομμάτια Σύντομου και Αργού Αναστασιματαρίου, τα οποία καλύπτουν μεγάλο μέρος του. Τα

κομμάτια αυτά έχουν Σύντομο Ειρμολογικό και Σύντομο Στιχηραρικό μέλος. Επίσης, υπάρχουν κομμάτια και από τους οκτώ ήχους της Βυζαντινής Μουσικής και είναι μάλιστα ισοκατανεμημένα.

Επίσης, χρησιμοποιήθηκαν ηχογραφήσεις από τον κύριο Θανάση Δασκαλοθανάση σε αργές καταβασίες όλου του χρόνου. Πρόκειται για κομμάτια Αργού Ειρμολογικού μέλους. Υπάρχουν κομμάτια και από τους οκτώ ήχους της Βυζαντινής Μουσικής.
Το σύνολο δεδομένων απαρτίζεται συνολικά από 403 ηχητικά αρχεία με συνολική διάρκεια 10 ώρες, 53 λεπτά και 40 δευτερόλεπτα. Η μέση διάρκεια του κάθε αρχείου είναι 1 λεπτό και 38 δευτερόλεπτα. Λόγω της μεγάλης διάρκειας των αρχείων και για λόγους απόδοσης του συστήματος, όλα τα κομμάτια διαμελίστηκαν σε πιο μικρά μέρη των 20 δευτερολέπτων το καθένα. Τα κομμάτια 20 δευτερολέπτων συνολικά είναι 9487.

Όλα τα κομμάτια που περιεγράφηκαν παραπάνω, είναι σε μορφή MP3 και ομαδοποιήθηκαν σε ένα σύνολο δεδομένων με βάση τον ήχο τους για να χρησιμοποιηθούν ως δεδομένα εκπαίδευσης του μοντέλου. Μαζί με την ομαδοποίησή τους, έχει δημιουργηθεί και ένα αρχείο τύπου CSV (comma-separated values) για τον κάθε ήχο, το οποίο συμπεριλαμβάνει το όνομα του κάθε ηχητικού μαζί με την ετικέτα του, που αντιστοιχεί στον ήχο του. Αυτή η διαδικασία έχει γίνει ούτως ώστε το σύστημα να μπορεί να εξάγει τα χαρακτηριστικά του κάθε αρχείου και να ξέρει ταυτόχρονα σε ποια κλάση-ήχο ανήκει.

Για καλύτερες επιδόσεις του μοντέλου, συνίσταται τα δεδομένα να είναι όσο το δυνατόν ισοκατανεμημένα στις διάφορες κλάσεις. Παρακάτω παρουσιάζεται η γραφική παράσταση Κλάσεων/Ηχων-Αριθμό δειγμάτων διάρκειας 20 δευτερολέπτων, όπου φαίνεται ότι σε μεγάλο βαθμό είναι ισοκατανεμημένα τα δεδομένα.



2.1.1 Γραφική παράσταση Ήχων-Αριθμού δειγμάτων διάρκειας 20 δευτερολέπτων

Το σύνολο που χρησιμοποιήθηκε παρά του ότι είναι περιορισμένο, είναι επαρκές για εξαγωγή αρκετών χαρακτηριστικών ούτως ώστε να έχουμε αρκετά καλά αποτελέσματα.

2.2 Χαρακτηριστικά Ηχητικών Σημάτων

Τα ηχητικά σήματα έχουν πολλά χαρακτηριστικά τα οποία μπορούν να εξαχθούν. Ανάλογα με το σκοπό που θέλουμε να πετύχουμε, κάποια χαρακτηριστικά έχουν αποδειχθεί να είναι πιο αποτελεσματικά από άλλα για συγκεκριμένες διαδικασίες. Για το σκοπό αυτό, έχει γίνει μια προμελέτη για την επιλογή των χαρακτηριστικών που θα μπορούσαν να φανούν αποδοτικά για την ταξινόμηση των ψαμλωδιών και έπειτα δοκιμάστηκαν με διάφορους τρόπους και συνδυασμούς για να βρεθούν τα αποδοτικότερα.

Κυριότερα χαρακτηριστικά που εξάγονται για επεξεργασία ανθρώπινης φωνής είναι τα Mel-Frequency Cepstral Coefficients (MFCC). Χρησιμοποιήθηκαν σε πολλές εφαρμογές για αναγνώριση του ομιλητή [34] [35], για ταξινόμηση του είδους ευρωπαικής μουσικής [18] και για ταξινόμηση μακάμ [32], και επέφεραν πολύ καλά αποτελέσματα. Η εργασία αυτή ασχολείται με ταξινόμηση ανθρώπινης φωνής, άρα θα χρησιμοποιηθούν τα MFCC.

Άλλα χαρακτηριστικά που χρησιμοποιούνται γενικότερα στην Ανάκτηση Μουσικής Πληροφορίας και πιο συγκεκριμένα σε πολλές εφαρμογές αναγνώρισης μακαμ, είναι τα Chroma Features. Έχουν γίνει έρευνες και δοκιμές σε μοντέλα ταξινόμησης μακάμ [16] , καθώς επίσης και σε αναγνώριση συγχορδιών [36] και κλιμάκων/κλειδιών [37], επιφέροντας και πάλι πολύ καλά αποτελέσματα. Παρόμοιες εφαρμογές με αυτές που αναλύθηκαν πιο πάνω μπορούν να χρησιμοποιηθούν για την αναγνώριση του ήχου μιας ψαλμωδίας, οπότε θα χρησιμοποιηθούν και τα Chroma Features.

Πολλές μελέτες ταξινόμησης είδους μουσικής χρησιμοποίησαν κατά κύριο χαρακτηριστικό είτε τα Mel-Frequency Cepstral Coefficients (MFCC) είτε τα Chroma Features. Εν τούτοις, αφού η τρέχουσα εργασία ασχολείται με επεξεργασία ηχητικού σήματος, παράληψη θα ήταν να μην χρησιμοποιηθούν και δεδομένα από το φάσμα του σήματος. Παρατηρήσαμε ότι σε έρευνες χρησιμοποιούνται τα Φασματικό Κέντρο (Spectral Centroid) [33] [38] , η Φασματική Διάθεση (Spectral Roll-Off) [33] [39] και το Φασματικό Εύρος Ζώνης (Spectral Bandwidth).

Τέλος, θα χρησιμοποιηθεί ο Ρυθμός Αλλαγής Πρόσημου (Zero Crossing Rate) και η Ρίζα Μέσης Τετραγωνικής Ενέργειας (Root Mean Square Energy RMSE).

Παρακάτω θα εξηγηθούν τα επιλεχθέντα χαρακτηριστικά.

2.2.1 Mel-frequency cepstral coefficients (MFCC)

Τα Mel-Frequency Cepstral Coefficients (MFCC) είναι οι συντελεστές που χρησιμοποιούνται στην προσπάθεια που γίνεται για την περιγραφή της ανθρώπινης αντίληψης για τον ήχο. Η ανθρώπινη ακοή δεν μπορεί να αντιληφθεί τη συχνότητα των ήχων με μια γραμμική σχέση, αλλά σύμφωνα με έρευνες που έχουν γίνει, μπορεί να την αντιληφθεί με μια λογαριθμική σχέση. Οι συντελεστές MFC βασίζονται στο φάσμα του σήματος και αντί για γραμμική κλίμακα χρησιμοποιούν την κλίμακα mel, η οποία είναι λογαριθμική και προσομοιάζει την ακουστική αντίληψη του ανθρώπινου συστήματος. Για τον υπολογισμό τους, ακολουθείται ο παρακάτω αλγόριθμος: [40]

1. Υπολογίζεται το φάσμα του σήματος χρησιμοποιώντας το Μετασχηματισμό Φουριέ (συνήθως διαμελίζεται το κομμάτι σε πλαίσια).
2. Μετατρέπεται το φάσμα σε κλίμακα mel.
3. Υπολογίζεται ο λογάριθμος των πιο πάνω.
4. Υπολογίζεται ο διακριτός μετασχηματισμός συνημίτονου .

Οι συντελεστές MFCC εν τούτοις δεν έχουν ανοχή στην παρουσία θορύβου, για αυτό και συνηθίζεται να γίνεται κανονικοποίηση των τιμών πριν χρησιμοποιηθούν.

Σήμερα, οι συντελεστές Mel-Frequency Cepstral χρησιμοποιούνται σε εφαρμογές αναγνώρισης ομιλίας, όπως για παράδειγμα συστήματα που αυτόματα αναγνωρίζουν τους αριθμούς που λέγονται σε ένα τηλέφωνο. Επίσης χρησιμοποιούνται όλο και περισσότερο σε εφαρμογές ανάκτησης μουσικής πληροφορίας όπως κατηγοριοποίησης του είδους της μουσικής και μετρήσεις ομοιότητα ήχου.

Αυτό που κάνει αυτούς τους συντελεστές τόσο αποδοτικούς σε τομείς αναγνώρισης ανθρώπινης φωνής, είναι η κλίμακα mel, και το γεγονός ότι προσομοιάζεται η ανθρώπινη ακοή, σε αντίθεση με τη γραμμικότητα που αυξάνεται η συχνότητα.

2.2.2 Chroma Feature ή Chromagram

Στη μουσική, σαν οκτάβα χαρακτηρίζεται η απόσταση μεταξύ δύο μουσικών φθόγγων των οποίων η συχνότητα είναι η μια διπλάσια από τον άλλο.
Στην Δυτική Μουσική, μια οκτάβα απαρτίζεται από 12 διαστήματα που έχουν ίση απόσταση μεταξύ τους και ονομάζονται ημιτόνια. Στη μουσική σημειογραφία, μπορούμε να γράψουμε τις νότες μιας οκτάβας ως το παρακάτω σύνολο, το οποίο ονομάζεται πίνακας chroma:

$$\{C, C\sharp, D, D\sharp, E , F, F\sharp, G, G\sharp, A, A\sharp, B\}$$

Οι παραπάνω νότες είναι όπως γράφονται στη Δυτική Σημειογραφία.
Επίσης ανάλογα με την οκτάβα που βρίσκονται, το τονικό τους ύψος, διαχωρίζουμε και τον κάθε φθόγγο ξεχωριστά, δηλαδή

$$\{..., C_{-2}, C_{-1}, C_0, C_1, C_2, C_3 ...\}$$

όπου ο κάθε φθόγγος αυτού του συνόλου έχει απόσταση μεταξύ του ίση με μια οκτάβα.

Το Chroma Feature ή Chromagram είναι η αναπαράσταση των τόνων ενός μουσικού αρχείου χρησιμοποιώντας το σύνολο με τις 12 νότες που απαρτίζουν μια οκτάβα και έχουν απόσταση μεταξύ τους ίση με ένα ημιτόνιο. Το Chromagram δε λαμβάνει υπόψη του τις διάφορες οκτάβες μεταξύ των νοτών, αλλά αναφέρεται σε όλες τις νότες με το ίδιο όνομα, ανεξάρτητα από το ποια οκτάβα ανήκουν. [41]

Το Chroma Feature έχουν μεγάλη ανοχή στην ποιότητα τόνου, στο ποιο όργανο αναπαράγει τον τόνο δηλαδή, και αυτό τα καθιστά πολύ ασφαλή και ακριβή σε θέματα ανάλυσης της μουσικής. Πολλές διεργασίες αναγνώρισης συγχορδιών χρησιμοποιούν το Chroma Feature. Επίσης το Chroma Feature χρησιμοποιείται θέματα ευθυγράμμισης μουσικής και συγχρονισμού, καθώς επίσης και σε ανάλυση της μουσικής δομής.

2.2.3 Φασματικό Κέντρο (Spectral Centroid)

Το φασματικό κέντρο δείχνει που βρίσκεται το κέντρο της μάζας του φάσματος ενός ηχητικού κομματιού. Υπολογίζεται ως ο σταθμισμένος μέσος όρος των συχνοτήτων του κομματιού που υπολογίζονται με το μετασχηματισμό Φουριέ.

$$Spectral\ Centroid\ = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)}$$

Το Φασματικό Κέντρο χρησιμοποιείται σε θέματα επεξεργασίας ήχου και μουσικής για να ληφθούν μετρήσεις σχετικά με την ποιότητα τόνου, το ποιο όργανο αναπαράγει τον τόνο δηλαδή, αφού είναι πολύ καλή μέτρηση για τη φωτεινότητα ενός ήχου.

2.2.4 Φασματική Διάθεση (Spectral Roll-Off)

Είναι μια μέτρηση του σχήματος του σήματος. Αποτελεί τη συχνότητα κάτω από την οποία βρίσκεται κάποιο συγκεκριμένο ποσοστό της συνολικής ενέργειας του σήματος.

2.2.5 Ρίζα Μέσης Τετραγωνικής Ενέργειας (Root Mean Square Energy RMSE)

Αποτελεί την τετραγωνική ρίζα του αριθμητικού μέσου του τετραγώνου της ενέργειας του σήματος.

2.2.6 Φασματικό Εύρος Ζώνης (Spectral Bandwidth)

Με τον όρο Φασματικό Εύρος Ζώνης εννοούμε τη διαφορά μεταξύ της υψηλότερης και της χαμηλότερης συχνότητας σε ένα χρονικό παράθυρο ενός ηχητικού αρχείου.

2.2.7 Ρυθμός Αλλαγής Πρόσημου (Zero Crossing Rate)

Ο Ρυθμός Αλλαγής Πρόσημου είναι ο αριθμός των φορών όπου το σήμα ενός ηχητικού αλλάζει πρόσημο, μεταβαίνει δηλαδή από θετικό σε αρνητικό και αντίστροφα. Χρησιμοποιείται ευρέως σε εφαρμογές αναγνώρισης φωνής και ανάκτησης μουσικής πληροφορίας.

2.3 Εξαγωγή Χαρακτηριστικών και Προεπεξεργασία Δεδομένων

Αφού μελετήθηκαν τα χαρακτηριστικά που θα χρησιμοποιηθούν, πρέπει να εξαχθούν από τα ηχητικά αρχεία και να έρθουν σε κατάλληλη μορφή για να μπορούν να περαστούν στο μοντέλο και να εκπαιδευτεί. Η εξαγωγή τους προϋποθέτει πολλές γνώσεις ανάλυσης σήματος και είναι μια χρονοβόρα διαδικασία λόγω του μεγάλου όγκου δεδομένων.
Στον τομέα της Ανάκτησης Μουσικής Πληροφορίας (MIR) χρησιμοποιείται ευρέως η βιβλιοθήκη librosa [42] της Python, η οποία παρέχει συναρτήσεις για σκοπούς επεξεργασίας ηχητικού σήματος και μουσικής. Αυτή χρησιμοποιήθηκε και σε αυτή την εργασία για την εξαγωγή των χαρακτηριστικών που αναλύθηκαν από τα μουσικά αρχεία.

Για σκοπούς βελτίωσης των αποτελεσμάτων της εκπαίδευσης του μοντέλου καθώς επίσης και για να βρίσκονται όλα τα δεδομένα σε μορφή τέτοια ώστε να μπορεί το μοντέλο να τα επεξεργαστεί, τα δεδομένα περνούν από ένα στάδιο επεξεργασίας. Το στάδιο αυτό αποτελείται από πολλά επίπεδα τα οποία δοκιμάστηκαν και τα περισσότερα επέφεραν βελτιωμένα αποτελέσματα. Τα στάδια αναλύονται παρακάτω.

2.3.1 Αύξησης Δεδομένων (Data Augmentation)

Όσο μεγαλύτερο είναι το σύνολο δεδομένων που χρησιμοποιείται για την εκπαίδευση του μοντέλου, τόσο καλύτερα αποτελέσματα έχουμε. Η επέκτασή του σε πολλές περιπτώσεις επιφέρει καλύτερα αποτελέσματα, με περισσότερη ανοχή σε θορύβους λόγω κακής ποιότητας ηχογράφησης ή στη διαφορετική χροιά μεταξύ των διαφόρων ψαλτών. Εν τούτοις δεν είναι πάντα εύκολο και εφικτό να επεκταθεί με δεδομένα που να καλύπτουν τις απαιτήσεις που πρέπει. Για το λόγο αυτό, έγινε χρήση της τεχνικής Αύξησης Δεδομένων (Data Augmentation) ούτως ώστε με το ήδη υπάρχων σύνολο δεδομένων, να δημιουργηθούν καινούρια ηχητικά αρχεία από τα ήδη υπάρχοντα, με ορισμένες παραλλαγές σε ορισμένα χαρακτηριστικά. Τα χαρακτηριστικά που μπορούν να παραλλαχθούν είναι πολλά, σε αυτή την εργασία αλλάχθηκαν ο τόνος του κομματιού (pitch shifting) και προστέθηκε θόρυβος. Για να μην αλλοιωθούν τα στοιχεία του κάθε ψαλμού, η αλλαγή του τόνου που έχει γίνει είναι 0.4 του τόνου (4.8 μόρια) και 0.5 του τόνου (6 μόρια, 1 ημιτόνιο) και ο θόρυβος είναι τυχαίος με συντελεστή 0.08. Επίσης το κάθε αρχείο έχει μετατοπιστεί κατά 5, 10 και 15 δευτερόλεπτα. Με αυτό τον τρόπο το μοντέλο εκπαιδεύτηκε καλύτερα, και η ακρίβειά του βελτιώθηκε στο βαθμό του 10% από τα προηγούμενα αποτελέσματα.

2.3.2 Διαίρεση κομματιού

Για να μπορεί το σύστημα να επεξεργαστεί τα δεδομένα που εξάχθηκαν από τα ηχητικά αρχεία του συνόλου δεδομένων, πρέπει να έχουν όλα τις ίδιες διαστάσεις. Για να επιτευχθεί αυτός ο σκοπός, τα αρχεία θα πρέπει να έχουν όλα την ίδια διάρκεια, για αυτό και το κάθε κομμάτι διαιρέθηκε σε περισσότερα ίσης διάρκειας. Το άρθρο [32] προτείνει διάρκεια κομματιού ίση με 6 δευτερόλεπτα. Εντούτοις, η διάρκεια που επιλέχθηκε για το κάθε κομμάτι του αρχείου μετά από δοκιμές ήταν τα 20 δευτερόλεπτα, που είναι επαρκές για την αναγνώριση στοιχείων του ήχου του, και επίσης περίπου τόσα δευτερόλεπτα διαρκεί μια μουσική φράση σε μια ψαλμωδία. Επίσης με τη μετατόπιση που έχει γίνει από την τεχνική της Αύξησης Δεδομένων, θα βρεθούν περισσότερες μουσικές φράσεις του κάθε ήχου, με αποτέλεσμα το σύστημα να έχει μεγαλύτερη ακρίβεια στις προβλέψεις του. Έχοντας τις παραπάνω ιδέες υπόψιν, δοκιμαστηκαν και οι διαιρέσεις κομματιού 15 και 25 δευτερολέπτων, τα οποία όμως είχαν ως αποτέλεσμα ο ταξινομητής μετά την εκπαίδευση να έχει χαμηλότερη ακρίβεια, οπότε δεν προτιμήθηκαν.

2.3.3 Κανονικοποίηση δεδομένων (Normalization)

Κατά την εκπαίδευση του, το μοντέλο χρησιμοποιεί τις τιμές που του περνάμε, μαζί με το βάρος του κάθε ενός, που διαμορφώνεται κατά την εκπαίδευση του. Για να μην χάνεται η αξία των βαρών που διαμορφώνει το μοντέλο, γίνεται κανονικοποίηση των τιμών στο διάστημα (-1,1) ή (0,1), ούτως ώστε όσο μεγάλη απόκλιση και να έχουν οι τιμές μεταξύ τους, να μην έχει μεγαλύτερη επήρεια ένα χαρακτηριστικό λόγω της τιμής του.

2.3.4 Παραγέμισμα (Padding)

Όπως προαναφέρθηκε, όλα τα δεδομένα που θα περαστούν σαν είσοδος στο μοντέλο πρέπει να έχουν τις ίδιες διαστάσεις. Πολλά από τα δεδομένα που εξάχθηκαν δεν έχουν τις ίδιες διαστάσεις. Ως εκ τούτου για να μπορέσουν να χρησιμοποιηθούν οι πληροφορίες από όλα τα χαρακτηριστικά των ηχητικών αρχείων, χρησιμοποιήθηκε η τεχνική του Παραγεμίσματος. Με αυτή την τεχνική, προστίθενται τιμές στα χαρακτηριστικά με τις πιο μικρές διαστάσεις μέχρι να έχουν όλα τα δεδομένα τις ίδιες διαστάσεις. Για να υπολογιστούν οι τιμές που προστίθενται υπάρχουν διάφορες τεχνικές. Δοκιμάστηκαν η προσθήκη του μέσου όρου όλων των δεδομένων του χαρακτηριστικού όσες φορές χρειαζόταν, η επανάληψη των ήδη υπάρχοντών τιμών μια-μια και τέλος η προσθήκη του μέσου όρου ανά δύο των τιμών αναμεταξύ τους. Με τον τρόπο αυτό, δε χάνεται η γενική μορφή του κάθε χαρακτηριστικού, αλλά αυξάνονται οι διαστάσεις του.

2.3.5 Προσθήκη πληροφορίας από το τέλος του κομματιού

Μετά από μελέτη της έρευνας ‘Automatic Makam recognition using chroma features’ [16], στην ενότητα 3.2.1 όπου αναλύονται τα χαρακτηριστικά που χρησιμοποιούνται για την κατηγοριοποίηση ενός κομματιού, χρησιμοποιήθηκε η κατάληξη κάθε κομματιού, αφού η κατάληξη των ψαλμωδιών σε πολλές περιπτώσεις είναι η ίδια σε κάθε ήχο. Έτσι, για κάθε κομμάτι 20 δευτερολέπτων από το κάθε ηχητικό αρχείο, εκτός από τα χαρακτηριστικά του, κρατάμε τα χαρακτηριστικά από τα τελευταία 20 δευτερόλεπτα του κομματιού. Η υλοποίηση της πιο πάνω τεχνικής επέφερε αύξηση της ακρίβειας του μοντέλου ύψους 10%.

3

Μοντέλα που θα δοκιμαστούν και Εξαγωγή Αποτελεσμάτων

3.1 Προτεινόμενα Μοντέλα

Στην ενότητα 1.1 αναλύθηκαν οι διαφορετικές προσεγγίσεις που έχουν αναπτυχθεί στον τομέα της μηχανικής μάθησης. Η κάθε προσέγγιση έχει επιφέρει σημαντικά αποτελέσματα σε διάφορα προβλήματα. Ορισμένες προσεγγίσεις όμως έχουν ξεχωρίσει με τις επιδόσεις τους σε συγκεκριμένα προβλήματα. Σε αυτή την κατηγορία ανήκουν τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks, CNN), τα οποία έχουν πάρα πολύ καλές επιδόσεις σε θέματα ανάλυσης και κατηγοριοποίησης εικόνων. Πολλά υπερσύγχρονα συστήματα ταξινόμησης εικόνων χρησιμοποιούν αρχιτεκτονική Συνελικτικών Νευρωνικών Δικτύων. [43] [44] [45] [46]. Παρόμοιας φύσης πρόβλημα με την κατηγοριοποίηση εικόνων είναι και το πρόβλημα κατηγοριοποίησης ηχητικών αρχείων, αφού όπως μια εικόνα χαρακτηρίζεται από ένα πολυδιάστατο πίνακα με αριθμούς που αντιπροσωπεύουν το χρώμα του κάθε pixel, έτσι και τα ηχητικά αρχεία χαρακτηρίζονται από πολυδιάστατους πίνακες με αριθμούς που αντιπροσωπεύουν χαρακτηριστικά τους κάθε χρονική στιγμή. Επίσης αναφορές υπάρχουν για τα Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (LSTM) ότι έχουν πολύ καλές επιδόσεις στον κλάδο της ανάκτησης μουσικής πληροφορίας [47], γι’ αυτό θα δοκιμαστούν και αυτά.

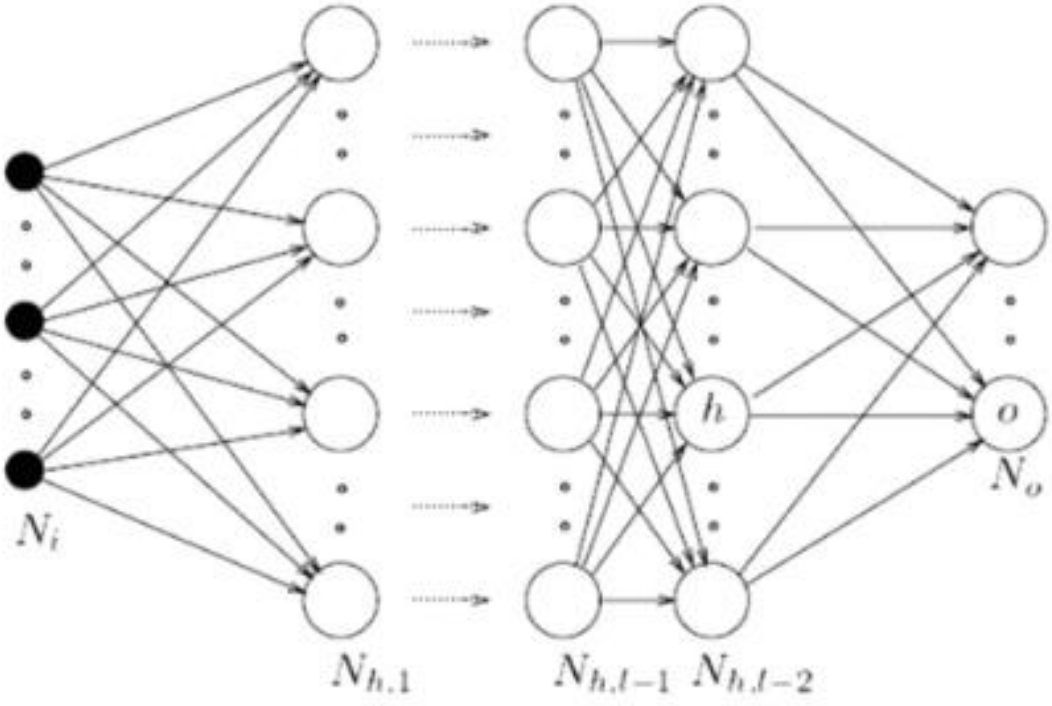
Όπως έχει αναφερθεί, σε αυτή την εργασία θα δοκιμαστούν μοντέλα με διάφορα ενδιάμεσα επίπεδα και θα βρεθεί η καλύτερη δυνατή προσέγγιση. Θα βελτιστοποιηθούν επίσης μερικοί παράγοντες που επηρεάζουν άμεσα την ακρίβεια του μοντέλου όπως η επιλογή των χαρακτηριστικών που θα περαστούν σαν είσοδος προς εκπαίδευση στο μοντέλο, ο αριθμός των κρυφών επιπέδων, το είδος των κρυφών επιπέδων, ο αριθμός των

συνδέσεων του κάθε κρυφού επιπέδου, ο τρόπος παραγεμίσματος των δεδομένων και ο βελτιστοποιητής. Σε αυτή την εργασία δεν έχει χρησιμοποιηθεί η τεχνική της μεταφοράς μάθησης, και κατ' επέκταση όλα τα μοντέλα που θα δοκιμαστούν έχουν σχεδιαστεί και εκπαιδευτεί από την αρχή.

Για την υλοποίηση των μοντέλων που θα περιγράφουν έχει χρησιμοποιηθεί η βιβλιοθήκη TensorFlow της Python. Η βιβλιοθήκη αυτή είναι εξειδικευμένη στον τομέα της μηχανικής μάθησης και παρέχει συναρτήσεις που απλοποιούν τη διαδικασία μάθησης. Για την ευκολότερη χρήση της βιβλιοθήκης, χρησιμοποιήθηκε η διεπαφή εφαρμογών προγραμματισμού (Application Programming Interface , API) Keras, η οποία παρέχει δομικά στοιχεία για την ανάπτυξη και αποστολή λύσεων μηχανικής εκμάθησης με υψηλές επιδόσεις. Όλα τα μοντέλα που δοκιμάστηκαν είναι υλοποιημένα από το TensorFlow2 και τη διεπαφή εφαρμογών προγραμματισμού Keras.

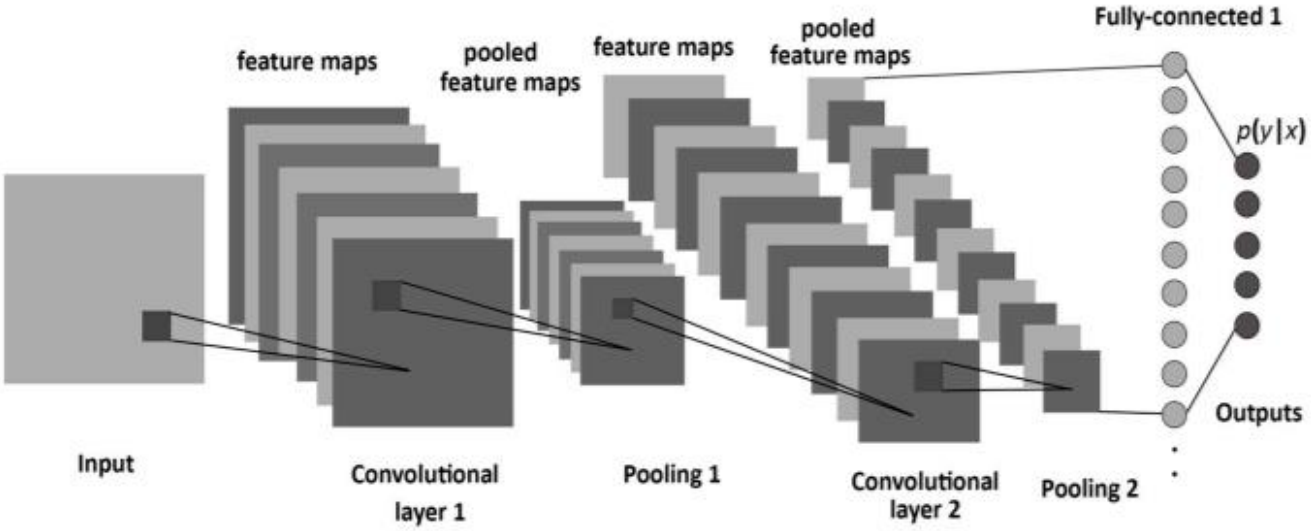
Τα μοντέλα που δοκιμάστηκαν χωρίζονται σε τέσσερις κατηγορίες, τα Πλήρως Συνδεδεμένα Νευρωνικά Δίκτυα (Fully Connected Neural Network), τα Συνελικτικά Νευρωνικά Δίκτυα (CNN), τα Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (LSTM) και ένα υβριδικό μοντέλο CNN + LSTM, το οποίο είναι ένα κανονικό Συνελικτικό Νευρωνικό Δίκτυο το οποίο στο τέλος του έχει ένα επίπεδο LSTM..

Τα Πλήρως Συδεδεμένα Νευρωνικά Δίκτυα αποτελούνται από το επίπεδο εισόδου, το επίπεδο εξόδου και τα ενδιάμεσα κρυμμένα επίπεδα. Το επίπεδο εξόδου αποτελείται από ένα πλήρως συνδεδεμένο επίπεδο Dense, το οποίο αποτελείται από τον ίδιο αριθμό κόμβων όσες είναι οι ομάδες που κατηγοριοποιεί το μοντέλο, στην περίπτωση αυτής της εργασίας 8. Το πρώτο κρυμμένο επίπεδο είναι ένα επίπεδο ισοπέδωσης Flatten, το οποίο απλά δέχεται τον πολυδιάστατο πίνακα εισόδου και τον μετατρέπει σε μονοδιάστατο πίνακα. Τα υπόλοιπα ενδιάμεσα επίπεδα αποτελούνται από μια σειρά πλήρως συνδεδεμένων επιπέδων Dense, με διάφορους αριθμούς κόμβων. Ο αριθμός των ενδιάμεσων επιπέδων, καθώς επίσης και ο αριθμός των κόμβων του κάθε επιπέδου αποτελεί μέρος της πειραματικής έρευνας της εργασίας αυτής, για αυτό και δοκιμάστηκαν διάφοροι συνδυασμοί.



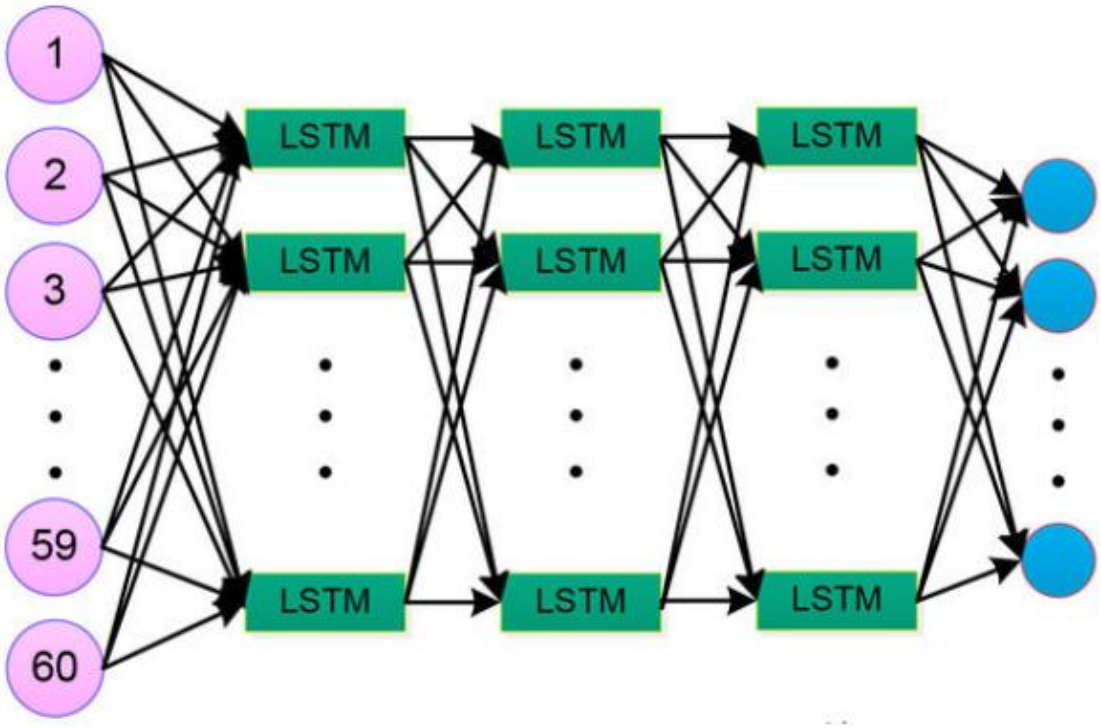
3.1.1 Αρχιτεκτονική Πλήρως Συνδεδεμένου Νευρωνικού Δικτύου [48]

Τα Συνελικτικά Νευρωνικά Δικτυα αποτελούνται όπως και τα Πλήρως Συνδεδεμένα Νευρωνικά Δίκτυα από το επίπεδο εισόδου, το επίπεδο εξόδου και τα ενδιάμεσα κρυμμένα επίπεδα. Εν τούτοις, τα κρυφά επίπεδα των Συνελικτικών Νευρωνικών δικτύων αποτελούνται από μια συστοιχία επιπέδων που επαναλαμβάνονται με συγκεκριμένη σειρά. το οποίο λαμβάνει τη μέγιστη τιμή από κάθε παράθυρο-πίνακα διαστάσεων που δηλώνονται. Ο αριθμός αυτών των συστοιχιών, καθώς και οι παράμετροι τους αποτελούν μέρος της πειραματικής έρευνας της εργασίας αυτής, για αυτό και δοκιμάστηκαν διάφοροι συνδυασμοί.



3.1.2 Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου [49]

Τα Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (LSTM) αποτελούνται όπως και τα προηγούμενα 2 Δίκτυα από το επίπεδο εισόδου, το επίπεδο εξόδου και τα ενδιάμεσα κρυμμένα επίπεδα. Το επίπεδο εξόδου τους είναι ένα πλήρως συνδεδεμένο επίπεδο Dense, το οποίο έχει τον ίδιο αριθμό κόμβων όσες είναι οι ομάδες που κατηγοριοποιεί το μοντέλο, στην περίπτωση αυτής της εργασίας 8. Τα ενδιάμεσα κρυμμένα επίπεδα αποτελούνται από ένα ή περισσότερα επίπεδα LSTM, με διάφορους αριθμούς κόμβων, ανάλογα με την εφαρμογή. Ο αριθμός αυτών των επιπέδων, καθώς και οι παράμετροι τους αποτελούν μέρος της πειραματικής έρευνας της εργασίας αυτής, για αυτό και δοκιμάστηκαν διάφοροι συνδυασμοί.



3.1.3 Αρχιτεκτονική LSTM Δικτύου

3.2 Δοκιμή Μοντέλων Ταξινόμησης Βαθιάς Μάθησης και Βελτιστοποίηση Υπερπαραμέτρων

Κρατώντας σταθερό ένα μοντέλο μαζί και όλες τις παραμέτρους του και αλλάζοντας ένα μόνο χαρακτηριστικό, μπορούμε να παρατηρήσουμε την επίδραση που έχει αυτό το χαρακτηριστικό στην ταξινόμηση του ήχου των ψαλμωδιών στην τρέχουσα εργασία. Αυτή ήταν και η τακτική που χρησιμοποιήθηκε για την εύρεση των βέλτιστων παραμέτρων.

Σε πρώτο στάδιο χρησιμοποιήθηκαν ένα-ένα τα χαρακτηριστικά που εξάχθηκαν από τα ηχητικά αρχεία, για να παρατηρηθεί η επίδραση, το βάρος που μπορεί να προσδώσει το κάθε χαρακτηριστικό κατά την ταξινόμηση.

Για να παρατηρηθεί η επίδραση του κάθε χαρακτηριστικού ξεχωριστά έχουν γίνει οι παρακάτω δοκιμές:
Τα χαρακτηριστικά που χρησιμοποιήθηκαν δεν συμπεριλάμβαναν πληροφορία από το τέλος του κάθε κομματιού.
Αριθμός εποχών: 60
Batch Size: 32
Βελτιστοποιητής: Adam (learning_rate = 0,001)

Συνελικτικό Νευρωνικό Δίκτυο:

Chroma	Spectral Centroid	Spectral Bandwidth	Spectral Rolloff	MFCC (20 coefficients)	RMSE	Zero Crossing Rate
61%	35%	28%	34%	66%	33%	27%

3.2.1 Πίνακας επιδόσεων Συνελικτικού Νευρωνικού Δικτύου για όλα τα χαρακτηριστικά που εξάχθηκαν

Πλήρως Συνδεδεμένο Νευρωνικό Δίκτυο:

Chroma	Spectral Centroid	Spectral Bandwidth	Spectral Rolloff	MFCC (20 coefficients)	RMSE	Zero Crossing Rate
43%	9%	16%	16%	55%	16%	16%

3.2.2 Πίνακας επιδόσεων Πλήρως Συνδεδεμένου μοντέλου για όλα τα χαρακτηριστικά που εξάχθηκαν

LSTM Δίκτυο:

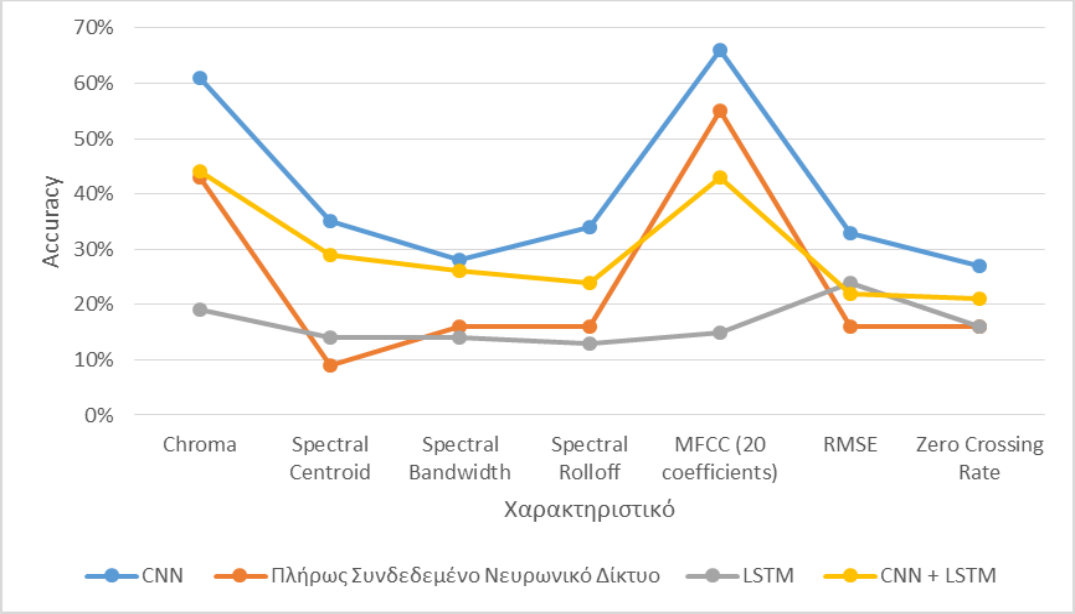
Chroma	Spectral Centroid	Spectral Bandwidth	Spectral Rolloff	MFCC (20 coefficients)	RMSE	Zero Crossing Rate
19%	14%	14%	13%	15%	24%	16%

3.2.3 Πίνακας επιδόσεων LSTM Δικτύου για όλα τα χαρακτηριστικά που εξάχθηκαν

CNN + LSTM Δίκτυο:

Chroma	Spectral Centroid	Spectral Bandwidth	Spectral Rolloff	MFCC (20 coefficients)	RMSE	Zero Crossing Rate
44%	29%	26%	24%	43%	22%	21%

3.2.4 Πίνακας επιδόσεων CNN + LSTM Δικτύου για όλα τα χαρακτηριστικά που εξάχθηκαν



3.2.1 Γραφική παράσταση Ακρίβεια-Χαρακτηριστικά για τα 2 μοντέλα που δοκιμάστηκαν, Πλήρως Συνδεδεμένο Νευρωνικό Μοντέλο και Συνελικτικό Νευρωνικό Μοντέλο

Όπως φαίνεται από τα πειραματικά αποτελέσματα ξεχωρίζουν οι επιδόσεις που είχαν τα μοντέλα CNN, CNN + LSTM και Πλήρως Συνδεδεμένο, όταν χρησιμοποιήθηκαν τα χαρακτηριστικά Chroma και MFCC. Κάτι που είναι επίσης άμεσα αντιληπτό είναι το γεγονός ότι το μοντέλο CNN έχει πολύ καλύτερες επιδόσεις σε σχέση με τα υπόλοιπα μοντέλα, σε όλα τα πειράματα που έχουν γίνει χρησιμοποιώντας ένα-ένα τα χαρακτηριστικά. Μάλιστα, μερικές μετρήσεις που λήφθηκαν από συγκεκριμένα χαρακτηριστικά, οδήγησαν ορισμένα μοντέλα στο να έχουν επιδόσεις της τάξης του τυχαίου ταξινομητή, οπότε ίσως αυτά τα χαρακτηριστικά να μην περιέχουν πολλή πληροφορία για να βελτιώσει τις επιδόσεις των μοντέλων της τρέχουσας εργασίας. Παρατηρείται επίσης ότι το μοντέλο LSTM έχει πολύ χαμηλές επιδόσεις με όλα τα χαρακτηριστικά. Για αυτό το λόγο δε χρησιμοποιήθηκε στα παρακάτω πειράματα βελτιστοποίησης.

Έχοντας υπόψιν τα παραπάνω και τις όσες τεχνικές αναφέρθηκαν στο προηγούμενο κεφάλαιο για καλύτερα αποτελέσματα, χρησιμοποιήθηκαν σε πρώτο στάδιο όλα τα χαρακτηριστικά που εξάχθηκαν. Για να επιτευχθεί όμως αυτό έπρεπε όλα τα χαρακτηριστικά να έχουν τις ίδιες διαστάσεις, ούτως ώστε να μπορούν να περαστούν στο εκάστοτε μοντέλο για εκπαίδευση. Για αυτό το λόγο έχει γίνει χρήση της τεχνικής του παραγεμίσματος (Padding) ούτως ώστε να έχουν όλα τα δεδομένα τις ίδιες διαστάσεις. Έγινε χρήση 3 τεχνικών παραγεμίσματος, η προσθήκη του μέσου όρου όλων των δεδομένων του χαρακτηριστικού όσες φορές χρειαζόταν, η επανάληψη των ήδη υπάρχοντών τιμών μια-μια αναμεταξύ τους και τέλος η προσθήκη του μέσου όρου ανά δύο των τιμών αναμεταξύ τους.

Δοκιμάζοντας τα παραπάνω, πάρθηκαν τα παρακάτω αποτελέσματα:
Χαρακτηριστικά που χρησιμοποιήθηκαν από τα ηχητικά αρχεία του συνόλου δεδομένων: Chroma Features, MFCC (12 components), μαζί με τα χαρακτηριστικά Chroma Features και MFCC (12 components) από του κάθε κομματιού.
Αριθμός εποχών: 60
Βελτιστοποιητής: Adam (learning_rate = 0,001)
Batch Size: 32

Συνελικτικό Νευρωνικό Δίκτυο:

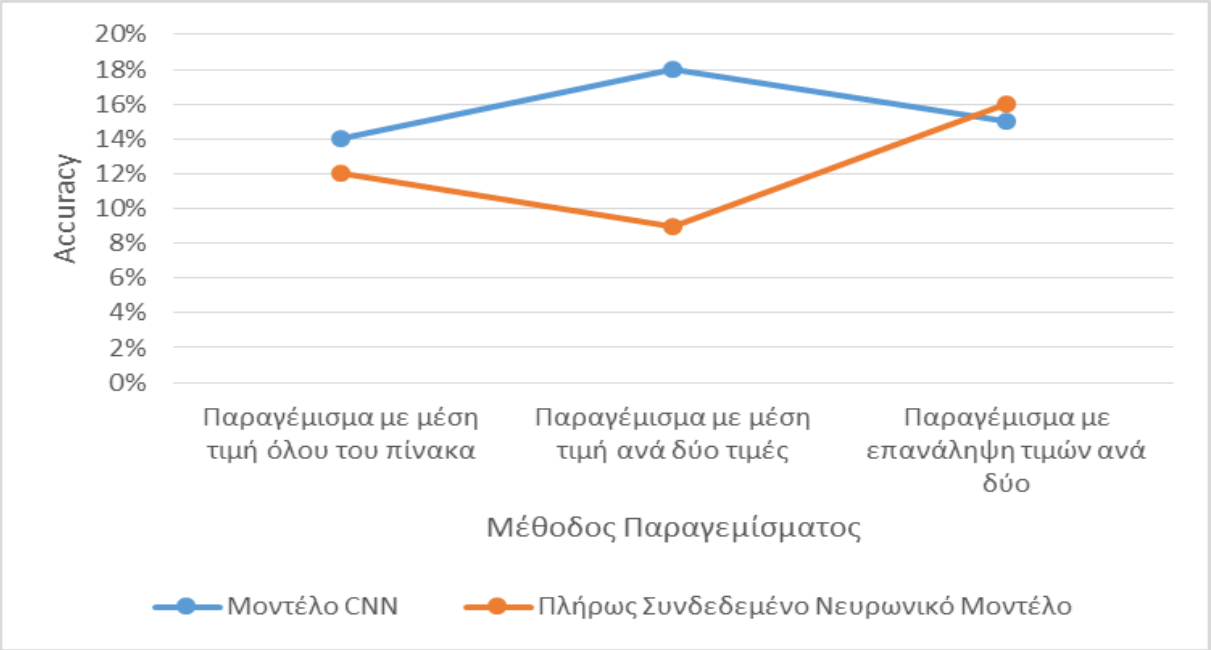
	Παραγέμισμα με μέση τιμή όλου του πίνακα	**Παραγέμισμα με μέση τιμή ανά δύο τιμές**	**Παραγέμισμα με επανάληψη τιμών ανά δύο**
Accuracy	14%	18%	15%
Loss	2.12	2.32	2.02

3.2.3 Πίνακας επιδόσεων Συνελικτικού Νευρωνικού Μοντέλου χρησιμοποιώντας όλα τα χαρακτηριστικά και της τεχνικής παραγεμίσματος

Πλήρως Συνδεδεμένο Νευρωνικό Μοντέλο:

	Παραγέμισμα με μέση τιμή όλου του πίνακα	**Παραγέμισμα με μέση τιμή ανά δύο τιμές**	**Παραγέμισμα με επανάληψη τιμών ανά δύο**
Accuracy	12%	9%	16%
Loss	2.01	2.09	2.16

3.2.4 Πίνακας επιδόσεων Πλήρως Συνδεδεμένου μοντέλου χρησιμοποιώντας όλα τα χαρακτηριστικά και της τεχνικής παραγεμίσματος



3.2.2 Γραφική παράσταση Ακρίβεια-Μέθοδος Παραγεμίσματος για τα 2 μοντέλα που δοκιμάστηκαν, Πλήρως Συνδεδεμένο Νευρωνικό Μοντέλο και Συνελικτικό Νευρωνικό Μοντέλο

Παρατηρώντας τα παραπάνω αποτελέσματα, φαίνεται ότι η τεχνική του παραγεμίσματος δεν επέφερε καλά αποτελέσματα, παρόλο που ουσιαστικά παρέχεται περισσότερη πληροφορία στο σύστημα, άρα αναμένεται περισσότερη ακρίβεια. Εν τούτοις, παρατηρείται ραγδαία πτώση της και αυτό μπορεί να οφείλεται στο γεγονός ότι τα 5 επιπρόσθετα χαρακτηριστικά που περνιούνται στο σύστημα (Spectral Centroid, Spectral Roll-Off, Spectral Bandwidth, Zero Crossing Rate, Root Mean Square Error) δεν προσδίδουν αρκετή πληροφορία, όπως φάνηκε και στο προηγούμενο πείραμα, οπόταν μειώνουν τη βαρύτητα των 2 ουσιαστικών χαρακτηριστικών (MFCC και Chroma Features) που προσδίδουν σημαντική πληροφορία ταξινόμησης.

Ως εκ τούτου θα χρησιμοποιηθούν τα χαρακτηριστικά Chroma και MFCC(12 components) τα οποία έχουν τις ίδιες διαστάσεις και μπορούν να χρησιμοποιηθούν αυτούσια, αλλά και επειδή προσδίδουν την περισσότερη πληροφορία που χρειάζεται αυτή η εργασία όπως φάνηκε σε προηγούμενο πείραμα. Θα χρησιμοποιηθούν αυτά τα 2 χαρακτηριστικά τόσο στην εξαγωγή πληροφορίας από το κομμάτι κάθε 20 δευτερολέπτων, όσο και στην εξαγωγή πληροφορίας από το τέλος του κάθε ηχητικού κομματιού. Θα χρησιμοποιηθούν ένα-ένα ξεχωριστά, αλλά και μαζί για να δούμε τις επιδόσεις τους.

Τα χαρακτηριστικά που χρησιμοποιήθηκαν δεν συμπεριλάμβαναν πληροφορία από το τέλος του κάθε κομματιού.
Αριθμός εποχών: 60
Βελτιστοποιητής: Adam (learning_rate = 0,001)
Batch Size: 32

Συνελικτικό Νευρωνικό Δίκτυο:

	Chroma Features	MFCC(12 components)	Chroma Features + MFCC (12 components)
Accuracy	67%	72%	75%
Loss	3.21	2.41	1.99

3.2.5 Πίνακας επιδόσεων Συνελικτικού Νευρωνικού Μοντέλου

Πλήρως Συνδεδεμένο Νευρωνικό Δίκτυο:

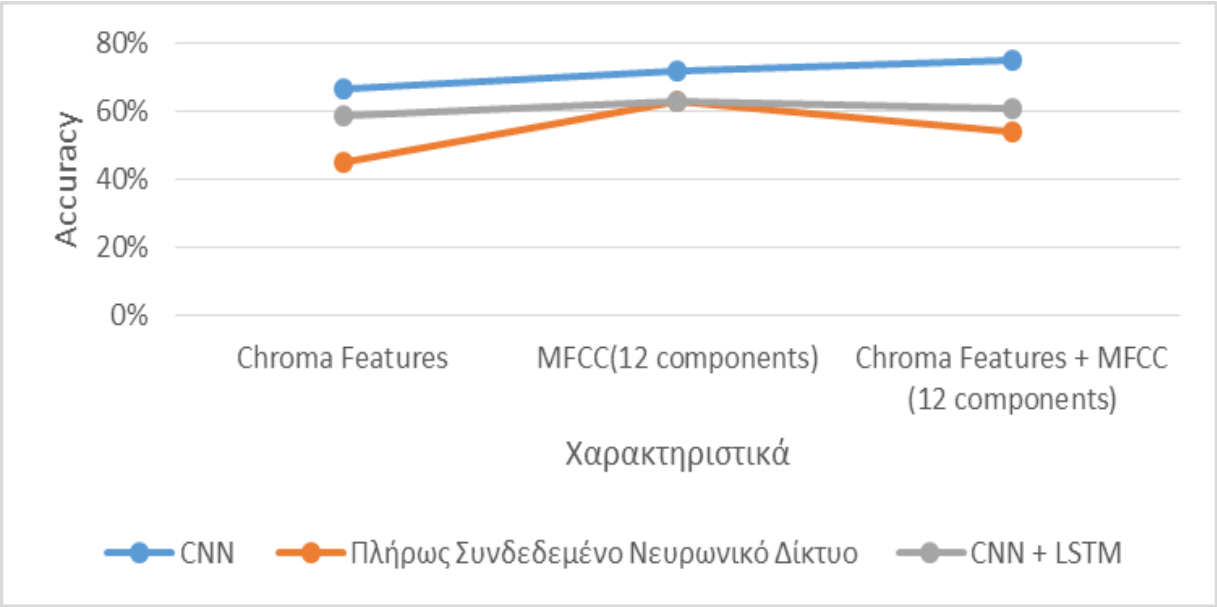
	Chroma Features	MFCC(12 components)	Chroma Features + MFCC (12 components)
Accuracy	45%	63%	54%
Loss	6.81	8.89	4.53

3.2.6 Πίνακας επιδόσεων Πλήρως Συνδεδεμένου Μοντέλου

CNN + LSTM Δίκτυο:

	Chroma Features	MFCC(12 components)	Chroma Features + MFCC (12 components)
Accuracy	59%	63%	61%
Loss	2.43	1.71	1.85

3.2.7 Πίνακας επιδόσεων CNN + LSTM Μοντέλου



3.2.3 Γραφική παράσταση Ακρίβεια-Χαρακτηριστικά για τα 3 μοντέλα που δοκιμάστηκαν

Από τα παραπάνω στοιχεία είναι φανερό ότι τα καλύτερα αποτελέσματα βγαίνουν όταν χρησιμοποιηθούν και τα 2 χαρακτηριστικά μαζί για το CNN μοντέλο, ενώ για τα υπόλοιπα 2 όταν χρησιμοποιήθηκαν μόνο τα MFCC.
Στη συνέχεια θα πραγματοποιηθούν πειράματα για την εύρεση των βέλτιστων παραμέτρων που μεγιστοποιούν την ακρίβεια του συστήματος. Αυτά περιλαμβάνουν την έυρεση του βέλτιστου Batch Size, του βέλτιστου αριθμού εποχών εκπαίδευσης, του βέλτιστου Βελτιστοποιητή, καθώς επίσης και του βέλτιστου αριθμού κρυφών επιπέδων, μαζί με τις παραμέτρους τους.

3.2.1 Βελτιστοποίηση Batch Size

Για τη βελτιστοποίηση του batch size έχουν γίνει οι παρακάτω δοκιμές:
Χαρακτηριστικά που χρησιμοποιήθηκαν από τα ηχητικά αρχεία του συνόλου δεδομένων: Chroma Features, MFCC (12 components), μαζί με τα χαρακτηριστικά Chroma Features και MFCC (12 components) από το τέλος του κάθε κομματιού.
Αριθμός εποχών: 60
Βελτιστοποιητής: Adam (learning_rate = 0,001)
Μοντέλο: Συνελικτικό Νευρωνικό Μοντέλο

Batch Size	Accuracy	Crossentropy Loss
16	73%	2.56
32	73%	2.31
64	72%	2.71
128	74%	2.36
256	73%	3.15

3.2.1.1 Πίνακας επιδόσεων Συνελικτικού Νευρωνικού Μοντέλου για διάφορες τιμές Batch Size

Χαρακτηριστικά που χρησιμοποιήθηκαν από τα ηχητικά αρχεία του συνόλου δεδομένων: MFCC (12 components)
Αριθμός εποχών: 60
Βελτιστοποιητής: Adam (learning_rate = 0,001)
Μοντέλο: Πλήρως Συνδεδεμένο Νευρωνικό Μοντέλο

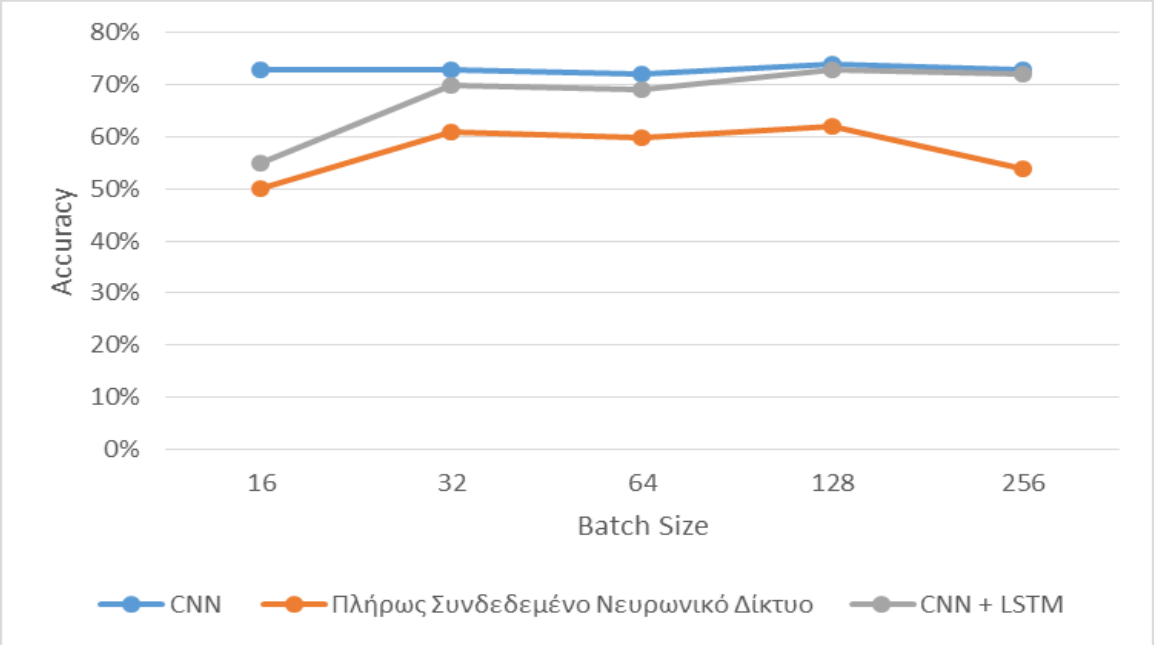
Batch Size	Accuracy	Crossentropy Loss
16	50%	1.93
32	61%	1.82
64	60%	2.71
128	62%	4.53
256	54%	4.12

3.2.1.2 Πίνακας επιδόσεων Πλήρως Συνδεδεμένου Νευρωνικού Μοντέλου για διάφορες τιμές Batch Size

Χαρακτηριστικά που χρησιμοποιήθηκαν από τα ηχητικά αρχεία του συνόλου δεδομένων:
MFCC (12 components)
Αριθμός εποχών: 60
Βελτιστοποιητής: Adam (learning_rate = 0,001)
Μοντέλο: CNN + LSTM Μοντέλο

Batch Size	Accuracy	Crossentropy Loss
16	55%	2.20
32	70%	1.90
64	69%	2.48
128	73%	2.39
256	72%	2.42

3.2.1.3 Πίνακας επιδόσεων CNN + LSTM Μοντέλου για διάφορες τιμές Batch Size



3.2.1.1 Γραφική παράσταση Accuracy-Batch Size

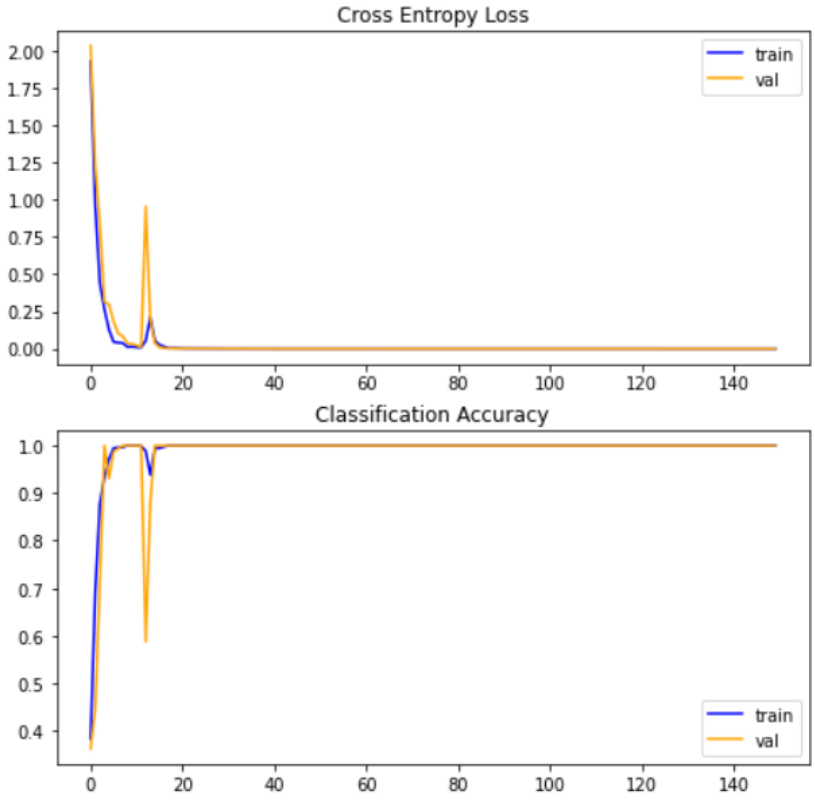
Όπως παρατηρούμε από τα αποτελέσματα των δοκιμών που έχουν γίνει, το καλύτερο αποτέλεσμα έρχεται και στις 3 αρχιτεκτονικές χρησιμοποιώντας Batch Size ίσο με 128, αν και τα αποτελέσματα είναι όλα πολύ κοντινά.

3.2.2 Βελτιστοποίηση Αριθμού εποχών εκπαίδευσης μοντέλου

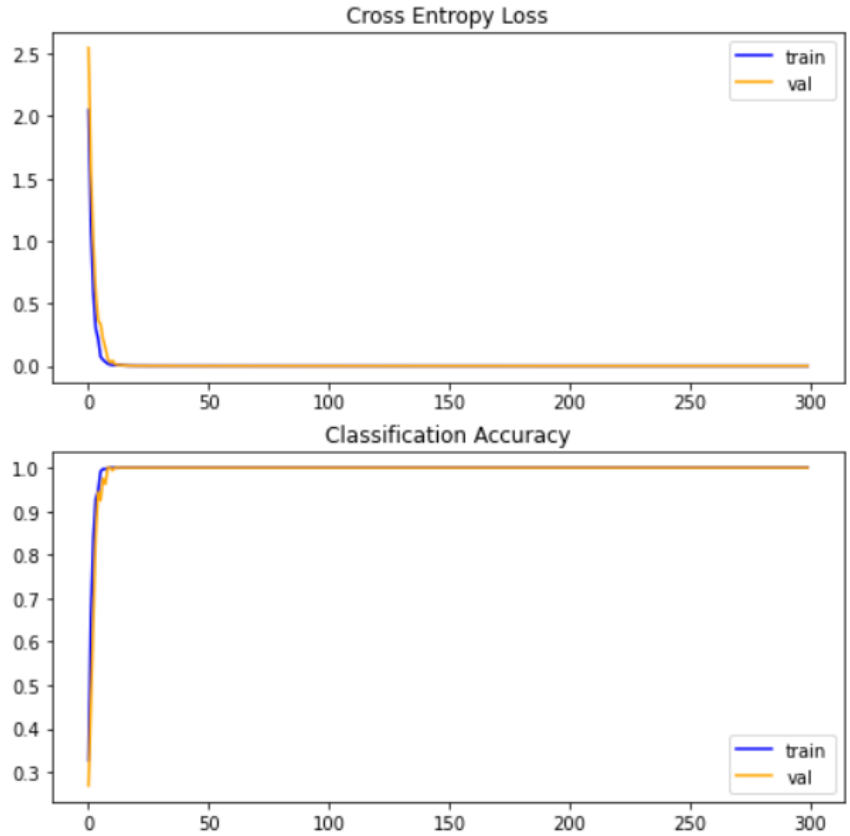
Για τη βελτιστοποίηση του αριθμού των εποχών για τις οποίες θα εκπαιδευτεί το μοντέλο έχουν γίνει οι παρακάτω δοκιμές:
Χαρακτηριστικά που χρησιμοποιήθηκαν από τα ηχητικά αρχεία του συνόλου δεδομένων: Chroma Features, MFCC (12 components), μαζί με τα χαρακτηριστικά Chroma Features και MFCC (12 components) από το τέλος του κάθε κομματιού.
Batch Size: 32
Βελτιστοποιητής: Adam (learning_rate = 0,001)
Μοντέλο: Συνελικτικό Νευρωνικό Μοντέλο

Αριθμός Εποχών	Accuracy	Crossentropy Loss
20	71%	2.14
30	71%	1.60
40	71%	1.80
50	73%	3.08
60	71%	2.34
70	74%	2.99
80	73%	2.07
90	73%	2.62
120	75%	2.01
150	74%	2.95
200	78%	2.47
250	73%	2.73
300	73%	3.21

3.2.2.1 Πίνακας επιδόσεων Συνελικτικού Νευρωνικού Μοντέλου για διάφορες τιμές αριθμού εποχών



3.2.2.1 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών για 150 εποχές

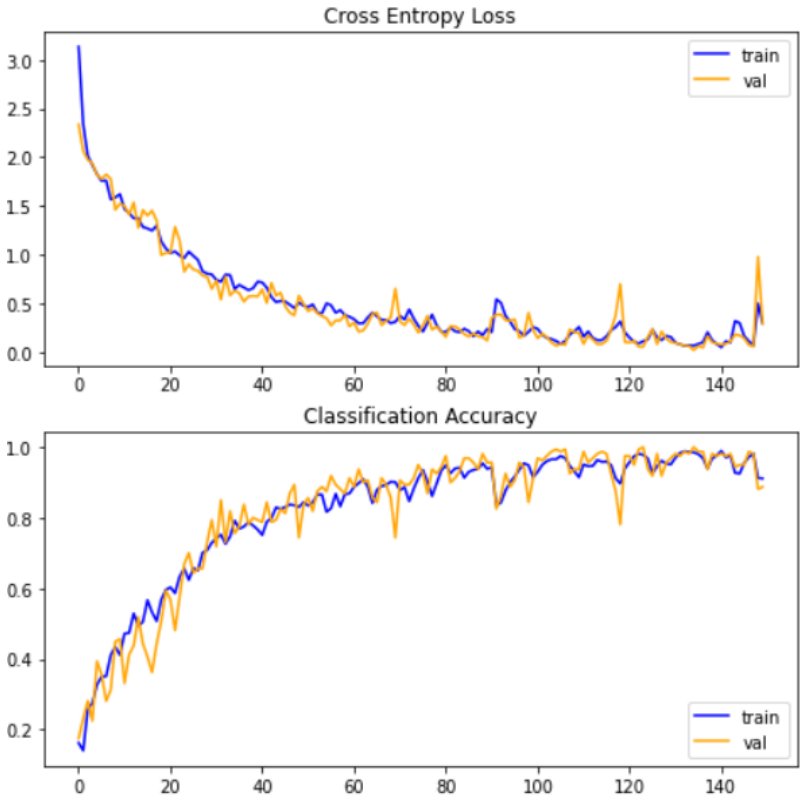


3.2.2.2 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών για 300 εποχές

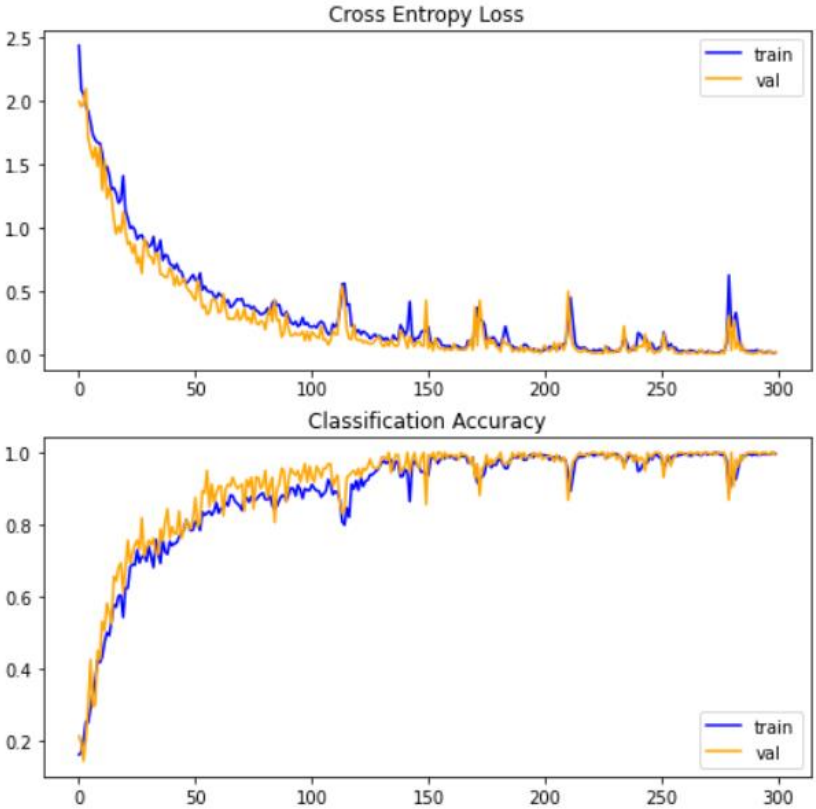
Χαρακτηριστικά που χρησιμοποιήθηκαν από τα ηχητικά αρχεία του συνόλου δεδομένων:
MFCC (12 components)
Batch Size: 32
Βελτιστοποιητής: Adam (learning_rate = 0,0001)
Μοντέλο: Πλήρως Συνδεδεμένο Νευρωνικό Μοντέλο

Αριθμός Εποχών	**Accuracy**	**Crossentropy Loss**
20	42%	2.69
30	55%	1.70
40	51%	2.08
50	56%	2.31
60	61%	2.18
70	57%	3.32
80	57%	3.29
90	54%	3.06
120	62%	4.18
150	58%	3.86
200	61%	4.85
250	63%	3.89
300	63%	6.79

3.2.2.2 Πίνακας επιδόσεων Πλήρως Συνδεδεμένου Νευρωνικού Μοντέλου για διάφορες τιμές αριθμού εποχών



3.2.2.3 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών για 150 εποχές

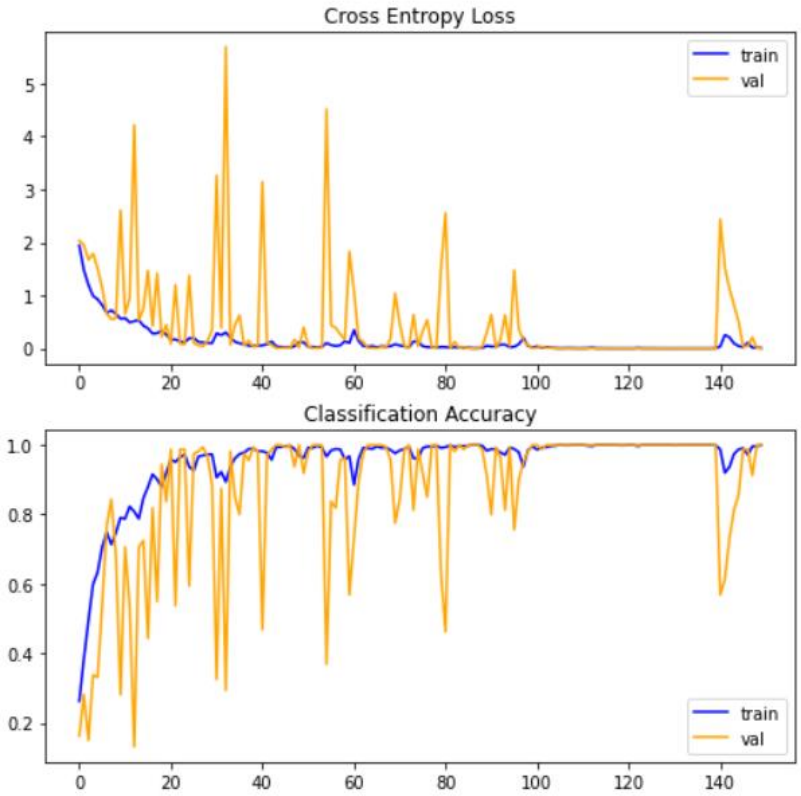


3.2.2.4Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών για 300 εποχές

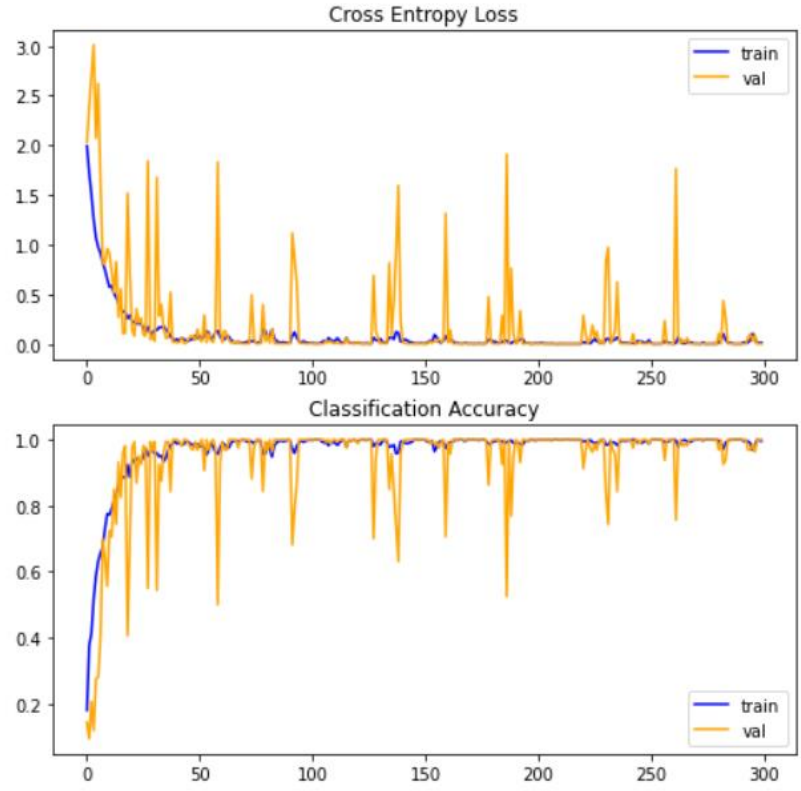
Χαρακτηριστικά που χρησιμοποιήθηκαν από τα ηχητικά αρχεία του συνόλου δεδομένων: MFCC (12 components), μαζί με τα χαρακτηριστικά MFCC (12 components) από το τέλος του κάθε κομματιού.
Batch Size: 32
Βελτιστοποιητής: Adam (learning_rate = 0,001)
Μοντέλο: CNN + LSTM Μοντέλο

Αριθμός Εποχών	**Accuracy**	**Crossentropy Loss**
20	31%	3.61
30	65%	1.58
40	54%	2.97
50	66%	2.18
60	70%	1.90
70	49%	3.99
80	76%	2.22
90	73%	1.77
120	69%	2.58
150	73%	2.26
200	63%	2.70
250	69%	2.97
300	68%	2.61

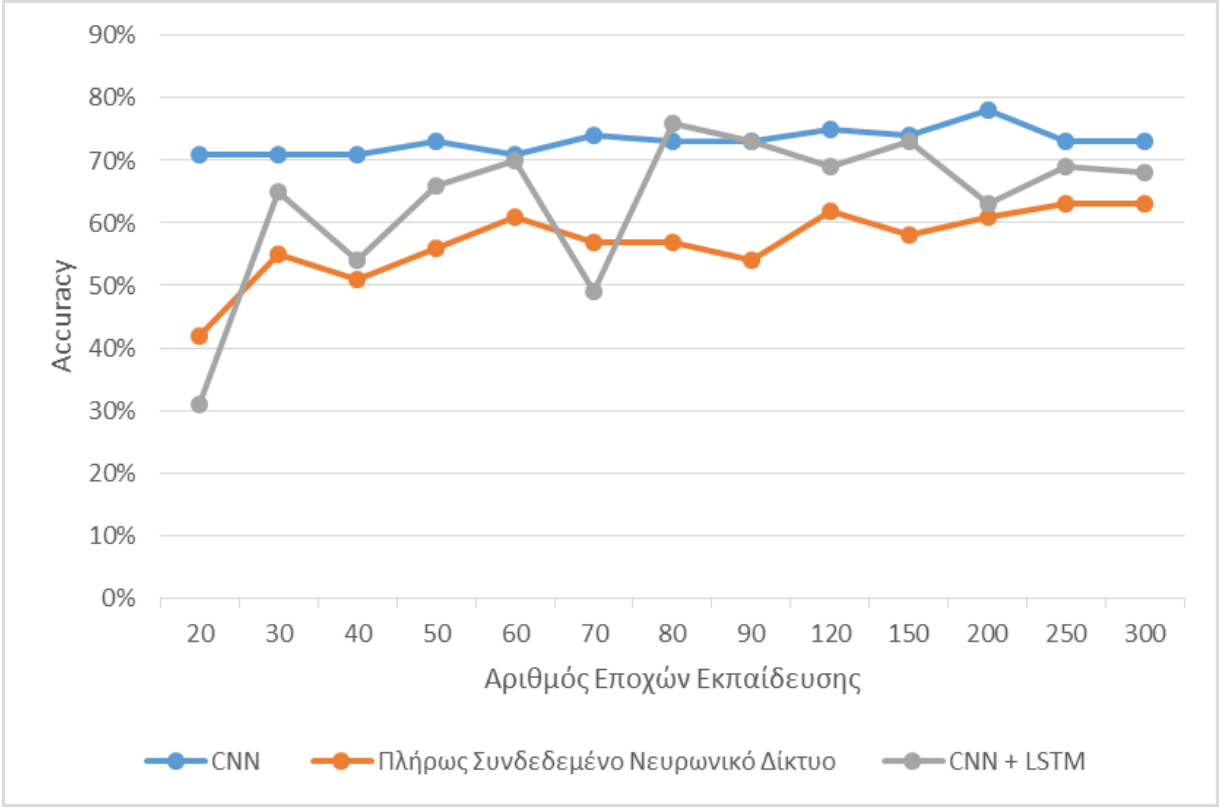
3.2.2.3 Πίνακας επιδόσεων CNN + LSTM Μοντέλου για διάφορες τιμές αριθμού εποχών



3.2.2.5 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών για 150 εποχές



3.2.2.6 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών για 300 εποχές



3.2.2.7 Γραφική παράσταση εκπαίδευσης μοντέλων Accuracy-Αριθμός εποχών εκπαίδευσης

Από τις παραπάνω δοκιμές, παρατηρούμε ότι η πιο υψηλή ακρίβεια στο Συνελικτικό Νευρωνικό Δίκτυο επετεύχθη όταν το μοντέλο εκπαιδεύτηκε για 200 εποχές, ενώ για λιγότερες και περισσότερες από 200 εποχές εκπαίδευσης έχουμε χαμηλότερη ακρίβεια ύψους 5%. Παρατηρούμε επίσης ότι η πιο υψηλή ακρίβεια στο Πλήρως Συνδεδεμένο Νευρωνικό Δίκτυο επετεύχθη όταν το μοντέλο εκπαιδεύτηκε για 250 και 300 εποχές, ενώ για πολύ λίγες εποχές εκπαίδευσης έχουμε χαμηλότερη ακρίβεια ύψους έως 20%. Τέλος, παρατηρούμε ότι η πιο υψηλή ακρίβεια στο CNN + LSTM Δίκτυο επετεύχθη όταν το μοντέλο εκπαιδεύτηκε για 80 εποχές, ενώ για 20 και 70 εποχές εκπαίδευσης έχουμε τις χαμηλότερες επιδόσεις, με πτώση ακρίβειας της τάξης του 30%. Παρόλο που από τις γραφικές παραστάσεις Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών μπορούμε να πούμε ότι τα μοντέλα υπερεκπαιδεύονται ήδη από πολύ μικρό αριθμό εποχών εκπαίδευσης, εν τούτοις από τα αποτελέσματα είναι φανερό ότι με την αύξηση του αριθμού των εποχών εκπαίδευσης, αυξάνεται και η ακρίβεια του μοντέλου.

3.2.3 Βελτιστοποίηση Βελτιστοποιητή και Ρυθμού Μάθησης

Για τη βελτιστοποίηση του βελτιστοποιητή και του ρυθμού μάθησης του μοντέλου έχουν γίνει οι παρακάτω δοκιμές:
Χαρακτηριστικά που χρησιμοποιήθηκαν από τα ηχητικά αρχεία του συνόλου δεδομένων: Chroma Features, MFCC (12 components), μαζί με τα χαρακτηριστικά Chroma Features και MFCC (12 components) από το τέλος του κάθε κομματιού.
Batch Size: 32
Αριθμός εποχών: 60
Μοντέλο: Συνελικτικό Νευρωνικό Μοντέλο

Βελτιστοποιητής	Accuracy	Crossentropy Loss
Adam(learning_rate = 0,0001)	70%	1.85
Adam(learning_rate = 0,0005)	70%	2.00
Adam(learning_rate = 0,001)	75%	2.14
Adam(learning_rate = 0,005)	72%	2.09
Adam(learning_rate = 0,01)	73%	1.91
Adam(learning_rate = 0,02)	72%	2.78

3.2.3.1 Πίνακας επιδόσεων Συνελικτικού Νευρωνικού Μοντέλου για διάφορους βελτιστοποιητές με διάφορους ρυθμούς μάθησης

Χαρακτηριστικά που χρησιμοποιήθηκαν από τα ηχητικά αρχεία του συνόλου δεδομένων: Chroma Features, MFCC (12 components)
Batch Size: 32
Αριθμός εποχών: 60
Μοντέλο: Πλήρως Συνδεδεμένο Νευρωνικό Μοντέλο

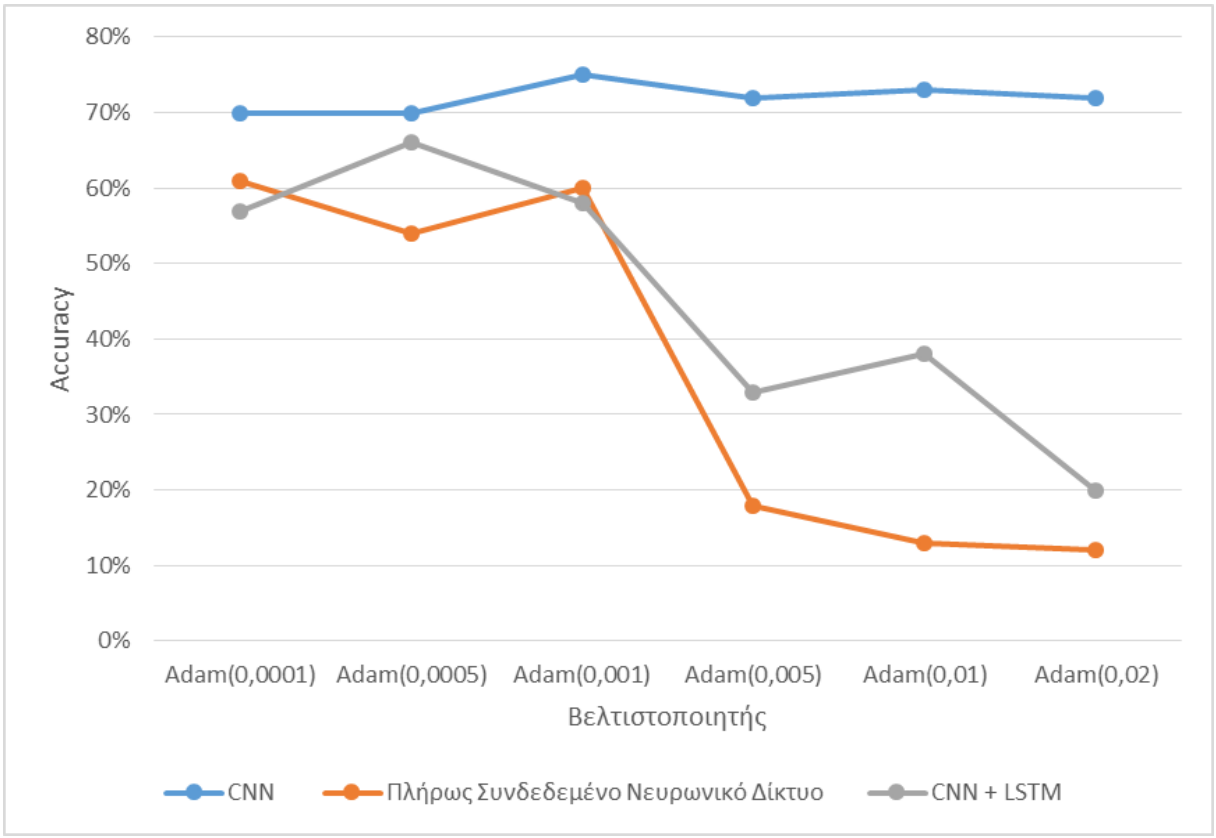
Βελτιστοποιητής	Accuracy	Crossentropy Loss
Adam(learning_rate = 0,0001)	61%	2.33
Adam(learning_rate = 0,0005)	54%	1.53
Adam(learning_rate = 0,001)	60%	2.43
Adam(learning_rate = 0,005)	18%	1.99
Adam(learning_rate = 0,01)	13%	2.16
Adam(learning_rate = 0,02)	12%	2.16

3.2.3.2 Πίνακας επιδόσεων Πλήρως Συνδεδεμένου Νευρωνικού Μοντέλου για διάφορους βελτιστοποιητές με διάφορους ρυθμούς μάθησης

Χαρακτηριστικά που χρησιμοποιήθηκαν από τα ηχητικά αρχεία του συνόλου δεδομένων: MFCC (12 components), μαζί με τα χαρακτηριστικά MFCC (12 components) από το τέλος του κάθε κομματιού.
Batch Size: 32
Αριθμός εποχών: 60
Μοντέλο CNN + LSTM Μοντέλο

Βελτιστοποιητής	Accuracy	Crossentropy Loss
Adam(learning_rate = 0,0001)	57%	2.11
Adam(learning_rate = 0,0005)	66%	2.78
Adam(learning_rate = 0,001)	58%	3.19
Adam(learning_rate = 0,005)	33%	3.23
Adam(learning_rate = 0,01)	38%	2.23
Adam(learning_rate = 0,02)	20%	2.41

3.2.3.1 Πίνακας επιδόσεων CNN + LSTM Μοντέλου για διάφορους βελτιστοποιητές με διάφορους ρυθμούς μάθησης



3.2.3.1 Γραφική παράσταση Accuracy- Optimizer

Όπως παρατηρούμε από τα αποτελέσματα των δοκιμών που έχουν γίνει, το καλύτερο αποτέλεσμα έρχεται στο Συνελικτικό Νευρωνικό Δίκτυο χρησιμοποιώντας το βελτιστοποιητή Adam με ρυθμό μάθησης 0.001, ενώ στο Πλήρως Συνδεδεμένο Νευρωνικό Δίκτυο χρησιμοποιώντας το βελτιστοποιητή Adam με ρυθμό μάθησης 0.0001 και στο Δίκτυο CNN + LSTM χρησιμοποιώντας το βελτιστοποιητή Adam με ρυθμό μάθησης 0.0005. Στο Συνελικτικό Νευρωνικό Δίκτυο παρατηρούμε ότι υπάρχει ανοχή σε διάφορες τιμές του ρυθμού μάθησης, σε αντίθεση με το Πλήρως Συνδεδεμένο Νευρωνικό Δίκτυο και το δίκτυο CNN + LSTM, όπου για τιμές μάθησης μικρότερες από 0.001 έχουμε ραγδαία πτώση της επίδοσης.

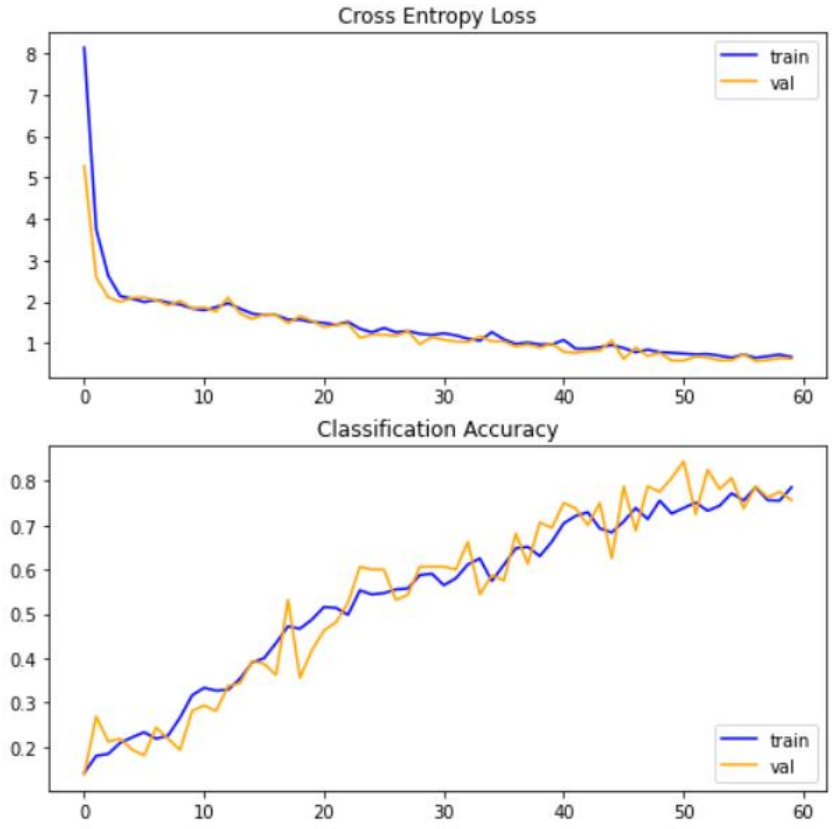
3.2.4 Βελτιστοποίηση Αριθμού Κρυφών Επιπέδων και Αριθμού Νευρώνων κάθε Κρυφού Επιπέδου

Η διαδικασία βελτιστοποίησης του αριθμού των κρυφών επιπέδων και του αριθμού των νευρώνων του κάθε κρυφού επιπέδου έγινε ξεχωριστά για τις δυο αρχιτεκτονικές που θα δοκιμαστούν, την αρχιτεκτονική του Πλήρως Συνδεδεμένου Νευρωνικού Δικτύου και την αρχιτεκτονική του Συνελικτικού Νευρωνικού Δικτύου.

Δοκιμές στο Πλήρως Συνδεδεμένο Νευρωνικό Δίκτυο

Σε όλες τις δοκιμές που έχουν γίνει με το Πλήρως Συνδεδεμένο Νευρωνικό Δίκτυο, τα χαρακτηριστικά που περάστηκαν σαν είσοδο είναι τα MFCC (12 components) από το κάθε κλιπ 20 δευτερολέπτων. Το Batch Size είναι 32, ο βελτιστοποιητής είναι ο Adam με ρυθμό μάθησης learning_rate = 0.0001 και ο αριθμός εποχών εκπαίδευσης είναι 60.

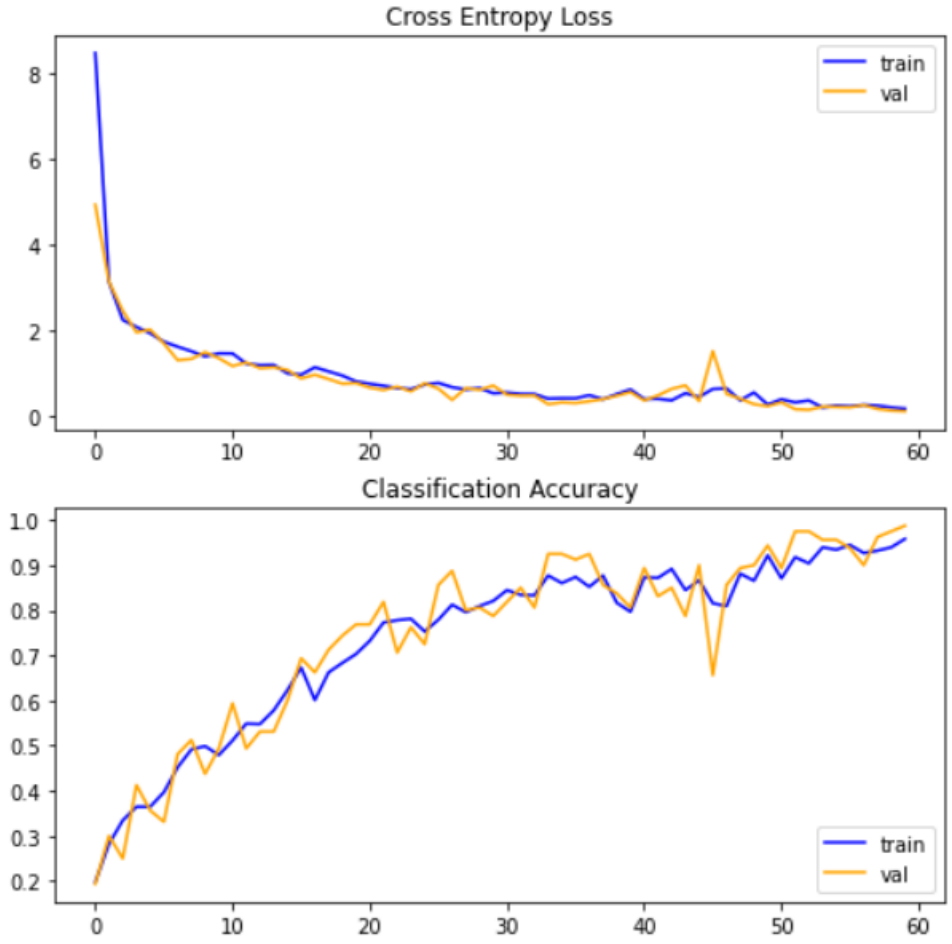
Επίπεδο	Παράμετροι
Flatten	-
Dense	Units = 128, activation = relu
Dense	Units = 64, activation = relu
Dense	Units = 32, activation = relu
Dense	Units = 16, activation = relu
Dense	Units = 8, activation = softmax



3.2.4.1 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών

Απόδοση στο Test Dataset:
Accuracy: 49%
Crossentropy Loss: 2.29

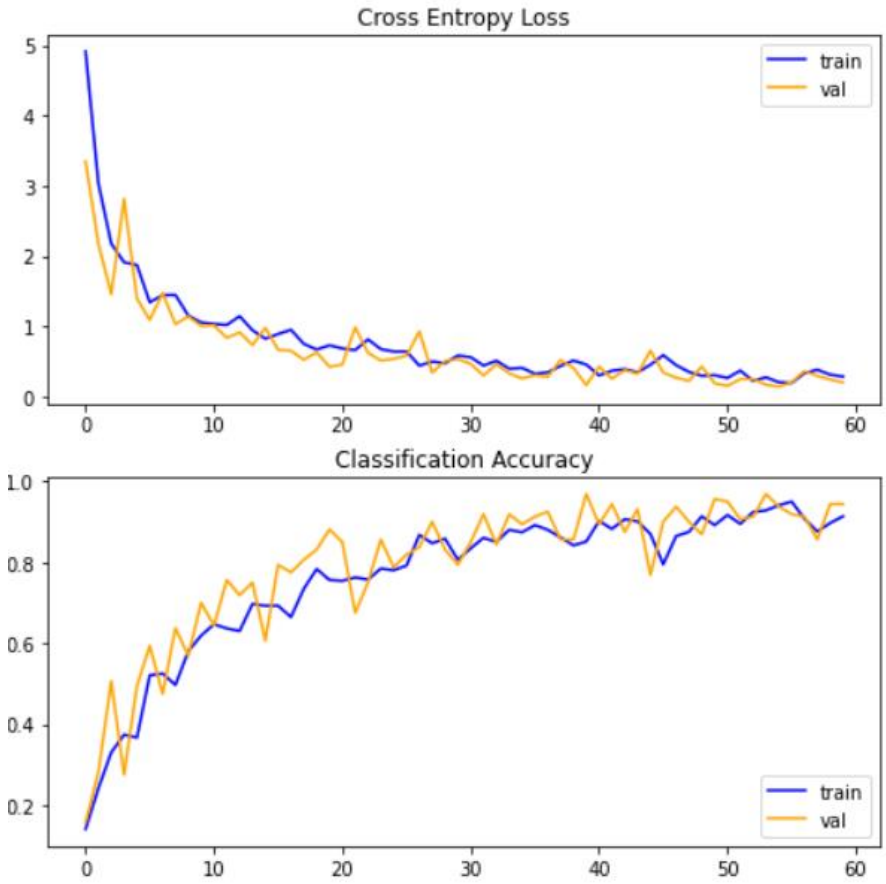
Επίπεδο	Παράμετροι
Flatten	-
Dense	Units = 512, activation = relu
Dense	Units = 256, activation = relu
Dense	Units = 128, activation = relu
Dense	Units = 64, activation = relu
Dense	Units = 32, activation = relu
Dense	Units = 16, activation = relu
Dense	Units = 8, activation = softmax



3.2.4.2 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών

Απόδοση στο Test Dataset:
Accuracy: 59%
Crossentropy Loss: 2.69

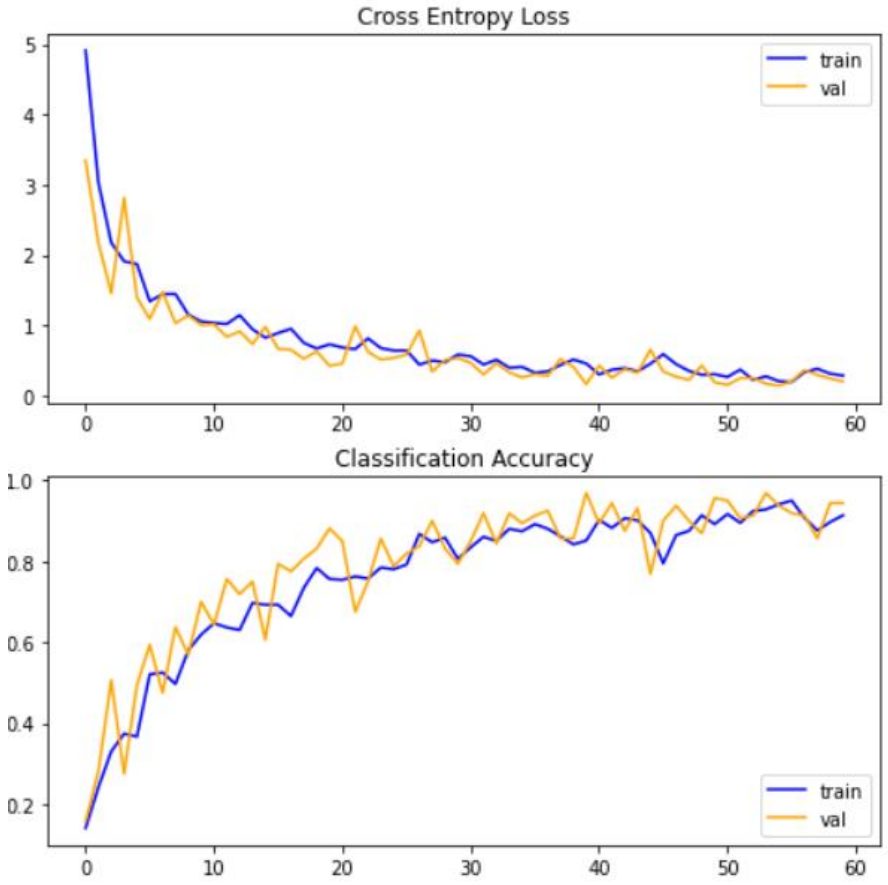
Επίπεδο	Παράμετροι
Flatten	-
Dense	Units = 1024, activation = relu
Dense	Units = 512, activation = relu
Dense	Units = 256, activation = relu
Dense	Units = 128, activation = relu
Dense	Units = 64, activation = relu
Dense	Units = 32, activation = relu
Dense	Units = 16, activation = relu
Dense	Units = 8, activation = softmax



3.2.4.3 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών

Απόδοση στο Test Dataset:
Accuracy: 65%
Crossentropy Loss: 2.70

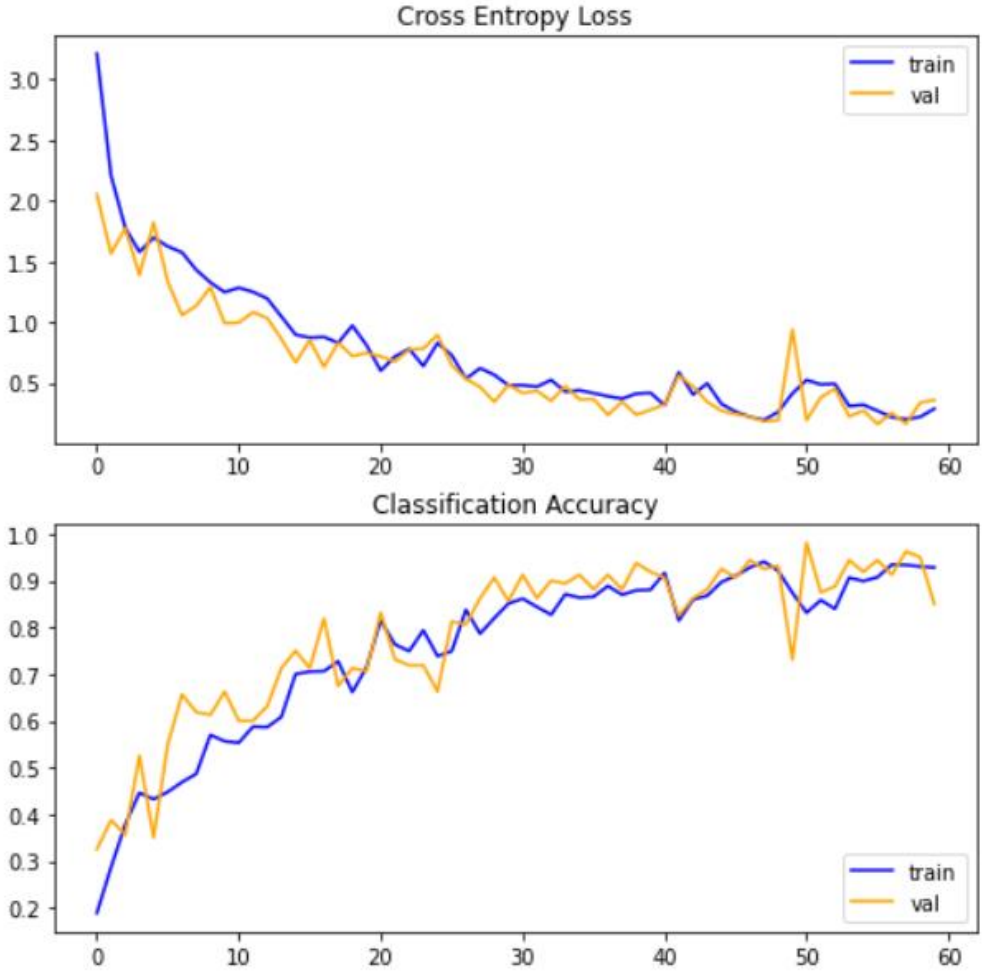
Επίπεδο	Παράμετροι
Flatten	-
Dense	Units = 1024, activation = relu
Dense	Units = 512, activation = relu
Dropout	Rate = 0.3
Dense	Units = 256, activation = relu
Dense	Units = 128, activation = relu
Dense	Units = 64, activation = relu
Dense	Units = 32, activation = relu
Dense	Units = 16, activation = relu
Dense	Units = 8, activation = softmax



3.2.4.4 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών

Απόδοση στο Test Dataset:
Accuracy: 52%
Crossentropy Loss: 1.58

Επίπεδο	Παράμετροι
Flatten	-
Dense	Units = 1024, activation = relu
Dense	Units = 512, activation = relu
Dense	Units = 512, activation = relu
Dense	Units = 256, activation = relu
Dense	Units = 128, activation = relu
Dense	Units = 64, activation = relu
Dense	Units = 32, activation = relu
Dense	Units = 16, activation = relu
Dense	Units = 8, activation = softmax



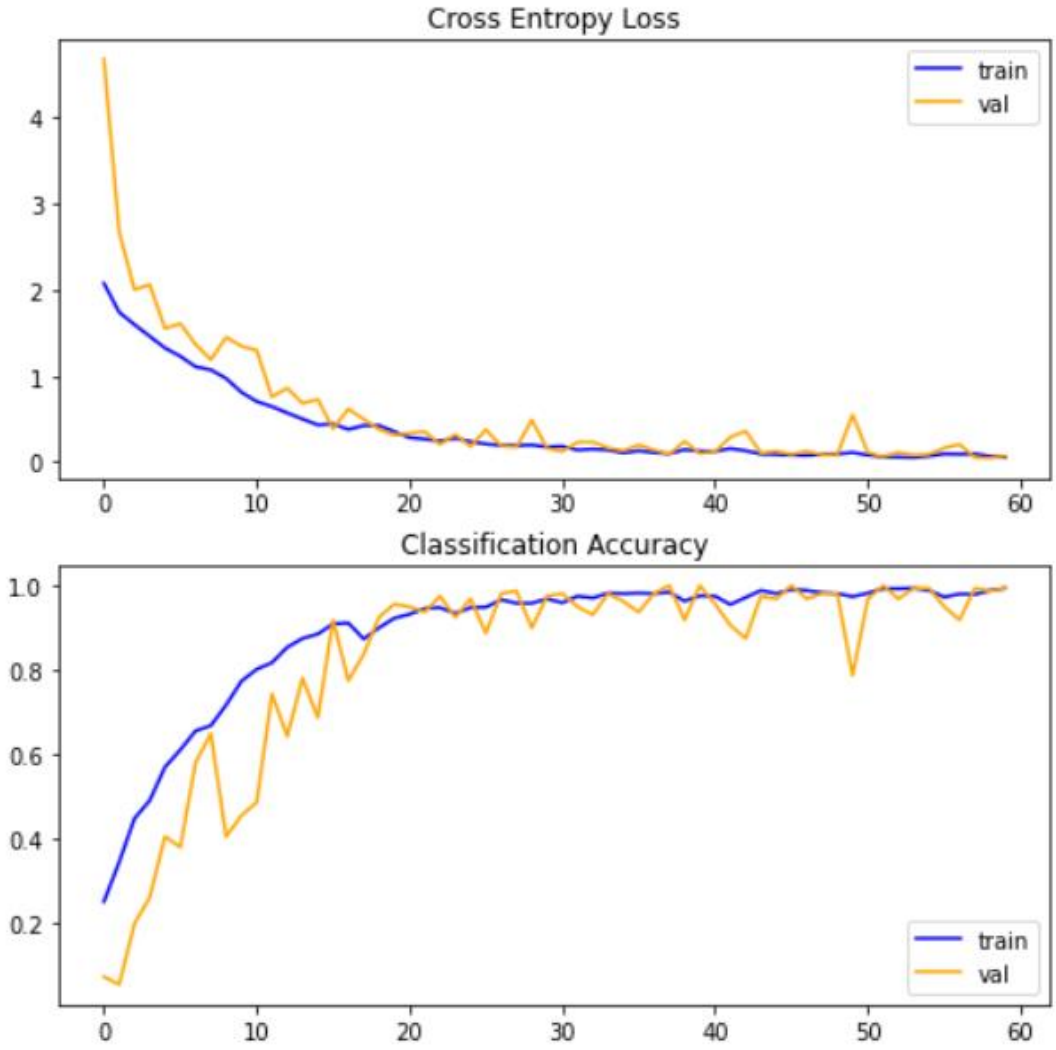
3.2.4.5 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών

Απόδοση στο Test Dataset:
Accuracy: 65%
Crossentropy Loss: 3.05

Δοκιμές στο Συνελικτικό Νευρωνικό Δίκτυο

Σε όλες τις δοκιμές που έχουν γίνει με το Συνελικτικό Νευρωνικό Δίκτυο, τα χαρακτηριστικά που περάστηκαν σαν είσοδο είναι τα Chroma Features και MFCC (12 components) από το κάθε κλιπ 20 δευτερολέπτων, μαζί με τα χαρακτηριστικά Chroma Features και MFCC (12 components) από το τέλος του κάθε κομματιού. Το Batch Size είναι 32, ο βελτιστοποιητής είναι ο Adam με ρυθμό μάθησης learning_rate = 0.001 και ο αριθμός εποχών εκπαίδευσης είναι 60.

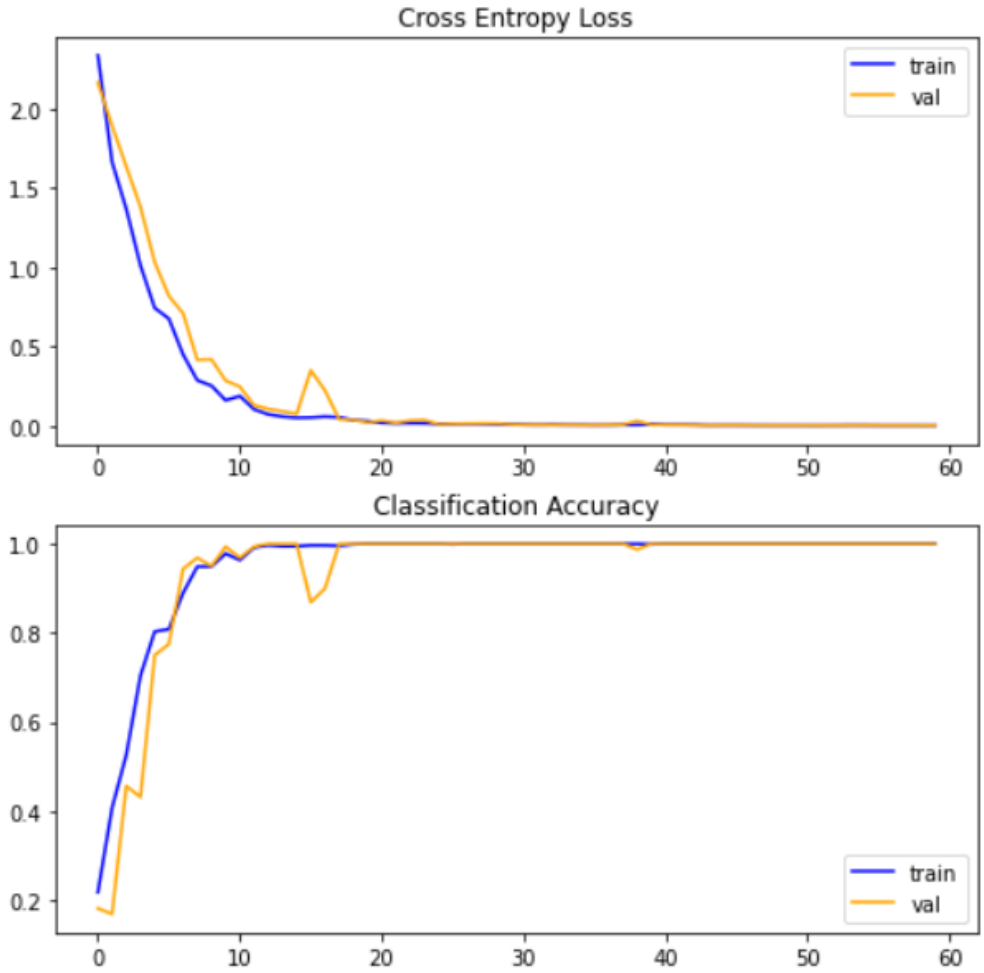
Επίπεδο	Παράμετροι
Conv2D	Filters = 24, kernel_size=(3,3), padding = same
BatchNormalization	-
Activation	relu
MaxPooling2D	Pool_size = (2,2)
Conv2D	Filters = 32, kernel_size=(3,3), padding = same
BatchNormalization	-
Activation	relu
GlobalMaxPooling2D	-
Dense	Units = 32
Dense	Units = 8, activation = softmax



3.2.4.6 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών

Απόδοση στο Test Dataset:
Accuracy: 63%
Crossentropy Loss: 3.02

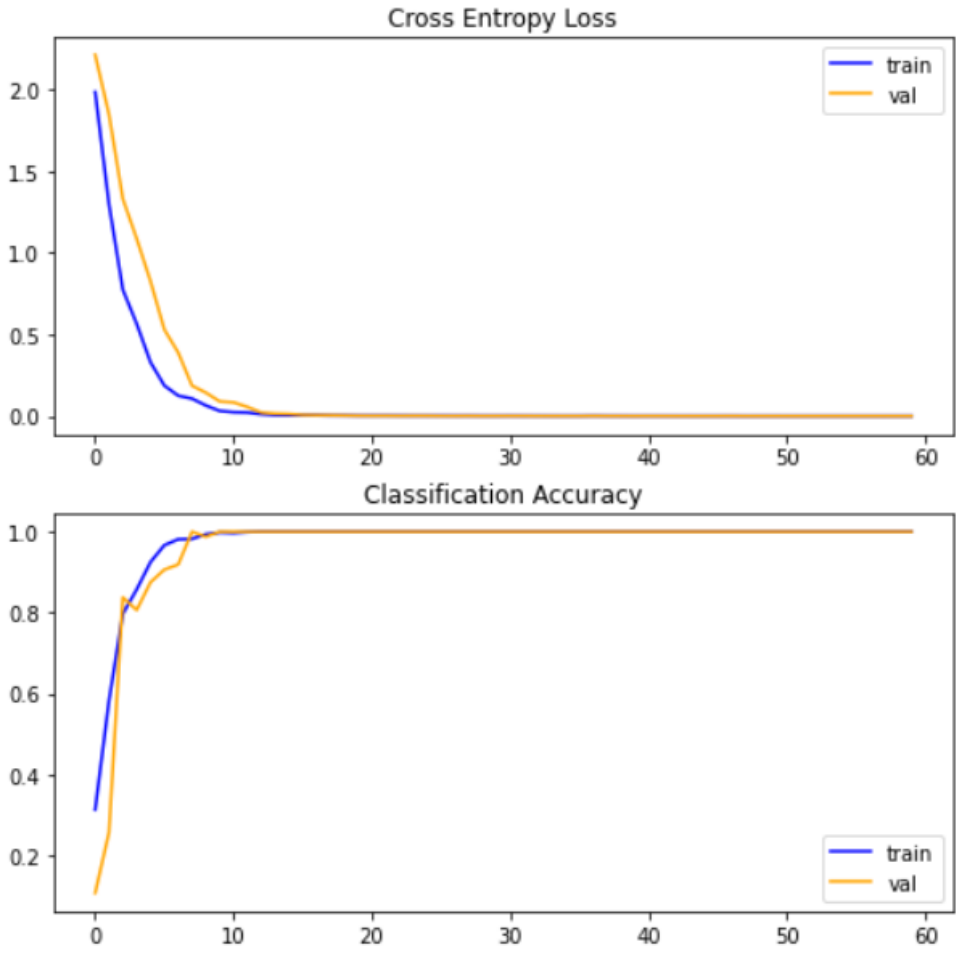
Επίπεδο	Παράμετροι
Conv2D	Filters = 24, kernel_size=(3,3), padding = same
BatchNormalization	-
Activation	relu
MaxPooling2D	Pool_size = (2,2)
Conv2D	Filters = 32, kernel_size=(3,3), padding = same
BatchNormalization	-
Activation	relu
MaxPooling2D	Pool_size = (2,2)
Conv2D	Filters = 64, kernel_size=(3,3), padding = same
BatchNormalization	-
Activation	relu
GlobalMaxPooling2D	-
Dense	Units = 64
Dense	Units = 8, activation = softmax



3.2.4.7 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών

Απόδοση στο Test Dataset:
Accuracy: 72%
Crossentropy Loss: 2.20

Επίπεδο	Παράμετροι
Conv2D	Filters = 24, kernel_size=(3,3), padding = same
BatchNormalization	-
Activation	relu
MaxPooling2D	Pool_size = (2,2)
Conv2D	Filters = 32, kernel_size=(3,3), padding = same
BatchNormalization	-
Activation	relu
MaxPooling2D	Pool_size = (2,2)
Conv2D	Filters = 64, kernel_size=(3,3), padding = same
BatchNormalization	-
Activation	relu
MaxPooling2D	Pool_size = (2,2)
Conv2D	Filters = 128, kernel_size=(3,3), padding = same
BatchNormalization	-
Activation	relu
GlobalMaxPooling2D	-
Dense	Units = 128
Dense	Units = 8, activation = softmax



3.2.4.8 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών

Απόδοση στο Test Dataset:
Accuracy: 77%
Crossentropy Loss: 1.73

3.2.5 Χρήση τεχνικής Cross Validation

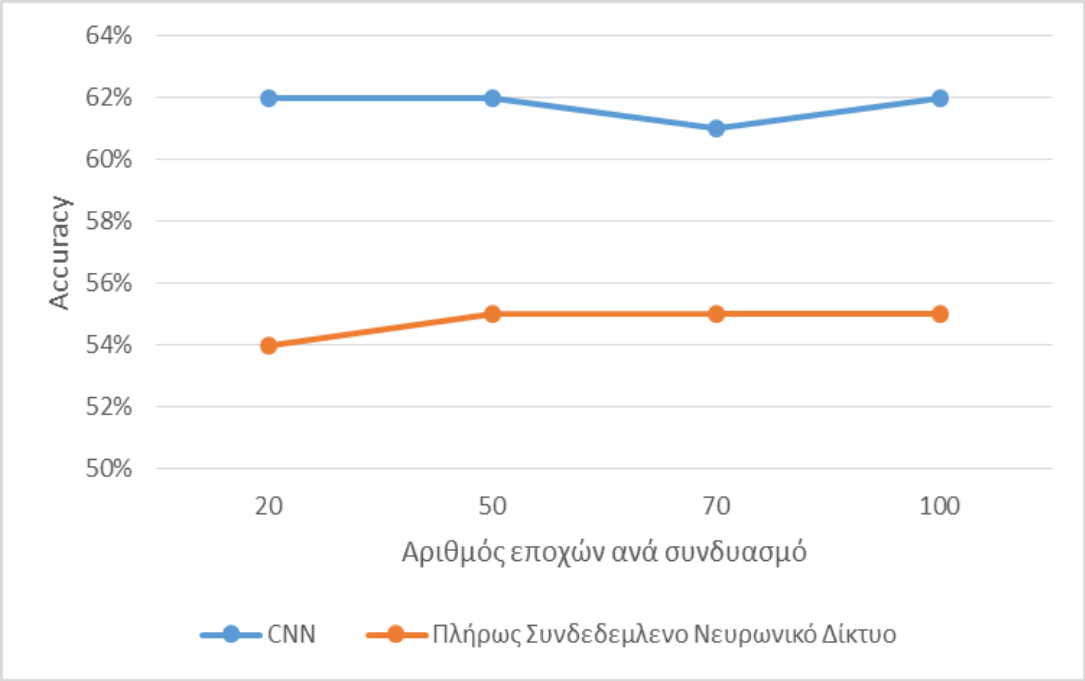
Η τεχνική Cross Validation χρησιμοποιείται πολύ συχνά κατά την εκπαίδευση νευρωνικών μοντέλων, ούτως ώστε να διασφαλιστεί ότι το μοντέλο έχει γενικεύσει τη γνώση που του προσφέρεται.

Κατά την τεχνική αυτή, τα δεδομένα εκπαίδευσης χωρίζονται, πτυχές τόσες όσες θα γίνει το Cross Validation. Σε κάθε επανάληψη της τεχνικής, η μια πτυχή των δεδομένων χρησιμοποιούνται για εκτίμηση των αποτελεσμάτων εκπαίδευσης, και οι υπόλοιπες πτυχές σαν δεδομένα εκπαίδευσης. Με αυτό τον τρόπο, το σύστημα εκπαιδεύεται καλύτερα, αποφεύγοντας έτσι το σύστημα από σενάρια υπερεκπαίδευσης.

Στο πειραματικό μέρος της εργασίας αυτής, έχει γίνει 8-Fold Cross Validation, ούτως ώστε να διασφαλιστεί ότι το μοντέλο εκπαιδεύεται σωστά. Η χρήση της τεχνικής Cross Validation έχει γίνει στις αρχιτεκτονικές Πλήρως Συνδεδεμένου Νευρωνικού Δικτύου και Συνελικτικού Νευρωνικού Δικτύου. Τα δεδομένα που χρησιμοποιήθηκαν είναι τα Chroma Features, MFCC (12 components), μαζί με τα χαρακτηριστικά Chroma Features και MFCC (12 components) από το τέλος του κάθε κομματιού και ο βελτιστοποιητής είναι ο Adam (learning_rate = 0,001).

Αριθμός Εποχών ανά συνδυασμό	**Συνελικτικό Νευρωνικό Δίκτυο**	**Πλήρως Συνδεδεμλενο Νευρωνικό Δίκτυο**
20	62%	54%
50	62%	55%
70	61%	55%
100	62%	55%

3.2.5.1 Πίνακας επιδόσεων Μοντέλων κατά το Cross Validation



3.2.5.1 Γραφική παράσταση Ακρίβεια-Αριθμός Εποχών ανά συνδυασμό για τα Μοντέλα

Από τα παραπάνω πειραματικά αποτελέσματα και από τη γραφική παράσταση, παρατηρείται ότι η τεχνική του Cross Validation έχει επιφέρει χαμηλότερες επιδόσεις και στα δύο μοντέλα της τάξης του 15%. Αυτό σημαίνει ότι η εκπαίδευση που έχει γίνει προηγουμένως δεν έχει οδηγήσει τα μοντέλα σε υπερεκπαίδευση.

4

Συμπεράσματα και Μελλοντικές Επεκτάσεις

4.1 Συμπεράσματα

Στον τομέα της Ανάκτησης Μουσικής Πληροφορίας, η αυτόματη κατηγοριοποίηση μουσικής είναι ανεπτυγμένη σε αρκετά καλό βαθμό. Εν τούτοις, η εφαρμογή της στον κλάδο της Βυζαντινής Μουσικής δεν έχει γίνει ακόμα. Οπότε η τρέχουσα εργασία αποτελεί μια καινοτόμο προσπάθεια στον κλάδο αυτό.

Στην τρέχουσα εργασία προσεγγίσθηκε το πρόβλημα της αυτόματης κατηγοριοποίησης ψαλμωδιών στους οκτώ ήχους, χρησιμοποιώντας σαν είσοδο αρχεία της μορφής mp3. Όλη η υλοποίηση βασίστηκε σε τεχνικές επιβλεπόμενης μάθησης και έγινε χρήση τεχνικών βαθιάς μηχανικής μάθησης. Αφού μελετήθηκαν όλες οι υπερσύγχρονες μέθοδοι που χρησιμοποιούνται για παρόμοιου είδους προβλήματα, εφαρμόστηκαν οι βασικές ιδέες τους και τέλος εκπαιδεύτηκαν τα διάφορα μοντέλα με σκοπό την εύρεση του καλύτερου ως προς την ακρίβεια πρόβλεψης.

Προτού ξεκινήσει η ανάπτυξη ενός συστήματος για τη λύση του προβλήματος της αυτόματης κατηγοριοποίησης μουσικής, αφιερώθηκε ένα μεγάλο χρονικό διάστημα για να πραγματοποιηθεί μελέτη για έρευνες που έχουν γίνει στον τομέα της αυτόματης κατηγοριοποίησης μουσικής και στα ήδη υπάρχοντα υπερσύγχρονα συστήματα κατηγοριοποίησης, καθώς επίσης και για έρευνες που έχουν γίνει στον τομέα της ψηφιακής επεξεργασίας σήματος και μουσικής. Μετά από αυτή τη μελέτη, αποφασίστηκαν τα διάφορα χαρακτηριστικά των ηχητικών αρχείων που θα εξαχθούν, καθώς επίσης και οι αρχιτεκτονικές των μοντέλων μηχανικής μάθησης που θα δοκιμαστούν. Συγκεκριμένα, επιλέχθηκαν τα χαρακτηριστικά Mel-Frequency Cepstral Coefficients (MFCC), Chroma Features, το Φασματικό Κέντρο (Spectral Centroid), η Φασματική Διάθεση (Spectral Roll-

Off), το Φασματικό Εύρος Ζώνης (Spectral Bandwidth), ο Ρυθμός Αλλαγής Πρόσημου (Zero Crossing Rate) και η Ρίζα Μέσης Τετραγωνικής Ενέργειας (Root Mean Square Energy, RMSE). Οι αρχιτεκτονικές που δοκιμάστηκαν είναι οι αρχιτεκτονικές Πλήρως Συνδεδεμένου Νευρωνικού Δικτύου και Συνελικτικού Νευρωνικού Δικτύου.

Το σύνολο δεδομένων που θα χρησιμοποιείτο έπρεπε να βρεθεί και να τοποθετηθούν ετικέτες σε όλα τα αρχεία σε πρώτο στάδιο. Δεν υπάρχει κάτι ήδη έτοιμο στον τομέα της Βυζαντινής Μουσικής, οπότε δημιουργήθηκε ένα για τους σκοπούς αυτής της εργασίας. Χρησιμοποιήθηκαν ηχογραφήσεις υψηλής ποιότητας για πιο εύκολη και ακριβή εξαγωγή χαρακτηριστικών και η ανάθεση ετικετών στα αρχεία έχει γίνει χειροκίνητα.

Στο στάδιο εξαγωγής χαρακτηριστικών και προεπεξεργασίας δεδομένων, εξάγονται όλα τα χαρακτηριστικά που προαναφέρονται και περνούν από ένα στάδιο επεξεργασίας για να έρθουν σε τέτοια μορφή ώστε να μπορεί το νευρωνικό μοντέλο να εκπαιδευτεί με καλύτερα αποτελέσματα. Το στάδιο αυτό της επεξεργασίας στην τρέχουσα εργασία αποτελείτο από 5 στάδια, την αύξηση δεδομένων (Data Augmentation) τη διαίρεση κομματιού σε ισομήκη κομμάτια μικρής διάρκειας, την κανονικοποίηση δεδομένων (Normalization), το παραγέμισμα (Padding) και τέλος την προσθήκη πληροφορίας από το τέλος του κομματιού. Όλα τα παραπάνω δοκιμάστηκαν με διάφορες παραμέτρους για να βρεθούν οι βέλτιστες τιμές που βοηθούν το μοντέλο να εκπαιδευτεί καλύτερα και να έχει την υψηλότερη δυνατή ακρίβεια πρόβλεψης.

Στο πειραματικό στάδιο, έχει γίνει πληθώρα δοκιμών ούτως ώστε να αυξηθεί η ακρίβεια πρόβλεψης του μοντέλου στο μέγιστο δυνατό. Δοκιμάστηκαν και βελτιστοποιήθηκαν οι παράμετροι Batch Size, Αριθμός εποχών εκπαίδευσης μοντέλου και Βελτιστοποιητής και Ρυθμού Μάθησης μοντέλου. Επίσης, σε αυτό το στάδιο έχει γίνει εκπαίδευση των διαφόρων μοντέλων χρησιμοποιώντας ένα-ένα τα χαρακτηριστικά των ηχητικών αρχείων καθώς επίσης και συνδυασμός αυτών και παρατηρήθηκε ότι ο συνδυασμός των Mel-Frequency Cepstral Coefficients (MFCC) και Chroma Features έχει επιφέρει τα καλύτερα αποτελέσματα. Τέλος, έχουν δοκιμαστεί πολλά μοντέλα Πλήρως Συνδεδεμένων και Συνελικτικών Νευρωνικών Δικτύων, με διάφορους συνδυασμούς κρυφών επιπέδων και παρατηρήθηκαν σημαντικές επιδράσεις που έχουν στην επίδοση του συστήματος η αρχιτεκτονική του, ο αριθμός των κρυφών επιπέδων του κάθε συστήματος, καθώς επίσης και ο αριθμός των νευρώνων-κόμβων του κάθε κρυφού επιπέδου.

Η μεγαλύτερη ακρίβεια που έχει επιτευχθεί σε αυτή την έρευνα είναι το **78%**, με τη χρήση των παρακάτω παραμέτρων:

Batch Size: 128
Αριθμός εποχών εκπαίδευσης: 200
Βελτιστοποιητής: Adam (learning_rate = 0,001)
Χαρακτηριστικά που χρησιμοποιήθηκαν από τα ηχητικά αρχεία του συνόλου δεδομένων:
Chroma Features, MFCC (12 components), μαζί με τα χαρακτηριστικά Chroma Features και MFCC (12 components) από το τέλος του κάθε κομματιού.
Αρχιτεκτονική Μοντέλου:

Επίπεδο	Παράμετροι
Conv2D	Filters = 24, kernel_size=(3,3), padding = same
BatchNormalization	-
Activation	relu
MaxPooling2D	Pool_size = (2,2)
Conv2D	Filters = 32, kernel_size=(3,3), padding = same
BatchNormalization	-
Activation	relu
MaxPooling2D	Pool_size = (2,2)
Conv2D	Filters = 64, kernel_size=(3,3), padding = same
BatchNormalization	-
Activation	relu
MaxPooling2D	Pool_size = (2,2)
Conv2D	Filters = 128, kernel_size=(3,3), padding = same
BatchNormalization	-
Activation	relu
GlobalMaxPooling2D	-
Dense	Units = 128
Dense	Units = 8, activation = softmax

4.2 Μελλοντικές Επεκτάσεις και Ανοικτά Πεδία

Η παρούσα εργασία ασχολείται με την εκπαίδευση ενός νευρωνικού δικτύου με τη χρήση Επιβλεπόμενης Μάθησης. Για να μπορεί ένα τέτοιο σύστημα να αποκομίσει όλα τα χαρακτηριστικά της κάθε κλάσης χωρίς να μάθει να ξεχωρίζει απλά τα δεδομένα εκπαίδευσης (υπερεκπαίδευση), θα πρέπει το σύνολο δεδομένων να είναι επαρκώς μεγάλο, και μάλιστα με διαφορετικούς ψάλτες και ήχους. Για αυτό το λόγο, ο εμπλουτισμός συνόλου δεδομένων του με νέα μουσικά ηχητικά αρχεία από διαφορετικούς μάλιστα ψάλτες, θεωρώ θα βοηθούσε στη γενίκευση των χαρακτηριστικών των οκτώ ήχων της Βυζαντινής Μουσικής και το σύστημα θα είχε πολύ μεγαλύτερη ανοχή σε θορύβους ηχογραφήσεων λόγω χρήσης μη επαγγελματικού εξοπλισμού, καθώς επίσης και σε ανοχή στις διάφορες χροιές των ψαλτών και διαφόρους ιδιωματισμούς που δημιουργούνται από περιοχή σε περιοχή.

Μια επέκταση αυτής της εργασίας μπορεί να η διαφοροποίηση των ήχων μεταξύ τους με βάση του είδους μελοποιίας, αφού όπως προαναφέρθηκε στην ενότητα 1.3, ο ίδιος ήχος σε διαφορετικό είδος μελοποιίας έχει διαφορετικά χαρακτηριστικά, αυξάνοντας με αυτό τον τρόπο τον αριθμό των κλάσεων που καλείται το σύστημα να κατηγοριοποιήσει το κάθε ηχητικό κομμάτι.

Η εργασία αυτή μπορεί να χρησιμοποιηθεί σαν μέρος συστημάτων προτάσεων, αφού μπορεί να ανιχνεύσει ψαλμούς που έχουν μεγάλη ομοιότητα μεταξύ τους. Η πληροφορία αυτή σε συνδυασμό με διάφορες άλλες μεταβλητές και παραμέτρους που χρησιμοποιούν τα συστήματα προτάσεων μπορούν να έχουν πολύ καλά αποτελέσματα.
Το σύστημα αυτό μπορεί επίσης με τις κατάλληλες τροποποιήσεις να χρησιμοποιηθεί και σαν σύστημα αναγνώρισης κλειδιού/κλίμακας ενός τραγουδιού. Προβλήματα εύρεσης κλειδιού/κλίμακας ενός τραγουδιού έχουν αναπτυχθεί τα τελευταία χρόνια και αυτή η εργασία μπορεί να θεωρηθεί μέρος αυτού του τομέα.

Επιπρόστθετα, το σύστημα αυτό μπορεί να χρησιμοποιηθεί σαν μέρος μιας πιο μεγάλης έρευνας η οποία θα αναπτύσσει ένα σύστημα που θα δέχεται σαν είσοδο ένα ηχητικό αρχείο ψαλμωδίας και θα το μεταγράφει σε νότες της βυζαντινής σημειογραφίας (transcribe). Ανάλογα με τον ήχο στον οποίο ανήκει μια ψαλμωδία, υπάρχουν προκαθορισμένες μουσικές φράσεις που χρησιμοποιούνται πολύ συχνά, άρα η γνώση του ήχου μιας ψαλμωδίας, βοηθά σε μεγάλο βαθμό σε προβλήματα μεταγραφής.

Τέλος, μια άμεση αλλά όχι τόσο απλή επέκταση του τρέχοντος συστήματος είναι η μετατροπή του σε σύστημα που αναγνωρίζει τον ήχο ενός κομματιού σε πραγματικό χρόνο. Η υλοποίηση αυτής της έρευνας θα ανοίξει το δρόμο σε πολλά άλλα συστήματα που τρέχουν σε πραγματικό χρόνο, όπως για παράδειγμα η αυτόματη μεταγραφή ψαλμωδιών σε νότες βυζαντινής σημειογραφίας.

Κατάλογος Σχημάτων

1.1.2.1 Πλήρως Συνδεδεμένο Επίπεδο 5
1.1.3.1 Πίνακας Χαρακτηριστικών 7
1.1.3.2 Συγκεντρωτικό Επίπεδο 8
1.1.4.1 RNN νευρώνας 9
1.1.5.1 LSTM νευρώνας 10
2.1.1 Γραφική παράσταση Ήχων-Αριθμού δειγμάτων διάρκειας 20 δευτερολέπτων 24
3.1.1 Αρχιτεκτονική Πλήρως Συνδεδεμένου Νευρωνικού Δικτύου 34
3.1.2 Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου 34
3.1.3 Αρχιτεκτονική LSTM Δικτύου 35
3.2.1 Γραφική παράσταση Ακρίβεια-Χαρακτηριστικά για τα 2 μοντέλα που δοκιμάστηκαν, Πλήρως Συνδεδεμένο Νευρωνικό Μοντέλο και Συνελικτικό Νευρωνικό Μοντέλο 37
3.2.2 Γραφική παράσταση Ακρίβεια-Μέθοδος Παραγεμίσματος για τα 2 μοντέλα που δοκιμάστηκαν, Πλήρως Συνδεδεμένο Νευρωνικό Μοντέλο και Συνελικτικό Νευρωνικό Μοντέλο 39
3.2.3 Γραφική παράσταση Ακρίβεια-Χαρακτηριστικά για τα 3 μοντέλα που δοκιμάστηκαν 41
3.2.1.1 Γραφική παράσταση Accuracy-Batch Size 43
3.2.2.1 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών για 150 εποχές 45
3.2.2.2 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών για 300 εποχές 45
3.2.2.3 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών για 150 εποχές 47
3.2.2.4 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών για 300 εποχές 47
3.2.2.5 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών για 150 εποχές 49
3.2.2.6 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών για 300 εποχές 49

3.2.2.7 Γραφική παράσταση εκπαίδευσης μοντέλων Accuracy-Αριθμός εποχών εκπαίδευσης 50
3.2.3.1 Γραφική παράσταση Accuracy-Optimizer 52
3.2.4.1 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών 54
3.2.4.2 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών 55
3.2.4.3 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών 56
3.2.4.4 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών 57
3.2.4.5 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών 58
3.2.4.6 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών 60
3.2.4.7 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών 62
3.2.4.8 Γραφική παράσταση εκπαίδευσης μοντέλου Cross Entropy Loss-Αριθμός Εποχών και Accuracy-Αριθμός Εποχών 64
3.2.5.1 Γραφική παράσταση Ακρίβεια-Αριθμός Εποχών ανά συνδυασμό για τα Μοντέλα 66

Κατάλογος Πινάκων

3.2.1 Πίνακας επιδόσεων Συνελικτικού Νευρωνικού Δικτύου για όλα τα χαρακτηριστικά που εξάχθηκαν 36
3.2.2 Πίνακας επιδόσεων Πλήρως Συνδεδεμένου Δικτύου για όλα τα χαρακτηριστικά που εξάχθηκαν 36
3.2.3 Πίνακας επιδόσεων LSTM Δικτύου για όλα τα χαρακτηριστικά που εξάχθηκαν . 36
3.2.4 Πίνακας επιδόσεων CNN + LSTM Δικτύου για όλα τα χαρακτηριστικά που εξάχθηκαν 36
3.2.3 Πίνακας επιδόσεων Συνελικτικού Νευρωνικού Μοντέλου χρησιμοποιώντας όλα τα χαρακτηριστικά και της τεχνικής παραγεμίσματος 38
3.2.4 Πίνακας επιδόσεων Πλήρως Συνδεδεμένου μοντέλου χρησιμοποιώντας όλα τα χαρακτηριστικά και της τεχνικής παραγεμίσματος 38
3.2.5 Πίνακας επιδόσεων Συνελικτικού Νευρωνικού Μοντέλου 40
3.2.6 Πίνακας επιδόσεων Πλήρως Συνδεδεμένου Μοντέλου 40
3.2.7 Πίνακας επιδόσεων CNN + LSTM Μοντέλου 40
3.2.1.1 Πίνακας επιδόσεων Συνελικτικού Νευρωνικού Μοντέλου για διάφορες τιμές Batch Size 42
3.2.1.2 Πίνακας επιδόσεων Πλήρως Συνδεδεμένου Νευρωνικού Μοντέλου για διάφορες τιμές Batch Size 42
3.2.1.3 Πίνακας επιδόσεων CNN + LSTM Μοντέλου για διάφορες τιμές Batch Size43
3.2.2.1 Πίνακας επιδόσεων Συνελικτικού Νευρωνικού Μοντέλου για διάφορες τιμές αριθμού εποχών 44
3.2.2.2 Πίνακας επιδόσεων Πλήρως Συνδεδεμένου Νευρωνικού Μοντέλου για διάφορες τιμές αριθμού εποχών 46
3.2.2.3 Πίνακας επιδόσεων CNN + LSTM Μοντέλου για διάφορες τιμές αριθμού εποχών 48
3.2.3.1 Πίνακας επιδόσεων Συνελικτικού Νευρωνικού Μοντέλου για διάφορους βελτιστοποιητές 51
3.2.3.2 Πίνακας επιδόσεων Πλήρως Συνδεδεμένου Νευρωνικού Μοντέλου για διάφορους βελτιστοποιητές 51
3.2.3.1 Πίνακας επιδόσεων CNN + LSTM Μοντέλου για διάφορους βελτιστοποιητές 52

3.2.5.1 Πίνακας επιδόσεων Μοντέλων κατά το Cross Validation 65

Βιβλιογραφία

[1] C. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag New York, 2006.

[2] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. The MIT Press, 2012.

[3] A. Kolesnikov *et al.*, "Big Transfer (BiT): General Visual Representation Learning," Dec. 2019, [Online]. Available: http://arxiv.org/abs/1912.11370.

[4] Chang-Hsing Lee, Jau-Ling Shih, Kun-Ming Yu, and Hwai-San Lin, "Automatic Music Genre Classification Based on Modulation Spectral Analysis of Spectral and Cepstral Features," *IEEE Trans. Multimed.*, vol. 11, no. 4, pp. 670–682, Jun. 2009, doi: 10.1109/TMM.2009.2017635.

[5] D. B. Fogel, T. J. Hays, S. L. Hahn, and J. Quon, "A self-learning evolutionary chess program," *Proc. IEEE*, vol. 92, no. 12, pp. 1947–1954, Dec. 2004, doi: 10.1109/JPROC.2004.837633.

[6] R. Collobert and J. Weston, "A unified architecture for natural language processing," in *Proceedings of the 25th international conference on Machine learning - ICML '08*, 2008, pp. 160–167, doi: 10.1145/1390156.1390177.

[7] A. van den Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation," in *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2013, pp. 2643–2651.

[8] Y. Song, S. Dixon, and M. Pearce, "A survey of music recommendation systems and future perspectives," in *9th International Symposium on Computer Music Modeling and Retrieval*, 2012, vol. 4, pp. 395–410.

[9] K. Yoshii, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno, “An Efficient Hybrid Music Recommender System Using an Incrementally Trainable Probabilistic Generative Model,” *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 16, no. 2, pp. 435–447, Feb. 2008, doi: 10.1109/TASL.2007.911503.

[10] T. Heittola, A. Klapuri, and T. Virtanen, “Musical Instrument Recognition in Polyphonic Audio Using Source-Filter Model for Sound Separation,” 2009.

[11] A. Eronen and A. Klapuri, “Musical instrument recognition using cepstral coefficients and temporal features,” in *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*, vol. 2, pp. II753–II756, doi: 10.1109/ICASSP.2000.859069.

[12] A. Belouchrani, K. Abed-Meraim, J.-. Cardoso, and E. Moulines, “A blind source separation technique using second-order statistics,” *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997, doi: 10.1109/78.554307.

[13] Z. Duan, Y. Zhang, C. Zhang, and Z. Shi, “Unsupervised Single-Channel Music Source Separation by Average Harmonic Structure Modeling,” *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 16, no. 4, pp. 766–778, May 2008, doi: 10.1109/TASL.2008.919073.

[14] O. Gillet and G. Richard, “Transcription and Separation of Drum Signals From Polyphonic Music,” *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 16, no. 3, pp. 529–540, Mar. 2008, doi: 10.1109/TASL.2007.914120.

[15] M. P. Ryynanen and A. Klapuri, “Polyphonic music transcription using note event modeling,” in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005.*, Oct. 2005, pp. 319–322, doi: 10.1109/ASPAA.2005.1540233.

[16] E. Demirel, B. s Bozkurt, and X. Serra, “Automatic makam recognition using chroma features,” 2018, [Online]. Available: https://doi.org/10.5281/zenodo.1239435.

[17] A. Meng, P. Ahrendt, J. Larsen, and L. K. Hansen, “Temporal Feature Integration for Music Genre Classification,” *IEEE Trans. Audio, Speech Lang. Process.*, vol. 15, no. 5, pp. 1654–1664, Jul. 2007, doi: 10.1109/TASL.2007.899293.

[18] G. Kour and N. Mehan, “Music genre classification using MFCC, SVM and BPNN,” *Int. J. Comput. Appl.*, vol. 112, no. 6, 2015.

[19] L.-C. Yang, S.-Y. Chou, and Y.-H. Yang, “MidiNet: A Convolutional Generative Adversarial Network for Symbolic-domain Music Generation,” Mar. 2017, [Online]. Available: http://arxiv.org/abs/1703.10847.

[20] Κ. Φλώρος, *Η ελληνική παράδοση στις μουσικές γραφές του μεσαίωνα*. Θεσσαλονίκη: Ζητη, 1998.

[21] Κ.Α.Ψάχου, *Η Παρασημαντική της Βυζαντινής Μουσικής*. Αθήνα: Διόνυσος, 1978.

[22] Χρύσανθος εκ Μαδύτων, *Θεωρητικόν Μέγα της Μουσικής*. Παρίσι: Κουλτούρα, 1821.

[23] Κ.Α.Ψάχου, *Το Οκτάηχον Σύστημα της Βυζαντινής Μουσικής*. Κρητη: Μιχαήλ Ι. Πολυχρονάκης, 1980.

[24] D. G. Panagiotopoulos, *Theory and Praxis of the Byzantine Ecclesiastical Music*. Athens, 1947.

[25] Θεώδορος Φωκαευς, *Κρηπις του θεωρητικού και πρακτικού της εκκλησιαστικήε μουδικής*. Αθήνα.

[26] V. G. Gezerlis and S. Theodoridis, “An optical music recognition system for the notation of the Orthodox Hellenic Byzantine Music,” in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, vol. 4, pp. 837–840, doi: 10.1109/ICPR.2000.903047.

[27] V. G. Gezerlis and S. Theodoridis, “Optical character recognition of the Orthodox Hellenic Byzantine Music notation,” *Pattern Recognit.*, vol. 35, no. 4, pp. 895–914,

2002, doi: https://doi.org/10.1016/S0031-3203(01)00098-X.

[28] C. Dalitz, G. K. Michalakis, and C. Pranzas, “Optical recognition of psaltic Byzantine chant notation,” *Int. J. Doc. Anal. Recognit.*, vol. 11, no. 3, pp. 143–158, Dec. 2008, doi: 10.1007/s10032-008-0074-4.

[29] G. Chrysochoidis, G. Kouroupetroglou, D. Delviniotis, and S. Theodoridis, *Formant Tuning in Byzantine Chant*. 2013.

[30] K. Kokkinidis, T. Mastoras, A. Tsagaris, and P. Fotaris, “An empirical comparison of machine learning techniques for chant classification,” in *2018 7th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, May 2018, pp. 1–4, doi: 10.1109/MOCAST.2018.8376596.

[31] B. Bozkurt, “Features for analysis of Makam music,” 2012.

[32] M. A. Kızrak, K. S. Bayram, and B. Bolat, “Classification of Classic Turkish Music Makams,” in *2014 IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA) Proceedings*, Jun. 2014, pp. 394–397, doi: 10.1109/INISTA.2014.6873650.

[33] G. Tzanetakis and P. Cook, “Musical genre classification of audio signals,” *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 293–302, Jul. 2002, doi: 10.1109/TSA.2002.800560.

[34] V. Tiwari, “MFCC and its applications in speaker recognition,” *Int. J. Emerg. Technol.*, vol. 1, no. 1, pp. 19–22, 2010.

[35] C. Ittichaichareon, S. Suksri, and T. Yingthawornsuk, “Speech recognition using MFCC,” in *International Conference on Computer Graphics, Simulation and Modeling*, 2012, pp. 135–138.

[36] N. Jiang, P. Grosche, V. Konz, and M. Müller, “Analyzing chroma feature types for automated chord recognition,” 2011.

[37] G. Peeters, “Chroma-based estimation of musical key from audio-signal analysis.,” in *ISMIR*, 2006, pp. 115–120.

[38] J. M. K. Kua, T. Thiruvaran, M. Nosratighods, E. Ambikairajah, and J. Epps, “Investigation of spectral centroid magnitude and frequency for speaker recognition.,” in *Odyssey*, 2010, p. 7.

[39] M. Kos, Z. Kačič, and D. Vlaj, “Acoustic classification and segmentation using modified spectral roll-off and variance-based features,” *Digit. Signal Process.*, vol. 23, no. 2, pp. 659–674, Mar. 2013, doi: 10.1016/j.dsp.2012.10.008.

[40] M. Sahidullah and G. Saha, “Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition,” *Speech Commun.*, vol. 54, no. 4, pp. 543–565, May 2012, doi: 10.1016/j.specom.2011.11.004.

[41] A. Shah, M. Kattel, A. Nepal, and D. Shrestha, *Chroma Feature Extraction*. 2019.

[42] B. McFee *et al.*, *librosa: Audio and Music Signal Analysis in Python*. 2015.

[43] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, “A Review on Deep Learning Techniques Applied to Semantic Segmentation,” Apr. 2017, [Online]. Available: http://arxiv.org/abs/1704.06857.

[44] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[45] C. Szegedy *et al.*, “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[46] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi:

10.1109/5.726791.

[47] N. Chen and S. Wang, “High-Level Music Descriptor Extraction Algorithm Based on Combination of Multi-Channel CNNs and LSTM,” 2017.

[48] H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, and A.-B. M. Salem, “Classification using deep learning neural networks for brain tumors,” *Futur. Comput. Informatics J.*, vol. 3, no. 1, pp. 68–71, Jun. 2018, doi: 10.1016/j.fcij.2017.12.001.

[49] S. Albelwi and A. Mahmood, “A Framework for Designing the Architectures of Deep Convolutional Neural Networks,” *Entropy*, vol. 19, Jun. 2017, doi: 10.3390/e19060242.