



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΥΠΟΛΟΓΙΣΤΩΝ

Διπλωματική Εργασία

Μουσική παραγωγή με τη χρήση Μηχανικής Μάθησης



Φοιτήτρια: Καραφίλη Αικατερίνη

ΑΜ: 711171056

Επιβλέπουσα Καθηγήτρια

Παναγιώτα Τσελέντη

ΑΘΗΝΑ, ΙΟΥΛΙΟΣ 2023



UNIVERSITY OF WEST ATTICA

FACULTY OF ENGINEERING

DEPARTMENT OF INFORMATICS & COMPUTER ENGINEERING

Diploma Thesis

Music generation using Machine Learning

Student: Karafili Aikaterini

Registration Number: 711171056

Supervisor

Panagiota Tselenti

ATHENS, JULY 2023

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μουσική παραγωγή με τη χρήση Μηχανικής Μάθησης

**Αικατερίνη Καραφίλη
Α.Μ. 711171056**

Εισηγητής:

Παναγιώτα Τσελέντη, ΕΔΙΠ

Εξεταστική Επιτροπή:

Χρήστος Τρούσσας, Επ. Καθηγητής

Ακριβή Κρούσκα, Μεταδιδακτορική Ερευνήτρια

Ημερομηνία εξέτασης :

Η κάτωθι υπογεγραμμένη Καραφίλη Αικατερίνη του Θεοδώρου, με αριθμό μητρώου 711171056 φοιτήτρια του Πανεπιστημίου Δυτικής Αττικής της Σχολής Μηχανικών του Τμήματος Μηχανικών Πληροφορικής και Υπολογιστών, δηλώνω υπεύθυνα ότι:

«Βεβαιώνω ότι είμαι συγγραφέας αυτής της Διπλωματικής εργασίας και κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος. Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Η δηλούσα

Αικατερίνη Καραφίλη



ΕΥΧΑΡΙΣΤΙΕΣ

Αρχικά, θα ήθελα να εκφράσω τις ευχαριστίες μου, την ειλικρινή μου εκτίμηση και ευγνωμοσύνη στην επιβλέπουσα καθηγήτρια μου κ. Παναγιώτα Τσελέντη για την αμέριστη υποστήριξη, την ανταπόκριση και το γνήσιο ενδιαφέρον της για την παρούσα διπλωματική εργασία. Η καθοδήγησή της έπαιξε καθοριστικό ρόλο, καθ' όλη τη διάρκεια της συγγραφής της εργασίας.

Ακόμη, θέλω να εκφράσω την βαθύτατη ευγνωμοσύνη και αγάπη μου στην οικογένειά μου, τους γονείς μου και τον θείο μου που με τη συνεχή ενθάρρυνση και πίστη τους σε μένα ήταν οι πυλώνες που με στήριξαν σε δύσκολες στιγμές και με ώθησαν προς την επιτυχία στο ακαδημαϊκό μου ταξίδι.

Τέλος, θα ήθελα να ευχαριστήσω το αγόρι μου, που ήταν ο «συνταξιδιώτης» μου σε αυτό το ταξίδι, στις δύσκολες στιγμές και στις χαρούμενες, και μια συνεχής πηγή δύναμης, υποστήριξης και ενθάρρυνσης σε κάθε εμπόδιο. Η υπομονή του, η κατανόηση και η προθυμία του να θυσιάσει το δικό του χρόνο έκανε τη διαφορά στην ικανότητά μου να επιμένω και να πετυχαίνω.

Περίληψη

Η μουσική, ιδίως η κλασική μουσική, είναι αναγκαία για τους ανθρώπους, αφού καταργεί τα πολιτισμικά σύνορα και αγγίζει τους ανθρώπους σε όλες τις πτυχές της ζωής τους εδώ και αιώνες. Στην εποχή μας, η σύνθεση μουσικής με τις μεθόδους της τεχνητής νοημοσύνης έχει κερδίσει τον δικό της χώρο ως επιστημονικός τομέας έχοντας τραβήξει το ενδιαφέρον πολλών ερευνητών. Η εφαρμογή που πραγματεύεται η εργασία βοηθάει τον καλλιτέχνη να μπορεί να προσδιορίσει μια ιδέα/κομμάτι του, με βάση έναν από τους συνθέτες Mozart, Bach, Haydn ή Chopin, και να παραχθεί ένα μουσικό κομμάτι βασισμένο στα χαρακτηριστικά που χαρακτηρίζουν τον συνθέτη, που είναι πιο κοντά στο στυλ του καλλιτέχνη, με σκοπό την ενσωμάτωση του στυλ του συνθέτη στην σύνθεση του καλλιτέχνη. Για την κατηγοριοποίηση χρησιμοποιούνται και αξιολογούνται διάφοροι αλγόριθμοι Μηχανικής Μάθησης όπως Decision Trees, Random Forest, Support Vector Machine κ.α., ενώ για την παραγωγή μουσικής χρησιμοποιούνται τα δυο μοντέλα βαθιάς μάθησης, το Long short-term memory (LSTM) και το WaveNet. Εμβαθύνοντας στα δυο μοντέλα, πραγματοποιείται διερεύνηση πάνω στην εκπαίδευση αυτών των μοντέλων σε τέσσερα διαφορετικά σύνολα δεδομένων, διαφόρων μεγεθών και μουσικών συνθέσεων. Με την εκτέλεση της παραγωγής κομματιών, παράγονται συνθέσεις για κάθε εκπαιδευμένο μοντέλο και αξιολογούνται. Τέλος, τονίζονται οι περιορισμοί, οι πιθανές βελτιώσεις αυτών και τα πεδία μελλοντικής έρευνας.

Λέξεις - Κλειδιά

παραγωγή μουσικής, κατηγοριοποίηση μουσικής, σύνθεση μουσικής, μηχανική μάθηση, βαθιά μάθηση, νευρωνικά δίκτυα, μοντέλα βαθιάς μάθησης

Abstract

Music, especially classical music, is necessary for people, as it breaks down cultural barriers and has touched people in all aspects of their lives for centuries. Nowadays, music composition using artificial intelligence methods has earned its own space as a scientific field having attracted the interest of many researchers. The application discussed in the paper helps the artist to be able to identify an idea/piece of music based on one of the composers Mozart, Bach, Haydn or Chopin, and produce a piece of music based on the characteristics of the composer that is closer to the artist's style, with the aim of incorporating the composer's style into the artist's composition. Various Machine Learning algorithms such as Decision Trees, Random Forest, Support Vector Machine etc. are used and evaluated for categorization, while the two deep learning models, Long short-term memory (LSTM) and WaveNet are used for music generation. Delving deeper into the two models, an investigation on the training of these models on four different datasets of different sizes and musical compositions is carried out. By performing track generation, compositions for each trained model are produced and evaluated. Finally, limitations, possible improvements of these and areas for future research are highlighted.

Keywords

music generation, music classification, music synthesis, machine learning, deep learning, neural networks, deep learning models

Περιεχόμενα

1	Εισαγωγή.....	11
1.1	Σύνθεση Μουσικής/Παραγωγή Μουσικής	11
1.2	Αλγοριθμική Σύνθεση.....	12
1.2.1	Μέθοδοι Αλγοριθμικής Σύνθεσης.....	13
1.3	Ιστορική Αναδρομή	18
1.3.1	Τεχνολογίες/Τρόποι παραγωγής μουσικής	18
2	Τεχνητή νοημοσύνη και μουσική	27
2.1	Μηχανική Μάθηση	28
2.1.1	Αλγόριθμοι Μηχανικής Μάθησης	30
2.2	Βαθιά Μάθηση.....	33
2.2.1	Αλγόριθμοι Βαθιάς Μάθησης.....	35
3	Βιβλιογραφική ανασκόπηση	43
4	Μεθοδολογία.....	46
4.1	Περιβάλλον υλοποίησης	46
4.2	Αρχεία MIDI και σύνολα δεδομένων	46
4.2.1	Αρχεία MIDI.....	46
4.2.2	Σύνολα δεδομένων	47
4.3	Βιβλιοθήκες	49
4.4	Κατηγοριοποίηση συνθετών μουσικής	50
4.4.1	Προ επεξεργασία δεδομένων MIDI αρχείων	51
4.4.2	Εφαρμογή των μοντέλων κατηγοριοποίησης.....	54
4.4.3	Αξιολόγηση αποτελεσμάτων	55
4.4.4	Αποτελέσματα.....	55
4.5	Παραγωγή μουσικής	60
4.5.1	Προ επεξεργασία MIDI αρχείων.....	61
4.5.2	Εφαρμογή μοντέλων παραγωγής	61
4.5.3	Αποτελέσματα.....	64
4.6	Ανάλυση αποτελεσμάτων	68
5	Συμπεράσματα και μελλοντικές προκλήσεις	72
6	Βιβλιογραφία	74

Εικόνες

Εικόνα 1 Ταξινόμηση των μεθόδων αλγοριθμικής σύνθεσης (Fernandez & Vico, 2013)	13
Εικόνα 2 Αλγοριθμική σύνθεση με Markov chains (ALGORITHMIC COMPOSER, 2010)	14
Εικόνα 3 Επισκόπηση του αλγορίθμου VNS (Herremans & Sørensen, Composing Fifth Species Counterpoint Music, 2012)	16
Εικόνα 4 Παράδειγμα μιας εξελικτικής μεθόδου, της διασταύρωσης (crossover) (Fox, 2006)	17
Εικόνα 5 (Smith, 2016)	18
Εικόνα 6 “In C”, Terry Riley 1964	19
Εικόνα 7 Reich’s Piano Phase (Classicfm, 2016)	20
Εικόνα 8 Βρόχος μαγνητικής ταινίας (Scoates, 2013)	21
Εικόνα 9 Η γραφική σημειογραφία του Brian Eno για το “Music for Airports”	21
Εικόνα 10 Παράδειγμα «γενετικής»(generative) γραμματικής (Woolhouse & Rohrmeier, 2014)	23
Εικόνα 11 Music Mouse (Computer Software by Laurie Spiegel)	24
Εικόνα 12 Markov Chains με μουσική σημειογραφία (Kim & Yeo, 2013)	25
Εικόνα 13 Ένα ζετυλιγμένο νευρωνικό δίκτυο RNN (Malik & Ek, 2017)	26
Εικόνα 14 Μετατροπή ήχου σε παρτιτούρα	28
Εικόνα 15 Τύποι μηχανικής μάθησης (Horn)	29
Εικόνα 16 Απεικόνιση του μαθηματικού τύπου που βασίζεται ο αλγόριθμος Naive Bayes	33
Εικόνα 17 Παράδειγμα ταξινόμησης μουσικών ειδών με χρήση τεχνικών βαθιάς μάθησης (Rai, 2021)	34
Εικόνα 18 Επαναλαμβανόμενα Νευρωνικά Δίκτυα (Rana, και συν., 2016)	35
Εικόνα 19 (a) Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (b) Περιφραγμένη Επαναλαμβανόμενη Μονάδα (Rana, και συν., 2016)	36
Εικόνα 20 Συνελκτικό Νευρωνικό Δίκτυο (Jordan, 2017)	37
Εικόνα 21 Pianoroll (Jazz Studies Online)	38
Εικόνα 22 Οπτικοποίηση μιας στοίβας διασταλμένων αιτιακών συνελκτικών στρωμάτων (Oord, Dieleman, Simonyan, Vinyals, & Graves, 2016)	39
Εικόνα 23 Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (Liu & Ramakrishnan, 2014)	40
Εικόνα 24 Αρχιτεκτονική υψηλού επιπέδου του Generative Adversarial Network (Almeida & Pinho, 2018)	42
Εικόνα 25 Ένα βασικό MIDI αρχείο και τα στοιχεία που το αποτελούν (Hartley, 2008)	47
Εικόνα 26 Το σύνολο δεδομένων Classical Music MIDI (Rakshit, 2019)	48
Εικόνα 27 Διάγραμμα της προτεινόμενης μεθοδολογίας κατηγοριοποίησης MIDI αρχείου με βάση τον συνθέτη	50
Εικόνα 28 Πίνακας σύγκρισης για τον αλγόριθμο decision trees	56
Εικόνα 29 Πίνακας σύγκρισης για τον αλγόριθμο knn	56
Εικόνα 30 Πίνακας σύγκρισης για τον αλγόριθμο logistic regression	57
Εικόνα 31 Πίνακας σύγκρισης για τον αλγόριθμο naive bayes	57
Εικόνα 32 Πίνακας σύγκρισης για τον αλγόριθμο neural nets	58
Εικόνα 33 Πίνακας σύγκρισης για τον αλγόριθμο random forest	58

Εικόνα 34 Πίνακας σύγκρισης για τον αλγόριθμο <i>stochastic gradient descent</i>	59
Εικόνα 35 Πίνακας σύγκρισης για τον αλγόριθμο <i>svm</i>	59
Εικόνα 36 Διάγραμμα της προτεινόμενης μεθοδολογίας παραγωγής μουσικής	60
Εικόνα 37 Οπτική αναπαράσταση της αρχιτεκτονικής του μοντέλου <i>LSTM</i>	62
Εικόνα 38 Οπτική αναπαράσταση της αρχιτεκτονικής του μοντέλου <i>WaveNet</i>	64

Γραφήματα

Γράφημα 1 Ιστόγραμμα τόνων του <i>bach</i>	51
Γράφημα 2 Ιστόγραμμα τόνων του <i>chorin</i>	52
Γράφημα 3 Ιστόγραμμα τόνων του <i>haydn</i>	53
Γράφημα 4 Ιστόγραμμα τόνων του <i>mozart</i>	53
Γράφημα 5 Απώλεια (<i>loss</i>) <i>Mozart</i>	65
Γράφημα 6 Απώλεια (<i>loss</i>) <i>Chopin</i>	65
Γράφημα 7 Απώλεια (<i>loss</i>) <i>Bach</i>	66
Γράφημα 8 Απώλεια (<i>loss</i>) <i>Haydn</i>	66
Γράφημα 9 <i>Mozart</i>	67
Γράφημα 10 <i>Bach</i>	67
Γράφημα 11 <i>Chopin</i>	67
Γράφημα 12 <i>Haydn</i>	68

Πίνακες

Πίνακας 1 Χαρακτηριστικά συναισθημάτων και ήχου μερικών δημοφιλών μυθιστορημάτων που είναι επεξεργασμένα από το εργαλείο <i>TransProse</i> (Davis & Mohammad, 2014)	14
Πίνακας 2 Αναφορές σε αλγοριθμική σύνθεση με γραμματικές (Fernandez & Vico, 2013)	15
Πίνακας 3 Εξαχθέντα χαρακτηριστικά από <i>MIDI</i> αρχεία	54
Πίνακας 4 Απόδοση του μοντέλου <i>decision trees</i>	56
Πίνακας 5 Απόδοση του μοντέλου <i>KNN</i>	57
Πίνακας 6 Απόδοση του μοντέλου <i>logistic regression</i>	57
Πίνακας 7 Απόδοση του μοντέλου <i>naive bayes</i>	58
Πίνακας 8 Απόδοση του μοντέλου <i>neural nets</i>	58
Πίνακας 9 Απόδοση του μοντέλου <i>random forest</i>	59
Πίνακας 10 Απόδοση του μοντέλου <i>stochastic gradient descent</i>	59
Πίνακας 11 Απόδοση του μοντέλου <i>svm</i>	60
Πίνακας 12 Απώλεια (<i>loss</i>) <i>min, max, mean</i> του μοντέλου <i>LSTM</i>	66
Πίνακας 13 Απώλεια (<i>loss</i>) <i>min, max, mean</i> του μοντέλου <i>WaveNet</i>	68

1 Εισαγωγή

1.1 Σύνθεση Μουσικής/Παραγωγή Μουσικής

Μουσική σύνθεση (ή δημιουργία) είναι η διαδικασία δομής ή σύνθεσης ενός νέου μουσικού κομματιού. Ένα νέο μουσικό κομμάτι ορίζεται γενικά ως μια διαδοχή τόνων ή ρυθμών, ή και των δύο, σε κάποια συγκεκριμένα μοτίβα. Ο όρος μουσική σύνθεση μπορεί επίσης να αναφέρεται σε ένα πρωτότυπο κομμάτι ή έργο μουσικής που αποτελείται είτε από φωνές είτε από όργανα. Η μουσική σύνθεση απαιτεί δημιουργικότητα η οποία είναι μια μοναδική ανθρώπινη ικανότητα να κατανοεί και να παράγει έναν απεριόριστα μεγάλο αριθμό προτάσεων σε μια γλώσσα, οι περισσότερες από τις οποίες δεν έχουν ποτέ ξανασυναντηθεί ή ειπωθεί.

Οι άνθρωποι με την ικανότητα αυτή ονομάζονται συνθέτες (ή τραγουδοποιοί), κυρίως ασχολούνται με τραγούδια και μελωδίες. Σε πολλούς πολιτισμούς, συμπεριλαμβανομένης της δυτικής κλασικής μουσικής, η πράξη της σύνθεσης περιλαμβάνει συνήθως τη δημιουργία μουσικής σημειογραφίας, όπως μια παρτιτούρα που αποτελείται από πεντάγραμμα γραμμένα με νότες, η οποία στη συνέχεια εκτελείται από τον συνθέτη ή από άλλους μουσικούς. Σε πολλά σύγχρονα είδη μουσικής και στην παραδοσιακή μουσική, η σύνθεση τραγουδιών μπορεί να περιλαμβάνει τη δημιουργία μιας βασικής ακολουθίας του τραγουδιού, που ονομάζεται “lead sheet”, το οποίο καθορίζει τη μελωδία, τους στίχους και την εξέλιξη των συγχορδιών.

Στην κλασική μουσική, η ενορχήστρωση σε μια ορχήστρα, δηλαδή, η επιλογή των οργάνων ενός μεγάλου μουσικού συνόλου, που θα παίξουν τα διάφορα μέρη της μουσικής, όπως η μελωδία, η συνοδεία, η αντίστιξη κ.α. γίνεται συνήθως από τον συνθέτη, αλλά στο μουσικό θέατρο και στην ποπ μουσική, οι τραγουδοποιοί μπορεί να προσλάβουν έναν μαέστρο για να κάνει την ενορχήστρωση. Σε ορισμένες περιπτώσεις, ένας ποπ ή παραδοσιακός τραγουδοποιός μπορεί να μη χρησιμοποιεί καθόλου γραπτή παρτιτούρα και αντ' αυτού να συνθέτει το τραγούδι στο μυαλό του και στη συνέχεια να το παίζει, να το τραγουδά ή να το ηχογραφεί από μνήμης. Παρ' όλ' αυτά, στην τζαζ και τη λαϊκή μουσική, οι ηχογραφήσεις που έχουν ξεχωρίσει και έχουν μείνει στην ιστορία, από σημαντικούς ερμηνευτές, χωρίς την συνοδεία κάποιας συγκεκριμένης παρτιτούρας αλλά μόνο με την δημιουργικότητά τους, έχουν την ίδια αξία με τις γραπτές ή τυπωμένες παρτιτούρες της κλασικής μουσικής.

Η δημιουργικότητα είναι μια πολύ σημαντική πτυχή της σύνθεσης που πρέπει να λαμβάνεται υπόψη κατά το σχεδιασμό ή την πρόταση ενός αλγορίθμου μουσικής σύνθεσης βασισμένου στην τεχνητή νοημοσύνη.

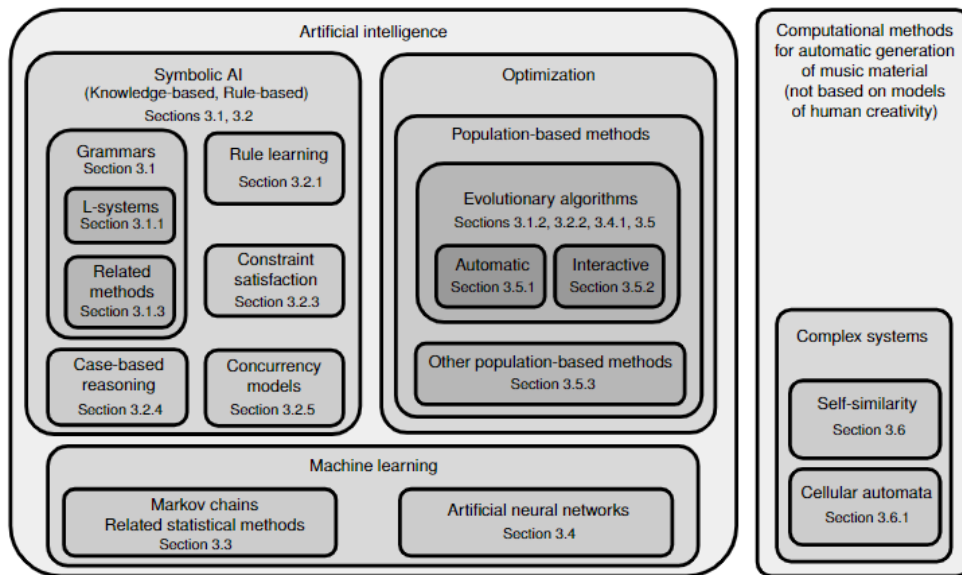
1.2 Αλγοριθμική Σύνθεση

Η αλγοριθμική σύνθεση είναι η τεχνική της χρήσης αλγορίθμων για τη δημιουργία μουσικής. Οι αλγόριθμοι χρησιμοποιούνται για τη σύνθεση μουσικής εδώ και χρόνια (Fernandez & Vico, 2013), ο όρος « αλγοριθμική σύνθεση » μπορεί να χρησιμοποιηθεί για να περιγράψει τεχνικές δημιουργίας μουσικής που εκτελούνται χωρίς συνεχή ανθρώπινη παρέμβαση, για παράδειγμα μέσω της εισαγωγής διαδικασιών τύχης (McLean & Dean, 2018) .

Ορισμένοι αλγόριθμοι (Jacob, 1996) ή δεδομένα που δεν έχουν άμεση μουσική σημασία χρησιμοποιούνται από συνθέτες ως δημιουργική έμπνευση για τη μουσική τους. Αλγόριθμοι όπως fractal (ατέρμονο μοτίβο), L-systems (μια συμβολοσειρά χαρακτήρων/συμβόλων ξαναγράφεται σε κάθε επανάληψη σύμφωνα με ορισμένους κανόνες αντικατάστασης) (Fernandez & Vico, 2013), ακόμη και στατιστικά μοντέλα (Ariza, 2011) έχουν χρησιμοποιηθεί ως πηγαίο υλικό.

Οι αλγόριθμοι σύνθεσης ταξινομούνται συνήθως με βάση τις συγκεκριμένες τεχνικές προγραμματισμού που χρησιμοποιούν. Τα αποτελέσματα της διαδικασίας μπορούν στη συνέχεια να διαχωριστούν σε μουσική που συντίθεται από υπολογιστή και σε μουσική που συντίθεται με τη βοήθεια υπολογιστή (Petit & Serrano, 2021). Οι αλγόριθμοι μπορούν είτε να παρέχουν σημειογραφικές πληροφορίες (παρτιτούρες ή MIDI) για άλλα όργανα είτε να παρέχουν έναν ανεξάρτητο τρόπο σύνθεσης ήχου (αναπαραγωγή της σύνθεσης από μόνη της). Υπάρχουν επίσης αλγόριθμοι που δημιουργούν τόσο σημειογραφικά δεδομένα όσο και σύνθεση ήχου.

Με βάση τους (Papadopoulos & Wiggins, 1999) η κατηγοριοποίηση των αλγορίθμων σύνθεσης γίνεται με βάση τη δομή τους και τον τρόπο επεξεργασίας των δεδομένων, μερικοί τύποι είναι η μετάφραση (translational models), τα μαθηματικά (mathematical models), οι γραμματικές (grammars), οι προσεγγίσεις βελτιστοποίησης (optimization approaches), οι εξελικτικές μέθοδοι (evolutionary methods) και τα συστήματα βασισμένα στη γνώση (knowledge-based systems).



Εικόνα 1 Ταξινόμηση των μεθόδων αλγοριθμικής σύνθεσης (Fernandez & Vico, 2013)

1.2.1 Μέθοδοι Αλγοριθμικής Σύνθεσης

Η μέθοδος της μετάφρασης

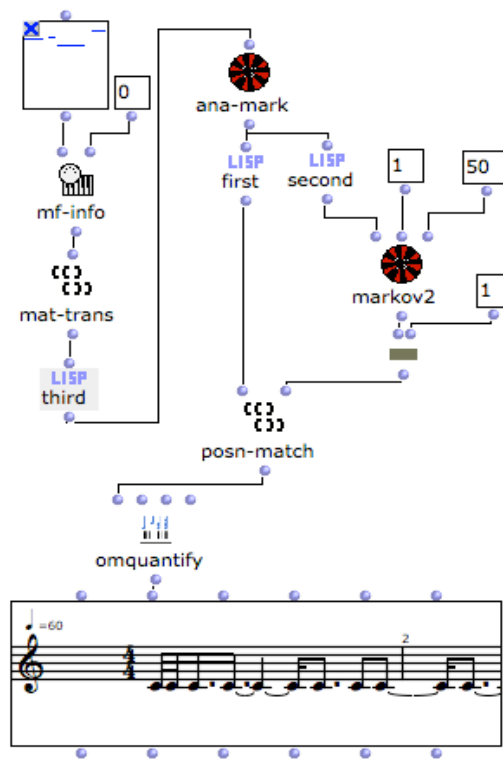
Πρόκειται για μια προσέγγιση στη σύνθεση μουσικής που περιλαμβάνει τη "μετάφραση" πληροφοριών από ένα υπάρχον μη μουσικό μέσο σε ένα νέο ήχο. Η μετάφραση μπορεί να είναι είτε βασισμένη σε κανόνες είτε στοχαστική. Για παράδειγμα, κατά τη μετάφραση μιας εικόνας σε ήχο, μια εικόνα jpeg με μια οριζόντια γραμμή μπορεί να ερμηνευτεί στον ήχο ως σταθερό ύψος, ενώ μια γραμμή με κλίση προς τα πάνω μπορεί να είναι μια αύξουσα κλίμακα. Ένα άλλο παράδειγμα είναι η μετάφραση κειμένου σε μουσική, η οποία μπορεί να προσεγγίσει τη σύνθεση εξάγοντας το συναίσθημα (θετικό ή αρνητικό) από το κείμενο χρησιμοποιώντας μεθόδους μηχανικής μάθησης, όπως η ανάλυση συναισθήματος, και αναπαριστά το συναίσθημα αυτό με όρους ποιότητας συγχοριδίων, όπως ελάσσονα (λυπημένη) ή μείζονα (χαρούμενη) συγχορδία στην παραγόμενη μουσική παραγωγή (Davis & Mohammad, 2014).

Book Title	Emotion 1	Emotion 2	Octave	Tempo	Pos/Neg	Key	Activity	Joy-Sad
<i>A Clockwork Orange</i>	Fear	Sadness	5	171	Negative	C Minor	0.009	-0.0007
<i>Alice in Wonderland</i>	Trust	Fear	5	150	Positive	C Major	0.007	-0.0002
<i>Anne of Green Gables</i>	Joy	Trust	6	180	Positive	C Major	0.010	0.0080
<i>Heart of Darkness</i>	Fear	Sadness	4	122	Negative	C Minor	0.005	-0.0060
<i>Little Prince, The</i>	Trust	Joy	5	133	Positive	C Major	0.006	0.0028
<i>Lord of The Flies</i>	Fear	Sadness	4	151	Negative	C Minor	0.008	-0.0053
<i>Peter Pan</i>	Trust	Joy	6	180	Positive	C Major	0.010	0.0040
<i>Road, The</i>	Sadness	Fear	4	42	Negative	C Minor	-0.002	-0.0080
<i>To Kill a Mockingbird</i>	Trust	Fear	5	132	Positive	C Major	0.006	-0.0013

Πίνακας 1 Χαρακτηριστικά συναισθημάτων και ήχου μερικών δημοφιλών μυθιστορημάτων που είναι επεξεργασμένα από το εργαλείο TransProse (Davis & Mohammad, 2014)

Στοχαστικά μοντέλα

Ο πιο συνηθισμένος τρόπος δημιουργίας συνθέσεων μέσω των μαθηματικών είναι οι στοχαστικές διαδικασίες. Στα στοχαστικά μοντέλα ένα μουσικό κομμάτι συντίθεται ως αποτέλεσμα μη ντετερμινιστικών μεθόδων. Η συνθετική διαδικασία ελέγχεται μόνο εν μέρει από τον συνθέτη με τη στάθμιση των πιθανοτήτων των τυχαίων γεγονότων. Εξέχοντα παραδείγματα στοχαστικών αλγορίθμων είναι οι αλυσίδες Markov και οι διάφορες χρήσεις των κατανομών Gauss. Οι στοχαστικοί αλγόριθμοι χρησιμοποιούνται συχνά μαζί με άλλους αλγορίθμους σε διάφορες διαδικασίες λήψης αποφάσεων. Η μουσική έχει επίσης συντεθεί μέσω φυσικών φαινομένων. Αυτά τα χαοτικά μοντέλα δημιουργούν συνθέσεις από τα αρμονικά και μη αρμονικά φαινόμενα της φύσης.



Εικόνα 2 Αλγοριθμική σύνθεση με Markov chains (ALGORITHMIC COMPOSER, 2010)

Η μέθοδος των γραμματικών

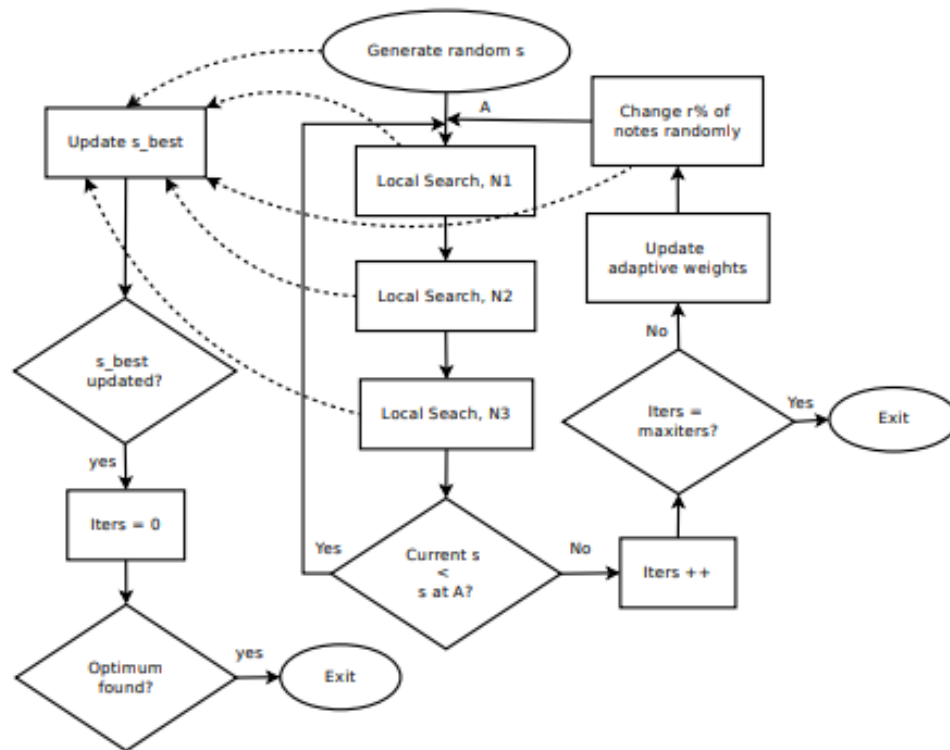
Η μουσική μπορεί επίσης να εξεταστεί ως μια γλώσσα με ένα διακριτό σύνολο γραμματικών. Οι συνθέσεις δημιουργούνται κατασκευάζοντας πρώτα μια μουσική γραμματική, η οποία στη συνέχεια χρησιμοποιείται για τη δημιουργία κατανοητών μουσικών κομματιών. Οι γραμματικές συχνά περιλαμβάνουν μουσικούς κανόνες, για παράδειγμα, με βάση την ευρωπαϊκή θεωρία μουσικής, όπως αρμονίες και ρυθμό, και όχι μεμονωμένες νότες (Fernandez & Vico, 2013).

Reference	Composition task	Comments
Lidov & Gabura, 1973	melody	early proposal
Rader, 1974	melody	early proposal, very detailed grammar
Ulrich, 1977	jazz chord identification	integrated in an <i>ad hoc</i> system (to produce jazz improvisations)
Baroni & Jacoboni, 1978	grammar for Bach chorales	early proposal
Leach & Fitch, 1995 (XComposer)	structure, rhythm and melody	uses chaotic non-linear systems (self-similarity)
Hamanaka et al., 2008	generate variations on two melodies (by altering the derivation tree)	inspired by Lerdahl et al.'s (1983) GTTM
Roads, 1977	structure, rhythm and melody	grammar compiler
Holtzman, 1981	structure, rhythm and melody	grammar compiler
Jones, 1980	structure	space grammars (uses the derivation tree)
Bel, 1992 (Bol Processor)	improvisation of <i>tabla</i> rhythms	tool for field research
Kippen & Bel, 1989	improvisation of <i>tabla</i> rhythms	grammatical inference
Cruz-Alcázar & Vidal-Ruiz, 1998	melody	grammatical inference
Gillick et al., 2009	jazz improvisation	grammatical inference. Implemented as an extension to Keller and Morrison's (2007) ImprovGenerator
Kitani & Koike, 2010 (ImprovGenerator)	real-time drum rhythm improvisation	online grammatical inference
Keller & Morrison, 2007 (Impro-Visor)	jazz improvisation	sophisticated GUI interface
Quick, 2010	classical three-voice counterpoint	integrated in a Schenkerian framework
Chemillier, 2004	jazz chord sequences	implemented in OpenMusic and MAX

Πίνακας 2 Αναφορές σε αλγοριθμική σύνθεση με γραμματικές (Fernandez & Vico, 2013)

Προσεγγίσεις βελτιστοποίησης

Κατά τη δημιουργία καλά καθορισμένων ειδών, η μουσική μπορεί να θεωρηθεί ως ένα πρόβλημα συνδυαστικής βελτιστοποίησης, όπου ο στόχος είναι να βρεθεί ο σωστός συνδυασμός νοτών έτσι ώστε να ελαχιστοποιηθεί η αντικειμενική συνάρτηση. Αυτή η αντικειμενική συνάρτηση τυπικά περιέχει κανόνες ενός συγκεκριμένου είδους, αλλά θα μπορούσε να μάθει χρησιμοποιώντας μεθόδους μηχανικής μάθησης, όπως τα μοντέλα Markov. Οι ερευνητές έχουν δημιουργήσει μουσική χρησιμοποιώντας μια πληθώρα διαφορετικών μεθόδων βελτιστοποίησης, συμπεριλαμβανομένου του ακέραιου προγραμματισμού, της αναζήτησης μεταβλητής γειτονιάς (Variable Neighborhood Search, VNS) (Herremans & Sörensen, Composing Fifth Species Counterpoint Music, 2012) και των εξελικτικών μεθόδων.



Εικόνα 3 Επισκόπηση του αλγορίθμου VNS (Herremans & Sörensen, Composing Fifth Species Counterpoint Music, 2012)

Εξελικτικές μέθοδοι

Οι εξελικτικές μέθοδοι σύνθεσης μουσικής βασίζονται σε γενετικούς αλγορίθμους. Μέσω της μετάλλαξης και της φυσικής επιλογής, διαφορετικές λύσεις εξελίσσονται προς ένα κατάλληλο μουσικό κομμάτι. Η επαναληπτική δράση του αλγορίθμου αποκόπτει τις κακές λύσεις και δημιουργεί νέες από αυτές που επιβιώνουν της διαδικασίας. Τα αποτελέσματα της διαδικασίας εποπτεύονται από τον κριτικό, ένα ζωτικό μέρος του αλγορίθμου που ελέγχει την ποιότητα των συνθέσεων που δημιουργούνται (Fox, 2006).

(1) Align parents:	
Parent1: RIFF => NOTE CHORUS => {(^5)RIFF} VERSE => {RIFF, (^3)RIFF} SONG => {VERSE, (&2)VERSE, CHORUS}	Parent2: TUNE1 => NOTE A => {TUNE1, (&5)TUNE1} B => {B, (^5)B} RONDO => {A, B, A, B, A}
(2) Convert to standard form and select genes:	
Parent1: 1 => NOTE 2 => {(^5)1} 3 => {1, (^3)1} 4 => {3, (&2)3, 2}	Parent2: 1 => NOTE 2 => {1, (&5)1} 3 => {3, (^5)3} 4 => {2, 3, 2, 3, 2}
(3) Create child from selected genes:	(4) Rename with original string names:
1 => NOTE 2 => {1, (&5)1} 3 => {3, (^5)3} 4 => {3, (&2)3, 2}	RIFF => NOTE A => {TUNE1, (&5)TUNE1} B => {B, (^5)B} SONG => {VERSE, (&2)VERSE, CHORUS}

Εικόνα 4 Παράδειγμα μιας εξελικτικής μεθόδου, της διασταύρωσης (crossover) (Fox, 2006)

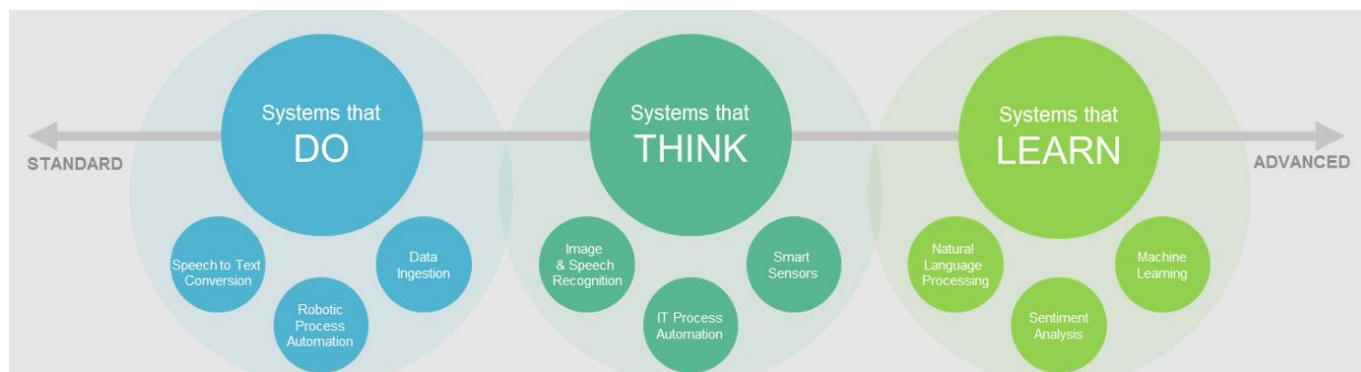
Η μέθοδος με την προσέγγιση Eno-Devo

Οι εξελικτικές μέθοδοι, σε συνδυασμό με τις αναπτυξιακές διαδικασίες, συνιστούν την προσέγγιση eno-devo για τη δημιουργία και βελτιστοποίηση πολύπλοκων δομών. Οι μέθοδοι αυτές έχουν εφαρμοστεί και στη μουσική σύνθεση, όπου η μουσική δομή προκύπτει από μια επαναληπτική διαδικασία που μετατρέπει μια πολύ απλή σύνθεση (αποτελούμενη από λίγες νότες) σε ένα σύνθετο ολοκληρωμένο κομμάτι (είτε πρόκειται για παρτιτούρα, είτε για αρχείο MIDI) (Fernandez & Vico, 2013).

Η μέθοδος με τα συστήματα μάθησης

Τα συστήματα εκμάθησης είναι προγράμματα που δεν έχουν καμία δεδομένη γνώση του είδους της μουσικής με το οποίο εργάζονται. Αντ' αυτού, συλλέγουν μόνα τους το μαθησιακό υλικό από το υλικό παραδείγματος που παρέχει ο χρήστης ή ο προγραμματιστής. Αυτή η μέθοδος αλγοριθμικής σύνθεσης συνδέεται στενά με την αλγοριθμική μοντελοποίηση του ύφους, (Dubnov, Assayag, Lartillot, & Bejerano, 2003) τον μηχανικό αυτοσχεδιασμό και μελέτες όπως η γνωστική επιστήμη και η μελέτη των νευρωνικών δικτύων. Οι Marchini και Purwins (Marchini & Purwins, 2011) παρουσίασαν ένα σύστημα που μαθαίνει

τη δομή μιας ηχητικής καταγραφής ενός ρυθμικού τμήματος κρουστών χρησιμοποιώντας μη επιβλεπόμενη ομαδοποίηση και αλυσίδες Markov μεταβλητού μήκους και που συνθέτει μουσικές παραλλαγές από αυτήν.



Εικόνα 5 (Smith, 2016)

1.3 Ιστορική Αναδρομή

1.3.1 Τεχνολογίες/Τρόποι παραγωγής μουσικής

Μέσα στο πέρασμα των χρόνων και καθώς αλλάζουν οι αιώνες ο τρόπος δημιουργίας της μουσικής εξελίσσεται. Κάθε καινούργια μέθοδος θεωρείται πρωτοπορία για την εκάστοτε χρονική περίοδο. Ειδικά, όταν παράλληλα εξελίσσεται η τεχνολογία αυτόματα εξελίσσονται οι διάφοροι τομείς που την περιτριγυρίζουν και έχουν την δυνατότητα να την χρησιμοποιήσουν με ποικίλους τρόπους. Συγκεκριμένα στην σύνθεση μουσικής τα συστήματα που παράγουν μουσική παίζουν τον πρωταγωνιστικό ρόλο. Το σύστημα είναι ένα σύνολο πραγμάτων που συνεργάζονται ως μέρη ενός μηχανισμού ή ενός διασυνδεδεμένου δικτύου, ένα πολύπλοκο σύνολο. Ακόμη μπορεί να θεωρηθεί ένα σύνολο αρχών ή διαδικασιών σύμφωνα με τις οποίες γίνεται κάτι- ένα οργανωμένο σύστημα ή μέθοδος.

Η αναδρομή ξεκινάει με το κομμάτι "In C" του Terry Riley το 1964 (Carl, 2009), είναι μια σύνθεση που εκφράζεται ως ένα σύνολο αρχών και διαδικασιών, ένα δημιουργικό κοινωνικό σύστημα. Αποτελείται από μια ακολουθία 53 σύντομων βρόχων και ένα σύνολο οδηγιών για την αναπαραγωγή τους. Κάθε εκτελεστής μπορεί να επαναλάβει κάθε βρόχο για όσο χρόνο θέλει, εφόσον συλλογικά η ομάδα παραμένει σε απόσταση λίγων βρόχων ο ένας από τον άλλον.

The image shows the musical score for "In C" by Terry Riley. It consists of 47 numbered patterns, each a short melodic phrase in C major. The score is presented on a page with a title "In C" and "Performing Directions". The page number "page 2" is visible in the top right corner. The score is written in a single staff with a treble clef and a common time signature. The patterns are numbered 1 through 47, with some patterns having multiple endings or variations. The score is accompanied by a page of instructions and a signature of Terry Riley.

Εικόνα 6 "In C", Terry Riley 1964

“Ήταν η απόλυτη διατάραξη του χρόνου όπως τον ξέραμε. Ήταν σαν να βρισκόμαστε σε μια χρονοκάψουλα και να αιωρούμαστε κάπου στο διάστημα περιμένοντας το επόμενο γεγονός να συμβεί. Και μου άρεσε αυτό το είδος της αναμονής.”, ανέφερε ο Terry Riley για το κομμάτι αυτό (Carl, 2009).

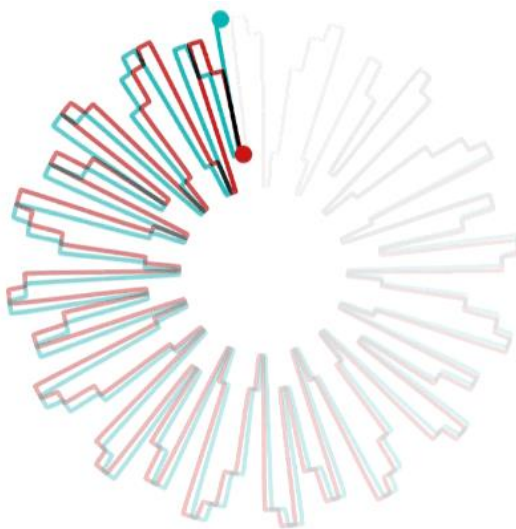
Είναι κάπως δύσκολο να το φανταστεί κανείς τώρα, αλλά στις αρχές της δεκαετίας του 1960 η όλη ιδέα των λουπών (loops), η τεχνική της επανάληψης ενός ήχου ξανά και ξανά, δεν ήταν ούτε κατά διάνοια τόσο διαδεδομένη όσο σήμερα. Αυτό οφειλόταν στο γεγονός ότι, η τεχνολογία για την κατασκευή βρόχων δεν υπήρχε για πολύ καιρό. Πραγματικά απογοιώθηκε μόνο όταν τα μαγνητόφωνα και οι μαγνητικές ταινίες έγιναν ευρέως διαδεδομένα μετά τον Β' Παγκόσμιο Πόλεμο. Οι βρόχοι μαγνητικής ταινίας, δηλαδή, η τροποποιημένη μαγνητική ταινία έτσι ώστε να παίζει σε έναν ατελείωτο βρόχο προς οποιαδήποτε κατεύθυνση, αντί μόνο προς τα εμπρός, αποδείχθηκαν ένα ισχυρό εργαλείο για πολλούς πειραματικούς μουσικούς (Carl, 2009). Ο Pierre Schaeffer και ο Pierre Henry τους χρησιμοποίησαν στα ηχητικά τους κολλάζ «musique concrete». Ο Karlheinz Stockhausen τους χρησιμοποίησε για να συνθέσει νέα ηχοχρώματα. Ο Steve Reich βρήκε σε αυτά τις σταδιακές διαδικασίες που θα διαμόρφωναν σε μεγάλο βαθμό τα πρώτα χρόνια της καριέρας του. Και οι Beatles με τον παραγωγό τους George Martin τα χρησιμοποίησαν στις καινοτομίες τους στο στούντιο που θα έφερναν επανάσταση σε ολόκληρη τη βιομηχανία ηχογραφήσεων.

Σε χρονολογική συνέχεια, έχουμε την σύνθεση του κομματιού «It's gonna rain» το 1965, όταν ο Steve Reich σχεδίασε το αφηγηματικό μηχανικό σύστημα, ήταν ένα σύστημα από μαγνητόφωνα. Ο Steve Reich είχε κάνει επιτόπιες ηχογραφήσεις ομιλίας σε κασέτα και έφτιαξε πάρα πολύ σύντομους πανομοιότυπους βρόχους για τις ηχογραφήσεις του, έναν για κάθε μαγνητόφωνο. Τα μαγνητόφωνα ήταν ρυθμισμένα σε

ελαφρώς διαφορετικές ταχύτητες. Το ένα μπορούσε να παίζει λίγο πιο γρήγορα από το άλλο. Αυτό ξεκίνησε μια διαδικασία phasing, ένα συνεχώς μεταβαλλόμενο ηχητικό τοπίο το οποίο δημιουργούνταν από τους σύντομους βρόχους.

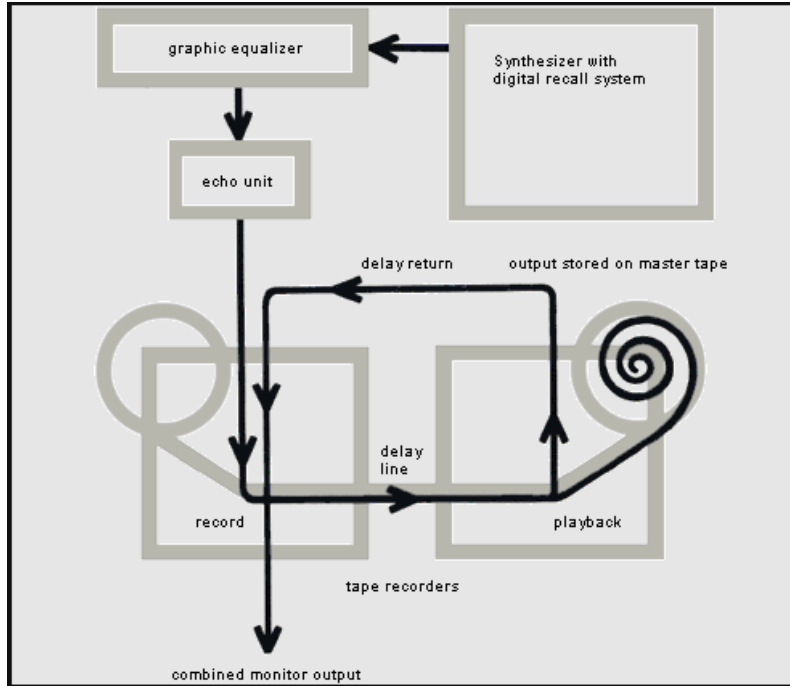
“Αν και μπορεί να έχω την ευχαρίστηση να ανακαλύπτω μουσικές διαδικασίες και να συνθέτω το μουσικό υλικό που θα τις διατρέξει, μόλις η διαδικασία ρυθμιστεί και φορτωθεί, τρέχει από μόνη της.” (Steve Reich, Music as a Gradual Process, 1968).

Μόλις ο Steve Reich ανακάλυψε αυτή την παραγωγική διαδικασία στην μαγνητική ταινία, την εφάρμοσε και σε άλλα μέσα. Για παράδειγμα το «Piano Phase», όπου είναι μια παράσταση και οπτικοποίηση του πρώτου τμήματος από το έργο «Piano Phase» του Steve Reich του 1967. Δύο πιανίστες επαναλαμβάνουν την ίδια ακολουθία δώδεκα νοτών, αλλά ο ένας επιταχύνει σταδιακά. Εδώ, τα μουσικά μοτίβα οπτικοποιούνται με τη σχεδίαση δύο γραμμών, μία ακολουθώντας κάθε πιανίστα. Παρακάτω απεικονίζεται η ροή των δυο πιανιστών με την διαφορά χρόνου.

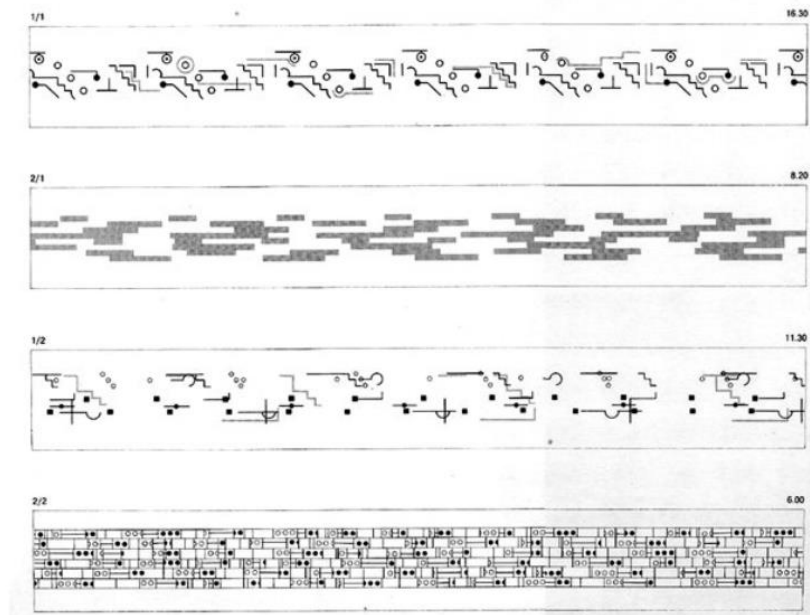


Εικόνα 7 Reich's Piano Phase (Classicfm, 2016)

Η διαδικασία phasing συνεχίζει με τον Brian Eno, ο οποίος εμπνεύστηκε από το «It's gonna rain» και χρησιμοποίησε την ίδια διαδικασία phasing σε μουσική για αεροδρόμιο το 1978. Το σχέδιό του είχε επτά μαγνητόφωνα αντί για δύο, η ταινία είχε ελαφρώς διαφορετικό μήκος, οι βρόχοι της ταινίας ήταν πολύ μεγαλύτεροι και ηχογράφησε μία μουσική νότα σε κάθε βρόχο. Επιπλέον, το άλμπουμ αποτελείται από τέσσερις συνθέσεις που δημιουργήθηκαν με τη διαστρωμάτωση βρόχων ταινίας διαφορετικού μήκους, και σχεδιάστηκε για συνεχή βρόγχο ως ηχητική εγκατάσταση, με σκοπό να εκτονώσει την τεταμένη, ανήσυχη ατμόσφαιρα ενός τερματικού σταθμού αεροδρομίου.



Εικόνα 8 Βρόχος μαγνητικής ταινίας (Scoates, 2013)



Brian Eno's graphic notation for Music for Airports Published on the back of the album sleeve

Εικόνα 9 Η γραφική σημειογραφία του Brian Eno για το "Music for Airports"

“Ήταν αρκετά σαφές με αυτά τα κομμάτια, για παράδειγμα το "In C", ότι ο συνθέτης δεν είχε μια εικόνα του τελικού κομματιού στο κεφάλι του όταν ξεκινούσε. Αυτό που είχε ο συνθέτης ήταν ένα είδος μενού, ένα πακέτο σπόρων, θα μπορούσαμε να πούμε. Και αυτοί οι μουσικοί σπόροι, μόλις φυτεύτηκαν, άνθισαν στο κομμάτι. Και άνθιζαν σε μια διαφορετική εκδοχή αυτού του κομματιού κάθε φορά” (Eno, 2011)

Το "It's Gonna Rain" και το "Music for Airports" δημιουργήθηκαν με τη χρήση παραγωγικών (generative) μεθόδων, αλλά τα αποτελέσματα δεν διατηρούν αυτόν τον παραγωγικό χαρακτήρα. Κάθε φορά που ακούγονται αυτά τα κομμάτια, ακούγεται το ίδιο πράγμα. Το "In C", αντίθετα, είναι ένα παραγωγικό προϊόν. Δεν δημιουργήθηκε μόνο μία φορά από τον συνθέτη, αλλά παράγεται εκ νέου κάθε φορά. Οποιοσδήποτε μπορεί να χρησιμοποιήσει την παρτιτούρα για να δημιουργήσει μια νέα έκδοση. Αυτό συμβαίνει κάπου στον κόσμο σχεδόν κάθε μέρα.

Αρκετά παραγωγικά συστήματα λογισμικού είναι ενσωματωμένα σε διάφορες ιστοσελίδες. Εκτός από την ευκολία παράδοσης τα συστήματα λογισμικού έχουν και μερικά άλλα πλεονεκτήματα. Ένα από αυτά είναι η ικανότητά τους να κάνουν πολλές τυχαίες επιλογές αποδοτικά, ένα άλλο είναι η τυχαιότητα, η οποία, είναι μια απίστευτα χρήσιμη παραγωγική τεχνική, για παράδειγμα μια μηχανή τυμπάνων που παράγει ένα τυχαίο μοτίβο τυμπάνων. Μπορούν να δημιουργηθούν αμέτρητα μοτίβα και ποτέ δεν θα παραχθεί το ίδιο μοτίβο δύο φορές, αυτό συμβαίνει επειδή υπάρχουν περίπου 12 τρισεκατομμύρια διαφορετικοί πιθανοί συνδυασμοί. Με κάποιο ρυθμιστικό αυξάνεται ακόμα περισσότερο η τυχαιότητα, έτσι γίνεται η εξερεύνηση των ορίων μεταξύ τάξης και χάους.

Ο John Cage ήταν πρωτοπόρος στη χρήση της τυχαιότητας στη μουσική σύνθεση. Η μέθοδός του από τη δεκαετία του 1950 ήταν να θέτει ερωτήματα αντί να κάνει επιλογές. Πρόκειται για ένα παραγωγικό σύστημα όπου θέτοντας μια ερώτηση δημιουργείτε η δυνατότητα να συμβούν όλες οι πιθανές απαντήσεις, π.χ. «τι διάρκεια πρέπει να έχει αυτή η νότα ;». Αφήνοντας στη συνέχεια μια τυχαία διαδικασία να απαντήσει στην ερώτηση, δημιουργείτε το αποτέλεσμα. Σχεδόν όλα τα παραγωγικά μουσικά συστήματα είναι επεξεργασίες αυτής της κεντρικής ιδέας. Ένα παράδειγμα είναι η μηχανή ντραμς, δηλαδή, ρωτάμε «τι μοτίβο τυμπάνων πρέπει να χρησιμοποιήσω;» και το σύστημα επιλέγει τυχαία μια απάντηση.

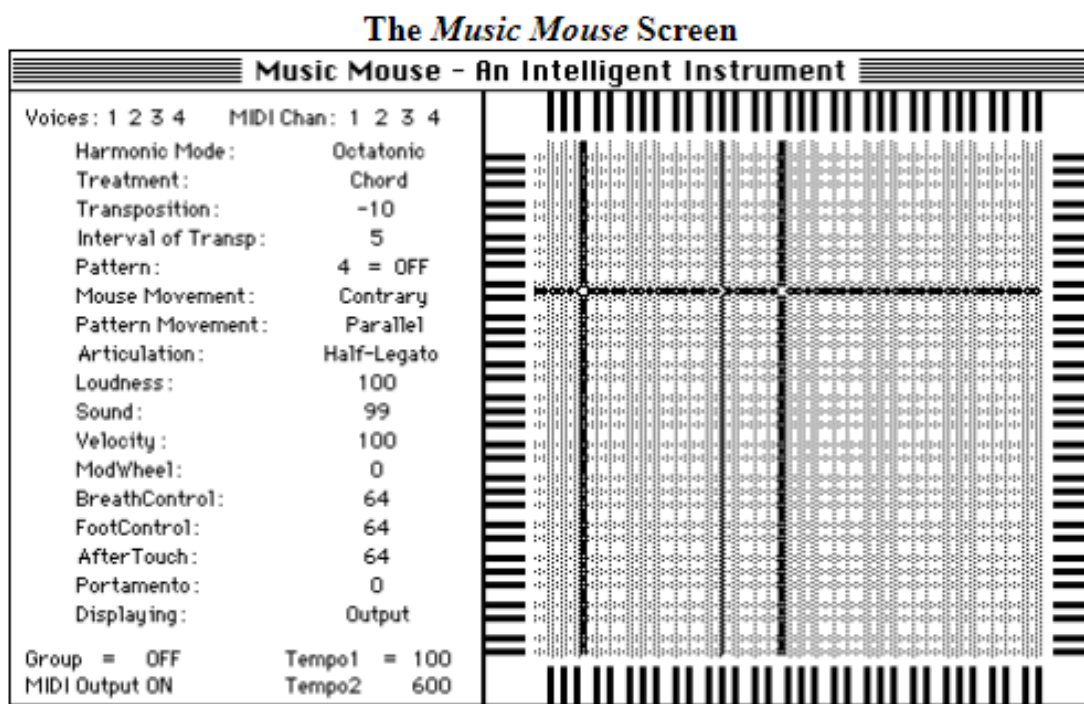
“Τότε οι απαντήσεις, αντί να προέρχονται από τις επιθυμίες μου, προέρχονται από τυχαίες λειτουργίες, και αυτό έχει ως αποτέλεσμα να μου ανοίγει πιθανότητες που δεν είχα σκεφτεί. Οι απαντήσεις που καθορίζονται από την τύχη θα ανοίξουν το μυαλό μου στον κόσμο γύρω μου.”

(John Cage, 1985)

Ένα άλλο πράγμα στο οποίο το λογισμικό είναι καλό είναι οι αλγόριθμοι. Δηλαδή, το να ακολουθείς ένα σαφές σχέδιο, να κάνεις υπολογισμούς και να κάνεις τυπικούς συλλογισμούς. Η ιδέα προέρχεται από τη γλωσσολογία, όπου αναπτύχθηκε για πρώτη φορά από τον Noam Chomsky τη δεκαετία του 1950. Για τον Noam Chomsky όλα αφορούσαν τη φυσική γλώσσα, πρόκειται για την τυποποίηση των κανόνων της γλώσσας σε ένα σύστημα που είναι ικανό να παράγει προτάσεις της γλώσσας. Μπορεί να εφαρμοστεί και στις μουσικές γλώσσες, ορίζοντας την παραγωγική γραμματική για τον ρυθμό. Για παράδειγμα ορίζουμε

συστημάτων, ένα παράδειγμα ηχογράφησης δεδομένων είναι μερικά παραγωγικά μουσικά συστήματα που βασίζονται σε γεγονότα ή δεδομένα που παράγονται από άλλα συστήματα. Τα συστήματα επιτρέπουν στον χρήστη να ολοκληρώσει τη μουσική αλλάζοντας τα όρια μεταξύ συνθέτη και χρήστη. Από τη μία πλευρά ο συνθέτης παραιτείται από τον έλεγχο χωρίς να λάβει καμία καλλιτεχνική απόφαση, και από την άλλη πλευρά, ο συνθέτης εξακολουθεί να έχει τον απόλυτο έλεγχο ως σχεδιαστής του συστήματος. Όπου υπάρχει περισσότερη ελευθερία, η αλληλεπίδραση γίνεται από κοινού δημιουργία του μουσικού και ενός παραγωγικού συστήματος.

Το «μουσικό ποντίκι» της Laurie Spiegel από το 1986, είναι ένα πρωτοποριακό παράδειγμα ενός τέτοιου συστήματος. Περιέχει αλγοριθμικές και τυχαίες διαδικασίες για τη δημιουργία και τον χρονισμό των νοτών. Ακόμη, επιτρέπει στον μουσικό να ελέγχει τη δομή της μουσικής, με αποτέλεσμα την συνεργασία μεταξύ του συνθέτη και του μουσικού ποντικιού.



Music Mouse parameter set as displayed in the new Atari ST version, and in the Dec. 1988 update to the Macintosh version. The Amiga version of Music Mouse features all the same live keyboard controls, but does not show them on-screen because it is an audioVISUAL instrument, with drawing modes, color faders, etc.

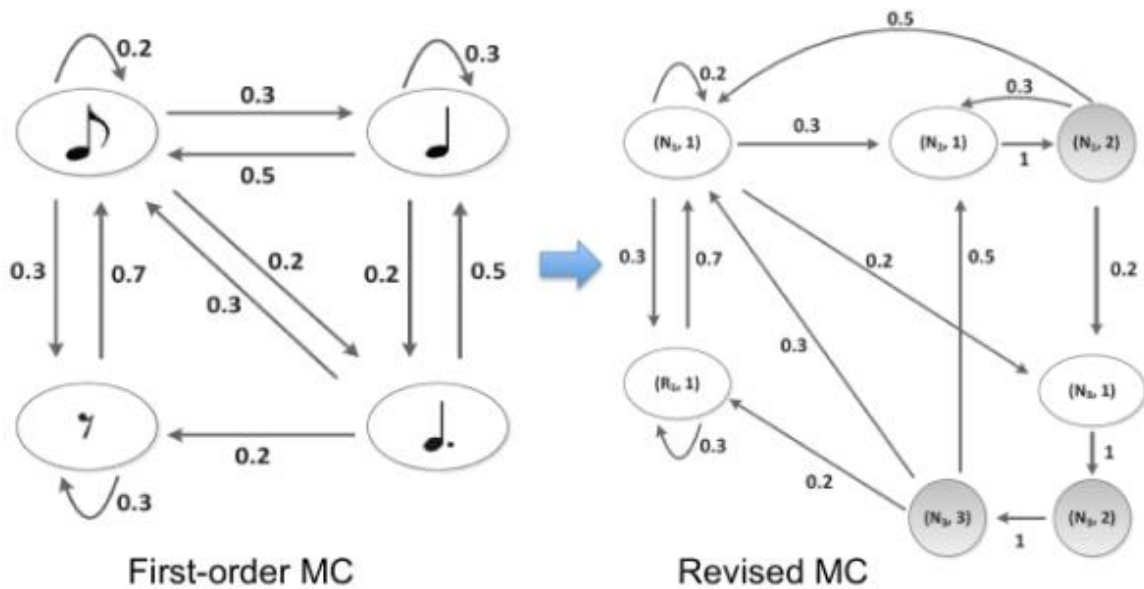
Εικόνα 11 Music Mouse (Computer Software by Laurie Spiegel)

“Ξέρω ότι πολλοί από τους τρόπους με τους οποίους η μουσική ενσωματώνεται στη ζωή μας αλλάζουν.

Μια από τις μεγάλες αλλαγές είναι ότι η μουσική γίνεται μια διαδικασία στην οποία οι άνθρωποι μπορούν να συμμετέχουν - αντί για ένα μάτσο σταθερές, πεπερασμένες οντότητες που ονομάζονται κομμάτια που μπορείς να ακούσεις και είναι ίδια κάθε φορά.” (Laurie Spiegel, 1987)

Μια άλλη μορφή αλληλεπίδρασης ανθρώπου-μηχανής είναι η εκπαίδευση των μηχανών για την εκτέλεση διαφόρων διαδικασιών. Αντί να σχεδιαστεί ένα σύστημα παραγωγής, μπορεί ένας υπολογιστής να μάθει ένα τέτοιο σύστημα, δείχνοντάς του παραδείγματα του είδους της μουσικής που θέλει κάποιος να παράγει, αυτό είναι μηχανική μάθηση.

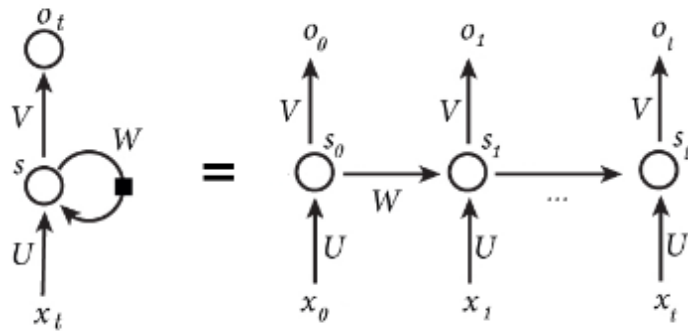
Οι αλυσίδες Markov είναι ένα παλιό, αλλά ακόμα αποτελεσματικό παράδειγμα της προσέγγισης της μηχανικής μάθησης. Μια αλυσίδα Markov είναι ένα μαθηματικό μοντέλο του που απεικονίζει την ακολουθία των γεγονότων με την πάροδο του χρόνου, όπως οι νότες σε μια μελωδία. Επιπροσθέτως, μια αλυσίδα Markov μπορεί να εκπαιδευτεί παίζοντάς της μουσική, έτσι μαθαίνει τη μουσική που παίζετε και δημιουργεί ένα γράφημα γι' αυτήν. Στη συνέχεια, η δημιουργία νέας μουσικής βάζοντας την αλυσίδα Markov να περπατήσει τυχαία στο γράφημά της, ακολουθώντας τις συνδέσεις, που έμαθε είναι δύσκολη, διότι είναι μια τόσο περιορισμένη μέθοδος που ένα τόσο απλό μοντέλο μπορεί να συλλάβει μόνο ένα περιορισμένο ποσό πληροφοριών. Σε αυτή την περίπτωση μαθαίνει μόνο για τη βραχυπρόθεσμη μελωδική εξέλιξη που μπορεί όμως να είναι ακόμα χρήσιμη.



Εικόνα 12 Markov Chains με μουσική σημειογραφία (Kim & Yeo, 2013)

Οι αλυσίδες Markov αποτελούν πρόγονο των σύγχρονων προσεγγίσεων μηχανικής μάθησης για τη δημιουργία μουσικής, όπως τα βαθιά νευρωνικά δίκτυα.

Τέλος, ο πιο σύγχρονος τρόπος για την εποχή μας είναι τα βαθιά νευρωνικά δίκτυα, τα οποία είναι εξαιρετικά πολύπλοκα στατιστικά μοντέλα και συνήθως εκπαιδεύονται χρησιμοποιώντας μεγάλο όγκο δεδομένων αλλά βασικά λειτουργούν ακριβώς όπως οι αλυσίδες Markov, η μηχανή μαθαίνει πώς λειτουργεί η μουσική δείχνοντάς της παραδείγματα. Δημιουργεί ένα μοντέλο της μουσικής και στη συνέχεια χρησιμοποιεί το μοντέλο για να παράγει νέα μουσική.



Εικόνα 13 Ένα ζευτελιγμένο νευρωνικό δίκτυο RNN (Malik & Ek, 2017)

2 Τεχνητή νοημοσύνη και μουσική

Ένας κοινός ερευνητικός στόχος του τομέα της Μουσικής και του τομέα της Τεχνητής Νοημοσύνης είναι η αυτοματοποίηση των μουσικών εργασιών και η ανάπτυξη αυτόνομων δημιουργικών συστημάτων, δηλαδή, συστήματα που είναι ικανά να «συνθέτουν» μουσική αυτόνομα ενός συγκεκριμένου είδους. Τέτοια συστήματα ασχολούνται με την προσομοίωση των συνθετών, με στόχο την παραγωγή αποτελεσμάτων που μπορούν να συγκριθούν με αυτά που παράγονται από τους ίδιους τους συνθέτες. Οι πρόσφατες εξελίξεις στη βαθιά μάθηση έχουν ως αποτέλεσμα, το καλωσόρισμα μιας νέας εποχής για τα συστήματα παραγωγής μουσικής, τα οποία μπορούν, πλέον, να ξεπεράσουν το πρόβλημα της μουσικής αναπαράστασης, μαθαίνοντας απευθείας από ακατέργαστα δεδομένα ήχου (δηλαδή κυματομορφές), με μια επιβάρυνση, της υψηλότερης υπολογιστικής ισχύος και του μεγαλύτερου χρόνου εκπαίδευσης. Τα προηγούμενα συστήματα αυτόματης σύνθεσης βασίζονταν σε αναπαραστάσεις όπως αρχεία MIDI ή ανάλυση μουσικών παρτιτούρων, οι εφαρμογές βαθιάς μάθησης (π.χ. Wavenet) μπορούν να μάθουν από μη δομημένα δεδομένα, δηλαδή, από μη επεξεργασμένο ήχο, και με αυτόν τον τρόπο απλοποιούν τη διαδικασία εκπαίδευσης, ενώ παράλληλα διευρύνουν θεωρητικά την περιοχή της εφαρμογής των συστημάτων αυτόματης σύνθεσης σε μουσικές διαλέκτους που θέτουν προκλήσεις για τη μουσική αναπαράσταση. Επιπλέον, τα αυτόνομα δημιουργικά συστήματα στερούν στους καλλιτέχνες την συνεργατικότητα και την διάδραση των δυνατοτήτων τους. Σκοπός μιας τέτοιας προσέγγισης είναι να ενισχύσει την ανθρώπινη δημιουργικότητα, ανοίγοντας νέες δημιουργικές δυνατότητες και προκαλώντας τις παραδοσιακές έννοιες της δημιουργίας και των ορισμών του μουσικού έργου (Miranda, 2021).

Με τα νέα εργαλεία «σύνθεσης» έρχονται νέες συνθετικές δυνατότητες και δυνητικά νέες καλλιτεχνικές έννοιες και πρακτικές. Τα εργαλεία ασκούν τότε έναν τύπο πράξης που έχει τη δυνατότητα να τροποποιεί την «κατάσταση των πραγμάτων» της καλλιτεχνικής παραγωγής, εμπνέοντας και συμβάλλοντας στην καλλιτεχνική καινοτομία. Υπό αυτή την έννοια, τα εργαλεία μπορούν να θεωρηθούν ως «δευτερεύοντες» βοηθοί, με την ιδιαιτερότητα του να διευρύνουν το χώρο των δημιουργικών δυνατοτήτων του συνθέτη και να επηρεάζουν τη μουσική του σκέψη. Για παράδειγμα, ένας μουσικός βοηθός μπορεί να χρησιμοποιήσει αλγορίθμους ομαδοποίησης για να ομαδοποιήσει ήχους με βάση την ομοιότητά τους, ή αλγορίθμους ταξινόμησης για την ταξινόμηση των εισερχόμενων ήχων και να παράγει τις κατάλληλες απαντήσεις.

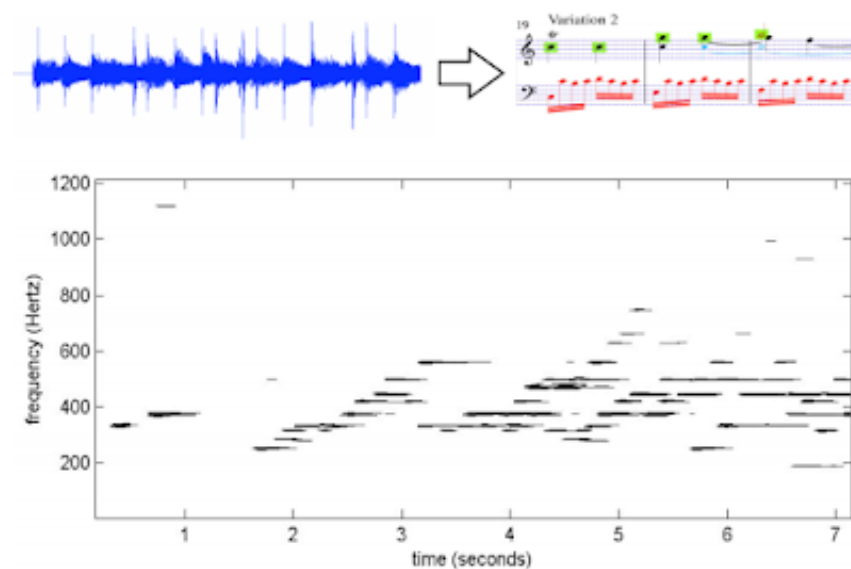
Τέτοιες ιδιότητες επιτρέπουν την αλληλεπίδραση λειτουργιών, όπως η αναγνώριση και η ανταπόκριση σε συγκεκριμένα ηχοχρώματα ή μουσικές χειρονομίες, ερμηνεύοντας ροές υψηλής διάστασης χαρακτηριστικών σε επίπεδο σήματος και σε πραγματικό χρόνο. Ακόμη, εφαρμογές, ίδιας κατηγορίας, αποδεικνύουν τις μοναδικές και ενδεχομένως μετασχηματιστικές δυνατότητες των εργαλείων τεχνητής νοημοσύνης για τη σύνθεση μουσικής. Συνεπώς, η τεχνητή νοημοσύνη θεωρείται ένας παράγοντας που

συμβάλλει στον συνδυασμό της δημιουργικότητας του ανθρώπου και της βοήθειας του υπολογιστή να εξελίξει τον άνθρωπο και όχι ως υποκατάστατο της ανθρώπινης δημιουργικότητας.

2.1 Μηχανική Μάθηση

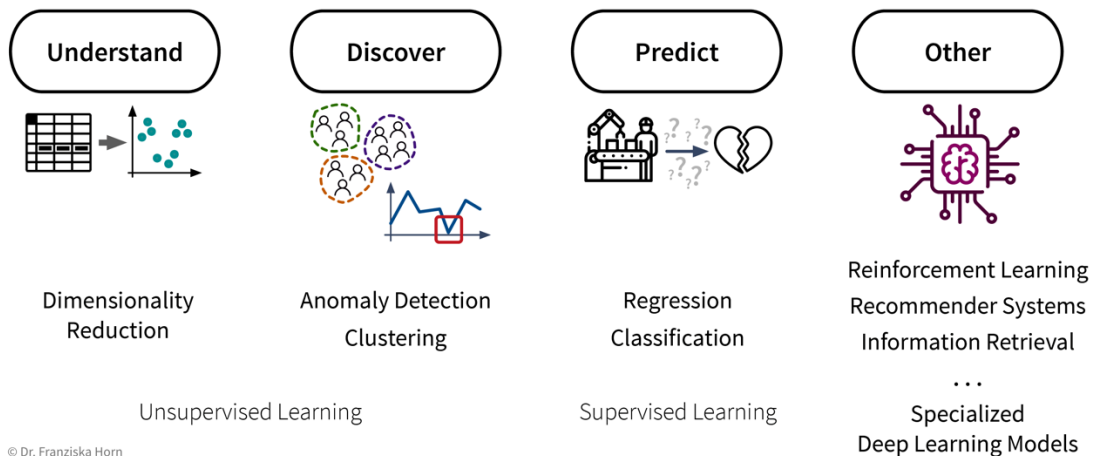
Στην σύνθεση μουσικής με την χρήση μηχανικής μάθησης είναι πιθανόν να συνδυαστούν διάφοροι τομείς της επιστήμης των υπολογιστών, όπως η ανάκτηση μουσικής πληροφορίας (Music Information Retrieval, MIR) όπου ασχολείται με την εξαγωγή πληροφοριών από τη μουσική. Αυτό μπορεί να περιλαμβάνει εργασίες όπως η μεταγραφή ήχου σε παρτιτούρες, ο προσδιορισμός του κλειδιού ενός κομματιού, η αναγνώριση του ρυθμού ενός τραγουδιού αλλά και η οργάνωση μιας μουσικής βιβλιοθήκης. Κάποιες από τις παραπάνω εργασίες συμβάλλουν σημαντικά στην μουσική σύνθεση.

Ο ρόλος της μηχανικής μάθησης στο MIR είναι η βελτίωση της ακρίβειας και της αποτελεσματικότητας αυτών των εργασιών. Για παράδειγμα, τα νευρωνικά δίκτυα συνελκτικής μάθησης (CNN) έχουν χρησιμοποιηθεί (Troxel, 2016) για τη μεταγραφή ήχου σε παρτιτούρες. Ένα άλλο παράδειγμα (Sigtia, et al., 2014) αλγορίθμων μηχανικής μάθησης είναι τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) που έχουν χρησιμοποιηθεί για τον εντοπισμό του κλειδιού και του ρυθμού ενός τραγουδιού. Επιπλέον, μοντέλα βασισμένα στη μηχανική μάθηση έχουν χρησιμοποιηθεί για τη δημιουργία μουσικής, όπως η μελωδία και η εξέλιξη των συγχορδιών. Σημαντικό γεγονός αποτελεί η κατανόηση της λειτουργίας των αλγορίθμων μηχανικής μάθησης, διότι συμβάλλει στην καλύτερη απόδοση και εξέλιξη του προγράμματος.



Εικόνα 14 Μετατροπή ήχου σε παρτιτούρα

Εμβαθύνοντας, οι αλγόριθμοι περιλαμβάνουν την μείωση της διάστασης, την ομαδοποίηση, την πρόβλεψη και την ενισχυτική μάθηση. Η μείωση της διάστασης είναι μια βασική τεχνική στο πλαίσιο της μάθησης χωρίς επίβλεψη, συμπιέζει τα δεδομένα με την εύρεση ενός μικρότερου, διαφορετικού συνόλου μεταβλητών που αποτυπώνουν ό,τι έχει μεγαλύτερη σημασία στα αρχικά χαρακτηριστικά, ελαχιστοποιώντας παράλληλα την απώλεια πληροφοριών (Tatsat, Puri, & Lookabaugh, 2020). Σε αντίθεση με τη μάθηση με επίβλεψη, η μάθηση χωρίς επίβλεψη περιλαμβάνει είσοδο αλλά όχι δεδομένα εξόδου. Το έργο του αλγορίθμου σε αυτή την περίπτωση είναι να βρει δομή στα δεδομένα. Για το παράδειγμα, οι αλγόριθμοι ομαδοποίησης μπορούν να χρησιμοποιηθούν για να χωρίσουν ένα σύνολο ήχων σε μικρότερες ομάδες (clusters) με βάση τη φασματική ομοιότητα. Μια άλλη εργασία μηχανικής μάθησης με εφαρμογές στη μουσική είναι η πρόβλεψη, η οποία περιλαμβάνει την πρόβλεψη μελλοντικών γεγονότων με βάση τα ιστορικά δεδομένα. Τα πιθανοτικά μοντέλα πρόβλεψης μπορούν να χρησιμοποιηθούν για να μάθουν και να παράγουν νέα δεδομένα, όπως το να παράγουν μελωδίες χρησιμοποιώντας υπό συνθήκη πιθανότητες (Rasmussen & Williams, 2006). Τέλος, η ενισχυτική μάθηση είναι μια τεχνική μηχανικής μάθησης που βασίζεται στην ανατροφοδότηση, στην οποία ένας βοηθός (μαθαίνει να συμπεριφέρεται σε ένα περιβάλλον εκτελώντας τις ενέργειες και βλέποντας τα αποτελέσματα των ενεργειών, για κάθε καλή ενέργεια, ο πράκτορας λαμβάνει θετική ανατροφοδότηση και για κάθε κακή ενέργεια, ο πράκτορας λαμβάνει αρνητική ανατροφοδότηση ή ποινή (Javatpoint).



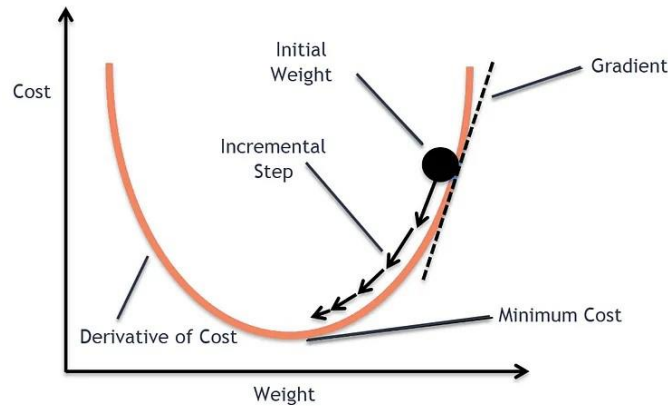
Εικόνα 15 Τύποι μηχανικής μάθησης (Horn)

2.1.1 Αλγόριθμοι Μηχανικής Μάθησης

Stochastic Gradient Descent (SGD)

Ο αλγόριθμος Gradient Descent είναι μια επαναληπτική διαδικασία βελτιστοποίησης που αναζητά για τη βέλτιστη τιμή μιας αντικειμενικής συνάρτησης (ελάχιστο/μέγιστο). Είναι μία από τις πλέον διαδεδομένες μεθόδους που χρησιμοποιείται για την αλλαγή των παραμέτρων ενός μοντέλου προκειμένου να μειώσει τον κόστους της πολυπλοκότητας σε έργα μηχανικής μάθησης. Ο πρωταρχικός στόχος του αλγορίθμου (GD) είναι ο εντοπισμός των παραμέτρων του μοντέλου που παρέχουν τη μέγιστη ακρίβεια τόσο στα σύνολα δεδομένων εκπαίδευσης όσο και στα σύνολα δεδομένων δοκιμής (Ketkar, 2017). Στον Gradient Descent, η κλίση είναι ένα διάνυσμα που δείχνει στη γενική κατεύθυνση της πιο απότομης ανόδου της συνάρτησης σε ένα συγκεκριμένο σημείο, δηλαδή, ο αλγόριθμος μπορεί να πέσει σταδιακά προς χαμηλότερες τιμές της συνάρτησης κινούμενος προς την αντίθετη κατεύθυνση της κλίσης, μέχρι να φτάσει στο ελάχιστο της συνάρτησης (Ketkar, 2017).

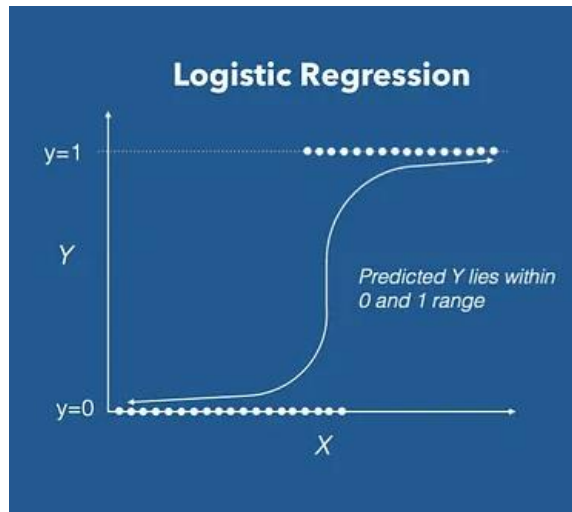
Ένας τύπος αλγορίθμου Gradient Descent είναι ο στοχαστικός (Stochastic) Gradient Descent (SGD) είναι μια παραλλαγή του Gradient Descent αλγορίθμου που χρησιμοποιείται για τη βελτιστοποίηση μοντέλων μηχανικής μάθησης. Αντιμετωπίζει την υπολογιστική αναποτελεσματικότητα του παραδοσιακού αλγορίθμου Gradient Descent κατά την επεξεργασία μεγάλων συνόλων δεδομένων σε έργα μηχανικής μάθησης (Ketkar, 2017). Στο SGD, αντί να χρησιμοποιείται ολόκληρο το σύνολο των δεδομένων για κάθε επανάληψη, μόνο ένα τυχαίο παράδειγμα εκπαίδευσης (ή μια μικρή παρτίδα) επιλέγεται για να υπολογισμός της κλίσης και την ενημέρωση των παραμέτρων του μοντέλου. Αυτή η τυχαία επιλογή εισάγει τυχαιότητα στη διαδικασία βελτιστοποίησης, ως εκ τούτου, προκύπτει και ο όρος "στοχαστική". Το πλεονέκτημα της χρήσης της SGD είναι ότι με τη χρήση ενός μόνο παραδείγματος ή μιας μικρής παρτίδας, το υπολογιστικό κόστος ανά επανάληψη μειώνεται σημαντικά σε σύγκριση με τις παραδοσιακές μεθόδους Gradient Descent που απαιτούν την επεξεργασία ολόκληρου του σύνολο δεδομένων (Bottou, 2010).



Εικόνα 18 Απεικόνιση συνάρτησης του αλγορίθμου *Stochastic Gradient Descent* (Dawar, 2020)

Logistic Regression

Η λογιστική παλινδρόμηση είναι ένας αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται για τα προβλήματα ταξινόμησης, είναι ένας αλγόριθμος προγνωστικής ανάλυσης και βασίζεται στην έννοια της πιθανότητας. Παρομοιάζεται με την γραμμική παλινδρόμηση, αλλά η *Logistic Regression* χρησιμοποιεί μια πιο σύνθετη συνάρτηση κόστους, αυτό το κόστος συνάρτηση μπορεί να οριστεί ως "σιγμοειδής συνάρτηση" ή επίσης γνωστή ως "λογιστική συνάρτηση" αντί για γραμμική συνάρτηση. Η υπόθεση της λογιστικής παλινδρόμησης τείνει να περιορίζει τη συνάρτηση κόστους μεταξύ 0 και 1. Ως εκ τούτου, οι γραμμικές συναρτήσεις αποτυγχάνουν να την αναπαραστήσουν, καθώς μπορεί να έχει τιμή μεγαλύτερη από 1 ή μικρότερη από 0, πράγμα που δεν είναι δυνατό σύμφωνα με την υπόθεση της λογιστικής παλινδρόμησης. Η λογιστική παλινδρόμηση είναι ένας δημοφιλής αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται για προβλήματα δυαδικής ταξινόμησης. Μοντελοποιεί τη σχέση μεταξύ ενός εξαρτημένης μεταβλητής (δυαδική) και μιας ή περισσότερων ανεξάρτητων μεταβλητών με εκτιμώντας τις πιθανότητες με τη χρήση μιας λογιστικής συνάρτησης. Είναι ευρέως διαδεδομένη χρησιμοποιείται ευρέως σε διάφορους τομείς, όπως η χρηματοδότηση, η υγειονομική περίθαλψη και η κοινωνική επιστήμες.



Εικόνα 18 Απεικόνιση συνάρτησης του αλγορίθμου Logistic Regression

Naive Bayes

Ο Naive Bayes είναι ένας αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται συνήθως για εργασίες ταξινόμησης. Βασίζεται στο θεώρημα του Bayes και υποθέτει ότι τα χαρακτηριστικά είναι, υπό όρους, ανεξάρτητα μεταξύ τους. Ο αλγόριθμος Naive Bayes έχει αποδειχθεί ότι είναι αποτελεσματικός σε πολλές εφαρμογές του πραγματικού κόσμου και χρησιμοποιείται ευρέως σε διάφορους τομείς, όπως η ταξινόμηση κειμένων, το φιλτράρισμα ανεπιθύμητων μηνυμάτων και το συναίσθημα (Atmadja, Uriawan, Pritisen, Maylawati, & Arbain, 2019).

Η θεμελιώδης ιδέα πίσω από τον αλγόριθμο Naive Bayes είναι ο υπολογισμός της πιθανότητας ενός δεδομένου περίπτωση να ανήκει σε μια συγκεκριμένη κλάση χρησιμοποιώντας την υπό συνθήκη πιθανότητα της κλάσης δεδομένων των χαρακτηριστικών της περίπτωσης. (Atmadja, Uriawan, Pritisen, Maylawati, & Arbain, 2019) Ο αλγόριθμος υπολογίζει αυτές τις πιθανότητες για κάθε κλάση και επιλέγει την κλάση με την υψηλότερη πιθανότητα, ως την προβλεπόμενη κλάση για την περίπτωση. Ο Naive Bayes έχει διάφορες παραλλαγές, συμπεριλαμβανομένων των Gaussian Naive Bayes, Multinomial Naive Bayes και Bernoulli Naive Bayes, κάθε ένας από τους οποίους είναι κατάλληλος για διαφορετικούς τύπους δεδομένων και παραδοχών, σχετικά με την κατανομή των χαρακτηριστικών.

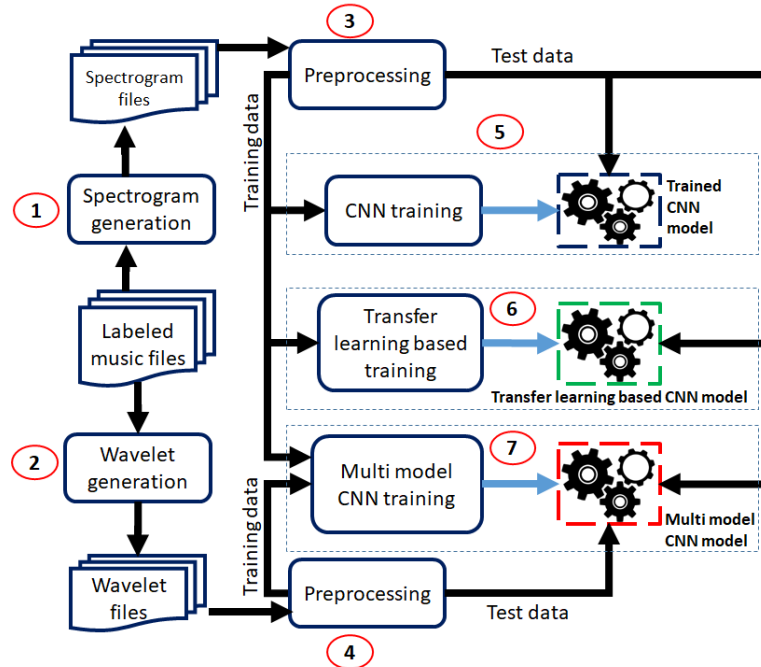
The diagram shows the Bayes' theorem equation:
$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$
 Four text boxes with arrows pointing to parts of the equation: 1. Top-left: "Η πιθανότητα των αποδείξεων ότι η υπόθεση είναι αληθινή" (The probability of evidence that the hypothesis is true) points to P(H|E). 2. Top-right: "Η προηγούμενη πιθανότητα της υπόθεσης" (The prior probability of the hypothesis) points to P(H). 3. Bottom-left: "Η μεταγενέστερη πιθανότητα της υπόθεσης με βάση τα στοιχεία είναι αληθινή" (The posterior probability of the hypothesis based on the evidence is true) points to P(H|E). 4. Bottom-right: "Η προηγούμενη πιθανότητα ότι τα στοιχεία είναι αληθινά" (The prior probability that the evidence is true) points to P(E).

Εικόνα 16 Απεικόνιση του μαθηματικού τύπου που βασίζεται ο αλγόριθμος Naive Bayes

2.2 Βαθιά Μάθηση

Η βαθιά μάθηση είναι ένα υποπεδίο της μηχανικής μάθησης που χρησιμοποιεί τεχνητά νευρωνικά δίκτυα για να μιμηθεί τη διαδικασία μάθησης του ανθρώπινου εγκεφάλου. Οι αλγόριθμοι βαθιάς μάθησης χρησιμοποιούν νευρωνικά δίκτυα με πολλαπλά επίπεδα για να μαθαίνουν από δεδομένα και να κάνουν προβλέψεις για αυτά. Αυτά τα νευρωνικά δίκτυα μαθαίνουν πολύπλοκα μοτίβα και μουσικές αναπαραστάσεις, γεγονός που τα καθιστά κατάλληλα για εργασίες όπως η αναγνώριση εικόνων και ήχων. Οι περισσότεροι αλγόριθμοι μηχανικής μάθησης λειτουργούν καλά σε σύνολα δεδομένων που έχουν έως και μερικές εκατοντάδες χαρακτηριστικά ή στήλες.

Οι αλγόριθμοι, που αναφέρθηκαν παραπάνω, ονομάζονται τεχνητά νευρωνικά δίκτυα (Artificial Neural Network, ANN), τα οποία έχουν σχεδιαστεί για να μαθαίνουν μοτίβα και σχέσεις από δεδομένα. Ακόμη, μπορούν να χρησιμοποιηθούν για ένα ευρύ φάσμα εργασιών, όπως η αναγνώριση εικόνων, η επεξεργασία φυσικής γλώσσας και η σύνθεση μουσικής. Συγκεκριμένα, η σύνθεση μουσικής με τη χρήση των αλγορίθμων βαθιάς μάθησης συμβαίνει συνήθως με την εκπαίδευση ενός τεχνητού νευρωνικού δικτύου (ANN) από ένα σύνολο δεδομένων μουσικής, όπως αρχεία MIDI ή ηχογραφήσεις, και στη συνέχεια το εκπαιδευμένο δίκτυο συμβάλλει στην δημιουργία νέας μουσικής. Οι διάφοροι τύποι αλγορίθμων βαθιάς μάθησης που έχουν χρησιμοποιηθεί για την παραγωγή μουσικής είναι: τα επαναλαμβανόμενα νευρωνικά δίκτυα (Recurrent Neural Network, RNN), τα δίκτυα μακράς βραχυπρόθεσμης μνήμης (Long Short-Term Memory, LSTM) και τα παραγωγικά αντιθετικά δίκτυα (Generative Adversarial Networks, GAN).



Εικόνα 17 Παράδειγμα ταξινόμησης μουσικών ειδών με χρήση τεχνικών βαθιάς μάθησης (Rai, 2021)

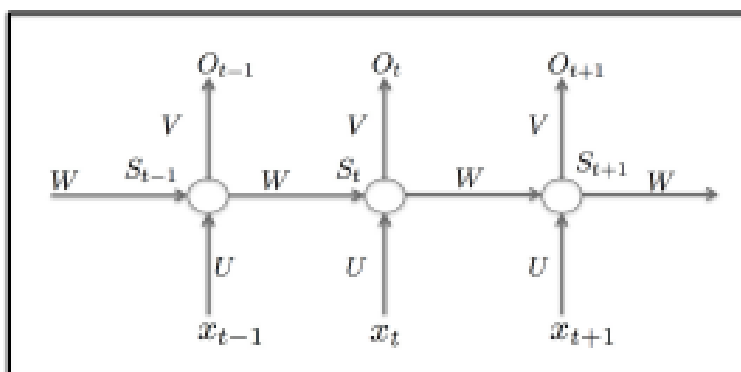
Προσεγγίζοντας την σύνθεση μουσικής η χρήση των RNN είναι κλασική μέθοδος, τα RNN είναι κατάλληλα για την επεξεργασία διαδοχικών δεδομένων όπως η μουσική. Επιπλέον, μπορούν να εκπαιδευτούν για να παράγουν μουσική που είναι παρόμοια με τα δεδομένα εκπαίδευσης και μπορούν επίσης να χρησιμοποιηθούν για να παράγουν νέες παραλλαγές υπάρχουσας μουσικής. Από την άλλη πλευρά έχουμε τα GAN, τα οποία αποτελούνται από δύο νευρωνικά δίκτυα, ένα δίκτυο γεννήτριας το οποίο παράγει νέα μουσική και ένα δίκτυο διάκρισης που προσπαθεί να διακρίνει την παραγόμενη μουσική από την πραγματική μουσική. Τα δύο δίκτυα εκπαιδεύονται παράλληλα, δηλαδή, το δίκτυο γεννήτριας προσπαθεί να παράγει μουσική που μπορεί να κάνει το δίκτυο διάκρισης να νομίζει ότι είναι πραγματική.

Συνοψίζοντας, η σύνθεση/παραγωγή μουσικής με τη χρήση της βαθιάς μάθησης είναι ένας ενεργός τομέας έρευνας και υπάρχουν πολλές διαφορετικές τεχνικές και μοντέλα που έχουν προταθεί. Το αποτέλεσμα, δηλαδή, η μουσική που παράγεται δεν είναι πάντα υψηλής ποιότητας, αλλά βελτιώνεται με την πάροδο του χρόνου καθώς η τεχνολογία και τα μοντέλα αναπτύσσονται ραγδαία στην εποχή μας.

2.2.1 Αλγόριθμοι Βαθιάς Μάθησης

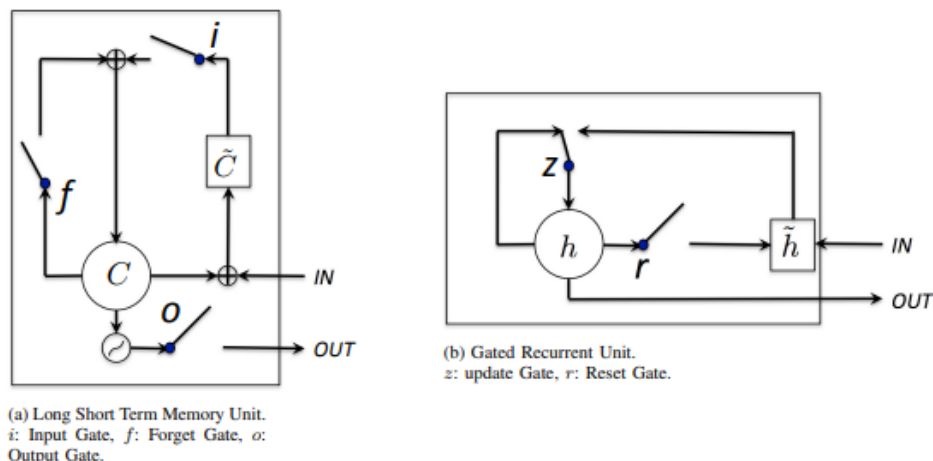
Recurrent Neural Networks (RNN)

Τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) είναι ένας τύπος νευρωνικού δικτύου που έχει σχεδιαστεί για να επεξεργάζεται διαδοχικά δεδομένα, όπως δεδομένα χρονοσειρών ή φυσική γλώσσα. Το βασικό χαρακτηριστικό των RNN είναι ότι διαθέτουν συνδέσεις ανατροφοδότησης, οι οποίες επιτρέπουν στο δίκτυο να διατηρεί μια κρυφή κατάσταση που μπορεί να αποθηκεύει πληροφορίες σχετικά με τις προηγούμενες εισόδους στην ακολουθία. Άρα, τα RNN μπορούν να ξετυλιχθούν στο χρόνο και να απεικονιστούν ως μια αλυσίδα επαναλαμβανόμενων μονάδων νευρωνικού δικτύου, όπου κάθε μονάδα λαμβάνει ως είσοδο την τρέχουσα είσοδο καθώς και την κρυφή κατάσταση από την προηγούμενη μονάδα (Liu & Ramakrishnan, 2014).



Εικόνα 18 Επαναλαμβανόμενα Νευρωνικά Δίκτυα (Rana, και συν., 2016)

Υπάρχουν διάφοροι τύποι RNN, ένας από αυτούς είναι ο vanilla RNN, αποτελεί την απλούστερη μορφή RNN, έχει ένα μόνο επίπεδο και απλές ενημερώσεις για την κρυφή κατάσταση. Ένας άλλος τύπος είναι ο Long Short-Term Memory (LSTM) RNN, έχει ένα πλήρως συνδεδεμένο κρυφό επίπεδο από μπλοκ μνήμης, το οποίο περιέχει μια ή περισσότερες μονάδες (κύτταρα μνήμης) και πύλες για τον έλεγχο της ροής των πληροφοριών στην κρυφή κατάσταση. Επιπλέον, ο τύπος Gated Recurrent Unit (GRU) RNN, είναι παρόμοιος με τον LSTM, αλλά έχει λιγότερες παραμέτρους και είναι υπολογιστικά λιγότερο δαπανηρός. Συνδυάζει τις πύλες μνήμης και εισόδους σε μια ενιαία "πύλη ενημέρωσης" και διαθέτει μια πρόσθετη "πύλη επαναφοράς". Το τελικό μοντέλο είναι απλούστερο από τα τυπικά μοντέλα LSTM και γίνεται όλο και πιο δημοφιλές (Rana, και συν., 2016).



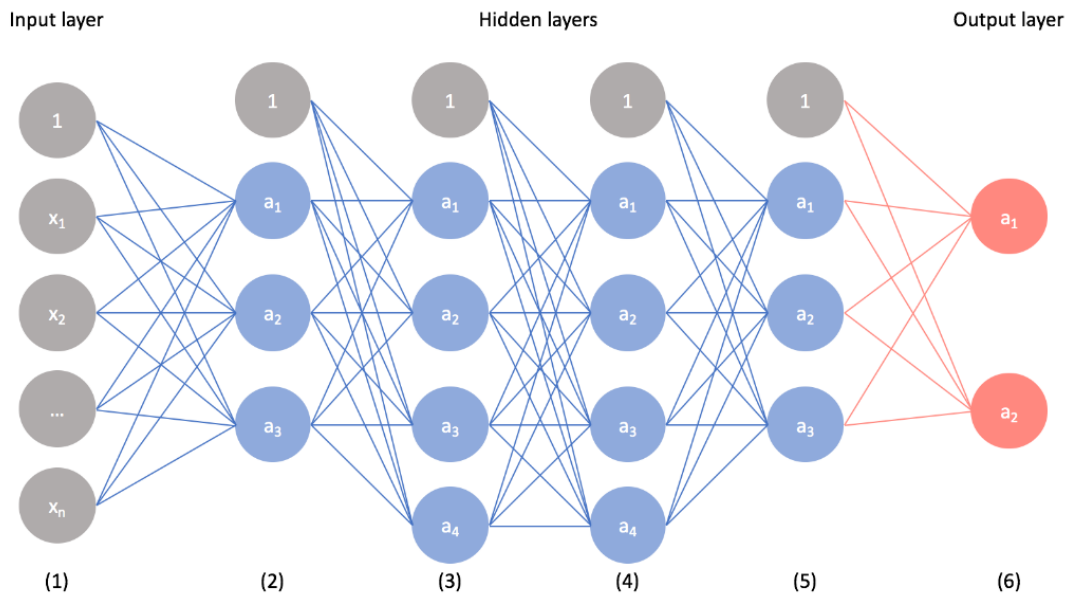
Εικόνα 19 (α) Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (β) Περιφραγμένη Επαναλαμβανόμενη Μονάδα (Rana, και συν., 2016)

Τα RNN είναι κατάλληλα για τη δημιουργία μουσικής, μπορούν να μάθουν μακροπρόθεσμες εξαρτήσεις σε μουσικές ακολουθίες, επιτρέποντάς τους να παράγουν μουσική που είναι συνεκτική και έχει δομή. Λόγω της παραπάνω ιδιότητας, τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN), ιδίως τα δίκτυα μακράς βραχυπρόθεσμης μνήμης (LSTM), έχουν χρησιμοποιηθεί ευρέως στον τομέα της μουσικής σύνθεσης για τη δημιουργία νέας μουσικής ή τη συνέχιση μιας μελωδίας. Στην παραγωγή μουσικής, τα RNN χρησιμοποιούνται συνήθως ως ένα είδος "γλωσσικού μοντέλου", το οποίο εκπαιδεύεται σε ένα σύνολο δεδομένων από αρχεία MIDI ή παρτιτούρες. Αφού εκπαιδευτεί το δίκτυο αυτό, το RNN μπορεί να παράγει νέα μουσική με δειγματοληψία από την ήδη εκπαιδευμένα δεδομένα.

Ένας τρόπος χρήσης ενός RNN είναι η πρόβλεψη της επόμενης νότας σε μια μελωδία, δεδομένων των προηγούμενων νοτών. Αυτό επιτυγχάνεται με την εκπαίδευση του RNN σε ένα σύνολο δεδομένων μελωδιών, όπου κάθε είσοδος στο RNN είναι μια ακολουθία νοτών και η αντίστοιχη έξοδος είναι η επόμενη νότα στην ακολουθία. Παρόμοια, μια άλλη αξιοποίηση του RNN είναι η δημιουργία ενός ολόκληρου μουσικού κομματιού χρησιμοποιώντας μια ακολουθία νοτών ως είσοδο και έξοδο, καθώς το δίκτυο εκπαιδεύεται για να προβλέπει την επόμενη ακολουθία νοτών δεδομένης της προηγούμενης ακολουθίας. Γενικά, τα RNN έχουν μεγάλες δυνατότητες πάνω στη σύνθεση μουσικής και παράγουν μουσική που είναι παρόμοια με την ανθρώπινη μουσική. Ωστόσο, είναι σημαντικό να σημειωθεί ότι η παραγόμενη μουσική μπορεί να εξακολουθεί να μην έχει τη δημιουργικότητα και την εκφραστικότητα της μουσικής που συνθέτει ο άνθρωπος.

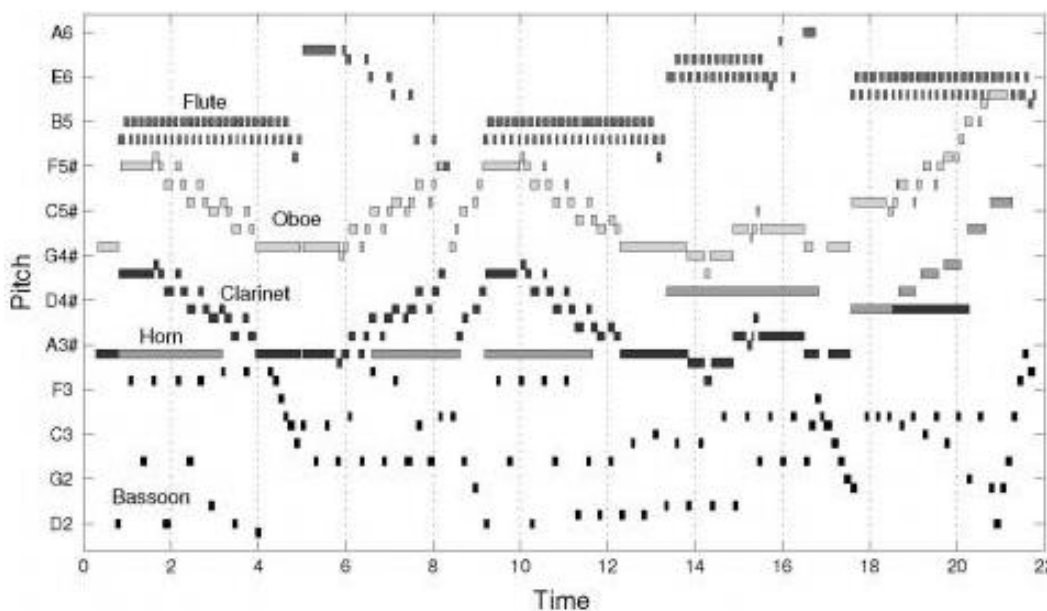
Convolutional Neural Networks (CNN)

Το βασικό χαρακτηριστικό των συνελκτικών νευρωνικών δικτύων (CNN) είναι η χρήση επιπέδων συνελίξεων, τα οποία έχουν σχεδιαστεί για να μαθαίνουν αυτόματα και προσαρμοστικά χαρακτηριστικά από τα δεδομένα εισόδου. Αυτά τα επίπεδα συνελίξεως εκτελούν μια μαθηματική πράξη που ονομάζεται συνέλιξη, η οποία περιλαμβάνει την εφαρμογή ενός συνόλου φίλτρων στα δεδομένα εισόδου για την εξαγωγή χαρακτηριστικών. Τα φίλτρα εκπαιδεύονται κατά τη διάρκεια της διαδικασίας εκπαίδευσης και η έξοδος της λειτουργίας συνελίξεως ονομάζεται χάρτης χαρακτηριστικών.



Εικόνα 20 Συνελκτικό Νευρωνικό Δίκτυο (Jordan, 2017)

Τα CNN έχουν χρησιμοποιηθεί στον τομέα της μουσικής σύνθεσης, αλλά όχι τόσο συχνά όσο τα RNN, (π.χ. LSTM). Αυτό οφείλεται στο γεγονός ότι τα CNN είναι καλύτερα προσαρμοσμένα για εργασίες που περιλαμβάνουν δεδομένα που μοιάζουν με πλέγμα, όπως οι εικόνες, ενώ οι μουσικές ακολουθίες αναπαρίστανται συνήθως ως μονοδιάστατες χρονοσειρές. Ωστόσο, ορισμένοι ερευνητές έχουν επιχειρήσει να εφαρμόσουν τα CNN στη δημιουργία μουσικής μετατρέποντας μουσικές ακολουθίες σε δισδιάστατες αναπαράστασεις, για παράδειγμα ένα ρολό πιάνου.



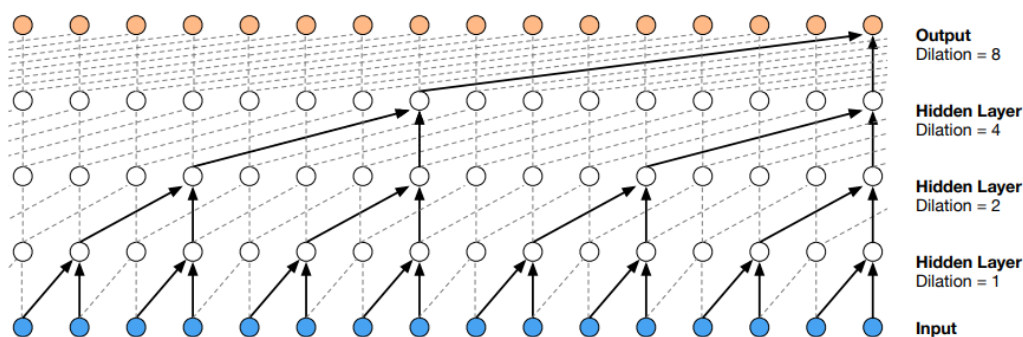
Εικόνα 21 Pianoroll (Jazz Studies Online)

Σε αυτή την αναπαράσταση, κάθε στήλη αναπαριστά ένα χρονικό βήμα και κάθε γραμμή αναπαριστά μια νότα, ενώ η τιμή σε κάθε κελί δείχνει αν η νότα παίζεται ή όχι. Όταν η μουσική ακολουθία αναπαρασταθεί με αυτόν τον τρόπο, ένα CNN μπορεί να εκπαιδευτεί με το αντίστοιχο σύνολο δεδομένων για να μάθει μοτίβα από τις μουσικές ακολουθίες. Το CNN μπορεί στη συνέχεια να χρησιμοποιηθεί για τη δημιουργία νέας μουσικής με δειγματοληψία από τα εκπαιδευμένα δεδομένα. Διαφορετικά ένα CNN μπορεί να εκτελέσει ένα είδος μείωσης των διαστάσεων στη μουσική ακολουθία και στη συνέχεια να χρησιμοποιήσει ένα RNN για τη δημιουργία νέας μουσικής. Σε αυτή την προσέγγιση, το CNN χρησιμοποιείται για να μάθει μια συμπαγή αναπαράσταση της μουσικής ακολουθίας που συλλαμβάνει τις πιο σημαντικές πληροφορίες και το RNN εκπαιδεύεται στη συνέχεια σε αυτή την αναπαράσταση για να παράγει νέα μουσική.

Συνοψίζοντας, ενώ τα CNN έχουν χρησιμοποιηθεί στη σύνθεση μουσικής, είναι λιγότερο διαδεδομένα από τα RNN, λόγω της μικρότερης αποτελεσματικότητάς τους στη σύλληψη των μακροπρόθεσμων εξαρτήσεων στις μουσικές ακολουθίες. Παρ' όλ' αυτά, κύρια πρόκληση στη χρήση των CNN για τη δημιουργία μουσικής αποτελεί η παραγόμενη μουσική και το γεγονός ότι μπορεί να είναι λιγότερο εκφραστική και δημιουργική από τη μουσική που συνθέτουν οι άνθρωποι. Εντούτοις, η προσπάθεια για τον συνδυασμό των πλεονεκτημάτων των CNN και των RNN για τη δημιουργία μουσικής δεν σταματά, όταν, μάλιστα, έχουν επιτευχθεί κάποια ελπιδοφόρα αποτελέσματα. Αξίζει να σημειωθεί ότι ο συνδυασμός CNN και RNN για τη δημιουργία μουσικής είναι ακόμη ενεργός τομέας έρευνας, αλλά η τεχνολογία δεν είναι ακόμη ώριμη και απαιτείται περισσότερη έρευνα για τη βελτίωση της ποιότητας και της εκφραστικότητας της παραγόμενης μουσικής.

WaveNet

Το Wavenet είναι ένα μοντέλο βαθιάς μάθησης που αναπτύχθηκε από την Google και είναι σε θέση να παράγει υψηλής ποιότητας κυματομορφές ήχου. Χρησιμοποιεί το συνελκτικό νευρωνικό δίκτυο (CNN) για την ανάλυση δειγμάτων ήχου και τη δημιουργία νέων δειγμάτων που είναι παρόμοια με αυτά.



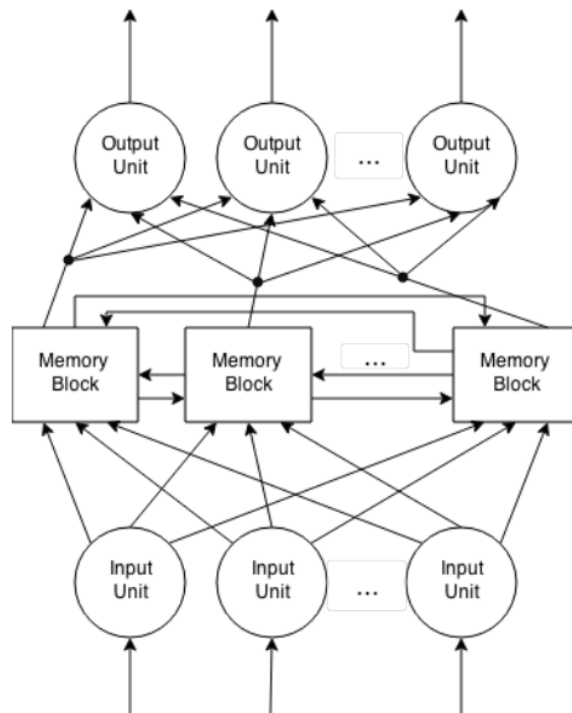
Εικόνα 22 Οπτικοποίηση μιας στρώβας διασταλμένων αιτιακών συνελκτικών στρωμάτων (Oord, Dieleman, Simonyan, Vinyals, & Graves, 2016)

Στο πλαίσιο της μουσικής σύνθεσης, το Wavenet έχει χρησιμοποιηθεί για τη δημιουργία νέας μουσικής εκπαιδύοντας το μοντέλο σε ένα σύνολο δεδομένων με υπάρχοντα μουσικά δείγματα. Αφού εκπαιδευτεί, το μοντέλο μπορεί να παράγει νέα μουσικά δείγματα που είναι παρόμοια με αυτά στα οποία εκπαιδεύτηκε. Τα παραγόμενα δείγματα μπορούν να χρησιμοποιηθούν ως αφετηρία για τη σύνθεση νέων τραγουδιών ή ως πηγή έμπνευσης για τους μουσικούς. Το WaveNet χρησιμοποιήθηκε σε πολλές πτυχές της ακατέργαστης παραγωγή ήχου ως αυτόματος κωδικοποιητής και συνθέτης ήχου. Τα πιο πειστικά αποτελέσματα επιτεύχθηκαν μέσω της προσθήκης συνθηκών ως πρόσθετη είσοδος. Για παράδειγμα, ο νευρωνικός συνθέτης ήχου με αυτόματους κωδικοποιητές WaveNet (Engel, et al., 2017) προσθέτει όρους για το ύψος του ήχου κατά τη διάρκεια της εκπαίδευσης.

Ένα από τα κύρια πλεονεκτήματα της χρήσης του Wavenet για τη σύνθεση μουσικής είναι η ικανότητά του να παράγει υψηλής ποιότητας ηχητικά δείγματα που είναι δύσκολο να διακριθούν από δείγματα που έχουν δημιουργηθεί από τον άνθρωπο. Το Wavenet μπορεί επίσης να εκπαιδευτεί ώστε να παράγει μουσική σε συγκεκριμένα στυλ ή είδη και μπορεί να ρυθμιστεί λεπτομερώς ώστε να παράγει μουσική που μοιάζει με την μουσική ενός συγκεκριμένου καλλιτέχνη ή συγκροτήματος. Τα παραγόμενα δείγματα ενδέχεται να μην έχουν το συναισθηματικό βάθος ή τις ερμηνευτικές αποχρώσεις που συνοδεύουν τη μουσική που παράγεται από τον άνθρωπο. Ωστόσο, τα παραγόμενα δείγματα μπορεί επίσης να απαιτούν ανθρώπινη επεξεργασία, ενορχήστρωση και μετα-επεξεργασία για να γίνουν ένα ολοκληρωμένο μουσικό κομμάτι.

Long Short-Term Memory (LSTM)

Τα δίκτυα μακράς βραχυπρόθεσμης μνήμης (LSTM) είναι ένας τύπος επαναλαμβανόμενου νευρωνικού δικτύου (RNN) που έχει σχεδιαστεί για να διαχειρίζεται το πρόβλημα των μακροπρόθεσμων εξαρτήσεων σε διαδοχικά δεδομένα. Αναλυτικά, οι μακροπρόθεσμες εξαρτήσεις αναφέρονται στην ικανότητα του μοντέλου να διατηρεί πληροφορίες από προηγούμενα χρονικά βήματα της ακολουθίας εισόδου όταν κάνει προβλέψεις για μεταγενέστερα χρονικά βήματα. Τα δίκτυα LSTM έχουν πιο σύνθετη δομή από τα παραδοσιακά RNN, αποτελούνται από κύτταρα μνήμης, τα οποία αποθηκεύουν πληροφορίες και πύλες, οι οποίες χρησιμοποιούνται για τον έλεγχο της ροής των πληροφοριών μέσα και έξω από τα κύτταρα. Οι πύλες υλοποιούνται ως τα ίδια τα νευρωνικά δίκτυα και μαθαίνουν πότε να αφήνουν τις πληροφορίες να περνούν και πότε να τις εμποδίζουν. Αξίζει να σημειωθεί ότι τα δίκτυα LSTM διατηρούν μια κρυφή κατάσταση που μπορεί να αποθηκεύει πληροφορίες για μεγάλες χρονικές περιόδους, γεγονός που τους επιτρέπει να μαθαίνουν και να κάνουν προβλέψεις με βάση μακροπρόθεσμα πρότυπα στα δεδομένα. Βασικές, είναι οι τρεις κύριες πύλες στα δίκτυα LSTM, η πύλη εισόδου, η οποία ελέγχει τη ροή των πληροφοριών στο κύτταρο μνήμης, η πύλη λήθης, η οποία ελέγχει τη ροή των πληροφοριών από το κύτταρο μνήμης, επιτρέποντάς του να ξεχνά επιλεκτικά ορισμένες πληροφορίες και η πύλη εξόδου που ελέγχει τη ροή πληροφοριών από το κύτταρο μνήμης στην έξοδο του δικτύου LSTM (Shewalkar, Nyavanandi, & Ludwig, 2019).



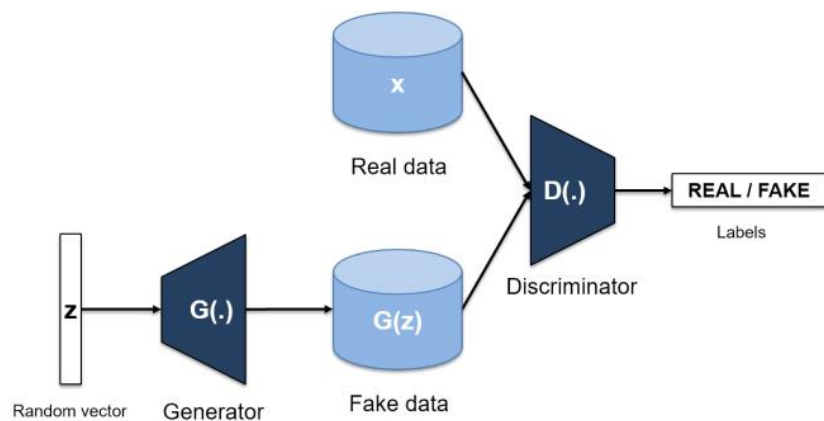
Εικόνα 23 Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (Liu & Ramakrishnan, 2014)

Τα δίκτυα LSTM, ως τύπος RNN, έχουν χρησιμοποιηθεί στον τομέα της μουσικής σύνθεσης για τη δημιουργία νέας μουσικής ή τη συνέχιση μιας μελωδίας, με βάση τα παραπάνω δεδομένα για τα δίκτυα RNN. Επίσης, υπάρχουν διάφοροι τρόποι για τη δημιουργία μουσικής με LSTM, οι προσεγγίσεις σε αυτό το κομμάτι μοιάζουν, κυρίως, με των RNN. Όμως, υπάρχουν διάφορες παραλλαγές των LSTM που έχουν χρησιμοποιηθεί στην παραγωγή μουσικής, όπως η χρήση αμφίδρομων δικτύων LSTM ή μηχανισμών προσοχής, οι οποίοι βοηθούν το δίκτυο να εστιάσει σε συγκεκριμένα μέρη της εισόδου και να βελτιώσει την παραγόμενο αποτέλεσμα, δηλαδή, την νέα μουσική, ωστόσο η δημιουργικότητα και η εκφραστικότητα δεν επιτυγχάνεται ούτε εδώ.

Generative Adversarial Networks (GAN)

Τα γενετικά αντιφατικά δίκτυα (GAN) είναι ένας τύπος νευρωνικού δικτύου που παρουσιάστηκε το 2014 από τον Ian Goodfellow (Goodfellow, et al., 2014) και τους συνεργάτες του, τα οποία έχουν εντοπιστεί και στον τομέα της μουσικής σύνθεσης για τη δημιουργία νέας μουσικής. Τα δίκτυα GAN αποτελούνται από δύο κύρια μέρη, το δίκτυο γεννήτριας που εκπαιδεύεται για τη δημιουργία νέας μουσικής, και το δίκτυο διάκρισης που εκπαιδεύεται για τη διάκριση μεταξύ της παραγόμενης μουσικής και της πραγματικής μουσικής. Πιο συγκεκριμένα, στην παραγωγή μουσικής, το δίκτυο γεννήτριας εκπαιδεύεται για να μάθει να αντιστοιχίζει ένα τυχαίο διάνυσμα θορύβου, που ονομάζεται λανθάνουσα μεταβλητή (latent variable), σε μια μουσική ακολουθία, ενώ το δίκτυο διάκρισης λαμβάνει ως είσοδο ένα δείγμα από την πραγματικής μουσική ή το δίκτυο γεννήτριας και εξάγει μια κλιμακωτή τιμή που αντιπροσωπεύει την πιθανότητα της εισόδου να περιέχει δείγμα από την πραγματικής μουσικής (Almeida & Pinho, 2018).

Κατά τη διάρκεια της διαδικασίας εκπαίδευσης, μπορεί να δυσκολευτεί το δίκτυο και να υπάρξουν ορισμένες προκλήσεις που πρέπει να αντιμετωπιστούν, όπως η κατάρρευση λειτουργίας, όπου η γεννήτρια παράγει μόνο ένα υποσύνολο των πιθανών εξόδων. Ένα άλλο ζήτημα είναι η σταθερότητα της διαδικασίας εκπαίδευσης, η οποία μπορεί να είναι ευαίσθητη στην επιλογή των υπερπαραμέτρων και στην αρχιτεκτονική των δικτύων γεννήτριας και διάκρισης. Τέλος, η παραπάνω διαδικασία είναι μια ανταγωνιστική διαδικασία, όπου τα δίκτυα γεννήτριας και διάκρισης ανταγωνίζονται μεταξύ τους προκειμένου να βελτιώσουν την απόδοσή τους.



Εικόνα 24 Αρχιτεκτονική υψηλού επιπέδου του Generative Adversarial Network (Almeida & Pinho, 2018)

Πρόσφατα, ορισμένοι ερευνητές προσπάθησαν να βελτιώσουν την απόδοση των δικτύων GAN εισάγοντας νέες αρχιτεκτονικές, όπως τα δίκτυα Wasserstein GAN (WGAN) (Gulrajani, Ahmed, Arjovsky, Dumoulin, & Courville, 2017) και τα δίκτυα Progressive GAN (ProGAN) (Karras, Aila, Laine, & Lehtinen, 2018), τα οποία προσπαθούν να αντιμετωπίσουν ορισμένους από τους περιορισμούς των αρχικών GAN και να εξελίξουν τα αρχικά χαρακτηριστικά τους. Ένα από τα πλεονεκτήματα της χρήσης των δικτύων GAN για τη δημιουργία μουσικής είναι ότι μπορούν να παράγουν πιο ποικίλη και ρεαλιστική μουσική, καθώς η γεννήτρια εκπαιδεύεται για να μιμείται την πραγματική μουσική. Τα δίκτυα GAN μπορούν επίσης να χρησιμοποιηθούν για τη δημιουργία μουσικής σε διαφορετικά είδη, εκπαιδεύοντας το δίκτυο γεννήτριας και το δίκτυο διάκρισης με βάση τα διαφορετικά σύνολα δεδομένων μουσικής που υπάρχουν.

3 Βιβλιογραφική ανασκόπηση

Η εργασία των Mai, Dongjin Lee, και Jae-Hun Jung (Tran, Lee, & Jung, 2022) διερευνά την εφαρμογή της Ανάλυσης Τοπολογικών Δεδομένων (TDA) με τη χρήση συνεχούς ομολογίας για την ανάλυση μουσικών κομματιών. Προτείνουν μια μέθοδο που περιλαμβάνει την εκπαίδευση ενός μοντέλου με τις αρχές της μουσικής σύνθεσης για ένα συγκεκριμένο είδος κορεατικής μουσικής αντί της χρήσης άμεσων μουσικών κομματιών ως δεδομένα εκπαίδευσης. Οι ερευνητές χρησιμοποιούν αρχικά την TDA για να αναλύσουν τους συνθετικούς τρόπους των μουσικών κομματιών με βάση καθιερωμένους μουσικούς κανόνες. Αυτή η ανάλυση αποτελεί βάση για τη δημιουργία ενός πίνακα επικάλυψης, ο οποίος αποτυπώνει την κατανομή και τη διαστρωμάτωση των πληροφοριών κατά τη διάρκεια ενός μουσικού κομματιού. Ο πίνακας επικάλυψης χρησιμεύει ως ένα σύνολο κανόνων σύνθεσης για το συγκεκριμένο είδος μουσικής και αντικαθιστά την ανάγκη για μεμονωμένα μουσικά κομμάτια σε εκπαίδευση της μηχανής. Δύο αλγόριθμοι, ο αλγόριθμος A και ο αλγόριθμος B, αναπτύσσονται με τη χρήση του πίνακα επικάλυψης. Ο αλγόριθμος A δεν ενσωματώνει τεχνητά νευρωνικά δίκτυα (ANN), ενώ ο αλγόριθμος B χρησιμοποιεί τον πίνακα επικάλυψης πίνακα για τη δημιουργία ενός συνόλου δεδομένων για την εκπαίδευση των ANN. Οι συγγραφείς σχεδιάζουν να μελετήσουν περαιτέρω τα αποτελέσματα και να εξετάσουν την ενσωμάτωση πρόσθετων χαρακτηριστικών όπως ο μετρητής και τα σύμβολα για να βελτιώσουν τα δεδομένα εκπαίδευσης, καθιστώντας τα πιο ολοκληρωμένα με μικρότερο αριθμό μουσικών κομματιών.

Μια άλλη εργασία των Goren, Nachmani και Wolf (Goren, Nachmani, & Wolf, 2021) με τίτλο “A-Muze-Net” παρουσιάζει μια νέα προσέγγιση για τη δημιουργία μουσικής με τη σύνθεση αρμονίας με βάση τις παραγόμενες μελωδίες. Ο σχεδιασμός του νευρωνικού δικτύου αποτελείται από δύο κύριες μονάδες: τη μονάδα παραγωγής μελωδίας και τη μονάδα παραγωγής αρμονίας. Η μονάδα παραγωγής αρμονίας λαμβάνει ως είσοδο την παραγόμενη μελωδία και παράγει την αντίστοιχη αρμονία, ενώ η μονάδα παραγωγής μελωδίας είναι υπεύθυνη για την παραγωγή της ίδιας της μελωδίας. Οι συγγραφείς αξιολόγησαν την απόδοση του μοντέλου συγκρίνοντας την παραγόμενη μουσική με ένα σύνολο δεδομένων από μουσική που έχει συνθέσει ο άνθρωπος και εκπαιδεύοντάς το σε μια συλλογή αρχείων MIDI, τα αποτελέσματα καταδεικνύουν ότι το προτεινόμενο μοντέλο είναι ικανό να παράγει μουσική σε διαφορετικά κλειδιά και στυλ, αναδεικνύοντας την ευελιξία του. Παράγει επίσης παραλλαγές ενός συγκεκριμένου τραγουδιού, γεγονός που υποδηλώνει την ικανότητά του να δημιουργεί μουσική υψηλής ποιότητας συγκρίσιμη με εκείνη που παράγεται από ανθρώπους. Η εργασία αναγνωρίζει ορισμένους περιορισμούς του μοντέλου, όπως ο περιορισμός του σε «ένα μόνο κλειδί» και η αδυναμία του να παράγει πολυφωνική μουσική ή διαφορετικά μουσικά στυλ. Οι συγγραφείς προτείνουν ότι η μελλοντική έρευνα θα μπορούσε

να αντιμετωπίσει αυτές τις ελλείψεις, ευρύνοντας τους ορίζοντες για τη δημιουργία μουσικής σε διάφορες μορφές.

Οι συγγραφείς Zhejiang Hu, Yan Liu, Gong Chen και Yongxu Liu (Hu, Liu, Chen, & Liu, 2022) στην εργασία τους με τίτλο "Can Machines Generate Personalized Music?" διερευνούν την πρόκληση της δημιουργίας εξατομικευμένης μουσικής λαμβάνοντας υπόψη τις ατομικές προτιμήσεις. Οι συγγραφείς προτείνουν μια υβριδική προσέγγιση που συνδυάζει τη μάθηση μεταφοράς και τη μάθηση με επίγνωση των προτιμήσεων στο πλαίσιο ενός μοντέλου δημιουργίας μουσικής. Η τεχνική χρησιμοποιεί το ιστορικό ακρόασης του χρήστη για τον εντοπισμό των προτιμώμενων μουσικών ειδών, τα οποία στη συνέχεια εφαρμόζονται σε ένα μοντέλο δημιουργίας μουσικής για την παραγωγή εξατομικευμένης μουσικής. Η μελέτη συγκρίνει αυτή την προσέγγιση με μια βασική στρατηγική που αγνοεί τις επιλογές του χρήστη και διαπιστώνει ότι η προτεινόμενη μέθοδος βελτιώνει σημαντικά την ικανότητα δημιουργίας εξατομικευμένης μουσικής με βάση τις προτιμήσεις, όπως υποδεικνύεται από μέτρα ακρίβειας όπως το F1-score και το accuracy. Επιπλέον, διεξήχθη μια έρευνα χρηστών, η οποία αποκάλυψε ότι οι περισσότεροι συμμετέχοντες αντιλαμβάνονται την παραγόμενη μουσική ως υψηλής ποιότητας και θυμίζει τα αγαπημένα τους τραγούδια. Οι συγγραφείς προτείνουν ότι η προσέγγιση αυτή έχει δυνατότητες εφαρμογής σε πραγματικά συστήματα σύστασης και παραγωγής μουσικής.

Η διερεύνηση που γίνεται στην εργασία με τίτλο "Theme Transformer" από τους Yi-Jen Shih, Shih-Lun Wu, Frank Zalkow, Meinard Muller και Yi-Hsuan Yang (Shih, Wu, Zalkow, Muller, & Yang, 2022) αφορά τη δημιουργία συμβολικής μουσικής με τη χρήση ενός μοντέλου μετασχηματιστή. Το προτεινόμενο μοντέλο, που ονομάζεται Theme Transformer, είναι ικανό να παράγει μουσική με βάση ένα συγκεκριμένο θέμα. Το θέμα κωδικοποιείται και παρέχεται ως είσοδος στον αποκωδικοποιητή του μετασχηματιστή, με αποτέλεσμα τη δημιουργία ποικίλης και συνεκτικής μουσικής που αποτυπώνει τα βασικά στοιχεία του θέματος. Η μελέτη εισάγει επίσης έναν μηχανισμό για τον έλεγχο του επιπέδου ακρίβειας του θέματος, επιτρέποντας στο μοντέλο να παράγει μουσική που είναι πιστή στο θέμα ή πιο δημιουργικά διαφοροποιημένη. Οι ερευνητές συγκρίνουν τον Theme Transformer με άλλα μοντέλα τελευταίας τεχνολογίας, αναδεικνύοντας την ανώτερη απόδοσή του όσον αφορά τη συνοχή και την ακρίβεια του θέματος. Επιπλέον, με την προ-εκπαίδευση του μοντέλου σε διάφορα μουσικά στυλ και τη βελτιστοποίησή του για κάθε στυλ, οι συγγραφείς αποδεικνύουν την ικανότητα του μοντέλου να παράγει μουσική σε πολλαπλά είδη, διατηρώντας παράλληλα τη συνέπεια και τη σταθερότητα του θέματος. Συνολικά, η έρευνα παρουσιάζει μια νέα προσέγγιση για τη δημιουργία θεματικά συμβατής μουσικής με τη χρήση ενός μετασχηματιστή και αναδεικνύει την αποτελεσματικότητα του μοντέλου σε διάφορα μουσικά είδη.

Η νέα προσέγγιση που παρουσιάζει ο Masahiro Suzuki (Suzuki, 2021) στην εργασία με τίτλο “Score Transformer” του πραγματεύεται τη δημιουργία μουσικών παρτιτούρων χρησιμοποιώντας ένα μοντέλο νευρωνικού δικτύου που βασίζεται σε μετασχηματιστή (Transformer). Η εργασία ξεκινά συζητώντας τους περιορισμούς των προηγούμενων τεχνικών παραγωγής παρτιτούρας, όπως οι μέθοδοι που βασίζονται σε κανόνες και τα μοντέλα Markov. Για να ξεπεραστούν αυτοί οι περιορισμοί, ο συγγραφέας προτείνει το μοντέλο Score Transformer, το οποίο βασίζεται στην αρχιτεκτονική του μετασχηματιστή και εκπαιδεύεται σε ένα σύνολο δεδομένων παρτιτούρας πιάνου MIDI. Η εργασία παρουσιάζει μια ανάλυση της απόδοσης του μοντέλου σε ένα σύνολο δεδομένων με παρτιτούρες πιάνου, έτσι γίνεται σύγκριση μεταξύ των παραγόμενων παρτιτούρων, των δεδομένων εκπαίδευσης και άλλων υφιστάμενων μεθόδων. Η αξιολόγηση καταδεικνύει ότι η μέθοδος Score Transformer υπερέρχει έναντι άλλων προσεγγίσεων όσον αφορά την ακρίβεια και την ομοιότητα με τα δεδομένα εκπαίδευσης. Με βάση τα αποτελέσματα της αξιολόγησης, ο Masahiro καταλήγει στο συμπέρασμα ότι το μοντέλο που βασίζεται στον μετασχηματιστή είναι ιδιαίτερα αποτελεσματικό στη δημιουργία νέων παρτιτούρων πιάνου που μοιάζουν πολύ με τα δεδομένα εκπαίδευσης. Αυτό αναδεικνύει τη χρησιμότητα των μοντέλων που βασίζονται σε μετασχηματιστές για τη δημιουργία παρτιτούρας και παρουσιάζει την ανώτερη απόδοσή τους σε σύγκριση με τις τρέχουσες προσεγγίσεις.

Το μοντέλο μηχανικής μάθησης για την εναρμόνιση μελωδιών με συγχορδίες, που παρουσιάζει η εργασία με τίτλο "CHORD-CONDITIONED MELODY HARMONIZATION WITH CONTROLLABLE HARMONICITY" (Wu, Li, & Sun, 2023) των Shangda Wu, Xiaobing Li και Maosong Sun, δημιουργεί εναρμονισμένες μελωδίες που είναι σύμφωνες με τις παρεχόμενες συγχορδίες χρησιμοποιώντας μια αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή μελωδίας υπό όρους συγχορδίας. Επιπλέον, οι συγγραφείς, προτείνουν ένα σύστημα μεταβλητής αρμονικότητας που θα επιτρέπει στους χρήστες να μεταβάλλουν την αρμονική πολυπλοκότητα των αρμονικών μελωδιών που δημιουργούνται. Στην εργασία αξιολογείται η ικανότητα του μοντέλου να ρυθμίζει την αρμονική πολυπλοκότητα των συνθεθειμένων εναρμονισμένων μελωδιών, αποδεικνύοντας ότι το μοντέλο είναι ικανό να παράγει εναρμονισμένες μελωδίες με διάφορα επίπεδα αρμονικής πολυπλοκότητας. Ακόμη, εξετάζεται η ικανότητα του μοντέλου να παράγει εναρμονισμένες μελωδίες που είναι συμβατές με τις συγχορδίες που παρέχονται από τα σύνολα δεδομένων και αποδεικνύουν ότι είναι ικανό να παράγει μελωδίες που είναι συγκρίσιμες με αυτές που συνθέτουν οι άνθρωποι. Το μοντέλο δοκιμάζεται, αφού εκπαιδευτεί σε ένα δείγμα λαϊκής μουσικής, και τα αποτελέσματα είναι ικανοποιητικά.

4 Μεθοδολογία

4.1 Περιβάλλον υλοποίησης

Η επιλογή του εργαλείου PyCharm και της γλώσσας προγραμματισμού Python για την ανάπτυξη μιας εφαρμογής τεχνητής νοημοσύνης προσφέρει αρκετά πλεονεκτήματα. Κάποια από αυτά είναι η ευχρηστία που τα χαρακτηρίζει και οι βιβλιοθήκες που παρέχονται. Ακόμη ένα, είναι η συμβατότητα που υπάρχει με πολλαπλά λειτουργικά συστήματα, επιτρέποντας να αναπτύσσονται εφαρμογές TN που λειτουργούν απρόσκοπτα σε διαφορετικά λειτουργικά συστήματα (Jetbrains). Έτσι δίνει τη δυνατότητα στους προγραμματιστές να δημιουργήσουν έξυπνες και καινοτόμες εφαρμογές TN.

Το PyCharm είναι ένα καλά εξοπλισμένο εργαλείο για την υποστήριξη και ανάπτυξη εφαρμογών τεχνητής νοημοσύνης. Ο επεξεργαστής κώδικα, που διαθέτει, παρέχει προηγμένα χαρακτηριστικά, όπως, η συμπλήρωση κώδικα, η ανίχνευση σφαλμάτων και η επισήμανση συντακτικού, όπου η χρήση τους είναι σημαντική κατά την ανάπτυξη αυτής της εργασίας (Jetbrains). Ακόμη, οι ιδιότητες του, επιτρέπει την αποτελεσματική δοκιμή και βελτιστοποίηση των μοντέλων TN. Επιπλέον, το PyCharm ενσωματώνει δημοφιλείς βιβλιοθήκες και πλαίσια TN, διευκολύνοντας την εισαγωγή μηχανικής μάθησης και λειτουργιών βαθιάς μάθησης σε οποιαδήποτε εφαρμογή.

Η Python είναι μια ευρέως υιοθετημένη γλώσσα προγραμματισμού για TN λόγω της απλότητας και της ευελιξίας της. Η εκτεταμένη συλλογή βιβλιοθηκών της Python ειδικά για την TN, με κάποιες βιβλιοθήκες να είναι η TensorFlow, η Keras, η PyTorch και η scikit-learn, παρέχουν προ-κατασκευασμένα εργαλεία, προ-εκπαιδευμένα μοντέλα, APIs υψηλού επιπέδου και αποτελεσματικές υπολογιστικές δυνατότητες, μειώνοντας την πολυπλοκότητα της ανάπτυξης εφαρμογών AI (The Python Standard Library).

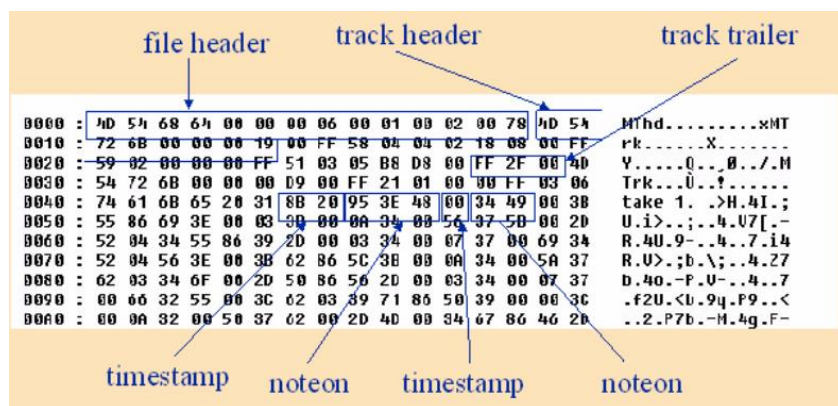
4.2 Αρχεία MIDI και σύνολα δεδομένων

4.2.1 Αρχεία MIDI

Η απόφαση για την χρήση MIDI αρχείων για την εξαγωγή και την επεξεργασία μουσικών πληροφοριών πάρθηκε με γνώμονα τις ιδιότητες που χαρακτηρίζουν αυτού του είδους αρχεία. MIDI σημαίνει Ψηφιακή Διεπαφή Μουσικών Οργάνων (Musical Instrument Digital Interface), τα αρχεία MIDI είναι μια τυποποιημένη μορφή που περιέχουν μουσικά δεδομένα που αντιπροσωπεύουν διάφορα στοιχεία μιας μουσικής σύνθεσης (Raffel & Ellis, 2016). Σε αντίθεση με τα αρχεία ήχου όπως τα αρχεία MP3 ή WAV που αποθηκεύουν πραγματικά ηχητικά κύματα, τα αρχεία MIDI περιέχουν οδηγίες και δεδομένα που περιγράφουν τον τρόπο αναπαραγωγής ενός μουσικού κομματιού. Ουσιαστικά είναι ένα σύνολο ψηφιακών

οδηγιών που περιγράφουν πώς να παράγουν ήχο στα συνθεσάιζερ, στα όργανα λογισμικού ή σε άλλες συσκευές συμβατές με MIDI αρχεία.

Βασικά χαρακτηριστικά μουσικής που αποτελούν ένα MIDI αρχείο είναι, για παράδειγμα, οι μουσικές νότες, τα όργανα που χρησιμοποιούνται για την αναπαραγωγή των νοτών, ο ρυθμός, η δυναμική αλλά και οι αποχρώσεις ενός μουσικού κομματιού (Raffel & Ellis, 2016). Ακόμη, σχετικά με τη μουσική δομή, άλλα χαρακτηριστικά είναι, τα τμήματα, τα μέτρα, οι χρονικές ιδιότητες και τα κλειδιά, στοιχεία που είναι απαραίτητα για τη σωστή ερμηνεία και αναπαραγωγή της σύνθεσης.



Εικόνα 25 Ένα βασικό MIDI αρχείο και τα στοιχεία που το αποτελούν (Hartley, 2008)

Στο πλαίσιο της ανάπτυξης μιας εφαρμογής τεχνητής νοημοσύνης για τη δημιουργία μουσικής, τα αρχεία MIDI χρησιμεύουν ως η κύρια πηγή δεδομένων εκπαίδευσης. Τροφοδοτώντας τα δεδομένα από τα αρχεία MIDI σε αλγορίθμους μηχανικής μάθησης, τα μοντέλα TN είναι ικανά να μάθουν τα μοτίβα, τις δομές και τα χαρακτηριστικά της μουσικής. Στη συνέχεια, το μοντέλο χρησιμοποιεί αυτή τη γνώση για να παράγει νέες μουσικές συνθέσεις με βάση τα εκπαιδευμένα μοτίβα. Τέλος, σημαντικό είναι να αναφερθεί η ευελιξία των αρχείων MIDI, διότι αποτελούν μια γέφυρα μεταξύ διαφορετικών μουσικών συσκευών και λογισμικού, επιτρέποντας την ανεμπόδιση επικοινωνία και ενσωμάτωση μεταξύ τους.

4.2.2 Σύνολα δεδομένων

Το Classical Music MIDI dataset (Rakshit, 2019) και το midi_classic_music (Fedorak) είναι τα σύνολα δεδομένων κλασικής μουσικής που χρησιμοποιούνται για την εξαγωγή της παρούσας προσέγγισης, είναι μια πολύτιμη πηγή για όποιον ενδιαφέρεται να μελετήσει ή να αναπτύξει εφαρμογές τεχνητής νοημοσύνης στον τομέα της κλασικής μουσικής. Αυτά τα σύνολα δεδομένων αποτελούνται από αρχεία MIDI που περιέχουν μουσικές συνθέσεις από διάφορα είδη κλασικής μουσικής, όπως συμφωνίες, κοντσέρτα,

σονάτες, όπερες και άλλα. Συγκεκριμένα παρέχουν μια ολοκληρωμένη συλλογή μουσικών παρτιτούρων. Ένα από τα βασικά πλεονεκτήματα των συνόλων δεδομένων MIDI κλασικής μουσικής είναι το πλούσιο και ποικίλο περιεχόμενό τους, διότι, περιλαμβάνουν έργα από διάσημους κλασικούς συνθέτες όπως ο Bach, ο Mozart, ο Haydn, ο Chopin και πολλοί άλλοι, οι οποίοι έχουν δημιουργήσει συνθέσεις που καλύπτουν διαφορετικές ιστορικές περιόδους, από την Μπαρόκ και την Κλασική εποχή έως τη Ρομαντική και τη Σύγχρονη εποχή.

Επιπλέον, τα σύνολα δεδομένων προσφέρουν ένα ευρύ φάσμα μουσικών στοιχείων προς ανάλυση και μοντελοποίηση, δηλαδή, παρέχουν λεπτομερείς μουσικές πληροφορίες, όπως μελωδίες, αρμονίες, συγχορδίες, ρυθμό, δυναμική και πολλά άλλα, που καθιστούν ικανή τη σε βάθος ανάλυση και ερμηνεία των δομών της κλασικής μουσικής, διευκολύνοντας την ανάπτυξη εξελιγμένων αλγορίθμων τεχνητής νοημοσύνης που μπορούν να παράγουν, να αναλύουν ή να ερμηνεύουν συνθέσεις κλασικής μουσικής. Συνεπώς, είτε ερευνητές είτε μουσικοί είτε προγραμματιστές ΤΝ έχουν την δυνατότητα να εξερευνήσουν την εξέλιξη των ειδών και των τεχνικών της κλασικής μουσικής συγκρίνοντας και αξιολογώντας την απόδοση διαφορετικών αλγορίθμων, τεχνικών και μοντέλων στον τομέα της δημιουργίας κλασικής μουσικής.

Data Explorer

Version 1 (7.22 MB)

- ▶ albeniz
- ▶ bach
- ▶ balakir
- ▶ beeth
- ▶ borodin
- ▶ brahms
- ▶ burgm
- ▶ chopin
- ▶ debussy
- ▶ granados
- ▶ grieg
- ▶ haydn
- ▶ liszt
- ▶ mendelssohn
- ▶ mozart
- ▶ muss
- ▶ schubert
- ▶ schumann
- ▶ tschai

Εικόνα 26 Το σύνολο δεδομένων Classical Music MIDI (Rakshit, 2019)

4.3 Βιβλιοθήκες

Οι βασικότερες βιβλιοθήκες που χρησιμοποιήθηκαν στην συγκεκριμένη εργασία είναι οι music21, pretty_midi, Scikit-learn, Keras, Seaborn και pandas.

Η music21 είναι μια βιβλιοθήκη πλούσια σε χαρακτηριστικά που προσφέρουν μια ολοκληρωμένη εργαλειοθήκη για τη μουσικολογία και τη σύνθεση μουσικής με τη βοήθεια υπολογιστή. Η βιβλιοθήκη αυτή παρέχει την δυνατότητα ανάλυσης, επεξεργασίας και δημιουργίας μουσικής σημειογραφίας σε διάφορες μορφές, όπως MIDI, MusicXML και άλλα (musicology). Με την βιβλιοθήκη music21, υπάρχει πρόσβαση σε αρχεία MIDI, και σε λεπτομερείς πληροφορίες που παρέχουν τα αρχεία MIDI, όπως νότες, συγχορδίες, κλειδιά, χρονισμοί κ.α. Ακόμη δίνετε η ευκαιρία είτε για την δημιουργία μουσικών μοτίβων, είτε για την δημιουργία σύνθετων μουσικών δομών. Με βάση τα παραπάνω η music21 χρησιμοποιείται ευρέως στη μουσική έρευνα, την εκπαίδευση και τη σύνθεση λόγω των εκτεταμένων δυνατοτήτων της και της ευελιξίας που διαθέτει σαν βιβλιοθήκη.

Η pretty_midi είναι μια βιβλιοθήκη για την εξαγωγή και τον χειρισμό μουσικών πληροφοριών από αρχεία MIDI (documentation). Με τις συναρτήσεις που περιέχονται στην pretty_midi, εκτιμάται το τέμπο, εξάγονται χρονικά χαρακτηριστικά και να υπολογίζονται ιστογράμματα ρυθμού και τέμπο. Ακόμη, η pretty_midi προσφέρει λειτουργικότητα για τη σύνθεση δεδομένων MIDI και τη μετατροπή τους σε ήχο. Η βιβλιοθήκη είναι γνωστή μεταξύ προγραμματιστών και ερευνητών που εργάζονται σε έργα που σχετίζονται με τη μουσική, όπως τη σύνθεση ήχου με τη χρήση μεθόδων μηχανικής μάθησης.

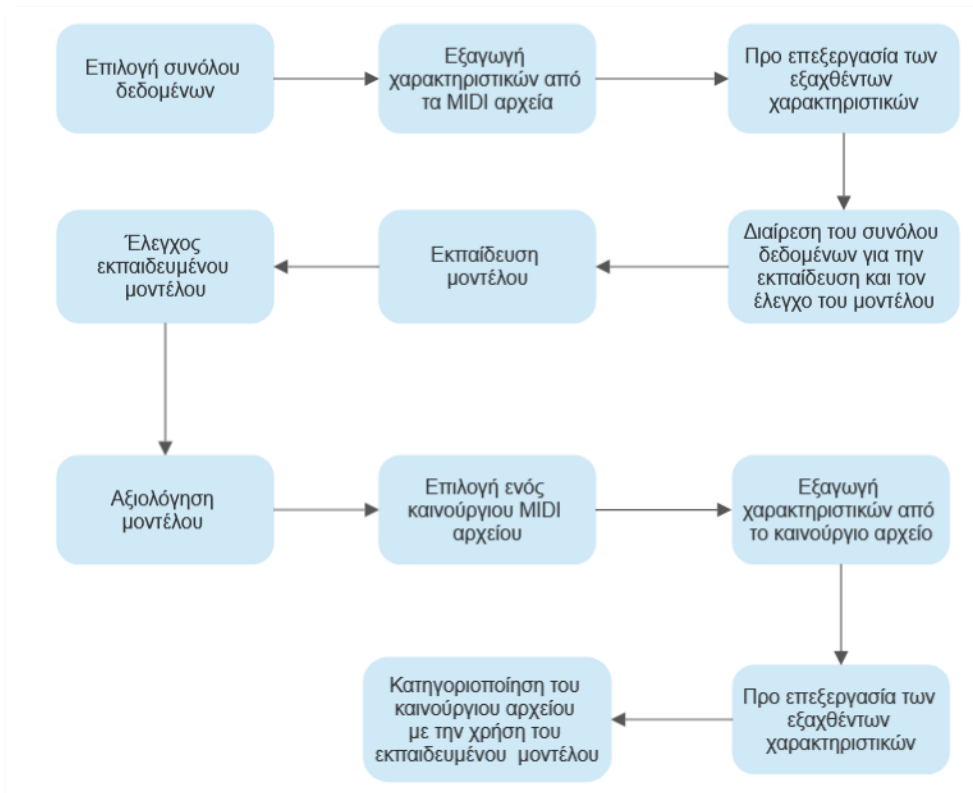
Η βιβλιοθήκες Scikit-learn (sklearn) και το Keras είναι δημοφιλείς βιβλιοθήκες στον τομέα της μηχανικής μάθησης και της βαθιάς μάθησης, αντίστοιχα (Ashish, 2019). Παρέχουν ισχυρά εργαλεία και λειτουργίες για τη δημιουργία και την εκπαίδευση μοντέλων μηχανικής μάθησης και βαθιάς μάθησης.

Η Seaborn είναι μια βιβλιοθήκη οπτικοποίησης δεδομένων Python που βασίζεται στην Matplotlib και παρέχει μια διεπαφή υψηλού επιπέδου για τη δημιουργία κατατοπιστικών στατιστικών γραφικών (Introduction to Seaborn – Python). Με τη Seaborn, δημιουργούνται γρήγορα διάφοροι τύποι γραφικών παραστάσεων, όπως bar plots, scatter plots, line plots, heatmaps και πολλά άλλα.

Η βιβλιοθήκη Pandas είναι μια θεμελιώδης βιβλιοθήκη χειρισμού και ανάλυσης δεδομένων στην Python, η οποία περιέχει μια ευέλικτη δομή δεδομένων που ονομάζεται DataFrame, η οποία δίνει την δυνατότητα να αποθηκεύονται και να γίνονται διαχειρίσιμα, αποτελεσματικά, τα δεδομένα σε μορφή πίνακα, αφού απλοποιεί τις εργασίες καθαρισμού, μετασχηματισμού και συνάθροισης δεδομένων (Pandas Introduction).

4.4 Κατηγοριοποίηση συνθετών μουσικής

Η κατηγοριοποίηση μουσικής με αρχεία MIDI βάσει κλασικών συνθετών περιλαμβάνει τη χρήση αλγορίθμων μηχανικής μάθησης για τον εντοπισμό και την κατηγοριοποίηση συνθέσεων σύμφωνα με το ύφος και τα χαρακτηριστικά συγκεκριμένων συνθετών. Η διαδικασία ξεκινά με τη συγκέντρωση ενός συνόλου δεδομένων από αρχεία MIDI που αντιπροσωπεύουν έργα διαφόρων κλασικών καλλιτεχνών, συγκεκριμένα στην εργασία χρησιμοποιήθηκαν σύνολα δεδομένων του Mozart, του Bach, του Chopin και του Haydn. Το σύνολο δεδομένων καλύπτει ένα ευρύ φάσμα συνθέσεων, συμπεριλαμβανομένων συμφωνιών, σονατών, κοντσέρτων και άλλων κλασικών μορφών. Στόχος είναι να δημιουργηθεί ένα μοντέλο μηχανικής μάθησης που να κατανοεί τα μοναδικά χαρακτηριστικά των κομματιών που διακρίνουν το έναν καλλιτέχνη από έναν άλλο. Αυτά τα χαρακτηριστικά περιλαμβάνουν ιστογράμματα τονικού ύψους, ρυθμό, διάρκειες νοτών, οπλισμός, μέτρο, αλλαγές στα μέτρα και άλλα, με αποτέλεσμα το μοντέλο να αναγνωρίζει τον συνθέτη ενός κομματιού που θα δώσει ο χρήστης.



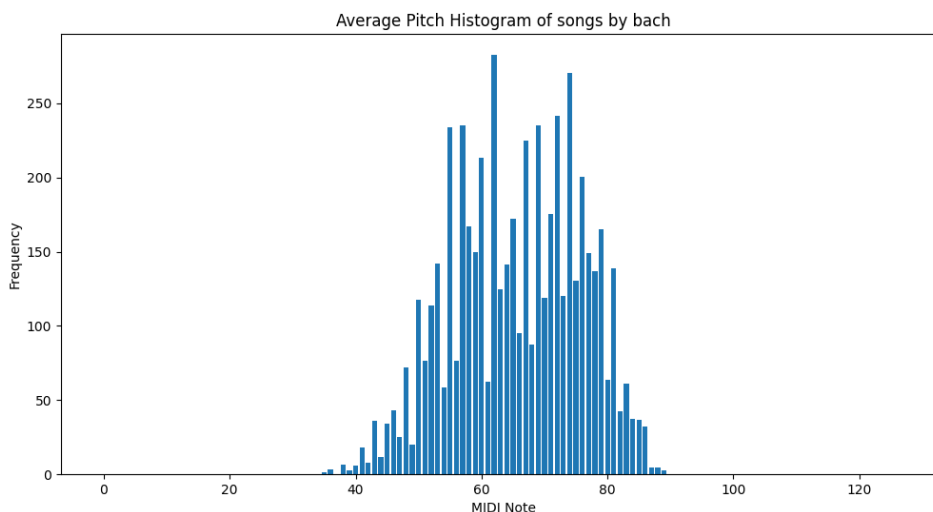
Εικόνα 27 Διάγραμμα της προτεινόμενης μεθοδολογίας κατηγοριοποίησης MIDI αρχείου με βάση τον συνθέτη

4.4.1 Προ επεξεργασία δεδομένων MIDI αρχείων

Σε αυτό το βήμα στο πρόγραμμα χρησιμοποιούνται συναρτήσεις, οι οποίες αφού επεξεργαστούν τα MIDI αρχεία και με την βοήθεια των ήδη υπάρχων συναρτήσεων από τις βιβλιοθήκες, επιστρέφουν τις μουσικές πληροφορίες που χρειάζεται το μοντέλο. Αρχικά, λαμβάνονται δεδομένα MIDI ως είσοδο και εξάγονται οι διάρκειες των μεμονωμένων νोटών. Επαναλαμβάνεται κάθε όργανο και νότα από τα δεδομένα MIDI και προστίθεται η διάρκεια της νότας σε μια λίστα που επιστρέφεται. Ακολούθως, μια συνάρτηση στον κώδικα δημιουργεί ένα ιστόγραμμα τόνων από τα δεδομένα MIDI, δηλαδή, η συνάρτηση επαναλαμβάνει όλα τα στοιχεία του αρχείου χρησιμοποιώντας τη μέθοδο `recurse()` και ελέγχει αν το κάθε στοιχείο είναι πραγματικά μια νότα, ώστε να εξάγει την τιμή του pitch και να την προσθέσει στην λίστα. Με τα στοιχεία που αποθηκεύονται στην λίστα δημιουργείται το ιστόγραμμα το οποίο αποτελείται από 128 bins και αυτό με την σειρά του αναπαριστά τη συχνότητα κάθε νότας MIDI (0 έως 127) που υπάρχει στο αρχείο MIDI. Δημιουργούνται τέσσερα ιστογράμματα, ένα για τον κάθε συνθέτη, σύμφωνα με το εκάστοτε σύνολο δεδομένων. Ακόμη, η διαδικασία της προ επεξεργασίας, εξάγει διάφορα σημαντικά χαρακτηριστικά για την εκπαίδευση του μοντέλου από τα δεδομένα MIDI, όπως το τέμπο, τον αριθμό των αλλαγών του μέτρου, το μέτρο, τον αριθμητή του μέτρου (π.χ. αν το μέτρο είναι $\frac{3}{4}$ τότε ο αριθμητής είναι το 3), τον παρονομαστή του μέτρου (π.χ. αν το μέτρο είναι $\frac{3}{4}$ τότε ο παρονομαστής είναι το 4) και τον αριθμό των οργάνων και τα επιστρέφει όλα σε μια λίστα.

Ένα παράδειγμα εξαγωγής χαρακτηριστικών είναι το ιστόγραμμα τόνου, το οποίο, συνήθως εμφανίζει τον αριθμό των εμφανίσεων κάθε κατηγορίας τόνων (ή συγκεκριμένων τόνων) σε ένα δεδομένο μουσικό κομμάτι ή σε μια συλλογή συνθέσεων. Για κάθε συνθέτη είναι διαφορετικό και τον χαρακτηρίζει.

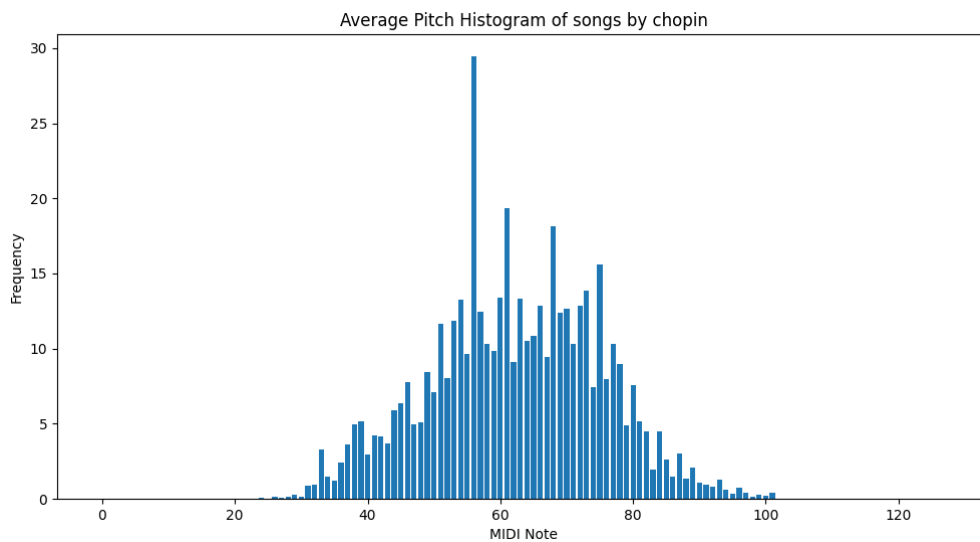
Bach



Γράφημα 1 Ιστόγραμμα τόνων του bach

Ο Bach συχνά συνέθετε στο πλαίσιο της τονικής αρμονίας, χρησιμοποιώντας τους τόνους της διατονικής κλίμακας (επταφωνική κλίμακα μέσα σε ένα δεδομένο κλειδί). Παρατηρείται ότι το συγκεκριμένο ιστόγραμμα των τόνων μπορεί να απεικονίζει μια υψηλότερη συχνότητα αυτών των διατονικών τόνων σε σύγκριση με των χρωματικών τόνων.

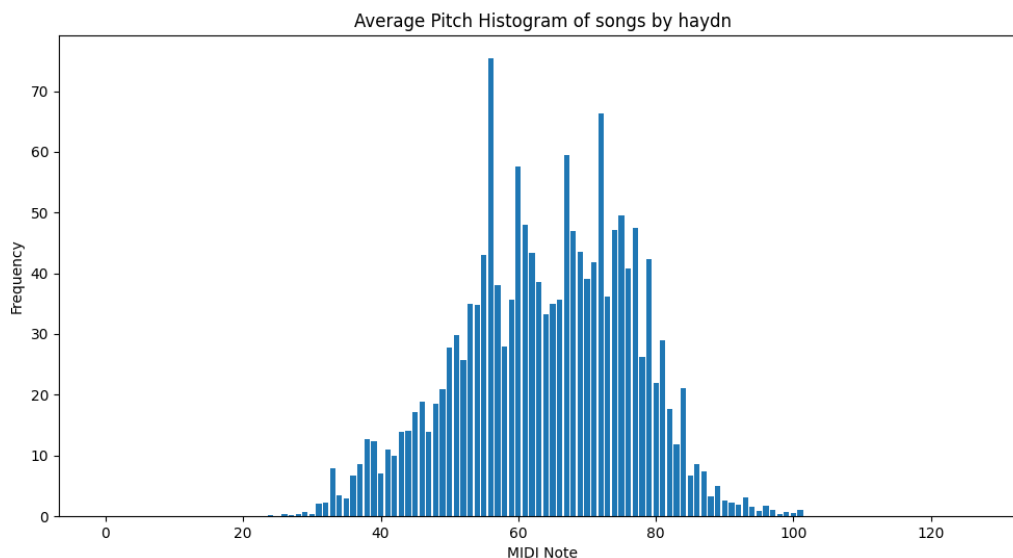
Chopin



Γράφημα 2 Ιστόγραμμα τόνων του chopin

Στην περίπτωση του Chopin Επιπλέον, τα έργα του Chopin συχνά γράφονταν σε συγκεκριμένα κλειδιά, όπως το E ελάσσονα, Λα μείζονα ή Φα ελάσσονα. Ως αποτέλεσμα, το ιστόγραμμα τόνων μπορεί να παρουσιάζει μεγαλύτερη συγκέντρωση νοτών που αντιστοιχούν σε αυτά τα συγκεκριμένα κλειδιά. Τα συγκεκριμένα κλειδιά έχουν χαμηλή συχνότητα, για τον λόγο αυτό παρατηρείται και στο ιστόγραμμα ότι η υψηλότερη συχνότητα νότας είναι κάτω από 30 Hz.

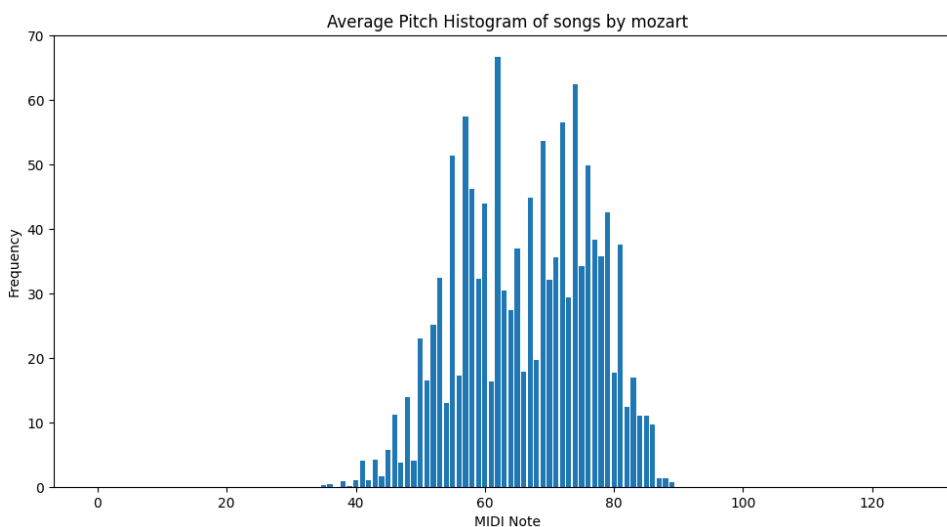
Haydn



Γράφημα 3 Ιστόγραμμα τόνων του haydn

Στα έργα του Haydn παρατηρείτε πιο ισορροπημένη και συμμετρική δομή. Το ιστόγραμμα δείχνει μια σχετικά ομοιόμορφη κατανομή των τόνων σε όλη τη σύνθεση, γεγονός που αντανακλά την ικανότητά του να δημιουργεί μια αίσθηση συμμετρίας και αναλογίας. Η ακριβή δεξιοτεχνία και το ευρύ φάσμα των συναισθημάτων που χαρακτηρίζουν την μουσική του Haydn, παρουσιάζουν συχνά ένα ισχυρό τονικό κέντρο, με μεγαλύτερη συχνότητα να έχουν οι τόνοι που σχετίζονται με το κλειδί.

Mozart



Γράφημα 4 Ιστόγραμμα τόνων του mozart

Η μουσική του Mozart τείνει να είναι σχετικά ισορροπημένη με έμφαση στα σαφή τονικά κέντρα και τις σαφώς καθορισμένες μελωδίες. Σε αυτό το ιστόγραμμα τόνων, παρατηρείται μια σχετικά ομοιόμορφη κατανομή των τόνων σε όλο το εύρος των διαθέσιμων νοτών. συστάδες τόνων μέσα σε συγκεκριμένες περιοχές ή μοτίβα. Επιβεβαιώνεται ότι είχε την ικανότητα να χρησιμοποιεί ένα ευρύ φάσμα τόνων σε διαφορετικές οκτάβες, εξερευνώντας τόσο τα υψηλά όσο και τα χαμηλά μητρώα διαφόρων οργάνων.

Στην συνέχεια, η κύρια συνάρτηση εκτελεί τις παραπάνω διαδικασίες για να συλλεχθούν τα χαρακτηριστικά και να αποθηκευτούν σε ένα πλαίσιο DataFrame, τα χαρακτηριστικά αυτά κανονικοποιούνται για να μπορέσουν να χρησιμοποιηθούν από τα μοντέλα, στη συνέχεια με τη εκτέλεση των min-max συμβαίνει η κανονικοποίηση στις αριθμητικές στήλες του DataFrame χρησιμοποιώντας τον StandardScaler από την βιβλιοθήκη scikit-learn. Τέλος, επιστρέφεται η λίστα all_features, η οποία περιέχει τα κανονικοποιημένα χαρακτηριστικά που εξάγονται από όλα τα αρχεία MIDI του συνόλου δεδομένων, μαζί με το όνομα του αντίστοιχου καλλιτέχνη.

Πίνακας 3 Εξαχθέντα χαρακτηριστικά από MIDI αρχεία

Path	Tempo	Num_of_changes	Resolution	Ts_1	Ts_2	Label
./data/mozart\mz_311_1.mid	0.3906679873	0.55	-0.1	0.125	0.125	mozart
./data/mozart\mz_331_3.mid	0.3591058788	0.55	-0.1	0.125	0.125	mozart
./data/bach\bach_846.mid	0.2544339563	0.55	-0.1	0.125	0.125	bach
./data/chopin\chpn-p11.mid	0.0938490866	0.55	-0.1	0.375	0.625	chopin
./data/chopin\chpn-p18.mid	0.2622188442	0.55	0.1	0.125	0.125	chopin
./data/chopin\chpn_op35_1.mid	0.1653730844	0.55	0.0	0.125	0.125	chopin
./data/chopin\chpn-p24.mid	0.2596304085	0.55	-0.1	0.375	0.625	chopin
./data/haydn\haydn_8_3.mid	0.0624024275	0.55	-0.1	0.125	0.625	haydn
./data/haydn\haydn_9_3.mid	0.2624035911	0.55	-0.1	0.125	0.125	haydn

4.4.2 Εφαρμογή των μοντέλων κατηγοριοποίησης

Το σημαντικότερο σημείο της ταξινόμησης στο πρόγραμμα είναι η συνάρτηση classification, που εκτελείται με το να δέχεται ένα σύνολο δεδομένων data ως είσοδο. Τα δεδομένα αυτά έχουν επεξεργαστεί κατάλληλα για να χρησιμοποιηθούν αποτελεσματικά, συγκεκριμένα, αφαιρούνται οι στήλες που αναγράφουν το path και το label. Στη συνέχεια, όλες οι στήλες, που παραμένουν, ανατίθενται στην μεταβλητή X, έπειτα κανονικοποιούνται ώστε όλα τα χαρακτηριστικά να βρίσκονται στην ίδια κλίμακα. Η κανονικοποιημένη λίστα X μετατρέπεται ξανά σε DataFrame και εκχωρείται ξανά στο X, ακολούθως χωρίζεται σε σύνολα εκπαίδευσης και δοκιμής. Το σύνολο εκπαίδευσης αποτελείται από το 70% του

συνόλου δεδομένων και η τυχαιότητα ορίζεται σε 42 τυχαίες καταστάσεις (`random_states`). Τέλος, δημιουργείται ένα αρχείο με όνομα "results.txt" σε κατάσταση εγγραφής για την αποθήκευση των αποτελεσμάτων.

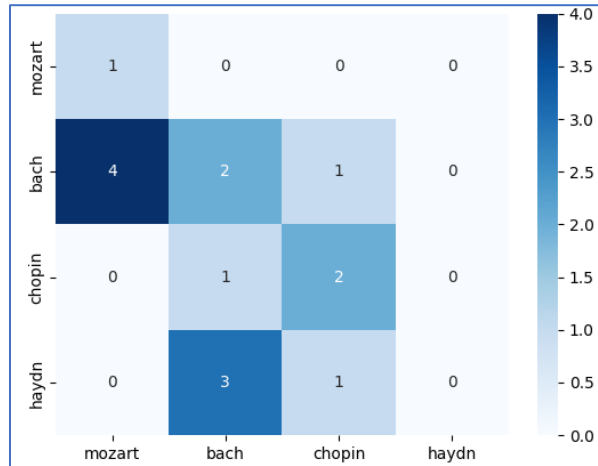
4.4.3 Αξιολόγηση αποτελεσμάτων

Στο πρόγραμμα ορίζεται μια συνάρτηση που αξιολογεί τα μοντέλα, αναλυτικά, λαμβάνει ένα μοντέλο ως είσοδο και αξιολογεί το μοντέλο στο σύνολο δοκιμών. Έπειτα, εκπαιδεύονται τα μοντέλα με τα δεδομένα εκπαίδευσης (`X_train`, `y_train`), και το μοντέλο κάνει προβλέψεις στα δεδομένα δοκιμής (`X_test`). Σε επόμενο βήμα, αξιολογούνται τα αποτελέσματα του μοντέλου με τη χρήση των `accuracy_score`, `precision_score`, `recall_score` και `f1_score` από το `scikit-learn`, η αξιολόγηση πραγματοποιείται για κάθε μοντέλο (Naive Bayes, Stochastic Gradient Descent, KNN, Decision Trees, Random Forest, Support Vector Machine, Logistic Regression και Neural Nets). Τα αποτελέσματα των αξιολογήσεων εγγράφονται, επίσης, στο αρχείο αποτελεσμάτων και υπολογίζεται ο πίνακας σύγχυσης με χρήση της `confusion_matrix()` του `scikit-learn`.

4.4.4 Αποτελέσματα

Η αξιολόγηση του μοντέλου της κατηγοριοποίησης είναι σημαντική, διότι συμβάλλει στην παραγωγή όσο το δυνατόν καλύτερων και αποδοτικότερων αποτελεσμάτων. Ένας τρόπος για να απεικονιστεί η απόδοση ενός μοντέλου ταξινόμησης είναι ο πίνακας σύγχυσης. Συνοψίζει τις προβλέψεις του μοντέλου συγκρίνοντας τις με τις πραγματικές ετικέτες κλάσης, αναλυτικά αποτελείται από τέσσερα κύρια στοιχεία: αληθή θετικά (TP), αληθή αρνητικά (TN), ψευδώς θετικά (FP) και ψευδώς αρνητικά (FN). Ακόμη, δημιουργείται ένα αρχείο "results.txt" που περιέχει τις αποδόσεις `accuracy_score()`, `precision_score()`, `recall_score()` και `f1_score()`, για κάθε μοντέλο, αυτό δίνει την δυνατότητα παρατήρησης και ελέγχου για το πόσο αποδοτικά είναι τα μοντέλα που έχουν δημιουργηθεί.

Decision Trees

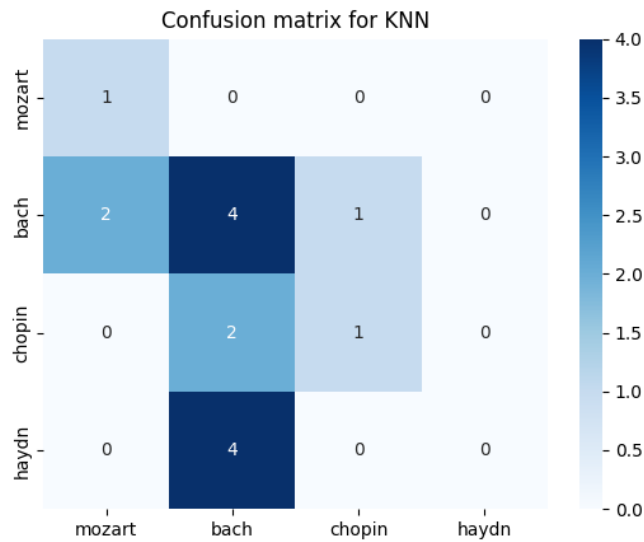


Εικόνα 28 Πίνακας σύγκρισης για τον αλγόριθμο decision trees

Πίνακας 4 Απόδοση του μοντέλου decision trees

	f1_score	precision	recall	accuracy
Decision trees	0.50939	0.50662	0.53333	0.53333

K-Nearest Neighbors Algorithm

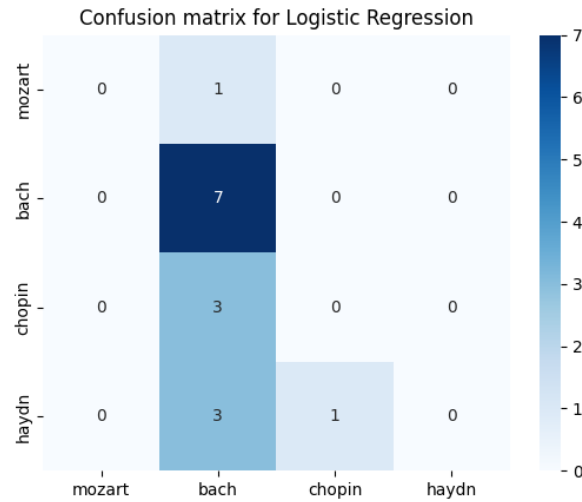


Εικόνα 29 Πίνακας σύγκρισης για τον αλγόριθμο knn

Πίνακας 5 Απόδοση του μοντέλου KNN

	f1_score	precision	recall	accuracy
KNN	0.50721	0.5006	0.51667	0.51667

Logistic Regression

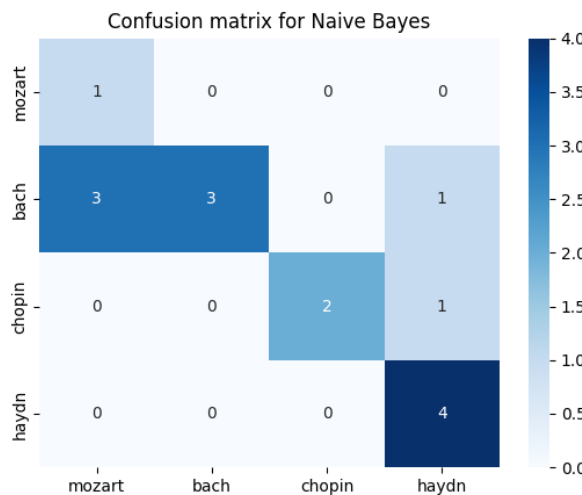


Εικόνα 30 Πίνακας σύγκρισης για τον αλγόριθμο logistic regression

Πίνακας 6 Απόδοση του μοντέλου logistic regression

	f1_score	precision	recall	accuracy
Logistic Regression	0.30038	0.22147	0.46667	0.46667

Naïve Bayes

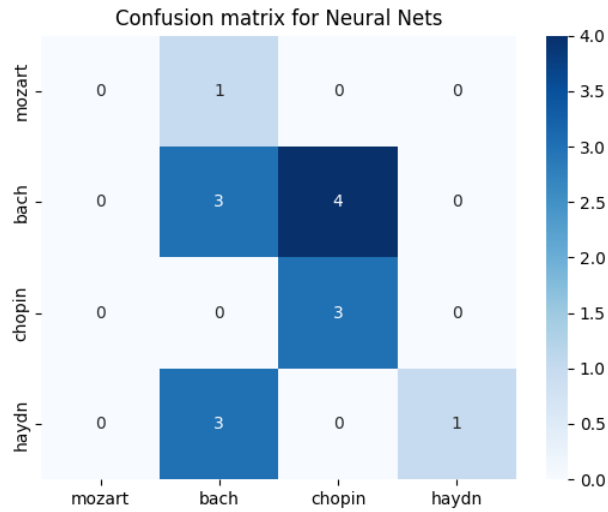


Εικόνα 31 Πίνακας σύγκρισης για τον αλγόριθμο naive bayes

Πίνακας 7 Απόδοση του μοντέλου naive bayes

	f1_score	precision	recall	accuracy
Naive Bayes	0.41809	0.45509	0.43333	0.43333

Neural Nets

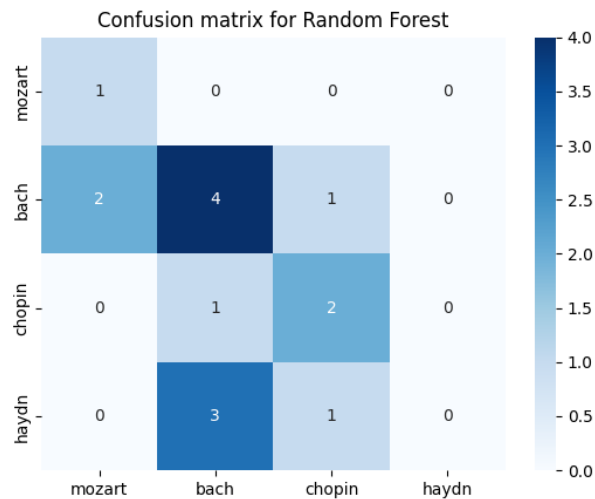


Εικόνα 32 Πίνακας σύγκρισης για τον αλγόριθμο neural nets

Πίνακας 8 Απόδοση του μοντέλου neural nets

	f1_score	precision	recall	accuracy
Neural Nets	0.5504	0.54907	0.58333	0.58333

Random Forest

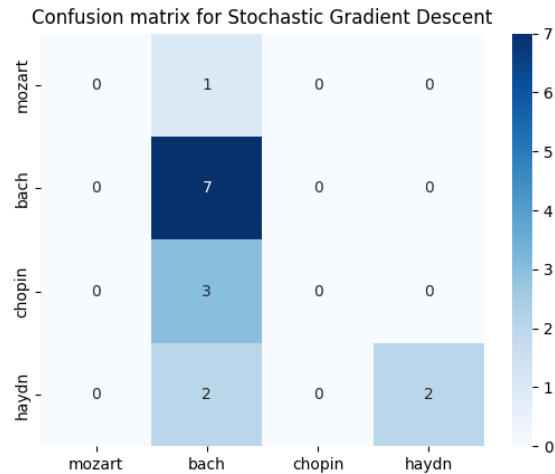


Εικόνα 33 Πίνακας σύγκρισης για τον αλγόριθμο random forest

Πίνακας 9 Απόδοση του μοντέλου random forest

	f1_score	precision	recall	accuracy
Random Forest	0.54078	0.53523	0.56667	0.56667

Stochastic Gradient Descent

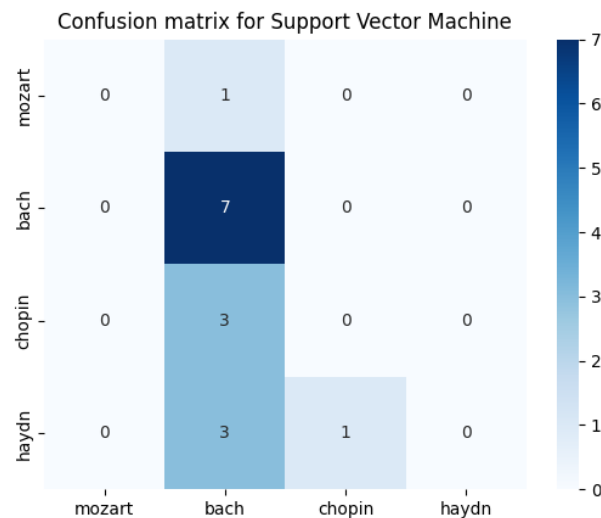


Εικόνα 34 Πίνακας σύγκρισης για τον αλγόριθμο stochastic gradient descent

Πίνακας 10 Απόδοση του μοντέλου stochastic gradient descent

	f1_score	precision	recall	accuracy
SGD	0.28549	0.21287	0.43333	0.43333

Support Vector Machine



Εικόνα 35 Πίνακας σύγκρισης για τον αλγόριθμο svm

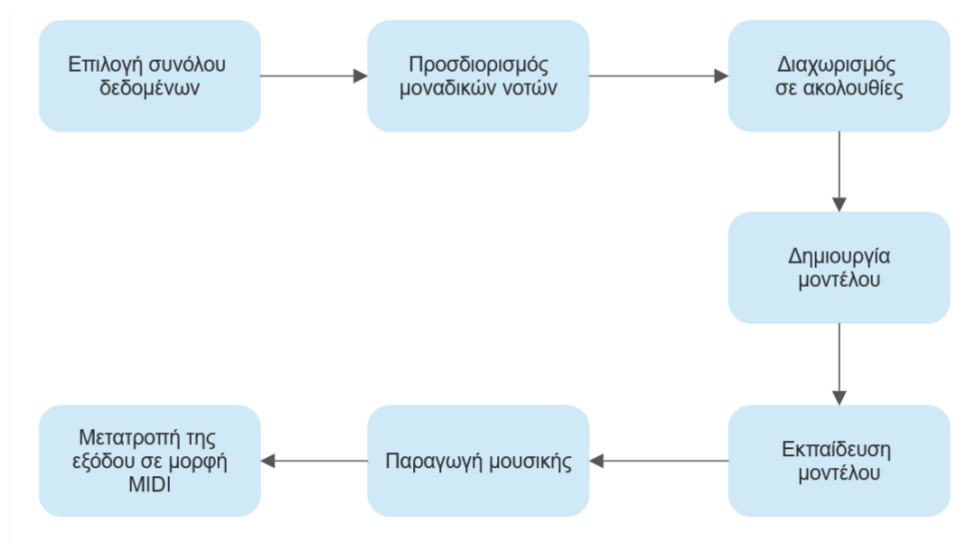
Πίνακας 11 Απόδοση του μοντέλου svm

	f1_score	precision	recall	accuracy
SVM	0.325	0.275	0.46667	0.46667

Από τα παραπάνω αποτελέσματα παρατηρείται ότι οι αποδόσεις των μοντέλων δεν είναι υψηλές. Τα καλύτερα μοντέλα κατηγοριοποίησης είναι τα Neural nets, Random forest και Decision trees με μικρή διαφορά να υπάρχει μεταξύ τους.

4.5 Παραγωγή μουσικής

Η παραγωγή μουσικής με αρχεία MIDI βάσει κλασικών συνθετών περιλαμβάνει τη χρήση αλγορίθμων μηχανικής μάθησης για την εκπαίδευση και την παραγωγή συνθέσεων σύμφωνα με το σύνολο δεδομένων ενός συγκεκριμένου συνθέτη που παίρνει από την κατηγοριοποίηση που έχει προηγηθεί με βάση το κομμάτι που έδωσε ο χρήστης/καλλιτέχνης. Η διαδικασία ξεκινά με τη συγκέντρωση ενός συνόλου δεδομένων από αρχεία MIDI που αντιπροσωπεύουν έργα ενός συνθέτη, συγκεκριμένα είτε του Mozart είτε του Bach είτε του Chopin είτε του Haydn. Στόχος είναι να εκπαιδευτούν τα μοντέλα μηχανικής μάθησης LSTM και Wavenet με τα μοναδικά χαρακτηριστικά των κομματιών που διακρίνουν τον συγκεκριμένο καλλιτέχνη και η παραγωγή ενός κομματιού που τα χαρακτηριστικά του να συνάδουν με του συνθέτη.



Εικόνα 36 Διάγραμμα της προτεινόμενης μεθοδολογίας παραγωγής μουσικής

4.5.1 Προ επεξεργασία MIDI αρχείων

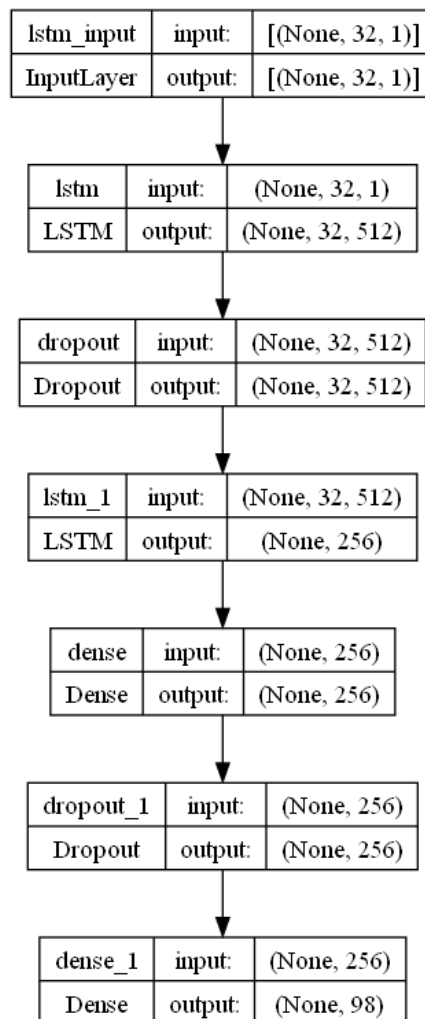
Η διαδικασία της προ επεξεργασίας των MIDI αρχείων προετοιμάζει δεδομένα MIDI για τη χρήση σε διαφορετικά μοντέλα μουσικής παραγωγής, όπως το WaveNet και το LSTM. Στην προετοιμασία περιλαμβάνεται, ο υπολογισμός των συχνοτήτων των νοτών, το φιλτράρισμα των σπάνιων νοτών, η επιλογή οργάνων, η εξαγωγή νοτών και συγχορδιών και στη συνέχεια η εκτέλεση κάποιου φιλτραρίσματος με στόχο τα δεδομένα να ικανοποιούν κατάλληλα τις απαιτήσεις κάθε μοντέλου. Αρχικά διαβάζεται αρχεία MIDI από το δοσμένο “path” του συνόλου δεδομένων και τα αποθηκεύει σε μια λίστα για περαιτέρω επεξεργασία. Στην συνέχεια, προεπεξεργάζεται τα δεδομένα MIDI για το μοντέλο WaveNet, δηλαδή μετατρέπει τον δισδιάστατο πίνακα νοτών που λαμβάνεται από τα αρχεία MIDI σε μονοδιάστατο πίνακα, υπολογίζει τη συχνότητα κάθε νότας, δημιουργεί ένα ιστόγραμμα για την οπτικοποίηση της κατανομής των νοτών, φιλτράρει τις σπάνιες νότες και δημιουργεί έναν νέο δισδιάστατο πίνακα που περιέχει μόνο τις συχνές νότες. Ακόμη, αναλύει ένα συγκεκριμένο αρχείο MIDI, εξάγει τις νότες από το πιάνο και τις μετατρέπει σε συμβολοσειρές, έτσι, οι αναλυμένες νότες επιστρέφονται ως πίνακας. Ομοίως, αναλύονται αρχεία MIDI για την προ επεξεργασία του μοντέλου LSTM, ξαναφορτώνει τα αρχεία MIDI, εξάγει τις νότες από το πιάνο, τις προσθέτει σε μια λίστα, εκτυπώνει τον συνολικό αριθμό των νοτών και η λίστα αποθηκεύεται σε ένα αρχείο.

4.5.2 Εφαρμογή μοντέλων παραγωγής

Το πρόγραμμα δέχεται ως είσοδο μερικές παραμέτρους, οι οποίες αφορούν όλα τα δεδομένα για να πραγματοποιηθεί η εκπαίδευση του μοντέλου, το οποίο μοντέλο θα χρησιμοποιηθεί μεταξύ του LSTM και WaveNet, με βάση ποιον σύνθετη έχει επιστρέψει η κατηγοριοποίηση για το κομμάτι που δόθηκε από τον χρήστη. Στην αρχή, αρχικοποιείται το dataset και πραγματοποιείται η φόρτωση των midi αρχείων. Έπειτα, πραγματοποιείται μια προ επεξεργασία των midi αρχείων ανάλογα το μοντέλο που έχει επιλεγεί όπου καταγράφονται οι νότες και το λεξιλόγιο όλων των αρχείων. Στη συνέχεια αρχικοποιείται είτε το μοντέλο LSTM είτε το μοντέλο Wavenet και δέχονται όλες τις νότες και το λεξιλόγιο ώστε να εκπαιδευτούν και να παράγουν μουσική. Κατά την εκπαίδευση των μοντέλων δημιουργούνται λεξικά με τις νότες και συγχορδίες και μέσω αυτών προετοιμάζονται οι ακολουθίες (sequences) του μοντέλου. Έτσι δημιουργούνται οι ακολουθίες εισόδου και οι αντίστοιχες εξόδου οι οποίες αναδιαμορφώνονται και κανονικοποιούνται σε μια μορφή που είναι συμβατή με τα επίπεδα των μοντέλων. Στη συνέχεια ορίζεται η δομή του LSTM ή Wavenet μοντέλου που θα εκπαιδευτούν για την παραγωγή νέου μουσικού κομματιού.

Παραγωγή με χρήση του μοντέλου LSTM

Η εκπαίδευση του LSTM μοντέλου, όπου χωρίζει τα δεδομένα σε σύνολα εκπαίδευσης και ελέγχου, αποθηκεύει τα δεδομένα ελέγχου, ρυθμίζει ένα σημείο ελέγχου μοντέλου για να αποθηκεύει τα καλύτερα βάρη (weights) κατά τη διάρκεια της εκπαίδευσης με βάση την απώλεια ελέγχου και εκπαιδεύει το μοντέλο LSTM χρησιμοποιώντας τις καθορισμένες παραμέτρους και επανακλήσεις (callbacks). Έπειτα, διαχωρίζονται τα δεδομένα σε υποσύνολα εκπαίδευσης (network_input) και ελέγχου (network_output). Τέλος, το μοντέλο εκπαιδεύεται λαμβάνοντας τα δεδομένα εισόδου, τα δεδομένα εξόδου και άλλες παραμέτρους όπως ο αριθμός των epochs (200), το batch size (128) και η λίστα επανάκλησης για την εξοικονόμηση των καλύτερων βαρών μοντέλου.



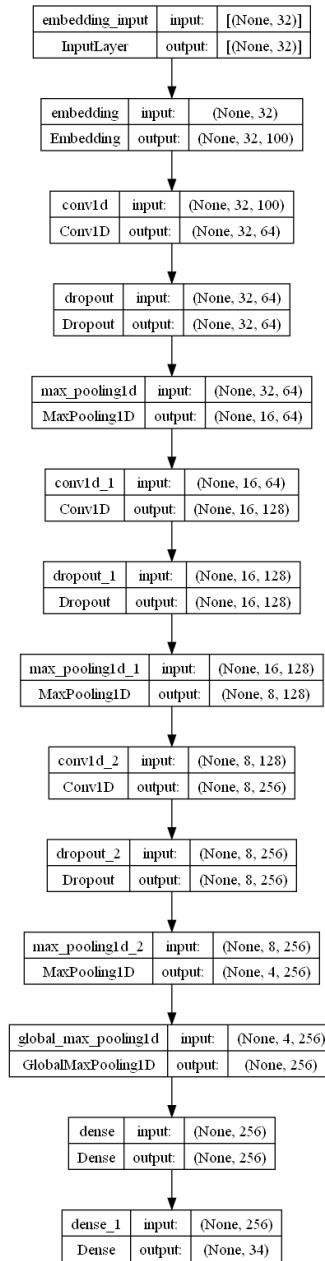
Εικόνα 37 Οπτική αναπαράσταση της αρχιτεκτονικής του μοντέλου LSTM

Η παραγωγή της μουσικής με βάση το εκπαιδευμένο μοντέλο γίνεται από δύο βασικές συναρτήσεις. Η πρώτη συνάρτηση χρησιμοποιεί το εκπαιδευμένο μοντέλο για να δημιουργήσει την ακολουθία από νότες και να τις μετατρέψει σε μελωδία χρησιμοποιώντας τη δεύτερη συνάρτηση η οποία μετατρέπει ένα απόσπασμα μουσικών στοιχείων σε μια μελωδία που αναπαρίσταται ως αρχείο MIDI. Αρχικά, καλείται η πρώτη συνάρτηση που δέχεται μια παράμετρος τον αριθμό των νοτών που θέλουμε να παράγουμε. Ξεκινά επιλέγοντας έναν τυχαίο σπόρο από ένα δεδομένο σύνολο και, στη συνέχεια, προβλέπει επαναληπτικά την επόμενη νότα της ακολουθίας με βάση τον προηγούμενο σπόρο, προσαρμόζοντας τις πιθανότητες πρόβλεψης χρησιμοποιώντας λογάριθμο και εκθετικότητα. Η προβλεπόμενη νότα προστίθεται στην ακολουθία και χρησιμοποιείται ως ο επόμενος σπόρος για την επόμενη επανάληψη. Οι νότες που δημιουργούνται στη συνέχεια μετατρέπονται σε μελωδία χρησιμοποιώντας μια καθορισμένη αντιστοίχιση και, τέλος, η μελωδία αποθηκεύεται ως αρχείο MIDI.

Παραγωγή με χρήση του μοντέλου Wavenet

Τα βήματα για την εκπαίδευση και παραγωγή μουσικής μέσω του Wavenet μοντέλου είναι όμοια με του LSTM. Η βασική διαφοροποίηση βρίσκεται στον ορισμό της δομής του μοντέλου. Σε αυτή την περίπτωση η διαδικασία δημιουργεί ένα διαδοχικό μοντέλο με ένα επίπεδο ενσωμάτωσης, ακολουθούμενο από πολλά 1D συνελκτικά επίπεδα με εγκατάλειψη και μέγιστη συγκέντρωση, και στη συνέχεια προσθέτει καθολική μέγιστη συγκέντρωση, πυκνά στρώματα και ένα στρώμα εξόδου SoftMax και συντάσσει το μοντέλο με την αραυή κατηγορική απώλεια διασταυρούμενης εντροπίας και Adam optimizer.

Έπειτα, δημιουργούνται οι ακολουθίες εισόδου-εξόδου, διαχωρίζονται σε δεδομένα εκπαίδευσης και ελέγχου (80/20), αποθηκεύει τα δεδομένα ελέγχου και ρυθμίζει ένα σημείο ελέγχου μοντέλου για να αποθηκεύει τα καλύτερα βάρη (weights) κατά τη διάρκεια της εκπαίδευσης με βάση την απώλεια ελέγχου και εκπαιδεύει το μοντέλο Wavenet χρησιμοποιώντας τις καθορισμένες παραμέτρους και επανακλήσεις (callbacks). Τέλος, η παραγωγή του τελικού αποτελέσματος χρησιμοποιεί το εκπαιδευμένο μοντέλο για να δημιουργήσει μια ακολουθία προβλεπόμενων νοτών ή συγχορδιών που βασίζονται σε μια τυχαία επιλεγμένη είσοδο από ένα σύνολο δεδομένων ελέγχου. Αυτές οι προβλεπόμενες τιμές μετατρέπονται σε αρχείο MIDI.



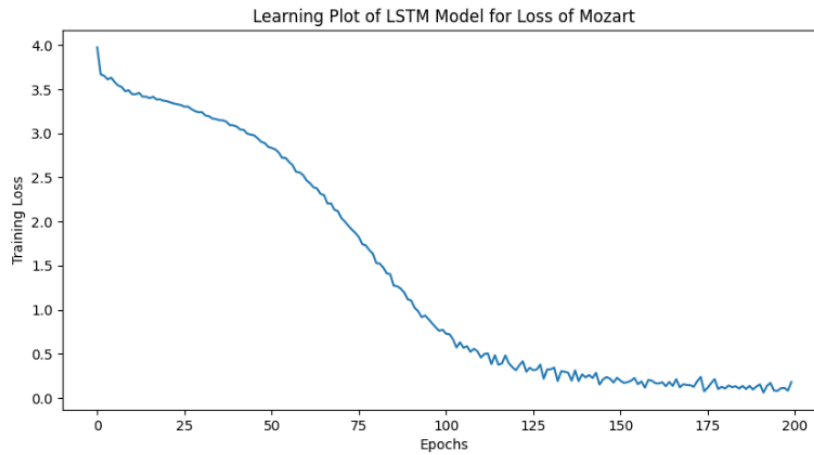
Εικόνα 38 Οπτική αναπαράσταση της αρχιτεκτονικής του μοντέλου WaveNet

4.5.3 Αποτελέσματα

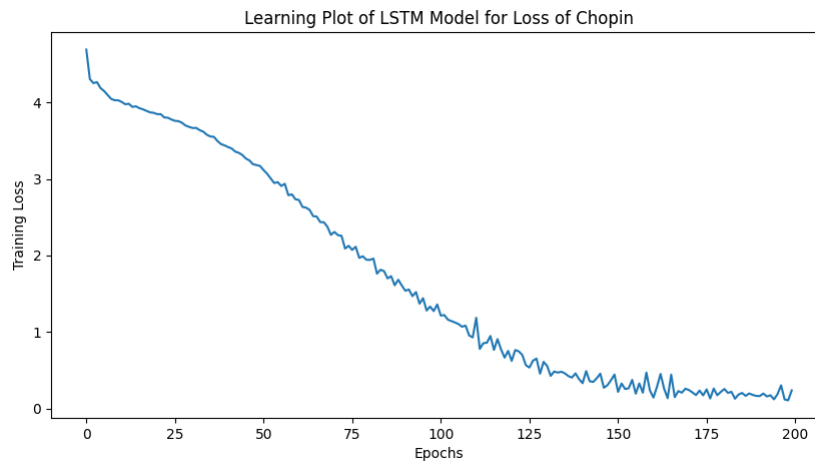
Η συνάρτηση απώλειας (loss), αν και αντικατοπτρίζει μόνο την ικανότητα του μοντέλου να εκπαιδεύεται στα δεδομένα, είναι ένας σημαντικός μετρικός τρόπος για την αξιολόγηση της απόδοσης των μοντέλων μηχανικής μάθησης κατά τη διάρκεια της εκπαίδευσης. Η μέση απώλεια υποδηλώνει τη μέση διαφορά μεταξύ της παραγόμενης μουσικής και των αρχικών συνθέσεων που χρησιμοποιήθηκαν για την εκπαίδευση. Μια χαμηλότερη μέση απώλεια υποδηλώνει ότι το μοντέλο μπορεί να αναπαράγει στενά τη

μουσική του συνθέτη στυλ, ενώ μια υψηλότερη μέση απώλεια μπορεί να υποδηλώνει περισσότερες παραλλαγές ή αποκλίσεις από τα δεδομένα εκπαίδευσης.

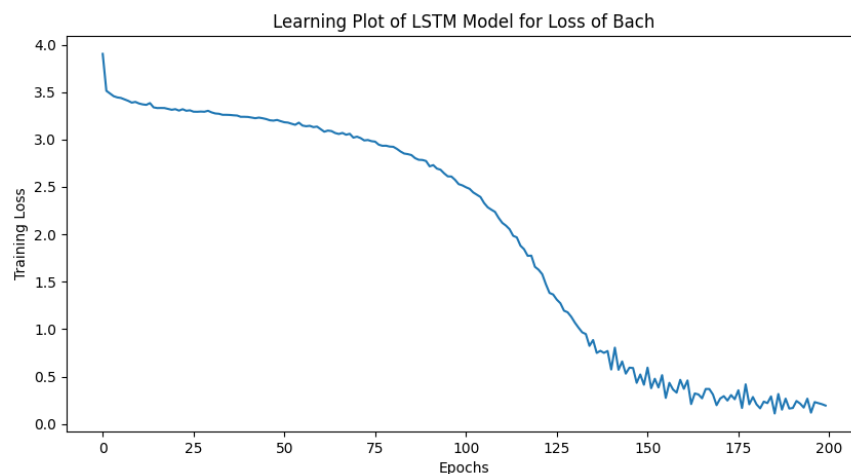
Αποτελέσματα εκπαίδευσης μοντέλου LSTM



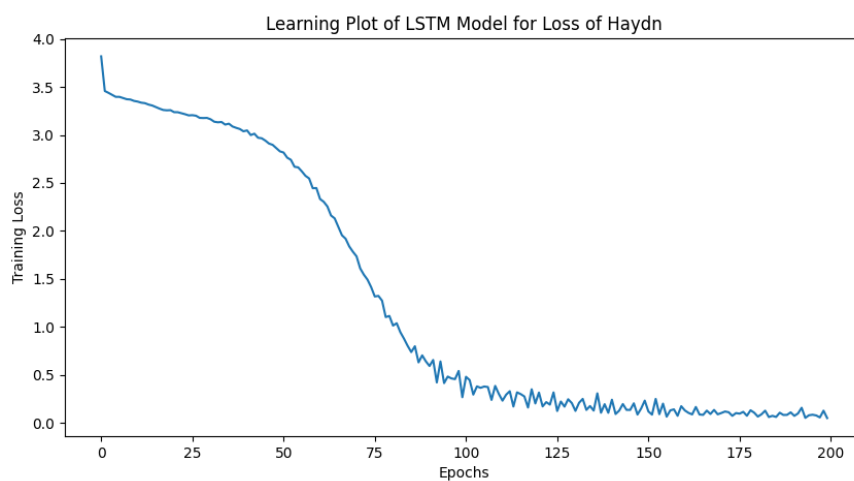
Γράφημα 5 Απώλεια (loss) Mozart



Γράφημα 6 Απώλεια (loss) Chopin



Γράφημα 7 Απώλεια (loss) Bach



Γράφημα 8 Απώλεια (loss) Haydn

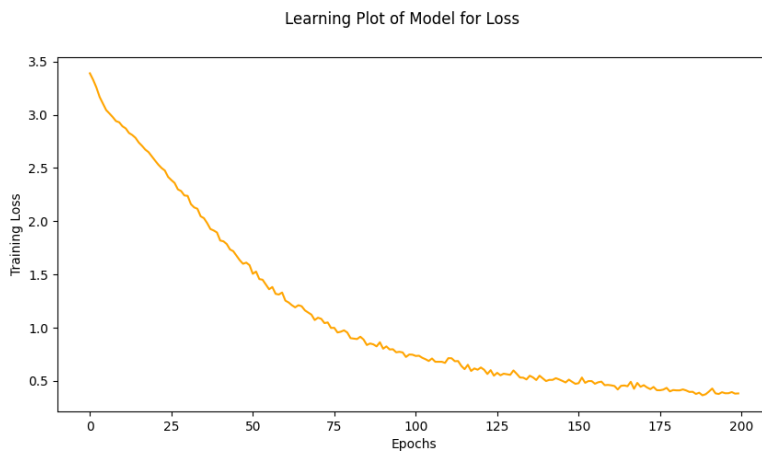
Πίνακας 12 Απώλεια (loss) min, max, mean του μοντέλου LSTM

COMPOSER	MIN	MAX	MEAN
MOZART	0.063189	3.975025	1.407450
BACH	0.110841	3.902534	1.989052
HAYDN	0.051192	3.818545	1.255480
CHOPIN	0.106205	4.689918	1.700122

Αποτελέσματα εκπαίδευσης μοντέλου Wavenet



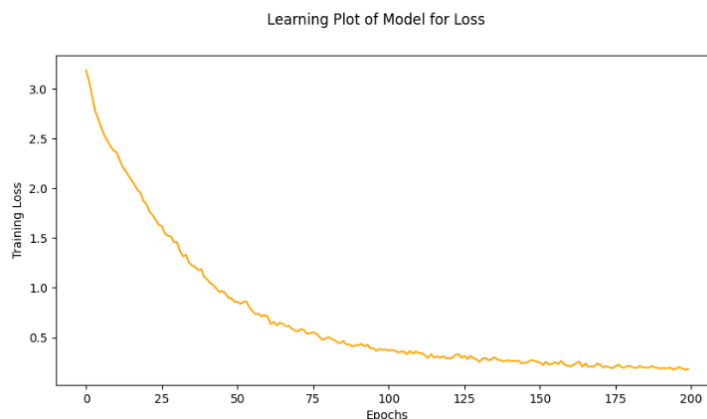
Γράφημα 9 Mozart



Γράφημα 10 Bach



Γράφημα 11 Chopin



Γράφημα 12 Haydn

Πίνακας 13 Απόλεια (loss) min, max, mean του μοντέλου WaveNet

COMPOSER	MIN	MAX	MEAN
MOZART	0.6453	3.4396	2.8791
BACH	2.3336	3.3934	1.9342
HAYDN	0.7529	3.1984	1.829
CHOPIN	1.7484	3.2960	2.4875

4.6 Ανάλυση αποτελεσμάτων

Αρχικά, η διαδικασία ταξινόμησης περιλαμβάνει την ανάλυση διαφόρων μουσικών μοτίβων και δομών μέσα στα δεδομένα MIDI για να κάνει μια τεκμηριωμένη εικασία σχετικά με τον συνθέτη. Στη συγκεκριμένη περίπτωση, η αξιολόγηση της ακρίβειας του προγράμματος με τη χρήση διαφόρων αλγορίθμων κατέληξε σε μια σχετικά χαμηλή ακρίβεια. Οι αλγόριθμοι που είχαν την μεγαλύτερη ακρίβεια ήταν οι Neural nets (0,58), Random forest (0,56) και Decision trees(0,53) με τον μέσο όρο της ακρίβειας να προσεγγίζει το 0,56 ή 56%.

Η επίτευξη υψηλής ακρίβειας στην ταξινόμηση αρχείων MIDI μπορεί να αποτελέσει πρόκληση λόγω της πολυπλοκότητας και της ποικιλομορφίας των μουσικών συνθέσεων από διαφορετικούς συνθέτες. Τα αρχεία MIDI περιέχουν μια αναπαράσταση των μουσικών νοτών, των χρονισμών και άλλων σχετικών πληροφοριών, αλλά στερούνται ορισμένων κρίσιμων στοιχείων, όπως η δυναμική, η έκφραση και οι ερμηνευτικές αποχρώσεις που είναι εγγενείς στην ανθρώπινη ερμηνεία.

Η περιορισμένη ακρίβεια του “0,56” υποδηλώνει ότι οι αλγόριθμοι που υλοποιούνται σε αυτό το πρόγραμμα μπορεί να δυσκολεύονται να συλλάβουν τα μοναδικά υφολογικά χαρακτηριστικά και τις αποχρώσεις που χαρακτηρίζουν κάθε συνθέτη. Είναι πιθανό ότι οι αλγόριθμοι από μόνοι τους μπορεί να μην έχουν αρκετό βάθος ή πολυπλοκότητα για να διακρίνουν αποτελεσματικά μεταξύ διαφορετικών συνθετών με βάση αποκλειστικά δεδομένα αρχείων MIDI.

Όσον αφορά τα αποτελέσματα του προγράμματος στο πεδίο της παραγωγής της μουσικής με μεθόδους βαθιάς μάθησης, η αξιολόγηση είναι μια ουσιαστική πτυχή της. Οι μετρικές χρησιμοποιούνται για τη μέτρηση της ποιότητας και της μουσικότητας της παραγόμενης μουσικής, δηλαδή, παρέχουν ένα ποσοτικό μέτρο της απόδοσης του μοντέλου. Η επιλογή της μετρικής εξαρτάται από τα εφαρμοζόμενα μοντέλα, τους στόχους της εργασίας, καθώς και από το μουσικό πεδίο στο οποίο η μουσική παραγωγή λειτουργεί. Μια από τις πιο συχνά χρησιμοποιούμενες μετρήσεις στη μηχανική μάθηση ως αντικειμενική συνάρτηση είναι η απώλεια (loss).

Στα αποτελέσματα της παραγωγής της μουσικής με τον μοντέλο του LSTM παρατηρείται η μέση απώλεια για κάθε συνθέτη ξεχωριστά. Συγκρίνοντας τις μέσες απώλειες διαφορετικών συνθετών, δηλαδή του Mozart, του Bach, του Haydn και του Chopin, μπορούν να γίνουν διάφορες παρατηρήσεις. Αυτές οι μέσες απώλειες παρέχουν ένα μέτρο της μέσης απόκλισης ή σφάλματος μεταξύ των προβλεπόμενων και των πραγματικών τιμών σε ένα δεδομένο πλαίσιο. Ξεκινώντας με τον Haydn, το μοντέλο LSTM ήταν πιο επιτυχημένο, αφού η μέση απώλεια είναι 1.255480 και υποδηλώνει ότι, κατά μέσο όρο, οι προβλέψεις που βασίζονται στις συνθέσεις του τείνουν να είναι πιο κοντά στις πραγματικές τιμές σε σύγκριση με τους άλλους συνθέτες. Ωστόσο, οι συνθέσεις του Haydn παρουσιάζουν ένα μέτριο επίπεδο απόκλισης από τις αναμενόμενες τιμές και είναι σημαντικό να σημειωθεί ότι τα έργα του Haydn χαρακτηρίζονται συχνά από τα καινοτόμα και απρόβλεπτα στοιχεία τους, τα οποία θα μπορούσαν να συμβάλουν σε αυτή την ελαφρώς υψηλότερη μέση απώλεια.

Προχωρώντας στον Mozart, του οποίου η μέση απώλεια είναι 1.407450, βλέπουμε μια ελαφρώς υψηλότερη τιμή σε σύγκριση με τον Haydn, αλλά εξακολουθεί να είναι χαμηλότερη από τις μέσες απώλειες του Bach και του Chopin. Η μουσική του Mozart φημίζεται για την πολυπλοκότητα και το συναισθηματικό της βάθος, γεγονός που μπορεί να συμβάλλει σε υψηλότερη μέση απώλεια σε σχέση με τον Haydn.

Ακολουθεί ο Chopin, του οποίου η μέση απώλεια είναι 1.700122, η οποία βρίσκεται μεταξύ του Mozart και του Bach, υποδεικνύοντας ότι οι συνθέσεις του τείνουν να αποκλίνουν περισσότερο από τις αναμενόμενες τιμές σε σύγκριση με τον Haydn, αλλά λιγότερο από τον Bach. Το ρομαντικό ύφος του Chopin περιλαμβάνει συχνά περίπλοκες αρμονίες και εκφραστικές μελωδίες, γεγονός που ενδέχεται να καθιστά πιο δύσκολη την ακριβή πρόβλεψη των αναμενόμενων τιμών.

Τέλος, ο Bach έχει την υψηλότερη μέση απώλεια μεταξύ των συνθετών που εξετάστηκαν, με τιμή 1.989052. Με αυτό παρατηρείται, κατά μέσο όρο, ότι υπάρχει σχετικά μεγαλύτερη απόκλιση μεταξύ των προβλεπόμενων και των πραγματικών τιμών στις συνθέσεις του Bach σε σύγκριση με τους άλλους συνθέτες. Το σχολαστικό και ακριβές ύφος του Bach στις συνθέσεις του είναι γνωστό ότι δύσκολα αποτυπώνεται, αυτό φαίνεται να αντανακλάται στην υψηλή μέση απώλεια.

Στα αποτελέσματα της παραγωγής της μουσικής με τον μοντέλο του Wavenet, ξεκινώντας με τον Mozart, το μοντέλο Wavenet πέτυχε μέση απώλεια 2,8971, παρατηρείται ότι η παραγόμενη μουσική για τις συνθέσεις του Mozart αποκλίνει ελαφρώς από τα πρωτότυπα, με υψηλότερο επίπεδο απόκλισης. Ενώ τα παραγόμενα κομμάτια μπορεί να εξακολουθούν να αποτυπώνουν κάποια ουσία του ύφους του Mozart, υπάρχουν αξιοσημείωτες παραλλαγές και αποκλίσεις από τα χαρακτηριστικά μοτίβα και τις αρμονίες του συνθέτη. Προχωρώντας στον Bach, διαπιστώνουμε χαμηλότερη μέση απώλεια 1,9342, οι παραγόμενες συνθέσεις πιθανόν να παρουσιάζουν μεγαλύτερη ομοιότητα με τα πρωτότυπα έργα του Bach, με υψηλότερο επίπεδο πιστότητας, συνεπώς το μοντέλο φαίνεται ότι έμαθε και αναπαρήγαγε αποτελεσματικά την περίπλοκη αντίστιξη και τις περίτεχνες αρμονίες του Bach.

Η μέση απώλεια 1,829 του Haydn υποδηλώνει ακόμη μεγαλύτερη ευθυγράμμιση μεταξύ της παραγόμενης μουσικής και των συνθέσεων του Haydn. Τα παραγόμενα κομμάτια είναι πιθανό να παρουσιάζουν υψηλό βαθμό ομοιότητας με τις αρχικές συνθέσεις του Haydn. Τέλος, έχουμε τον Chopin, με μέση απώλεια 2,4875, ενώ η τιμή αυτή είναι υψηλότερη σε σύγκριση με τον Haydn και τον Bach, εξακολουθεί να φαίνεται ένα λογικό επίπεδο ομοιότητας μεταξύ της παραγόμενης μουσικής και των συνθέσεων του Chopin. Το μοντέλο ενδέχεται να δυσκολεύεται να συλλάβει ορισμένες από τις μοναδικές συναισθηματικές αποχρώσεις και τα δεξιοτεχνικά στοιχεία που χαρακτηρίζουν τη μουσική του Chopin, οδηγώντας σε ελαφρώς υψηλότερες αποκλίσεις στα παραγόμενα κομμάτια.

Με βάση τα αποτελέσματα του μοντέλου Wavenet, το μοντέλο αυτό φαίνεται να υπερέχει στην αποτύπωση της ουσίας του Bach και του Haydn, ενώ τα αποτελέσματα του Mozart και τους Chopin παρουσιάζουν ελαφρώς υψηλότερα επίπεδα απόκλισης. Αυτές οι διαφορές θα μπορούσαν να αποδοθούν στις διαφορετικές πολυπλοκότητες και ιδιομορφίες που υπάρχουν στη μουσική κάθε συνθέτη, καθώς και στους πιθανούς περιορισμούς του μοντέλου Wavenet στην ακριβή αναπαραγωγή ορισμένων στυλ ή στοιχείων.

Συνεπώς, φαίνεται ότι, η σύγκριση των μέσων απωλειών μεταξύ του μοντέλου LSTM και του μοντέλου Wavenet παρέχει πληροφορίες σχετικά με τις σχετικές επιδόσεις τους στην αποτύπωση των μουσικών στυλ διαφορετικών συνθετών. Ενώ και τα δύο μοντέλα επιδεικνύουν πλεονεκτήματα και αδυναμίες, το μοντέλο LSTM παρουσιάζει ελαφρώς βελτιωμένη απόδοση για τον Bach σε σύγκριση με το μοντέλο Wavenet, ενώ το WaveNet δυσκολεύεται περισσότερο με τις συνθέσεις του Mozart και του Chopin. Ωστόσο, είναι σημαντικό να σημειωθεί ότι η μέση απώλεια από μόνη της δεν αποτυπώνει την πλήρη πολυπλοκότητα και

ποιότητα των μουσικών συνθέσεων και απαιτείται περαιτέρω αξιολόγηση για να εκτιμηθεί η συνολική αποτελεσματικότητα των μοντέλων στη δημιουργία μουσικής που αποτυπώνει την ουσία αυτών των διάσημων συνθετών.

Συνοπτικά, οι μέσες απώλειες των τεσσάρων συνθετών παρέχουν πληροφορίες σχετικά με τη συνολική ακρίβεια και προβλεψιμότητα των έργων των συνθετών, αναδεικνύοντας τα μοναδικά χαρακτηριστικά των μουσικών τους στυλ. Ωστόσο, ενώ η απώλεια είναι μια σημαντική μετρική για την αξιολόγηση της απόδοσης των μοντέλων μηχανικής μάθησης κατά τη διάρκεια της εκπαίδευσης, αντανακλά μόνο την ικανότητα του μοντέλου να επεξεργάζεται δεδομένα και όχι την ποιότητα της παραγόμενης μουσικής. Καταληκτικά, η πρόσθετη επιλογή άλλων μουσικά σημαντικών μετρικών είναι ζωτικής σημασίας για την αξιολόγηση της ποιότητας της παραγόμενης μουσικής μετά την ολοκλήρωση της εκπαίδευσης των μοντέλων.

5 Συμπεράσματα και μελλοντικές προκλήσεις

Η περιορισμένη διαθεσιμότητα των συνόλων δεδομένων αποτελεί πρόκληση για την πρόοδο των πολύπλοκων μοντέλων στον τομέα μας. Η ανάγκη για πιο ολοκληρωμένα σύνολα δεδομένων αποτελεί πρόκληση για την βέλτιστη υλοποίηση της εργασίας με σκοπό την ανάπτυξη καλύτερων μοντέλων δημιουργίας μουσικής. Η έλλειψη ποικίλων και ολοκληρωμένων συνόλων δεδομένων καθιστά δύσκολη την εκπαίδευση μοντέλων που μπορούν να αποτυπώσουν την πολυπλοκότητα και τις αποχρώσεις των μουσικών εκτελέσεων, οδηγώντας στο περιορισμό της ικανότητας των μοντέλων να παράγουν καλύτερη ποιότητα μουσικής. Ως εκ τούτου, υπάρχει ανάγκη δημιουργίας περισσότερων συνόλων δεδομένων.

Ένας άλλος πρωταρχικός περιορισμός στον τομέα της δημιουργίας μουσικής είναι η ποιότητα των συνόλων δεδομένων εκπαίδευσης. Όπως αναφέρεται στην εργασία από τα συμπεράσματά, τα σύνολα δεδομένων εκπαίδευσης που χρησιμοποιήθηκαν ήταν γενικά πρότυπα. Έτσι παράγονται χαμηλότερης ποιότητας αποτελέσματα, κάτι που μπορεί να φανεί σε πολλές αξιολογήσεις μετρικών. Ένας παράγοντας που μπορεί να επηρεάσει σημαντικά την ποιότητα αυτών των συνόλων δεδομένων είναι η παρουσία ελλিপών και ανοργάνωτων κομματιών με ασυνέχειες, μεγάλες παύσεις και θόρυβο. Λόγω των παραπάνω, είναι απαραίτητη η προσεκτική επιμέλεια και προ επεξεργασία των συνόλων δεδομένων εκπαίδευσης με σχολιασμούς υψηλής ποιότητας.

Μια άλλη πιθανή οδός βελτίωσης είναι η ενσωμάτωση πρόσθετων χαρακτηριστικών πέραν των ακατέργαστων δεδομένων MIDI. Η εξαγωγή χαρακτηριστικών υψηλότερου επιπέδου, όπως μελωδικά μοτίβα, ακολουθίες συγχορδιών ή ρυθμικά μοτίβα, θα μπορούσε να παρέχει περισσότερες διακριτικές πληροφορίες για την ταξινόμηση. Επίσης, η ανάπτυξη μεθόδων για την ερμηνεία και την κατανόηση των αποφάσεων που λαμβάνει το μοντέλο μπορεί να ενισχύσει την εμπιστοσύνη και να επιτρέψει την καλύτερη ανάλυση των αποτελεσμάτων. Τεχνικές όπως η ανάλυση της σημασίας των χαρακτηριστικών ή οι μηχανισμοί προσοχής μπορούν να είναι χρήσιμες από αυτή την άποψη. Επιπλέον, η ενσωμάτωση εξωτερικής γνώσης, όπως ειδικοί στιλιστικοί κανόνες συνθέτη ή ιστορικό πλαίσιο, θα μπορούσε να βελτιώσει περαιτέρω την ακρίβεια του προγράμματος.

Σε μελλοντικό χρόνο θα μπορούσαν να γίνουν πολλές βελτιώσεις με σκοπό την μεγαλύτερη ακρίβεια. Η ενίσχυση δεδομένων, δηλαδή, η επέκταση του συνόλου δεδομένων μέσω διαφόρων τεχνικών, όπως η σύνθεση δεδομένων, η αύξηση της δειγματοληψίας ή η δημιουργία νέων παραδειγμάτων, μπορεί να βοηθήσει στη βελτίωση της απόδοσης και της γενίκευσης του μοντέλου. Επιπρόσθετα, η προηγμένη προ επεξεργασία, όπως η βελτίωση των τεχνικών προ επεξεργασίας δεδομένων, με άλλα λόγια, η εξέλιξη της κανονικοποίησης, της κλιμάκωσης χαρακτηριστικών ή του χειρισμού ελλিপών τιμών, μπορεί να οδηγήσει σε βελτιωμένη απόδοση του μοντέλου. Η εξερεύνηση πιο προηγμένων αρχιτεκτονικών ή προσαρμοσμένων

σχεδίων δικτύων μπορεί ενδεχομένως να βελτιώσει την ικανότητα του μοντέλου ώστε να έχει την ικανότητα να συλλαμβάνει σύνθετα μοτίβα και σχέσεις στα δεδομένα. Ακόμη, και η προσαρμογή της αρχιτεκτονικής του μοντέλου και της διαδικασίας εκπαίδευσης στα ειδικά χαρακτηριστικά του προβληματικού τομέα μπορεί ενδεχομένως να οδηγήσει σε σημαντικές βελτιώσεις στην απόδοση. Σημαντικοί παράγοντες για την μελλοντική βελτιστοποίηση του προγράμματος είναι η μείωση της μεροληψίας, η βελτίωση της ακρίβειας και η αύξηση της ευρωστίας, αυτά μπορούν να επιτευχθούν με τη δημιουργία μοντέλων συνόλου με συνδυασμό πολλαπλών μοντέλων ή προβλέψεων. Άλλη μια προσέγγιση είναι η εφαρμογή τεχνικών που επιτρέπουν στο μοντέλο να μαθαίνει από νέα δεδομένα με την πάροδο του χρόνου, προσαρμόζοντας τις γνώσεις του και ενημερώνοντας τις παραμέτρους του σταδιακά, αυτή η αυξητική εκπαίδευση μπορεί να βοηθήσει στη βελτίωση των επιδόσεων καθώς γίνονται διαθέσιμες νέες πληροφορίες.

Για τα μοντέλα που πραγματεύεται η εργασία υπάρχουν βλέψεις βελτιώσεις στις τεχνικές εκπαίδευσης, επειδή η εκπαίδευση βαθιών μοντέλων όπως τα LSTM και Wavenet μπορεί να είναι υπολογιστικά δαπανηρή και χρονοβόρα, οι μελλοντικές εργασίες μπορούν να επικεντρωθούν στην ανάπτυξη πιο αποτελεσματικών τεχνικών εκπαίδευσης, όπως προηγμένοι αλγόριθμοι βελτιστοποίησης, μέθοδοι συμπίεσης μοντέλων ή στρατηγικές κατανεμημένης εκπαίδευσης, για την επιτάχυνση της διαδικασίας εκπαίδευσης χωρίς να διακυβεύεται η απόδοση. Επίσης, η κατανόηση της διαδικασίας λήψης αποφάσεων των μοντέλων LSTM και Wavenet, είναι δύσκολη, κάτι που δίνει το έναυσμα για μελλοντική έρευνα που να επικεντρώνεται στην ανάπτυξη τεχνικών για την ερμηνεία και την επεξήγηση των προβλέψεων και των εσωτερικών αναπαραστάσεων αυτών των μοντέλων. Αυτό θα μπορούσε να περιλαμβάνει μεθόδους όπως η οπτικοποίηση της προσοχής, η δημιουργία εξηγήσεων κατανοητών από τον άνθρωπο για την ενίσχυση της διαφάνειας και την οικοδόμηση εμπιστοσύνης στα αποτελέσματα του μοντέλου.

Σημαντική θα αποτελούσε η μελλοντική διερεύνηση εναλλακτικών μοντέλων. Μπορεί το LSTM και το Wavenet να έχουν δείξει πολλά υποσχόμενα αποτελέσματα, αλλά υπάρχουν διάφορες άλλες αρχιτεκτονικές που αξίζει να διερευνηθούν. Το μοντέλο των Transformers συγκεντρώνει ενδιαφέρον γύρω από την παραγωγή μουσικής και η εξερεύνηση του αλλά και άλλων εναλλακτικών μοντέλων μπορεί να οδηγήσει σε νέες ανακαλύψεις και ιδέες στη μοντελοποίηση διαδοχικών δεδομένων.

Συνοψίζοντας, είναι σημαντικό να σημειωθεί ότι η αποτελεσματικότητα αυτών των προσεγγίσεων που αναφέρθηκαν στην συγκεκριμένη εργασία μπορεί να ποικίλλει, ο πειραματισμός και η εμπειρική αξιολόγηση είναι ζωτικής σημασίας για οποιαδήποτε έρευνα, υλοποίηση και βελτιώσεις σε ένα δεδομένο πλαίσιο.

6 Βιβλιογραφία

- (2005). Ανάκτηση από Internet Archive Way Back Machine:
<http://www.carolinaclassical.com/articles/haydn.html>
- (2015). Ανάκτηση από Pandas Introduction: https://www.w3schools.com/python/pandas/pandas_intro.asp
- (2018). Ανάκτηση από Jetbrains: <https://www.jetbrains.com/pycharm/features/>
- (2019). Ανάκτηση από The Python Standard Library: <https://docs.python.org/3/library/index.html>
- (2021). Ανάκτηση από Introduction to Seaborn – Python: <https://www.geeksforgeeks.org/introduction-to-seaborn-python/>
- (2023). Ανάκτηση από ti-einai.gr: <https://ti-einai.gr/synthesi/>
- ALGORITHMIC COMPOSER.* (2010). Ανάκτηση από <http://www.algorithmiccomposer.com/>
- Almeida, D. d., & Pinho, M. (2018). *Music Generation Using Generative Adversarial.*
- Ariza, C. (2011). Two Pioneering Projects. *Computer Music Journal* .
- Ashish. (2019). 15 Most Important Features of Scikit-Learn!
- Atmadja, A. R., Uriawan, W., Pritisen, F., Maylawati, D. S., & Arbain, A. (2019). *Comparison of Naive Bayes and K-nearest neighbours for online transportation using sentiment analysis in social media.*
- Bottou, L. (2010). *Large-Scale Machine Learning with Stochastic Gradient Descent.*
- Carl, R. (2009). *Terry Riley's In C (Studies in Musical Genesis, Structure, and Interpretation).*
- Classicfm.* (2016). Ανάκτηση από <https://www.classicfm.com/artists/steve-reich/music/piano-phase-visualisation/>
- Computer Software by Laurie Spiegel.* (2012). Ανάκτηση από <http://www.musicmouse.com/>
- Davis, H., & Mohammad, S. M. (2014). *Generating Music from Literature.* arXiv.org.
- Dawar, H. (2020). Stochastic Gradient Descent. *Analytics Vidhya.*
- documentation, p. 0. (2019). Ανάκτηση από <https://craffel.github.io/pretty-midi/>
- Dubnov, S., Assayag, G., Lartillot, O., & Bejerano, G. (2003). *Using Machine-Learning Methods for Musical Style Modeling.*

- Engel, J., Resnick, C., Roberts, A., Dieleman, S., Norouzi, M., Eck, D., & Simonyan, K. (2017). *Neural Audio Synthesis of Musical Notes*.
- Eno, B. (2011). *Composers as Gardeners*.
- Fedorak, B. (2020). Ανάκτηση από <https://www.kaggle.com/datasets/blanderbuss/midi-classic-music>
- Fernandez, J., & Vico, F. (2013). *AI Methods in Algorithmic Composition: A Comprehensive Survey*.
- Fox, C. (2006). *Genetic Hierarchical Music Structures*.
- Gamal, B. (2020). Exploring Naive Bayes: Mathematics, How it works, Pros & Cons, and Applications. *Analytics Vidhya*.
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2014). *Generative Adversarial Nets*.
- Goren, O., Nachmani, E., & Wolf, L. (2021). *A-Muze-Net: Music Generation by Composing*.
- Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). *Improved Training of Wasserstein GANs*.
- Hartley, R. (2008). Ανάκτηση από <https://www.cs.nmsu.edu/~rth/research/mt/MT2/sld006.htm>
- Herremans, D., & Sörensen, K. (2012). *Composing Fifth Species Counterpoint Music*.
- Herremans, D., Weisser, S., Sörensen, K., & Conklin, D. (2014). *Generating structured music using quality metrics*. Research Papers in Economics - RePEc.
- Horn, F. (2020). *A Practitioner's Guide to Machine Learning*.
- Hu, Z., Liu, Y., Chen, G., & Liu, Y. (2022). *Can Machines Generate Personalized Music? A Hybrid Favorite-aware Method for User Preference Music Transfer*.
- Jacob, B. L. (1996). *Algorithmic Composition as a Model of Creativity*.
- Javatpoint. (2022). Ανάκτηση από <https://www.javatpoint.com/reinforcement-learning>
- Jazz Studies Online. (2020). Ανάκτηση από <https://www.jazzstudiesonline.org/resource/graphic-visualization-rich-piano-roll-score>
- Jordan, J. (2017). Convolutional neural networks.
- Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2018). *PROGRESSIVE GROWING OF GANS FOR IMPROVED*.
- Ketkar, N. (2017). Stochastic Gradient Descent. Στο *Deep Learning with Python*. Berkeley, CA: Apress.

- Kim, B., & Yeo, W. S. (2013). *Probabilistic prediction of rhythmic characteristics in markov chain-based melodic sequences*.
- Liu, I.-T., & Ramakrishnan, B. (2014). *BACH IN 2014: MUSIC COMPOSITION WITH RECURRENT NEURAL NETWORK*.
- Malik, I., & Ek, C. H. (2017). *Neural Translation of Musical Style*. Ανάκτηση από Medium.
- Marchini, M., & Purwins, H. (2011). *Unsupervised Analysis and Generation of Audio Percussion Sequences*.
- McLean, A., & Dean, R. T. (2018). *The Oxford Handbook of Algorithmic Music*.
- Miranda, E. R. (2021). *Handbook of Artificial Intelligence for Music*. Springer Cham.
- musicology, m. a.-a. (2021). Ανάκτηση από <http://web.mit.edu/music21/>
- Oord, A. v., Dieleman, S., Simonyan, K., Vinyals, O., & Graves, A. (2016). *WAVENET: A GENERATIVE MODEL FOR RAW AUDIO*.
- Papadopoulos, G., & Wiggins, G. (1999). *AI Methods for Algorithmic Composition: A Survey, a Critical*.
- Petit, B., & Serrano, M. (2021). *Generative Music Using Reactive Programming*.
- Poirier, D. (2017). Python Tool Review: Using PyCharm for Python Development - and More.
- Raffel, C., & Ellis, D. P. (2016). *EXTRACTING GROUND TRUTH INFORMATION FROM MIDI FILES*.: New York.
- Rai, S. (2021). Music Genres Classification using Deep learning techniques. *Analytics Vidhya*.
- Rakshit, S. (2019). Ανάκτηση από <https://www.kaggle.com/datasets/soumikrakshit/classical-music-midi>
- Rakshit, S. (2021). Ανάκτηση από <https://www.kaggle.com/datasets/soumikrakshit/classical-music-midi>
- Rana, R., Epps, J., Jurdak, R., Li, X., Goeck, R., Brereton, M., & Soar, J. (2016). *Gated Recurrent Unit (GRU) for Emotion*.
- Rasmussen, C. E., & Williams, C. K. (2006). *Gaussian Processes for Machine Learning*.
- Scoates, C. (2013). *Brian Eno: Visual Music*. Ανάκτηση από <https://knowuh.github.io/IDV2/frankenbits/ps-02/D/d.html>
- Shewalkar, A., Nyavanandi, D., & Ludwig, S. A. (2019). *PERFORMANCE EVALUATION OF DEEP NEURAL NETWORKS APPLIED TO SPEECH RECOGNITION: RNN, LSTM AND GRU*.

Shih, Y.-J., Wu, S.-L., Zalkow, F., Muller, M., & Yang, Y.-H. (2022). *Theme Transformer: Symbolic Music Generation with Theme-Conditioned Transformer.*

Sigtia, S., Benetos, E., Boulanger-Lewandowski, N., Weyde, T., Garcez, A. S., & Dixon, S. (2014). *A HYBRID RECURRENT NEURAL NETWORK FOR MUSIC TRANSCRIPTION.*

Smith, M. (2016). *Medium.* Ανάκτηση από https://medium.com/@_MatthewJSmith/beyond-rpa-getting-to-know-systems-that-think-and-learn-ef65d3c02193

Suzuki, M. (2021). *Score Transformer: Generating Musical Score from Note-level Representation.*

Tatsat, H., Puri, S., & Lookabaugh, B. (2020). *Machine Learning and Data Science Blueprints for Finance.*

TORO, M., RUEDA, C., AGÓN, C., & ASSAYAG, G. (2016). *GELISP: A FRAMEWORK TO REPRESENT MUSICAL.* JATIT & LLS.

Tran, M. L., Lee, D., & Jung, J.-H. (2022). *Machine Composition of Korean Music via Topological Data Analysis.*

Troxel, D. (2016). *MUSIC TRANSCRIPTION WITH A CONVOLUTIONAL NEURAL NETWORK.*

Woolhouse, M., & Rohrmeier, M. (2014). *Is there a relationship between pitch attraction and.*