Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

Αναγνώριση είδους μουσικής από συμβολικά δεδομένα (MIDI) με τεχνικές μηχανικής μάθησης

Διπλωματικη Εργασια

του

ΝΙΚΟΛΑΟΥ Δ. ΜΑΚΑΡΗ

Επιβλέπων: Γεώργιος Στάμου
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Εργαστηριο Συστηματων Τεχνητης Νοημοσυνης και Μαθησης
Αθήνα, Νοέμβριος 2020

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών
Εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης

Αναγνώριση είδους μουσικής από συμβολικά δεδομένα (MIDI) με τεχνικές μηχανικής μάθησης

Διπλωματικη Εργασια

του

ΝΙΚΟΛΑΟΥ Δ. ΜΑΚΑΡΗ

Επιβλέπων: Γεώργιος Στάμου
Αναπληρωτής Καθηγητής Ε.Μ.Π.

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

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Γεώργιος Στάμου	Ανδρέας Σταφυλοπάτης	Νικόλαος Παπασπύρου
Αναπλ. Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2020

(Υπογραφή)

.....

ΜΑΚΑΡΗΣ ΝΙΚΟΛΑΟΣ

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

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών
Εργαστήριο Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης

Περίληψη

Το θέμα της παρούσας διπλωματικής εργασίας είναι η αναγνώριση είδους μουσικής με ανάλυση μουσικών κομματιών από συμβολικά δεδομένα, δηλαδή δεδομένα που είναι κωδικοποιημένα σε MIDI (Musical Instrument Digital Interface) μορφή, βασισμένη σε αρχιτεκτονικές βαθιάς μηχανικής μάθησης (Deep Learning). Το θέμα αναγνώρισης είδους μουσικής (MGR - Music Genre Recognition) αποτελεί ένα ενεργό πρόβλημα στον τομέα της άντληση πληροφορίας από μουσική (MIR - Music Information Retrieval) και συνδέεται με πολλές ερευνητικές μελέτες τα τελευταία χρόνια.

Θα χρησιμοποιηθεί επιβλεπόμενη μηχανική μάθηση και πιο συγκεκριμένα, συνελικτικά νευρωνικά δίκτυα (CNN) για την ταξινόμηση κομματιών σε συγκεκριμένες κατηγορίες. Επιλέχθηκαν διάφορα σύνολα δεδομένων για τα πειράματά μας, όπως τα Million Song Dataset, Tagtraum, Lastfm, τα οποία είναι ευρέως διαδεδομένα στον συγκεκριμένο τομέα της MIR.

Περιλαμβάνεται επίσης μια συζήτηση περί κάποιων θεωρητικών θεμάτων που σχετίζεται με τα είδη της μουσικής, δηλαδή οι μηχανισμοι που χρησιμοποιούν οι άνθρωποι για να κατηγοριοποιήσουν τη μουσική ανά είδος και το εάν μπορούν να δημιουργηθούν αντικειμενικές ταξινομήσεις είδους μουσικής.

Λέξεις Κλειδιά

Αναγνώριση Είδους Μουσικής από συμβολικά δεδομένα, Επεξεργασία MIDI δεδομένων, Νευρωνικά Δίκτυα, Συνελικτικά Νευρωνικά Δίκτυα, Αντίληψη Μουσικής Έκφρασης, Μηχανική Μάθηση, Βαθιά Μηχανική Μάθηση, δεδομένα pianoroll.

Abstract

Subject of this diploma thesis is Music Genre Classification by analyzing musical pieces that are in MIDI (Musical Instrument Digital Interface) format utilizing Deep Learning techniques. The subject of automatic music genre recognition (MGR) from MIDI music tracks has been an active problem in the field of MIR (Music Information Retrieval) and it is associated with a lot of research studies in the recent years.

Supervised machine learning techniques and more specifically, convolutional neural networks (CNN) will be used to classify musical pieces into specific categories.

Various data sets were selected for our experiments, such as Million Song Dataset, Tagtraum, Lastfm, which are widely used in this field of MIR.

Also included is a discussion of some theoretical issues related to music genres, ie the mechanisms that people use to categorize music by genre and whether objective classifications of a genre of music can be created.

Keywords

Music MIDI Genre Classification, Music MIDI Genre Recognition, MIDI Data Processing, Neural Networks, Convolutional Neural Networks, Perception of Musical Phrases, Machine Learning, Deep Learning, Pianoroll

Ευχαριστίες

Ευχαριστώ τον καθηγητή Γεώργιο Στάμου και τα μέλη του εργαστηρίου Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης για την ευκαιρία που μου δόθηκε να εργαστώ στο συγκεκριμένο θέμα της διπλωματικής εργασίας μου.

Ευχαριστώ ιδιαίτερα τον διδακτορικό Έντι Δερβάκο, που μου προσέφερε την κάθε δυνατή βοήθεια και σωστή καθοδήγηση κατά την εκπόνηση της διπλωματικής μου εργασίας όπως επίσης και το μάθημα Νευρωνικών Δικτύων που αποτέλεσε βασικό γνωσιακό υπόβαθρο για την εκτέλεση της παρούσας εργασίας.

Επίσης, θα ήθελα να ευχαριστήσω τους καθηγητές Ανδρέα-Γεώργιο Σταφυλοπάτη και Νικόλαο Παπασπύρου που συμμετέχουν στην τριμελή επιτροπή.

Τέλος, θέλω να ευχαριστήσω πολύ τους φίλους μου για την υποστήριξη που μου προσέφεραν καθ΄ όλη τη διάρκεια εκπόνησης της διπλωματικής μου εργασίας και την οικογένειά μου που είναι πάντα δίπλα μου.

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1

Εισαγωγή

1.1 Αντικείμενο της διπλωματικής εργασίας

Η ταξινόμηση μουσικής είναι ένας ευρύς και διεπιστημονικός τομέας έρευνας που προσφέρει σημαντικά οφέλη τόσο από ακαδημαϊκή όσο και από εμπορική άποψη. Το αντικείμενο της παρούσας διπλωματικής εργασίας είναι η προσπάθια προσέγγισης του προβλήματος της κατηγοριοποίησης είδους μουσικής (music genre recognition - MGR) από συμβολικά δεδομένα (MIDI) με τεχνικές μηχανικής μάθησης (Machine Learning).

Ουσιαστικά, ο πρωταρχικός στόχος αυτής της διπλωματικής εργασίας είναι η παραγωγή ενός αποτελεσματικού και εύχρηστου συστήματος λογισμικού που θα μπορούσε να ταξινομήσει αυτόματα τα μουσικά κομμάτια από συμβολικά δεδομένα (MIDI δεδομένα) σε είδος μουσικής. Αυτό θα γίνει χρησιμοποιώντας νευρωνικά δίκτυα (Neural Networks), αφού έχει επισημανθεί με μια δεδομένη κατηγοροποίηση είδους και εκπαιδεύτεί σε ένα σύνολο δεδομένων. Προτού επιτευχθεί αυτό, φυσικά, υπάρχουν ορισμένες ενδιάμεσες εργασίες που πρέπει να ολοκληρωθούν, η καθεμία με διαφορετικό βαθμό δυσκολίας και της δικής της ερευνητικής αξίας από μόνη της.

Η κατηγοριοποίηση σε είδος μουσικής χρησιμοποιείται από μουσικούς συνθέτες, μουσικές βιβλιοθήκες και άτομα γενικά ως πρωταρχικό μέσο οργάνωσης μουσικής. Δεν υπάρχει αμφιβολία ότι το είδος είναι ένα από τα πιο σημαντικά μέσα που είναι διαθέσιμα για την ταξινόμηση και την οργάνωση της μουσικής.

Η αυτόματη αναγνώριση είδους μουσικής αποτελεί ένα από τα πιο ενεργά πεδία έρευνας MIR. Ωστόσο, η περισσότερη έρευνα σε αυτό το κομμάτι γίνεται χρησιμοποιώντας ηχητικά δεδομένα (audio data), δηλαδή γίνεται προσπάθια κατηγοριοποίηση του είδους μέσω επεξεργασία σήματος είτε με εξαγωγή χαρακτηριστικών (feature extraction) και εκμάθηση χαρακτηριστικών (feature learning) μέσω νευρωνικών δικτύων. Επίσης γίνεται έρευνα για την ίδια προσπάθεια, αναλύοντας στίχους των κομματιών και εξάγωντας συμπεράσματα από αυτά.

Στη παρούσα εργασία, αντιθέτως, θα χρησιμοποιήσουμε συμβολικά δεδομένα, δηλαδή δεδομένα MIDI για την ταξινόμηση σε είδη μουσικής.

1.2 Παρεμφερείς εργασίες

Πολλές εργασίες εστιάζουν στην αναγνώριση μουσικού είδους (Music Genre Recognition - MGR) από ηχογραφήσεις, συμβολικά δεδομένα και άλλους τρόπους. Ο Sturm (2012)[46] περιγράφει μία βιβλιογραφία από τέτοιες ερευνητικές εργασίες. Θα αναφέρουμε κάποιες τις οποίες μελετήσαμε και αποτέλεσαν έμπνευση για την κατασκευή του δικού μας συστήματος αναγνώρισης είδους μουσικής.

Ένα από τα πρώτα έργα σχετικά με το θέμα της ταξινόμησης του είδους δημοσιεύτηκε από τον Gabura (1965) [18]. Αυτό το άρθρο ασχολείται αποκλειστικά με την κλασική μουσική, δυστυχώς, κάτι που περιορίζει την εφαρμογή του. Παρά αυτού του ελαττώματος και παρά την ηλικία του, ωστόσο, αυτό το άρθρο προσφέρει μερικές ενδιαφέρουσες ιδέες που φαίνεται να έχουν παραβλεφθεί σε πολλές μεταγενέστερες δημοσιεύσεις, ιδίως όσον αφορά τη χρήση σχετικά περίπλοκων στατιστικών και θεωρητικών μοντέλων για την εξαγωγή χαρακτηριστικών.

Οι Shan, Kuo (2003)[43] δημοσίευσαν ένα από τα λίγα άρθρα που ασχολούνται άμεσα με την ταξινόμηση του είδους των ηχογραφήσεων MIDI. Εξήγαγαν χαρακτηριστικά βασισμένα αποκλειστικά σε μελωδίες και συγχορδίες, και απέκτησαν ποσοστά επιτυχίας μεταξύ 64 % και 84 % για αμφίδρομες ταξινομήσεις. Όλες οι ηχογραφήσεις ανήκαν σε μία από τις τέσσερις κατηγορίες (Enya, Beatles, Chinese folk, Japanese folk), με 38 έως 55 αρχεία να χρησιμοποιούνται για κάθε κατηγορία. Αυτή η έρευνα είναι ιδιαίτερα πολύτιμη όσον αφορά τους τρόπους με τους οποίους εξήχθησαν μελωδικά χαρακτηριστικά, και θα ήταν ενδιαφέρον να δούμε πόσο καλά θα είχε λειτουργήσει το σύστημα με μεγαλύτερη ποικιλία χαρακτηριστικών και μεγαλύτερο αριθμό κατηγοριών.

Οι Chai, Vercoe (2001)[9] χρησιμοποίησαν κρυφά μοντέλα Markov για να ταξινομήσουν τις μονοφωνικές μελωδίες που ανήκουν σε έναν από τους τρεις διαφορετικούς τύπους δυτικής λαϊκής μουσικής (Αυστριακή, Γερμανική και Ιρλανδική). Κατάφεραν να επιτύχουν 63 % ακρίβεια σε ταξινομήσεις που χρησιμοποίησαν μόνο μελωδικά χαρακτηριστικά. Είναι ενδιαφέρον ότι διαπίστωσαν ότι ο αριθμός των κρυφών καταστάσεων είχε μόνο σχετικά μικρή επίδραση στα ποσοστά επιτυχίας και ότι τα απλά μοντέλα Markov ξεπέρασαν τα πιο πολύπλοκα μοντέλα.

Οι Ponce de Leon και Inesta (2002) [23] δημιούργησαν ένα σύστημα που εξήγαγε και τμηματοποίησε μονοφωνικά τζαζ και κλασικά κομμάτιαMIDI προκειμένου να εξαγάγει μελωδικά, αρμονικά και ρυθμικά χαρακτηριστικά. Το σύστημα στη συνέχεια χρησιμοποίησε αυτές τις δυνατότητες για να διαμορφώσει διακριτές κατηγορίες χρησιμοποιώντας αυτο-οργανωμένους χάρτες. Περίπου το 77% των κομματιών ταξινομήθηκαν σωστά ως ανήκουν σε μια ομάδα που αντιστοιχούσε περίπου στην τζαζ ή μια ομάδα που αντιστοιχούσε περίπου στην κλασική μουσική.

Οι Lartillot et al. (2001) [29] συζήτησαν δύο εναλλακτικές μεθόδους μη εποπτευόμενης μάθησης, δηλαδή μια βελτιωμένη μέθοδος σταδιακής ανάλυσης και δέντρα επίθημα πρόβλεψης, με σκοπό την ταξινόμηση των ηχογραφήσεων με βάση το μουσικό στυλ. Αυτό έγινε χρησιμοποιώντας αναλύσεις μουσικών ακολουθιών από άποψη ρυθμού, μελωδικού περιγράμματος και πολυφωνικών σχέσεων.

Οι Basili, Serafini, Stellato (2004) [5] έκαναν μια εξερεύνηση στη συμβολική ανάλυση

των χαρακτηριστικών μουσικής και έκαναν ταξινόμηση είδους σε ένα σύνολο δεδομένων που περιλαμβάνει περίπου 300 MIDI αρχεία που συλλέχθηκαν από το διαδίκτυο. Τα τραγούδια συγκεντρώνονται σε έξι διαφορετικά μουσικά είδη και η ταξινόμηση έγινε μέσω κατασκευής χαρακτηριστικών και δημιουργία νευρωνικών δικτύων που χρησιμοποιούν αυτά τα χαρακτηριστικά.

Στα πλαίσα του διαγωνισμού MIREX 2005 για ταξινόμηση είδους μουσικής από MIDI ηχογραφήσεις (MIREX 2005 symbolic genre classification contest), διάφορες τεχνικές χρησιμοποιήθακαν από τους συμμετέχοντες:

Το σύστημα των McKay, Fujinaga [33] εξάγει χαρακτηριστικά (feature extraction), χρησιμοποιεί δύο τεχνικές ταξινόμησης ως βασικές μονάδες στο σχήμα ταξινόμησής του: εμπρόσθια τροφοδοτούμενα νευρωνικά δίκτυα και ταξινομητή k-nearest neighbour. Επίσης χρησιμοποιεί συνδυασμούς (ensembling) των νευρωνικών μοντέλων, δημιουργώντας ένα τελικό μοντέλο.

Οι Basili, Serafini, Stellato [6] έφτιαξαν δύο συστήματα. Αρχικά το σύστημα εξάγει χαρακτηριστικά (feature extraction), και στη συνέχεια χρησιμοποιήθηκαν ο ταξινομητής Naive Bayes στην 1η περίπτωση και δέντρα αποφάσεων decision trees στην άλλη.

Οι Leon, Iniesta [30] χρησιμοποιούν στατιστικά χαρακτηριστικά των κομματιών MIDI για να φτιάξουν διαφορετικά μοντέλα για τις κατηγορίες. Ως ταξινομητής χρησιμοποιήθηκε ένωση ensembling του k-nearest-neighbour και του bayesian classifier.

Οι Kong, Choi, Yang (2020) [27] παραθέτουν συστήματα ταξινόμησης συνθετών μεγάλης κλίμακας με βάση MIDI κλασικών σόλο έργων πιάνου. Έχουν αρκετά καλή ακρίβεια ταξινόμησης έως και 73,9% και 48,9% σε ταξινόμηση 10 συνθετών και 100 συνθετών αντίστοιχα.

Ίσως η **πιο επιτυχημένη και σύγχρονη προσσέγιση** στο πρόβλημα της ταξινόμηση είδους μεγάλης κλίμακας σε συμβολικά κωδικοποιημένη μουσική είναι αυτή των Ferraro και Lemstroem (2019) [15] με αυτόματη αναγνώριση επαναλαμβανόμενων μοτίβων. Σε αυτή υπάρχει ακρίβεια ταξινόμησης της τάξεως του **45% για ταξινόμηση σε 25 κατηγορίες** και **70% σε ταξινόμηση 12 κατηγοριών** για **25 χιλιάδες και 35 χιλιάδες** πλήθος κομματιών αντίστοιχα. Χρησιμοποιείται τα σύνολα δεδομένων MASD και TopMAGD από το Million song dataset [7]

1.3 Οργάνωση του εγγράφου

Το πρώτο βήμα ήταν να μελετήσουμε και να εξετάσουμε το είδος της μουσικής από θεωρητικές και ψυχολογικές προσεγγίσεις, προκειμένου να επιτευχθεί μια ευρύτερη κατανόηση των σχετικών θεμάτων. Αυτό ήταν χρήσιμο για την απόκτηση πληροφοριών σχετικά με τον τρόπο εφαρμογής της ταξινόμησης και για την κατανόηση των υποθέσεων που θα μπορούσαμε να κάνουμε ώστε να είναι λογικές και τι είδους υποθέσεις πρέπει να αποφεύγονται. Τα αποτελέσματα αυτής της μελέτης παρουσιάζονται στο Κεφάλαιο 3.

Το επόμενο βήμα ήταν να αναλύσουμε την πρόσφατη έρευνα στην ταξινόμηση μουσικών ειδών προκειμένου να ενσωματώσουμε προηγούμενες εργασίες σε αυτό το έργο, για να μάθουμε πληροφορίες από αυτές και να δούμε πώς θα μπορούσαμε να βασιστούμε σε αυτές. Αναφέραμε

συνοπτικά για αυτές στο 1.2.

Ήταν επίσης σημαντικό να οικοδομήσουμε ένα σταθερό τεχνικό υπόβαθρο με την εξέταση σχετικών πληροφοριών σχετικά με το MIDI και την μετατροπή του σε μια άλλη, πιο προσιτή μορφή, αυτή του pianoroll. Αυτά τα θέματα καλύπτονται στο Κεφάλαιο 4.

Η επόμενη εργασία ήταν η εύρεση συνόλων δεδομένων που θα μπορούν να χρησιμοποιηθούν για την κατηγοριοποίηση είδους μουσικής. Η συλλογή μιας βιβλιοθήκης δυνατοτήτων και το τεχνικό υπόβαθρο για την δυνατότητα κατασκευής των δεδομένων που θα χρησιμοποιήσουμε για την κατηγοριοποίηση των δεδομένων. Αναφέρουμε προβλήματα που προέκυψαν από επισημάνσεις δεδομένων (ετικέτες). Περιγράφουμε την μετατροπή των κομμάτια από MIDI δεδομένα σε κατάλληλη μορφή ώστε να γίνουν είσοδοι του νευρωνικού δικτύου. Αυτά τα θέματα καλύπτονται στο Κεφάλαιο 5.

Στη συνέχεια εφαρμόστηκε μια ποικιλία μεθοδολογιών ταξινόμησης, με βάση την αναγνώριση στατιστικών προτύπων και τη μηχανική μάθηση, και δημιουργήθηκε ένα σύστημα για το συντονισμό των ταξινομητών και τη βελτίωση της συλλογικής τους απόδοσης. Στην προσπάθεια της ανάπτυξης ενός συστήματος Μηχανικής Μάθησης που θα αντιμετωπίζει το πρόβλημα της κατηγοριοποίησης κομματιών κρίθηκε απαραίτητη η μελέτη και η εφαρμογή μεθόδων Μηχανικής Μάθησης και πιο συγκεκριμένα, η μελέτη συνελικτικών νευρωνικών δικτύων (convolutional neural networks - CNN) που χρησιμοποιούνται ευρέως στην επεξεργασία εικόνας και κατηγοριοποίηση αντικειμένων από μία εικόνα (image object recognition). Με παρόμοιο τρόπο έγινε η κατασκευή των συνελικτικών νευρωνικών δικτύων για την κατηγοριοποίηση είδους μουσικής. Το θεωρητικό υπόβαθρο για τα συνελικτικά νευρωνικά δίκτυα και γενικά για τη μηχανική μάθηση και την τεχνητή νοημοσύνη αναφέρθηκαν στο κεφάλαιο 2.

Τέλος, πραγματοποιήθηκαν διάφορες δοκιμές ταξινόμησης προκειμένου να αξιολογηθεί το σύστημα και να κριθεί η απόδοσή του σε διάφορες διαστάσεις. Το Κεφάλαιο 6 εξηγεί τις δοκιμές και παρουσιάζει τα αποτελέσματα. Το Κεφάλαιο 7 συνοψίζει τα αποτελέσματα, συγκρίνει την απόδοση του συστήματος με τα υπάρχοντα συστήματα, συζητά την έννοια των αποτελεσμάτων, περιγράφει τις αρχικές ερευνητικές συνεισφορές αυτής της διατριβής και παρουσιάζει ορισμένους τομείς για μελλοντική έρευνα.

Κεφάλαιο 2

Θεωρητικό υπόβαθρο

2.1 Τεχνητή Νοημοσύνη

Το πεδίο της τεχνητής νοημοσύνης (Artificial Intelligence ή AI) επιχειρεί να κατανοήσει και να κατασκευάσει νοήμονες οντότητες. Ασχολείται με την σχεδίαση ευφυών (νοημόνων) υπολογιστικών συστημάτων, δηλαδή συστημάτων που επιδεικνύουν χαρακτηριστικά που σχετίζουμε με την νοημοσύνη στην ανθρώπινη συμπεριφορά,κατά τους Barr και Feigenbaum [48] ενώ κατά τον Bellman είναι η αυτοματοποίηση των δραστηριοτήτων που συσχετίζουμε με την ανθρώπινη σκέψη, όπως η λήψη αποφάσεων, η επίλυση προβλημάτων, η μάθηση...[40]

Η τεχνητή νοημοσύνη ανήκει στον τομέα της επιστήμης των υπολογιστών (computer science). Όμως, η ανάπτυξη συστημάτων με νοημοσύνη προϋποθέτει την μελέτη και άλλων επιστημών, όπως για παράδειγμα της ψυχολογίας, της ιατρικής και της γλωσσολογίας συνθέτοντας έτσι ένα σημείο τομής μεταξύ πληροφορικής και των επιστημών αυτών. Η τεχνητή νοημοσύνη επεκτείνεται σε πολλές κατηγορίες όπως η επεξεργασία φυσικής γλώσσας (NTLK), η αναπαράσταση γνώσης και συλλογιστική (Knowledge representation and reasoning), η ρομποτική (Robotics), η όραση υπολογιστών (Computer Vision), η μηχανική μάθηση (Machine Learning), η εξόρυξη δεδομένων (Data Mining) κ.α. [40]

Υποστηρίζεται ότι η τενχητή νοημοσύνη θα μπορούσε να φέρει επανάσταση σημαντικές αλλαγές σε κάθε μέρος της ζωής μας, στις επιχειρήσεις και στην κοινωνία γενικότερα. Ίσως προκαλέσει μία νέα βιομηχανική επανάσταση. [32]

2.2 Μηχανική Μάθηση

Πρίν ορίσουμε την μηχανική μάθηση οφείλουμε να ορίσουμε τί είναι η μάθηση. Σύμφωνα με τους Witten & Frank κάτι μαθαίνει όταν αλλάζει τη συμπεριφορά του κατά τέτοιο τρόπο ώστε να αποδίδει καλύτερα στο μέλλον. [48] Η μηχανική μάθηση (Machine Learning - ML) αποτελεί υποκατηγορία της τεχνητής νοημοσύνης με σκοπό την εκμάθηση υπολογιστικών συστημάτων. Ουσιαστικά πρόκειται για την δημιουργία μοντέλων ή προτύπων από ένα σύνολο δεδομένων από υπολογιστές. Ως μοντέλο θεωρείται μία απλοποιημένη (αφαιρετική) εκδοχή του περιβάλλοντος, ενώ ως πρότυπο θεωρείται μια δομή από συσχετιζόμενες ή οργανωμένες

εμπιρείες.

Ουσιαστικά, οι αλγόριθμοι μηχανικής μάθησης κατασκευάζουν ένα μαθηματικό μοντέλο με την χρήση κάποιων δεδομένων (δεδομένα εκπαίδευσης). Έτσι, χρησιμοποιώντας το κατασκευασμένο μοντέλο, μπορούν να προβλέψουν και να πάρουν αποφάσεις για κάποια άλλα δεδομένα (δεδομένα ελέγχου), κάποιες φορές με πολύ μεγάλη ακρίβεια. Με το μοντέλο, δηλαδή, υπάρχει η δυνατότητα της γενίκευσης. Εάν έχει "μάθει" από τα δεδομένα εκπαίδευσης, τότε εάν τα δεδομένα ελέγχου έχουν κάποια ομοιότητα ή σχέση με τα προηγούμενα τότε μπορεί να τα αναγνωρίσει.[22]

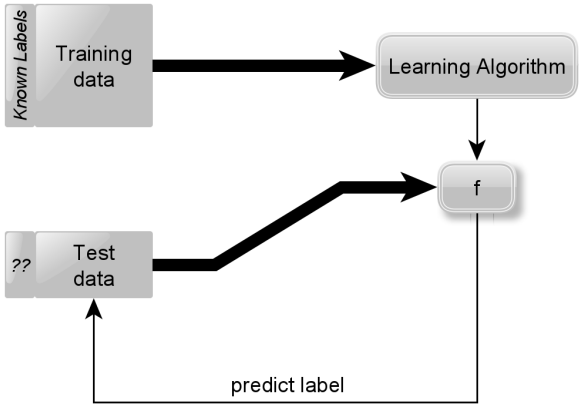
Οι αλγόριθμοι μηχανικής μάθησης μπορούν να ταξινομηθούν σε τρεις κύριες κατηγορίες (είδη) βάσει του τρόπου εκμάθησης, την επιβλεπόμενη μάθηση (Supervised learning), τη μη επιβλεπόμενη μάθηση (Unsupervised learning) και την ενισχυμένη μάθηση (Reinforcement learning).

2.2.1 Επιβλεπόμενη Μάθηση

Στην επιβλεπόμενη μάθηση, ή αλλιώς επιτηρούμενη μάθηση, η εκμάθηση γίνεται μέσω παραδειγμάτων στα οποία ξέρουμε το επιθυμητό αποτέλεσμα (τί θέλουμε να προβλέψουμε). Πιο συγκεκριμένα,το σύνολο των δεδομένων (dataset) που χρησιμοποιεί ο εκάστοτε αλγόριθμος αποτελείται από την εισόδο (input X) και την εξόδο (output Y) που καλείται και ως ετικέτες (labels).Το έργο του αλγορίθμου είναι να μάθει την αντιστοίχιση εισόδου-εξόδου. [4]

Για να το εξηγήσουμε καλύτερα, το σχήμα 2.1 μας δείχνει το γενικό πλαίσιο της κατασκευής του αλγορίθμου. Μας έχουν δοθεί τα δεδομένα εκπαίδευσης (traing data) με τα οποία ο αλγόριθμος αναμένεται να μάθει. Με βάση τα δεδομένα εκπαίδευσης, το εκπαιδευόμενο μοντέλο κατασκευάζει μια συνάρτηση στόχος f (target function), που θα αντιστοιχεί μία ετικέτα σε κάθε δεδομένο εισόδου και αποτελεί έκφραση του μοντέλου που περιγράφει τα δεδομένα.[22]

Ο στόχος είναι να προσεγγίσει την αντιστοίχιση τόσο καλά ώστε όταν θα κάνει προβλέψεις σε μια συλλογή από δεδομένα ελέγχου (test data)θα μπορεί να προβλέπει τις ετικέτες εξόδου για αυτά τα δεδομένα. Ο αλγόριθμος μηχανικής μάθησης θα έχει επιτύχει εάν η απόδοσή του στα δεδομένα ελέγχου είναι υψηλή. [22, 48]



Σχήμα 2.1: Η γενική προσέγγιση της επιβλεπόμενης μηχανικής μάθησης: Ένας αλγόριθμος που εκπαιδεύεται διαβάζει από τα δεδομένα εισόδου και δημιουργεί μία συνάρτηση f. Αυτή η συνάρτηση μπορεί μετά να δώσει ετικέτες (labels) στα δεδομένα ελέγχου (test data)

Τα προβλήματα εκπαίδευσης με επιβλεπόμενη μηχανική μάθηση μπορούν να ομαδοποιηθούν σε προβλήματα ταξινόμησης και παλινδρόμησης.

- Κατηγοριοποίηση (Classification) Τα προβλήματα ταξινόμησης ή αλλιώς κατηγοριοποίησης διαφορετικών κατηγοριών (classification), όταν αφορά δημιουργία μοντέλων πρόβλεψης διακριτών τάξεων(κλάσσεων/κατηγοριών), δηλαδή η έξοδος Y είναι μία κατηγορία, για παράδειγμα για την πρόβλεψη είδους μουσικής "κλασική" ή "ροκ" ή πρόβλεψη για το εάν υπάρχει όγκος με βάση την ανάλυση μιας ιατρικής εικόνας.

- Παλινδρόμηση (Regression) Τα προβλήματα παλινδρόμησης ή παρεμβολής (Rergression) αφορά τη δημιουργία μοντέλων πρόβλεψης αριθμητικών τιμών, δηλαδή έχουν έξοδο μία πραγματική τιμή, για παράδειγμα η τιμή της θερμοκρασίας κατά τη διάρκεια μιας ημέρας, έχοντας ως δεδομένα εκπαίδευσης θερμοκρασία άλλων ημερών ή η πρόβλεψη βάρους ενός ατόμου ξέροντας κάποια άλλα χαρακτηριστικά για αυτό.

2.2.2 Μη Επιβλεπόμενη Μάθηση

Η μη επιβλεπόμενη μάθηση περιλαμβάνει όλα τα είδη μηχανικής μάθησης, στα οποία δεν έχουμε κάποια γνωστή έξοδο (ή ετικέτα), κανέναν "δάσκαλο" για να διδάξει τον αλγόριθμο μάθησης. Στην μη επιβλεπόμενη μάθηση,τα δεδομένα εισόδου εισάγονται στο σύστημα μάθησης το οποίο καλείται να εξάγει γνώση από αυτά.

Τα προβλήματα μη επιβλεπόμενης μάθησης μπορούν να ομαδοποιηθούν σε δύο διαφορετικά είδη: μη επιβλεπόμενος μετασχηματισμός του σετ δεδομένων (Unsupervised transformations) και η συσταδοποίηση (Clustering)

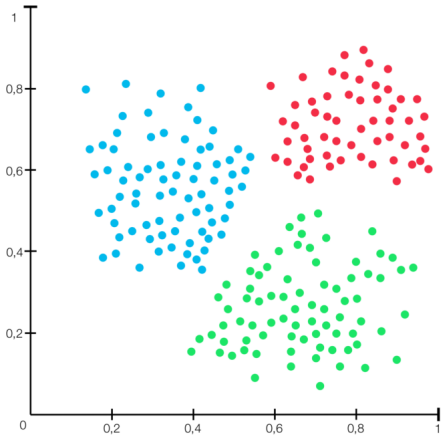
- **Μη επιβλεπόμενοι μετασχηματισμοί των δεδομένων**

  Οι μη επιβλεπόμενοι μετασχηματισμοί των σετ δεδομένων είναι αλγόριθμοι που δημιουργούν μία νέα αναπαράσταση των δεδομένων, η οποία ίσως είναι ευκολότερη για τους

ανθρώπους ή για κάποιο άλλο αλγόριθμο μηχανικής μάθησης να καταλάβουν σε σχέση με την αρχική αναπαράσταση των δεδομένων. Μία συνηθισμένη εφαρμογή της μη επιβλεπόμενης μάθησης είναι η μείωση των διαστάσεων (dimensionality reduction), η οποία λαμβάνει μία πολυδιάστατη αναπαράσταση των δεδομένων που αποτελούνται από πολλά γνωρίσματα (features) και βρίσκει έναν νέο τρόπο να αναπαραστήσει αυτά τα δεδομένα συνοψίζοντας τα απαραίτητα χαρακτηριστικά με λιγότερα γνωρίσματα. Μία συνήθης υλοποίηση είναι η ελλάτωση των διαστάσεων σε δύο διαστάσεις με σκοπό την εμφάνισή τους σε σχήματα ή εικόνες.

- **Συσταδοποίηση**

  Οι αλγόριθμοι συσταδοποίησης Clustering, αντιθέτως, διαχωρίζουν τα δεδομένα σε διακριτές ομάδες με παρόμοια αντικείμενα. Ας θεωρήσουμε το παράδειγμα των ανεβασμένων φωτογραφιών στα κοινωνικά δίκτυα. Η ιστοσελίδα μπορεί να ομαδοποιεί μαζί φωτογραφίες από το ίδιο άτομο. Σε αυτό το παράδειγμα το σύστημα δεν γνωρίζει ποια είναι τα άτομα στις φωτογραφίες, ούτε ξέρει πόσα διαφορετικά άτομα εμφανίζονται σε αυτές. Προσπαθεί να εξάγει όλα τα πρόσωπα από όλες τις φωτογραφίες και να τα συσταδοποιήσει σε ομάδες με πρόσωπα που φαίνονται όμοια. Σε άλλα προβλήματα, θα μπορούσαμε να ξέρουμε τον αριθμό των ομάδων που θα πρέπει ο αλγόριθμος να διαχωρίσει.



Πηγή: https://rocketloop.de/en/clustering-with-machine-learning/
Σχήμα 2.2: Συσταδοποίηση(Χωρισμός σε διαφορετικές ομάδες)

2.2.3 Ημι-Επιβλεπόμενη Μάθηση

Όπως προδίδει και το όνομα, ημι-επιβλεπόμενη μάθηση (semi-supervised learning) είναι η μάθηση κάπου ανάμεσα στην μη επιβλεπόμενη και στην επιβλεπόμενη. Στην πραγματικότητα, οι περισσότερες στρατηγικές για την ημιεπιβλεπόμενη μάθηση είναι βασισμένα στην επέκταση είτε της μη επιβλεπόμενης ή της επιβλεπόμενης μάθησης για να συμπεριλάβουν επιπρόσθετη πληροφορία που είναι τυπικά αναγνωρίσιμη από την άλλο πρότυπο. Πιο συγκεκριμένα, η μη επιβλεπόμενη μάθηση περιλαμβάνει πολλές διαφορετικές τεχνικές. Θα δούμε τις δύο πιο

συνηθισμένες:

- **Ημιεπιβλεπόμενη κατηγοριοποίηση (semi-supervised classification)**

  Γνωστή και ως κατηγοριοποίηση με δεδομένα με ετικέτες και δεδομένα χωρίς ετικέτες. Ουσιαστικά πρόκειται για την επέκταση της επιβλεπόμενης κατηγοριοποίησης. Τα δεδομένα εκπαίδευσης αποτελούνται από αμφότερα δεδομένα με ετικέτες (xi,yi) και δεδομένα χωρίς ετικέτες (xj). Συνήθως υποθέτουμε ότι τα δεδόμένα χωρίς καμία ετικέτα είναι πολύ περισσότερα από τα άλλα. Ο στόχος ενός τέτοιου συστήματος είναι να εκπαιδεύσει τον ταξινομητή κάνοντας χρήση και των δύο ειδών δεδομένων εκπαίδευσης, ώστε να γίνει καλύτερη εκπαίδεση από ότι σε έναν ταξινομητή με επίβλεψη κάνοντας χρήση μόνο των δεδομένων με ετικέτες.

- **Συσταδοποίηση με περιορισμούς (Constrained clustering)**

  Είναι επέκταση της συσταδοποίησης στην μη επιβλεπόμενη μάθηση. Τα δεδομένα εκπαίδευσης αποτελούνται από στιγμιότυπα χωρίς ετικέτες (xi) και από "επιβλεπόμενη πληροφόρηση" για τις συστάδες. Για παράδειγμα, τέτοια πληροφόρηση μπορεί να είναι *περιορισμοί υποχρεωτικής σύνδεσης*, δηλαδή ότι δύο δεδομένα xi, xj πρέπει να ανήκουν στην ίδια συστάδα είτε *περιορισμοί απαγορευτικής σύνδεσης*, δηλαδή ότι ότι δύο δεδομένα xi, xj πρέπει να ανήκουν σε διαφορετική συστάδα. Άλλος περιορισμός θα μπορούσε να ήταν το μέγεθος των συστάδων. Ο στόχος ενός συστήματος συσταδοποίησης με περιορισμούς είναι να δίνει καλύτερη συσταδοποίηση από μία απλή συσταδοποίση δεδομένων χωρίς ετικέτες.[51]

2.2.4 Ενισχυτική Μάθηση

Η τεχνική της ενισχυτικής Μάθησης (reinforcement learning) ασχολείται με το πρόβλημα της εύρεσης κατάλληλων ενεργειών για μια δεδομένη κατάσταση με σκοπό την μεγιστοποίηση μιας ανταμοιβής. Σε αυτήν την περίπτωση, δεν έχουν δοθεί στον αλγόριθμο εκμάθησης παραδείγματα με τις εξόδους τους, σε αντίθεση με την επιβλεπόμενη μάθηση, αλλά ο αλγόριθμος θα πρέπει να τα ανακαλύψει με μία διαδικασία δοκιμής και σφάλματος. Συνήθως υπάρχει μια ακολουθία καταστάσεων και ενεργειών στις οποίες το σύστημα αλληλεπιδρά με το περιβάλλον του. Σε πολλές περιπτώσεις, η τρέχουσα ενέργεια δεν επηρεάζει μόνο την άμεση ανταμοιβή σε αυτό το βήμα, αλλά έχει αντίκτυπο στις ανταμοιβές όλων των επόμενων χρονικών βημάτων. [8]

Για παράδειγμα, χρησιμοποιώντας κατάλληλη τεχνική στην ενισχυτική εκμάθηση ένα νευρωνικό δίκτυο μπορεί να μάθει να παίζει το παιχνίδι τάβλι σε ένα υψηλό επίπεδο.[47] Σε αυτό το παράδειγμα, το νευρωνικό δίκτυο, πρέπει να μάθει να δέχεται ως δεδομένο εισόδου τις θέσεις από όλα τα πούλια στο ταμπλό μαζί με το αποτέλεσμα της ρίψης ζαριών και να παράξει μία καλή ή έξυπνη κίνηση ως έξοδο. Αυτό συμβαίνει, έχοντας το δίκτυο να κάνει δοκιμαστικά παιχνίδια για κάποιες πιθανές κινήσεις, μπορεί και για 1000 διαφορετικά παιχνίδια. Μία μείζουσα πρόκληση είναι ότι το τάβλι μπορεί να περιλαμβάνει δεκάδες κινήσεις αλλά η μόνη ανταμοιβή που υπάρχει στο παιχνίδι είναι στο τέλος του παιχνιδιού σε περίπτωση που κερδίσει.

Η ανταμοιβή πρέπει να αποδοθεί κατάλληλα σε όλες τις κινήσεις οι οποίες οδήγησαν σε αυτή, παρόλο που κάποιες κινήσεις θα ήταν καλές ή έξυπνες και κάποιες άλλες λιγότερο καλές. Αυτό είναι ένα παράδειγμα προβλήματος ανάθεσης πίστωσης (credit assignment problem).

Ένα γενικό χαρακτηριστικό της ενισχυτικής μάθησης είναι ο συμβιβασμός ή αντιστάθμισμα μεταξύ της εξερεύνησης (exploration), στην οποία το σύστημα δοκιμάζει νέα είδη ενεργειών για να δει πόσο αποτελεσματικά είναι και της εκμετάλευσης (exploitation), στην οποία το σύστημα επιλέγει να κάνει χρήση ενεργειών που είναι γνωστές για την υψηλή ανταμοιβή τους. Αν εστιάσει υπερβολικά στην εξερεύνηση ή στην εκμετάλεσυ, τότε δεν θα μπορεί να εκπαιδευτεί και θα έχει κακή απόδοση.

2.3 Νευρωνικά Δίκτυα

Πρωτού ορίσουμε τα τεχνητά νευρωνικά δίκτυα που χρησιμοποιούνται για μηχανική μάθηση, θα εξηγήσουμε συνοπτικά τα βιολογικά νευρωνικά δίκτυα και αυτό διότι τα τεχνητά νευρωνικά δίκτυα αντλούν μεγάλο μέρος της έμπνευσής τους από το βιολογικό νευρικό σύστημα. Είναι επομένως πολύ χρήσιμο να γνωρίζουμε πώς οργανώνεται αυτό το σύστημα.

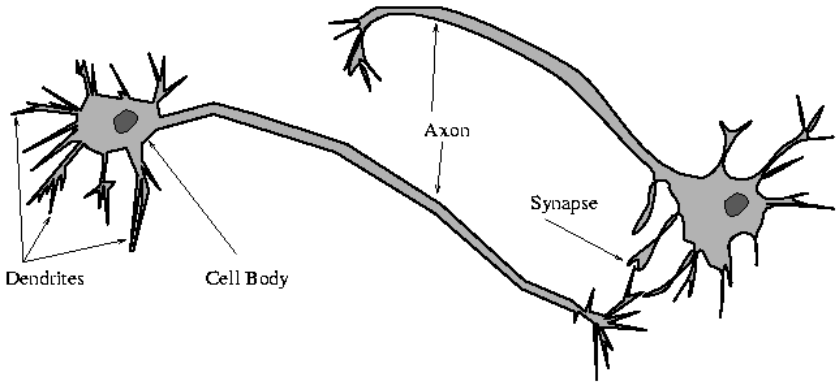
2.3.1 Βιολογικά Νευρωνικά δίκτυα

Τα περισσότερα ζωντανά πλάσματα, χρειάζονται μια μονάδα ελέγχου που είναι ικανή να μάθει και να προσαρμόζεται διαρκώς σε ένα μεταβαλλόμενο περιβάλλον. Ο εγκέφαλος των πιο αναπτυγμένων ζώων και των ανθρώπων χρησιμοποιεί πολύπλοκα δίκτυα νευρώνων για την εκτέλεση αυτού του έργου.

Η μονάδα ελέγχου (εγκέφαλος) μπορεί να χωριστεί σε διαφορετικές ανατομικές και λειτουργικές υπομονάδες, καθεμία από τις οποίες έχει συγκεκριμένα καθήκοντα όπως η όραση, η ακοή, η κίνηση και ο έλεγχος των αισθήσεων. Ο εγκέφαλος συνδέεται με νεύρα με τους αισθητήρες και τα μέρη του υπόλοιπου σώματος.

Ο εγκέφαλος αποτελείται από έναν πολύ μεγάλο αριθμό νευρώνων, περίπου 10^{11} κατά μέσο όρο. Αυτά μπορούν να θεωρηθούν ως τα βασικά δομικά τούβλα για το κεντρικό νευρικό σύστημα (ΚΝΣ). Οι νευρώνες διασυνδέονται σε σημεία που ονομάζονται συνάψεις. Η πολυπλοκότητα του εγκεφάλου οφείλεται στον τεράστιο αριθμό πολύ διασυνδεδεμένων απλών μονάδων που λειτουργούν παράλληλα, με έναν μεμονωμένο νευρώνα να λαμβάνει είσοδο από έως και 10000 άλλες.

Δομικά, ένας νευρώνας μπορεί να διαχωριστεί σε 3 μέρη: στο **κυτταρικό σώμα**, στους **δενδρίτες** και στον **άξονα**. Στο σχήμα 2.3 φαίνεται μία απλουστεμένη αναπαράσταση του βιολογικού νευρώνα.



Πηγή: A Modular Neural Network Architecture (Schmidt2000) [41]
Σχήμα 2.3: Η δομή ενός βιολογικού νευρώνα

Οι δενδρίτες είναι λεπτές και ευρέως διακλαδισμένες ίνες, φτάνοντας σε διαφορετικές κατευθύνσεις για να κάνουν συνδέσεις με μεγαλύτερο αριθμό κυττάρων εντός του συμπλέγματος κυττάρων. Μία σύνδεση γίνεται μέσω των αξόνων των άλλων κυττάρων στους δενδρίτες ή κατευθείαν στο σώμα του κυτάρου. Αυτή είναι γνωστή και ως σύναψη. Υπάρχει μόνο ένας άξονας ανά νευρώνα. Ο άξονας είναι μία πολύ λεπτή και μακριά ίνα, η οποία μεταφέρει ένα σήμα εξόδου ως ηλεκρικούς παλμούς κατά το μήκος του. Το άκρο του άξονα μπορεί να χωριστεί σε πολλούς κλάδους, οι οποίοι στη συνέχεια συνδέονται με άλλα κύτταρα. Οι κλάδοι έχουν τη λειτουργία να δίνουν το σήμα σε πολλές άλλες εισόδους.

Οι νευρώνες εκτελούν βασικά την ακόλουθη **λειτουργία**: Στο κύτταρο γίνεται **άθροισμα όλων των εισόδων**, οι οποίες μπορεί να διαφέρουν ανάλογα με την ισχύ της σύνδεσης ή τη συχνότητα του εισερχόμενου σήματος. Το άθροισμα των εισόδων υποβάλλεται σε επεξεργασία με **συνάρτηση κατωφλίου (threshold function)** και παράγει **σήμα εξόδου**.

Ο εγκέφαλος λειτουργεί **παράλληλα** και **σειριακά**. Η παράλληλη και σειριακή φύση του εγκεφάλου είναι εμφανής από τη φυσική ανατομία του νευρικού συστήματος. Ότι υπάρχει σειριακή και παράλληλη επεξεργασία μπορεί να φανεί εύκολα από το χρόνο που απαιτείται για την εκτέλεση εργασιών.

Ακόμα, τα βιολογικά νευρικά συστήματα έχουν συνήθως πολύ υψηλή ανοχή σφαλμάτων. Πειράματα με άτομα με εγκεφαλικά τραύματα έδειξαν ότι η βλάβη των νευρώνων σε ένα ορισμένο επίπεδο δεν επηρεάζει απαραίτητα την απόδοση του συστήματος, αν και σύνθετες λειτουργίες όπως η γραφή ή η ομιλία μπορεί να διδαχθούν ξανά. Αυτό μπορεί να θεωρηθεί ως **επανεκπαίδευση του νευρωνικού δικτύου(re-training the network).**

Τελικά, στα τεχνητά νευρωνικά δίκτυα δεν θα γίνει μοντελοποίηση συγκεκριμένου τμήματος ή λειτουργίας του εγκεφάλου. Αντίθετα, θα εφαρμοστούν τα **θεμελιώδη χαρακτηριστικά** του εγκεφάλου, δηλαδή τα χαρακτηριστικά του **παραλληλισμού** και της **ανοχής σφαλμάτων**. [41]

2.3.2 Τεχνητά Νευρωνικά δίκτυα

Στη βιβλιογραφία υπάρχει μια ευρεία ποικιλία ορισμών και εξηγήσεων για τους όρο Τεχνητά Νευρωνικά δίκτυα (Artificial Neural Network) ή (ANS) και για τον όρο υπολογισμός με νευρώνες. Κατά τον Ιγκόρ Αλεξάντερ "Ο υπολογισμός με νευρώνες είναι η μελέτη δικτύων προσαρμόσιμων κόμβων τα οποία, μέσω μιας διαδικασίας εκμάθησης από παραδείγματα εργασιών, αποθηκεύουν πειραματικές γνώσεις και το καθιστούν διαθέσιμο για χρήση."[2]

Ένα τεχνητό νευρικό δίκτυο είναι ένα σύστημα επεξεργασίας πληροφοριών που έχει ορισμένα χαρακτηριστικά απόδοσης κοινά με τα βιολογικά νευρωνικά δίκτυα. Τα τεχνητά νευρωνικά δίκτυα, προσπαθούν να συνθέσουν ένα απλοποιημένο μοντέλο των βιολογικών και όχι να τα αντιγράψουν.

Τα τεχνητά νευρωνικά δίκτυα έχουν αναπτυχθεί ως γενικεύσεις μαθηματικών μοντέλων ανθρώπινης γνώσης σύμφωνα με τους ακόλουθους κανόνες:

- Η επεξεργασία πληροφοριών πραγματοποιείται σε πολλά απλά στοιχεία που ονομάζονται νευρώνες.
- Τα σήματα διαβιβάζονται μεταξύ νευρώνων μέσω των συνδέσμων.
- Κάθε σύνδεσμος έχει ένα σχετικό βάρος, το οποίο, σε ένα τυπικό δίκτυο νευρώνων, πολλαπλασιάζει το μεταδιδόμενο σήμα.
- Κάθε νευρώνας εφαρμόζει μια λειτουργία ενεργοποίησης (συνήθως μη γραμμική) στην αρχική του είσοδο (άθροισμα σταθμισμένων σημάτων εισόδου) για να προσδιορίσει το σήμα εξόδου του.

[14]

2.3.3 Πλεονεκτήματα των νευρωνικών δικτύων

Ένα νευρωνικό δίκτυο οφείλει την υπολογιστική ισχύ του κατά πρώτον στην παράλληλη, κατανεμημένη δομή του και κατά δεύτερον στην ικανότητά του να μαθαίνει και, ως εκ τούτου, να γενικεύει. Τα νευρωνικά δίκτυα προσφέρουν πολλές χρήσιμες ιδιότητες και δυνατότητες. Μερικές από αυτές είναι:

- **Μη γραμμικότητα**

  Ένα νευρωνικό δικτυο αποτελούμενο από διασυνδεδεμένους μη γραμμικούς νευρώνες είναι, από τη φύση του μη γραμμικό. Αυτή η μη γραμμικότητα μπορεί να είναι κατανεμημένη σε όλη την έκταση του δικτύου. Αυτό μπορεί να είναι χρήσιμο διότι πολλές φορές η είσοδος στο νευρωνικό δίκτυο δεν είναι γραμμική.

- **Αντιστοίχιση εισόδου-εξόδου**

  Το δίκτυο τροποποιεί κατάλληλα τα συναπτικά βάρη ενός νευρωνικού δικτύου εφαρμόζοντας ένα σύνολο δειγμάτων εκπαίδευσης. Κάθε παράδειγμα αποτελείται από ένα μοναδικό σήμα εισόδου και μια αντιστοίχιση *επιθυμητή* απόκριση (στόχος).

- **Προσαρμοστικότητα**

  Τα νευρωνικά δίκτυα έχουν την δυνατότητα να *προσαρμοστούν* αλλάζοντας τα συναπτικά βάρη τους ανάλογα με τις μεταβολές που γίνονται στο περιβάλλον τους.

- **Ενδεικτική Απόκριση**

  Ένα νευρωνικό δίκτυο μπορεί να παρέχει το βαθμό εμπιστοσύνης για την έξοδό του. Για παράδειγμα για ένα σύστημα ταξινόμησης προτύπων το νευρωνικό δίκτυο θα επιλέξει ένα πρότυπο με ένα βαθμό επιστοσύνης.

- **Πληροφορία Σχετική με το Περιεχόμενο**

  Το νευρωνικό δίκτυο χειρίζεται με φυσικό τρόπο τη σχετική με το περιεχόμενο πληροφορία (contextual information). Αυτό σημαίνει ότι κάθε νευρώνας επηρεάζεται από τους άλλους νευρώνες και τις δραστηριότητες αυτών.

- **Ανοχή σε Βλάβες**

  Ένα νευρωνικό δίκτυο είνα *εύρωστο (ανεκτικό σε βλάβες)* υπό αντίξοες συνθήκες λειτουργίας,π.χ. σε καταστροφή κάποιον συνδέσεων. [21]

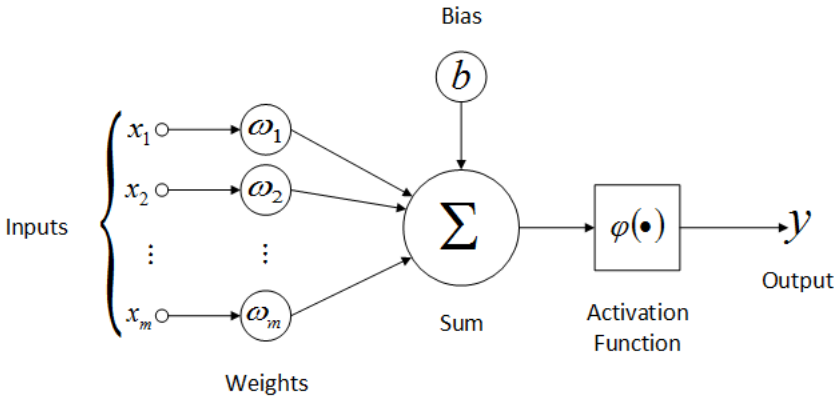
2.3.4 Ο αλγόριθμος Αντίληπτρο (Perceptron)

Το αντίληπτρο (perceptron) είναι η απλούστερη δυνατή μορφή ενός νευρωνικού δικτύου που χρησιμοποιείται για την ταξινόμηση γραμμικά διαχωρίσιμων προτύπων(δηλαδή προτύπων που βρίσκονται σε αντίθετες πλευρές ενός υπερεπιπέδου). Το perceptron του Rosenblatt (1958) βασίζεται στο μοντέλο ενός νευρώνα των McCulloch-Pitts (1943).Ένα perceptron παίρνει ένα διάνυσμα εισόδων πραγματικής αξίας, υπολογίζει ένα γραμμικό συνδυασμό αυτών των εισόδων και στη συνέχεια εξάγει 1 εάν το αποτέλεσμα είναι μεγαλύτερο από κάποιο κατώφλι και -1 διαφορετικά. Πιο συγκεκριμένα, δεδομένες εισόδους x_l έως x_n η έξοδος $y(x_1, .., x_n)$ που υπολογίζεται από το perceptron είναι:

$$y(x_1, .., x_n) = \begin{cases} 1 & \text{εάν } w_0 + w_l * x_l + w_2 * x_2 + ... + w_n * x_n > 0 \\ -1 & \text{σε άλλη περίπτωση} \end{cases}$$

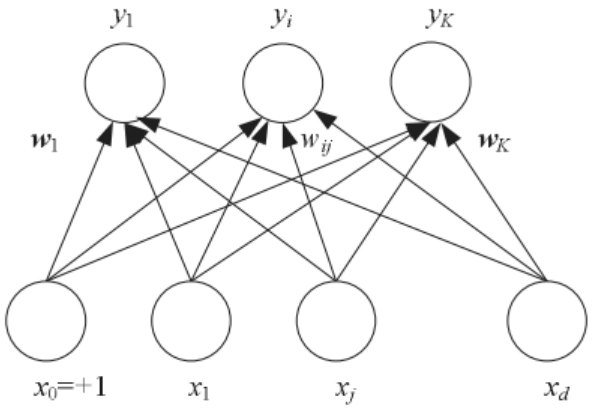
όπου κάθε w_i είναι το συναπτικό βάρος (weight), μια σταθερά πραγματικής τιμής, που καθορίζει τη συμβολή της εισόδου x_i στην έξοδο του. [36]

Αν υποθέσουμε ότι έχουμε ένα χαρακτηριστικό feature δηλαδή μία από τις εισόδους, που έχει μηδενικό βάρος τότε η έξοδος του συστήματος θα είναι η ίδια ανεξάρτητα από την τιμή αυτού του χαρακτηριστικού. Έτσι, οι είσοδοι με μηδενικό βάρος αγνοούνται. Επίσης το w_0 λέγεται «προδιάθεση» ή «πόλωση» (bias). Η συνάρτηση ενεργοποίησης (activation function) $\phi(x)$ περιορίζει το πλάτος του σήματος εξόδου στον νευρώνα και είναι αυτή που καθορίζει τις τιμές εξόδου 1 ή -1. Σε κάποιες άλλες περιπτώσεις μπορεί να έχουμε τιμές εξόδου 0 ή 1. Αυτές οι έξοδοι μπορεί να αντιπροσοπεύουν ότι δύο κλάσσεις, δηλαδή η έξοδος -1 αντιπροσοπεύει την κλάση C_1 ενώ η έξοδος +1 την κλάση C_2. [4]



Πηγή: researchgate.net/figure/Single-Perceptron-Haykin-1999_fig1_294583744
Σχήμα 2.4: Η δομή ενός νευρώνα Perceptron

Σε περίπτωση που έχουμε παραπάνω από $K > 2$ κλάσεις τις οποίες θέλουμε να ταξινομήσουμε τότε μπορούμε να χρησιμοποιήσουμε Κ αντίληπτρα, τα οποία έχουν ένα πίνακα από βάρη. (βλέπε 2.5)



Πηγή: Introduction to Machine Learing σελ.239 [4]
Σχήμα 2.5: Κ παράλληλα Perceptron x_j, όπου $j = 0, ..., d$ είναι οι είσοδοι και y_i, όπου $i = 1, ..., K$ είναι οι έξοδοι. Κάθε έξοδος είναι ένα ζυγισμένο άθροισμα των εισόδων. Ως έξοδοι είναι είτε η πιθανότητες για κάθε κλάση για μία είσοδο ή εάν τα περάσουμε από κάποια συνάρτηση ενεργοποίσης τότε θα βρούμε το max από τις πιθανότητες το οποίο θα είναι η πρόβλεψή μας.

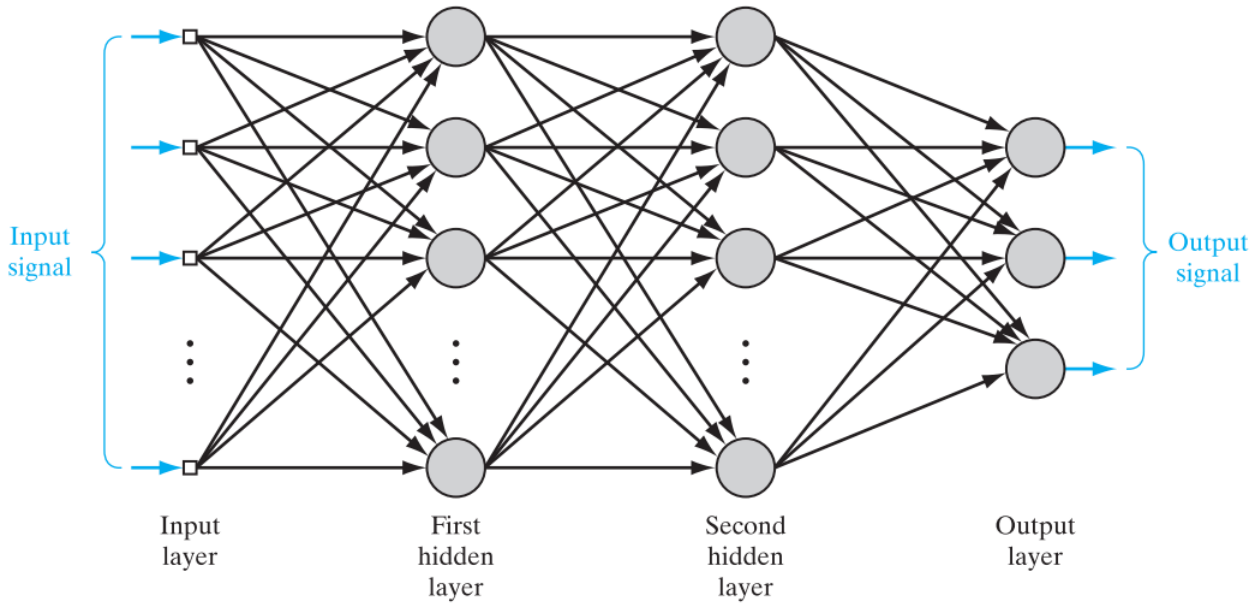
Όπως είπαμε και πιο πριν, ο νευρώνας Perceptron ανήκει στην κατηγορία των γραμμικών ταξινομητών (linear classifier), έτσι μπορεί να διαχωρίσει γραμμικά τα δεδομένα σε δύο κλάσεις. Πολλές φορές όμως τα δεδομένα δεν είναι γραμμικά διαχωρίσιμα, οπότε ο συγκεκριμένος νευρώνας δεν θα μπορέσει να ταξινομήσει τα δεδομένα σωστά. Ένα από τα πιο χαρακτηριστικά παραδείγματα μη γραμμικών διαχωριζόμενων προβλημάτων είναι το αποκλειστικό OR (XOR). Τότε χρειάζονται περισσότερα επίπεδα νευρώνων για να μπορούν να ταξινομηθούν σε

δύο κλάσεις. Τα πολυεπίπεδα νευρωνικά δίκτυα θα περιγραφούν με λεπτομέρεια στη συνέχεια (2.3.5).

2.3.5 Πολυεπίπεδα Νευρωνικά Δίκτυα

Όπως ειπώθηκε, ένας νευρώνας Perceptron μπορεί να προσεγγίσει μόνο γραμμικές συναρτήσεις, μη γραμμικά προβλήματα απαιτούν την χρήση μη γραμμικών συναρτήσεων, οι οποίες μπορούν να προσεγγιστούν από Perceptron πολλών επιπέδων νευρώνων. Θα ασχοληθούμε σε αυτό το σημείο με μία άλλη κατηγορία νευρωνικών δικτύων,τα πολυεπίπεδα νευρωνικά δίκτυα. Αυτά χαρακτηρίζονται από την παρουσία ενδιάμεσων ή κρυμμένων επίπεδων μεταξύ των επιπέδων εισόδου και εξόδου. Εάν χρησιμοποιηθούν για ταξινόμηση, τα *πολυστρωματικά perceptrons (Multilayer Perceptrons ή MLP)* μπορούν να εφαρμόσουν μη γραμμικούς διαχωρισμούς, ενώ εάν χρησιμοποιηθούν για παλινδρόμηση, μπορούν να προσεγγίσουν μη γραμμικές συναρτήσεις εισόδου.

Ουσιαστικά, ενα πολυεπίπεδο perceptron είναι απλώς μία μαθηματική συνάρτηση που αντιστοιχεί κάποιο συνόλο τιμών εισόδου σε τιμές εξόδου. Η συνάρτηση αυτή διαμορφώνεται συνθέτοντας πολλές απλούστερες συναρτήσεις. Μπορούμε να σκεφτούμε την εφαρμογή της κάθε απλής μαθηματικής συνάρτησης ως την δημιουργία μιας νέας αναπαράστασης της εισόδου.



Πηγή: Haykin Neural Networks and learning machines, σελ. 124 [21]
Σχήμα 2.6: Τοπολογία πολυεπίπεδου νευρωνικού δικτύου Perceptron (MLP)

Το σήμα εισόδου input signal τροφοδοτείται στο επίπεδο εισόδου(input layer), συμπεριλαμβανομένης της προδιάθεσης(bias), και διαδίδεται προς τα εμπρός (νευρώνα προς νευρώνα). Κάθε κρυφή μονάδα είναι ένα perceptron και δέχεται τις εισόδους, βρίσκει το σταθμισμένο

άθροισμα αυτών και εφαρμόζει τη συνάρτηση ενεργοποίησης. Η κάθε έξοδος y_i περνάει στο επόμενο επίπεδο ως είσοδος του επομένου κρυφού επιπέδου hidden layer. Τέλος, φτάνοντας στο τελεταίο επίπεδο, που αποκαλείται επίπεδο εξόδου (output layer), υπολογίζεται το τελικό σήμα εξόδου output signal.

Τα πολυεπίπεδα Peceptron (MLP) αποκαλούνται και εμπρόσθια τροφοδοτούμενα νευρωνικά δίκτυα (feedforward neural networks) ή και εμπρόσθια τροφοδοτούμενα δίκτυα βαθιάς μάθησης (Deep feedforward networks).

Αυτά τα μοντέλα ονομάζονται εμπρόσθια τροφοδοτούμενα επειδή οι πληροφορίες ρέουν μέσω της συνάρτησης που υπολογίζεται από την είσοδο x, μέσω των ενδιάμεσων υπολογισμών που χρησιμοποιούνται για τον προσδιορισμό της συνάρτησης f, και τελικά στην έξοδο y. Δεν υπάρχουν συνδέσεις ανατροφοδότησης στις οποίες οι έξοδοι του μοντέλου να ανατροφοδοτούνται, δηλαδή να γίνονται είσοδοι για νευρώνες του ίδιου επιπέδου ή προηγούμενου επιπέδου.

Όταν τα εμπρόσθια τροφοδοτούμενα νευρωνικά δίκτυα επεκτείνονται ώστε να περιλαμβάνουν συνδέσεις ανατροφοδότησης, τότε ονομάζονται επαναλαμβανόμενα νευρωνικά δίκτυα (Recurrent neural networksή RNN), τα οποία είναι χρήσιμα σε πολλέ εφαρμογές επεξεργασίας φυσικής γλώσσας natural language processing.

Ενώ, μια ειδική περίπτωση των πολυεπίπεδων peceptrons είναι και τα συνελικτικά νευρωνικά δίκτυα (convolutional neural networks), τα οποία χρησιμοποιούνται ευρέως για την αναγνώριση αντικειμένων στις φωτογραφίες και είναι αυτά που θα χρησιμοποιήσουμε για το αναγνώριση είδους μουσικής.

Τα πολυστρωματικά perceptrons ή αλλιώς εμπρόσθια τροφοδοτούμενα νευρωνικά δίκτυα είναι τα βασικά μοντέλα της βαθιάς μηχανικής μάθησης (Deep Learing).

Η σύγχρονη βαθιά μάθηση παρέχει ένα πολύ ισχυρό πλαίσιο για την εποπτευόμενη μάθηση. Προσθέτοντας περισσότερα επίπεδα και περισσότερες μονάδες μέσα σε ένα επίπεδο, ένα βαθύ νευρωνικό δίκτυο μπορεί να αντιπροσωπεύει λειτουργίες αυξανόμενης πολυπλοκότητας. Οι περισσότερες εργασίες που συνίστανται στη αντιστοίχιση ενός διανύσματος εισόδου σε έναν διάνυσμα εξόδου, και που είναι εύκολο για ένα άτομο να τα κάνει γρήγορα, μπορούν να επιτευχθούν μέσω βαθιάς μάθησης, δεδομένου ότι έχουμε αρκετά μεγάλα μοντέλα και αρκετά μεγάλα σύνολα δεδομένων με επισημασμένα παραδείγματα εκπαίδευσης. Άλλες εργασίες, που δεν μπορούν να περιγραφούν ως συσχέτιση ενός διανύσματος εισόδου με ένα εξόδου, ή που είναι αρκετά δύσκολα ώστε ένα άτομο να χρειάζεται χρόνο για να σκεφτεί και να ανταποκριθεί για να ολοκληρώσει την εργασία, παραμένει πέρα από το πεδίο της βαθιάς μάθησης προς το παρόν. [19]

Τα πολυεπίπεδα νευρωνικά δίκτυα στην διαδικασία της μάθησης, δηλαδή όταν εκπαιδεύονται και έτσι ανανεώνουν τα βάρη τους, κάνουν χρήση μιας τεχνικής επιβλεπόμενης μάθησης που ονομάζεται οπισθοδιάδοση (Backpropagation). Θα μιλήσουμε σε επόμενο κεφάλαιο για αυτή.(2.3.6.2)

2.3.6 Αλγόριθμοι Εκπαίδευσης και Βασικές Συναρτήσεις

2.3.6.1 Συνάρτηση Κόστους

Η συνάρτηση κόστους (Cost function) είναι μία μετρική της επίδοσης των συστημάτων μηχανικής μάθησης. Η επίδοση στα συστήματα μηχανικής μάθησης συνήθως μετράται ως τη διαφορά που έχουν οι προβλεπόμενες τιμές του συστήματος σε σχέση με τις αναμενόμενες τιμές. Σκοπός λοιπόν της συνάρτησης κόστους είναι να υπολογίσει το σφάλμα μεταξύ προβλεπόμενων και πραγματικών τιμών. Οι τιμές που αποκτά το σφάλμα συνηθίζεται να είναι πραγματικοί αριθμοί. Ο σκοπός της συνάρτησης κόστους είναι τις περισσότερες φορές να ελαχιστοποιήσει την τιμή της, άλλα ενδέχεται να πρέπει και να την μεγιστοποιήσει, ανάλογα με τη φύση του προβλήματος. Για παράδειγμα, αν η συνάρτηση κόστους υπολογίζει μία ανταμοιβή για το σύστημα τότε επιθυμητό θα ήταν να μεγιστοποιήσει την τιμή της. Ενώ σε αντίθετη περίπτωση, αν η συνάρτηση κόστους υπολογίζει ένα κόστος του συστήματος, όπως δηλώνει και το όνομά της, τότε η τιμή της θα πρέπει να ελαχιστοποιηθεί.

Ένα παράδειγμα συνάρτησης που προσπαθεί να ελαχιστοποιήσει την τιμή της είναι η συνάρτηση μέσου τετραγωνικού σφάλματος που περιγράφεται με την σχέση 2.1. Στη συνάρτηση μέσου τετραγωνικού σφάλματος (Mean Squarred Error) μετράται το τετραγωνικό σφάλμα μεταξύ της προβλεπόμενης ($f(x_i|\vartheta)$) και της αναμενόμενης (y_i) τιμής της εξόδου για N δεδομένα εισόδου (x_i) και παραμέτρους δικτύου (ϑ).

$$MSE(\vartheta) = \frac{1}{N}\sum_{i=1}^{N}(f(x_i|\vartheta) - y_i)^2 \tag{2.1}$$

Πρέπει να γίνει κατανοητό ότι η επιτυχία ενός νευρωνικού δικτύου συνδέεται με την ικανότητα του να γενικεύσει. Δηλαδή, για ένα σύνολο δεδομένων εκπαίδευσης το δίκτυο θα πρέπει να είναι σε θέση να προβλέψει με επιτυχία και για το σύνολο των δεδομένων δοκιμής. Η τιμή της συνάρτησης κόστους μειώνεται όσο το δίκτυο εκπαιδεύεται με τα δεδομένα εκπαίδευσης. Η εκτεταμένη εκπαίδευση του δικτύου συχνά μπορεί να οδηγήσει και στο πρόβλημα της υπερπροσαρμογής (overfitting), αυτό ουσιαστικά σημαίνει ότι το δίκτυο μαθαίνει να προβλέπει πολύ καλά για τα ΄γνωστά΄ δεδομένα εκπαίδευσης ενώ αδυναμεί στις προβλέψεις των ΄άγνωστων΄ δεδομένων δοκιμής. Σκοπός του δικτύου,όπως αναφέρθηκε, είναι να μπορέσει να γενικεύσει να μπορεί δηλαδή να προβλέψει με μεγάλη ακρίβεια για άγνωστα, στο δίκτυο, δεδομένα και όχι να προσεγγίσει την συνάρτηση που διέπει τα δεδομένα εκπαίδευσης.

2.3.6.2 Αλγόριθμος Backpropagation

Ο αλγόριθμος Backpropagation είναι ένα από τα σημαντικότερα δομικά στοιχεία ενός νευρωνικού δικτύου. Ο όρος backpropagation και η γενική χρήση του στα νευρωνικά δίκτυα αρχικά εδραιώθηκε το 1986 [10] από τους Rumelhart, Hinton και Williams, ενώ η ιδέα για την διαδικασία που περιγράφει ο αλγόριθμος υπήρχε ήδη από τη δεκαετία του 60. Ο αλγόριθμος ουσιαστικά χρησιμοποιείται για να εκπαιδεύσει ένα δίκτυο κάνοντας χρήση του κανόνα της αλυσίδας. Μια απλή περιγραφή είναι ότι μετά από κάθε εμπρόσθιο πέρασμα (forward pass) στο δίκτυο ο αλγόριθμος πραγματοποιεί ένα πέρασμα με αντίθετη φορά (backward pass)

υπολογίζοντας τις μερικές παραγώγους των παραμέτρων του δικτύου, δηλαδή τα βάρη και τις πολώσεις. Παρακάτω γίνεται μια μαθηματική περιγραφή της διαδικασίας αυτής.

Βασικός στόχος του αλγορίθμου είναι να υπολογίσει τη μερική παράγωγο μιας συνάρτησης κόστους, έστω J αναλυτική αναφορά γίνεται στην υποενότητα 2.3.6.1, ως προς τις παραμέτρους του δικτύου. Έστω ένας κόμβος με βάρος $w^l_{j,k}$ και πόλωση b^l_j, όπου τα j,k δηλώνουν την θέση του στοιχείου στο αντίστοιχο διάνυσμα ενώ το l τον αριθμό του επιπέδου στο οποίο αναφερόμαστε. Η μερική παράγωγος του βάρους $w^l_{j,k}$ υπολογίζεται με τον κανόνα της αλυσίδας ως:

$$\frac{\partial J}{\partial w^l_{j,k}} = \frac{\partial J}{\partial z^l_j}\frac{\partial z^l_j}{\partial w^l_{j,k}} \tag{2.2}$$

Εξ ορισμού, έστω m ο αριθμός των κόμβων στο επίπεδο $l-1$:

$$z^l_j = \sum_{k=1}^{m} w^l_{j,k} a^{l-1}_k + b^l_j \tag{2.3}$$

λόγω παραγώγισης προκύπτει:

$$\frac{\partial J}{\partial w^l_{j,k}} = a^{l-1}_k \tag{2.4}$$

Άρα η τελική τιμή της μερικής παραγώγου είναι:

$$\frac{\partial J}{\partial w^l_{j,k}} = \frac{\partial J}{z^l_j} a^{l-1}_k \tag{2.5}$$

Η ίδια διαδικασία ακολουθείται και για τον υπολογισμό της μερικής παραγώγου της πόλωσης b^l_j:

$$\frac{\partial J}{\partial b^l_j} = \frac{\partial J}{\partial z^l_j}\frac{\partial z^l_j}{\partial b^l_j} \tag{2.6}$$

Λόγω της εξίσωσης 2.3 προκύπτει:

$$\frac{\partial J}{\partial b^l_j} = 1 \tag{2.7}$$

και τελικά σε αυτή την περίπτωση η τιμή της μερικής παραγώγου είναι:

$$\frac{\partial J}{\partial b^l_j} = \frac{\partial J}{z^l_j} \tag{2.8}$$

Στη συνέχεια, οι τιμές των μερικών παραγώγων που υπολογίστηκαν από τον αλγόριθμο backpropagation θα χρησιμοποιηθούν από τον αλγόριθμο κατάβασης κλήσης για την ανανέωση των παραμέτρων του δικτύου. Η διαδικασία αυτή αναλύεται στην επόμενη ενότητα 2.3.7.1.

2.3.7 Θεώρημα Καθολικής Προσέγγισης

Η σημαντικότητα του θεωρήματος καθολικής προσέγγισης (Universal Approximation Theorem) συνδέεται άμεσα με την ισχύ των νευρωνικών δικτύων, συγκεκριμένα των εμπρόσθια τροφοδοτούμενων δικτύων (feedforward networks). Ένα εμπρόσθια τροφοδοτούμενο δίκτυο

είναι ένα νευρωνικό δίκτυο στο οποίο οι συνδέσεις μεταξύ των νευρώνων δεν σχηματίζουν κύκλο. Το θεώρημα καθολικής προσέγγισης αυτό που αποδεικνύει είναι ότι ένα εμπρόσθια τροφοδοτούμενο δίκτυο με μόνο ένα κρυφό επίπεδο, πέρα από τα επίπεδα της εισόδου και της εξόδου, και πεπερασμένο αριθμό νευρώνων μπορεί να προσεγγίσει οποιαδήποτε συνεχή συνάρτηση ορισμένη σε ένα κλειστό σύνολο των πραγματικών αριθμών, με οσοδήποτε μικρό σφάλμα.

Μία πιο επίσημη έκφραση του θεωρήματος καθολικής προσέγγισης διατυπωμένη με μαθηματικούς όρους είναι η παρακάτω. Έστω μια συνάρτηση $\varphi : \mathbb{R} \to \mathbb{R}$ η οποία είναι μη σταθερή, συνεχής και φραγμένη, γνωστή ως συνάρτηση ενεργοποίησης. Συμβολίζουμε με I_m τον m-διάστατο υπερκύβο στο χώρο $[0,1]^m$ και με $C(I_m)$ την κλάση όλων των πραγματικών συνεχών συναρτήσεων στο I_m. Τότε, δεδομένου οποιοδήποτε αριθμού $\varepsilon > 0$ και οποιαδήποτε συνάρτηση $f \in I_m$, υπάρχει ένας ακέραιος N, σταθερές r_i, $b_i \in \mathbb{R}$ και διανύσματα $w_i \in \mathbb{R}^m$ για $i = 1, 2, ..., N$ τέτοια ώστε να οριστεί η συνάρτηση $F(x)$:

$$F(x) = \sum_{i=1}^{N} r_i \varphi({w_i}^T x + b_i) \tag{2.9}$$

ως μια προσέγγιση της συνάρτησης f, τέτοια ώστε να ισχύει:

$$|F(x) - f(x)| < \varepsilon \tag{2.10}$$

για όλα τα $x \in I_m$.

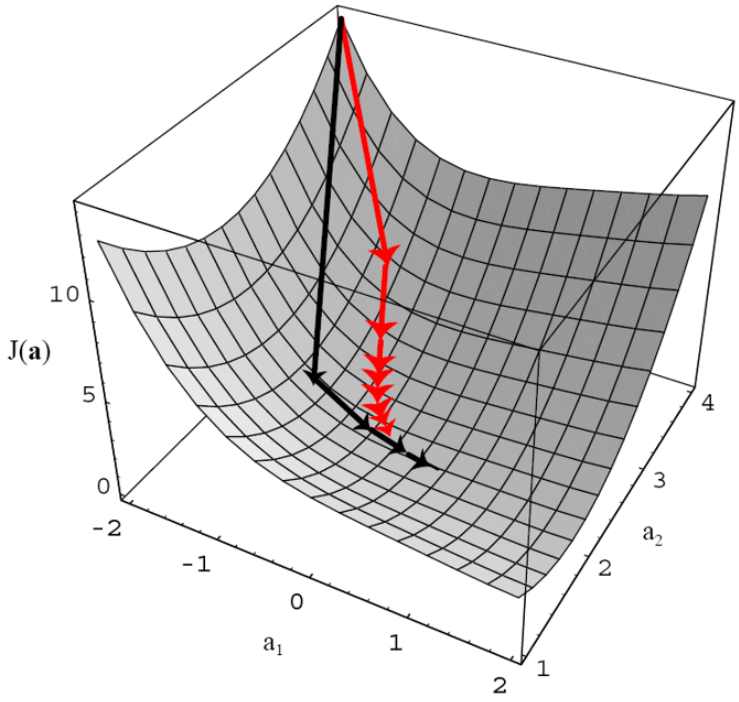
2.3.7.1 Αλγόριθμος Κατάβασης Κλίσης

Ο αλγόριθμος κατάβασης κλίσης (Gradient Descent Algorithm) είναι ένας αλγόριθμος βελτιστοποίησης και στον τομέα της μηχανικής μάθησης η χρήση του είναι πολύ συχνή. Πιο συγκεκριμένα, ο αλγόριθμος προσπαθεί να ελαχιστοποιήσει μια συνάρτηση κόστους, έστω $J(\vartheta)$ όπου ϑ είναι το διάνυσμα που αντιπροσωπεύει τις παραμέτρους του δικτύου, εν προκειμένω τα βάρη και τις πολώσεις. Ο αλγόριθμός κατάβασης κλίσης είναι επαναληπτικός και σε κάθε επανάληψη αφαιρεί μια μικρή ποσότητα από τις παραμέτρους του συστήματος. Την ποσότητα αυτή συνθέτει το γινόμενο της μερικής παραγώγου της συνάρτησης κόστους ως προς την παράμετρο που ανανεώνεται, όπως υπολογίστηκε από τον αλγόριθμο Backpropagation, επί την παράμετρο r που ονομάζεται ρυθμός μάθησης (learning rate) και καθορίζει το μέγεθος του βήματος εκμάθησης που θα κάνει κάθε φόρα ο αλγόριθμος. Με την παρακάτω σχέση 2.11 περιγράφεται πως γίνεται η ανανέωση της παραμέτρου θ_j τη χρονική στιγμή $t+1$

$$\vartheta_{j,t+1} = \vartheta_{j,t} - r \frac{\partial}{\partial \vartheta_{j,t}} J(\vartheta) \tag{2.11}$$

Στην παρακάτω εικόνα 2.7 γίνεται παρουσίαση ενός παραδείγματος της διαδρομής που ακολουθεί ο αλγόριθμος κατάβασης κλίσης προσεγγίζοντας την ελάχιστη τιμή της συνάρτησης κόστους.

Πολλές φορές ο αλγόριθμος κατάβασης κλίσης συναντάται και σε διάφορες εκδοχές οι οποίες, σε αντίθεση με την γενική περίπτωση (Batch Gradient Descent - BGD), χρησιμοποιούν



Πηγή: Duda et al. (2000) Pattern Classification [13]
Σχήμα 2.7: Γραφική απεικόνιση του αλγόριθμου κατάβασης κλίσης

ένα μέρος των δεδομένων εισόδου κάθε φορά και όχι το σύνολό τους. Συγκεκριμένα, αυτές οι εκδοχές του αλγόριθμου, χωρίζουν τα δεδομένα εισόδου σε τεμάχια (batches), όπου κάθε τέτοιο τεμάχιο χρησιμοποιείται με την σειρά για την ανανέωση των παραμέτρων του δικτύου.

Ο Batch Gradient Descent θα υπολογίσει την παράγωγο σε όλα τα δεδομένα και θα πραγματοποιήσει μια ανανέωση. Για το λόγο αυτό μπορεί να είναι πολύ αργός και για μεγάλα δεδομένα εισόδου που δεν χωρούν στη μνήμη ο έλεγχος να προκύψει πολύ δύσκολος.

Μια εκδοχή του αλγόριθμου κατάβασης κλήσης είναι ο Stohastic Gradient Descent (SGD), ο οποίος σε αντίθεση με τον προηγούμενο πραγματοποιεί ανανέωση παραμέτρων για κάθε δεδομένο εκπαίδευσης. Λόγω των συχνών ανανεώσεων, η ενημέρωση των παραμέτρων παρουσιάζει υψηλή διακύμανση και με την σειρά της η συνάρτηση κόστους λαμβάνει κυμαινόμενες τιμές. Αυτό είναι εξαιρετικά χρήσιμο, επειδή επιτρέπει στην συνάρτηση κόστους να εντοπίσει νέα τοπικά ελάχιστα. Όμως, οι συνεχείς διακυμάνσεις μπορούν να προκαλέσουν την συνεχή προσπέραση του ολικού ελαχίστου, έτσι τελικά η σύγκλιση του αλγορίθμου να προκύψει αρκετά αργή.

Μια άλλα εκδοχή του αλγορίθμου που αντιμετωπίζει το πρόβλημα αυτό είναι ο Mini Batch Gradient Descent (MBGD), που χρησιμοποιεί τα προτερήματα των δύο προηγούμενων μεθόδων. Σύμφωνα με αυτή την τεχνική τα δεδομένα εκπαίδευσης χωρίζονται σε μικρά τεμάχια που τροφοδοτούνται στο δίκτυο με τη σειρά και η ανανέωση των παραμέτρων γίνεται για κάθε τεμάχιο. Αυτή η τεχνική προκαλεί μικρότερες διακυμάνσεις στην ανανέωση των παραμέτρων με αποτέλεσμα η σύγκλιση του αλγόριθμου να είναι αρκετά πιο γρήγορη και αποτελεσματική.

2.3.7.2 Αλγόριθμοι Βελτιστοποίησης

Αντικείμενο της βελτιστοποίησης (Optimization) στα συστήματα μηχανικής μάθησης είναι η ελαχιστοποίηση, πιο σπάνια η μεγιστοποίηση, μιας συνάρτησης στόχου. Παραδείγματος χάρη ο αλγόριθμος κατάβασης κλίσης που παρουσιάστηκε στην προηγούμενη υποενότητα 2.3.7.1 είναι και αυτός ένας αλγόριθμος βελτιστοποίησης που ελαχιστοποιεί μια συνάρτηση κόστους. Για την αποδοτική και αποτελεσματική εκπαίδευση του συστήματος ο σωστός υπολογισμός και η ανανέωση των εσωτερικών παραμέτρων παίζει πολύ σημαντικό ρόλο. Για τον λόγο αυτό στρατηγικές και αλγόριθμοι βελτιστοποίησης εφαρμόζονται για τον υπολογισμό των παραμέτρων του συστήματος που θα επηρεάσουν την διαδικασία εκπαίδευσης και την έξοδό του.

Οι αλγόριθμοι βελτιστοποίησης χωρίζονται σε δύο κατηγορίες, με βάση την τάξη της παραγώγου που χρησιμοποιούν. Οι αλγόριθμοι βελτιστοποίησης *Πρώτης Τάξης (First Order Optimization Algorithms)* υπολογίζουν και χρησιμοποιούν την πρώτη παράγωγο της συνάρτησης στόχου με σκοπό την βελτιστοποίηση της. Η παράγωγος πρώτης τάξης μιας συνάρτησης σε ένα σημείο ουσιαστικά υπολογίζει την τιμή της κλίσης της στο σημείο αυτό, δηλαδή αν η συνάρτηση έχει την τάση να αυξήσει ή να μειώσει την τιμή της στο επόμενο σημείο. Ένας αλγόριθμός βελτιστοποίησης πρώτης τάξης είναι για παράδειγμα ο αλγόριθμος κατάβασης κλίσης.

Η δεύτερη κατηγορία αλγόριθμων βελτιστοποίησης είναι οι *Αλγόριθμοι Βελτιστοποίησης Δεύτερης Τάξης (Second Order Optimization Algorithms)*. Όπως είναι αναμενόμενο, οι αλγόριθμοι αυτοί υπολογίζουν και χρησιμοποιούν τις παραγώγους δεύτερης τάξης της συνάρτησης στόχου. Η δεύτερη παράγωγος μιας συνάρτησης σε ένα σημείο εκφράζει την κλίση της πρώτης παραγώγου στο σημείο αυτό, δηλαδή την κυρτότητα της συνάρτησης. Όμως ο υπολογισμός της δεύτερης παραγώγου είναι υπολογιστικά πολύ πιο δαπανηρός, έτσι οι αλγόριθμοι αυτοί δεν χρησιμοποιούνται συχνά στην πράξη. Το πλεονέκτημα των αλγόριθμων δεύτερης τάξης είναι ότι είναι αποτελεσματικότεροι καθώς δεν αγνοούν την καμπυλότητα της επιφάνειας.

Στη συνέχεια, με σκοπό την αντιμετώπιση των έντονων ταλαντώσεων που εμφανίζει ο αλγόριθμος κατάβασης κλίσης και κατά συνέπεια την δυσκολία σύγκλισης, παρουσιάζεται μια τεχνική βελτιστοποίησης που ονομάζεται *Ορμή (Momentum)* . Η τεχνική αυτή εισάγει έναν όρο γ στην 2.11 με αποτέλεσμα να επιταχύνει την διαδικασία σύγκλισης οδηγώντας τις παραμέτρους προς την σχετική κατεύθυνση και εξομαλύνοντας τις ταλαντώσεις στις άσχετες κατευθύνσεις. Όταν χρησιμοποιείται και η ορμή για την ανανέωση παραμέτρων ϑ η σχέση που τις ανανεώνει δίνεται από τις εξισώσεις:

$$V(t) = \gamma V(t-1) + r\nabla J(\vartheta) \tag{2.12}$$

$$\vartheta = \vartheta - V(t) \tag{2.13}$$

, μία τυπική τιμή για την ορμή είναι $\gamma = 0.9$.

Τρεις από τους πιο συνηθισμένους αλγόριθμους βελτιστοποίησης που χρησιμοποιούν την τεχνική της ορμής είναι οι παρακάτω:

- *Adagrad*

  Ο Adagrad [31] επιτρέπει στον ρυθμό μάθησης r να προσαρμόζεται σε κάθε παράμετρο βασιζόμενος στις παρελθοντικές κλίσεις (gradients) του. Έτσι για παραμέτρους που ανανεώνουν τις τιμές τους συχνά, ο αλγόριθμος κάνει μικρά βήματα στην ενημέρωση αυτών των παραμέτρων. Ενώ σε αντίθετη περίπτωση κάνει μεγάλα βήματα για παραμέτρους που οι τιμές τους ανανεώνονται λιγότερο συχνά. Ένα πλεονέκτημα του αλγόριθμου είναι ότι δεν απαιτείται να γίνει χειροκίνητη ρύθμιση του ρυθμού εκμάθησης, ενώ μειονέκτημα είναι ότι ο ρυθμός εκμάθησης παίρνει συνεχώς μικρότερες τιμές περιορίζοντας σημαντικά τη διαδικασία εκμάθησης.

- *AdaDelta*

  Ο αλγόριθμος AdaDelta [50] είναι μία βελτίωση του Adagrad που αντιμετωπίζει το πρόβλημα του φθίνοντος ρυθμού εκμάθησης. Χρησιμοποιεί μόνο ένα τμήμα των προηγούμενων τιμών της κλίσης της συνάρτησης για την ανανέωση των παραμέτρων σε αντίθεση με τον Adagrad που χρειαζόταν όλες τις προηγούμενες τιμές της κλίσης. Ένα ακόμη προτέρημα του αλγόριθμου είναι ότι δεν χρειάζεται προκαθορισμένη τιμή για το ρυθμό εκμάθησης.

- *Adam*

  Ο αλγόριθμός Adam (Adaptive Moment Estimation) [25] είναι και αυτός μία μέθοδος προσαρμοστική ως προς το ρυθμό εκμάθησης. Ο αλγόριθμός χρησιμοποιεί τις τιμές των ροπών πρώτης (μέσος όρος) και δεύτερης (διακύμανσης) τάξης για την ανανέωση των παραμέτρων. Σε πολλά σύγχρονα συστήματα μηχανικής μάθησης γίνεται χρήση του αλγόριθμου Adam λόγω της ικανότητας του να συγκλίνει πολύ γρήγορα και να αντιμετωπίζει τα προβλήματα που εμφανίζουν οι προηγούμενες τεχνικές.

2.3.7.3 Συνάρτηση Ενεργοποίησης

Η συνάρτηση ενεργοποίησης (Activation function) είναι η συνάρτηση που σε ένα τεχνητό νευρωνικό δίκτυο θα εφαρμοστεί στην έξοδο κάθε κόμβου και η τιμή της θα τροφοδοτηθεί στο επόμενο επίπεδο του δικτύου. Η χρήση της συνάρτησης ενεργοποίησης στα νευρωνικά δίκτυα είναι απαραίτητη καθώς εισάγει μη γραμμικότητα στο σύστημα αλλά και κανονικοποιεί την έξοδο των νευρώνων σε ένα κλειστό σύνολο.

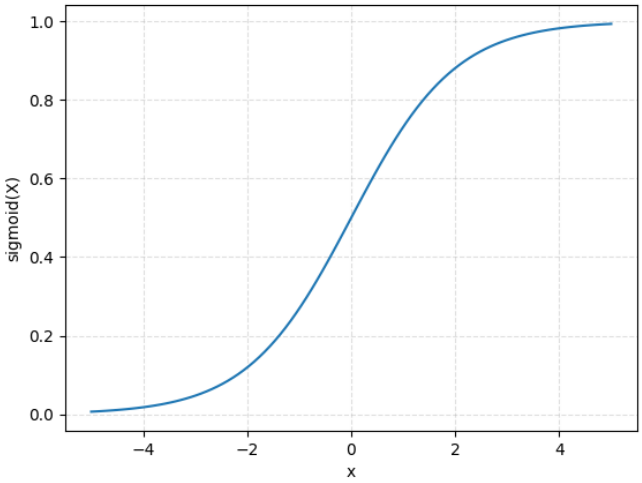
Για να γίνει πιο κατανοητή η σημασία της κανονικοποίησης της εξόδου, ας υποθέσουμε ότι έχουμε ένα νευρωνικό δίκτυο και εξετάζουμε έναν κόμβο του δικτύου με είσοδο το διάνυσμα x, βάρη w και πόλωση b. Η τιμή της εξόδου y, όπως φαίνεται στη σχέση 2.14, μπορεί να παίρνει τιμές στο $(-\infty, +\infty)$. Με την εφαρμογή της συνάρτησης ενεργοποίησης 2.15, έστω f καταφέρνουμε να αντιστοιχίσουμε την τιμή της εξόδου του νευρώνα z στο επιθυμητό σύνολο, συνήθως στο $[0, 1]$. Αυτή η διαδικασία είναι απαραίτητη ώστε ο νευρώνας να μπορέσει να διαχωρίσει τις περιπτώσεις που θα πρέπει να μεταδώσει την τιμή στα επόμενα επίπεδα ή να την αποκόψει. Τιμές κοντά στο ένα έχουν αποτέλεσμα την πυροδότηση (ενεργοποίηση) του νευρώνα ενώ τιμές κοντά στο μηδέν την αποκοπή του.

$$y = w \cdot x + b, y \in (-\infty, +\infty) \tag{2.14}$$

$$z = f(y) = f(w \cdot x + b), z \in [0, 1] \tag{2.15}$$

Μια άλλη σκοπιμότητα της συνάρτησης ενεργοποίησης είναι η εισαγωγή μη γραμμικότητας στο σύστημα, όπως αναφέρθηκε. Η μη γραμμικότητα σε ένα σύστημα μηχανικής μάθησης είναι αναγκαία ώστε να μπορέσει να προσεγγίσει μη γραμμικές συναρτήσεις. Η φύση πολλών συναρτήσεων ενεργοποίησης ως μη γραμμικές εξυπηρετούν τη σκοπιμότητα αυτή. Στη συνέχεια γίνεται μια σύντομη παρουσίαση των συνηθέστερων συναρτήσεων ενεργοποίησης:

- *Σιγμοειδής Συνάρτηση*

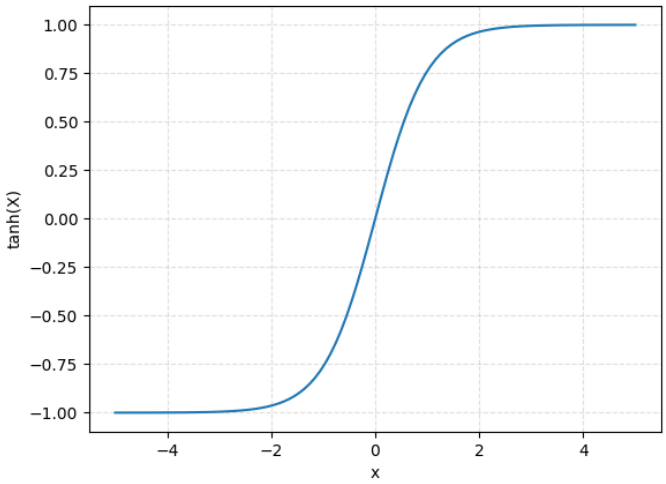


Σχήμα 2.8: Σιγμοειδής συνάρτηση

Η γραφική αναπαράσταση της σιγμοειδούς συνάρτησης Sigmoid function φαίνεται στο σχήμα 2.12, γίνεται εύκολα κατανοητή αλλά στην πράξη δεν χρησιμοποιείται συχνά. Η σιγμοειδής έχει πεδίο τιμών το $(0, 1)$ αλλά αυτό δημιουργεί πρόβλημα γιατί δεν είναι κεντραρισμένο στο μηδέν. Δύο ακόμα προβλήματα που εμφανίζει είναι η εξαφάνιση της κλίσης (Vanishing gradient problem) όπως επίσης και η αργή σύγκλιση που την χαρακτηρίζει στο στάδιο της εκπαίδευσης.

$$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.16}$$

- *Υπερβολική εφαπτομένη*

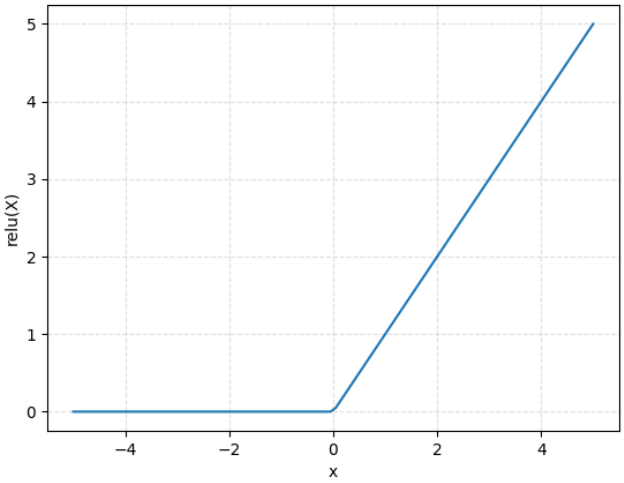


Σχήμα 2.9: Υπερβολική εφαπτομένη

Η υπερβολική εφαπτομένη (Hyperbolic tangent) 2.9 μοιάζει αρκετά με την σιγμοειδή συνάρτηση. Υπερτερεί σε σχέση με την σιγμοειδή καθώς το πεδίο τιμών της παίρνει τιμές στο $(-1, 1)$ και είναι κεντραρισμένο στο μηδέν. Ένα ακόμα προτέρημα είναι η μεγαλύτερη κλίση της που βοήθα στην ταχύτερη σύγκλιση. Όπως και η σιγμοειδής δεν καταφέρνει να αντιμετωπίσει το πρόβλημα της εξαφάνισης κλίσης.

$$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1 = 2\sigma(2x) - 1 \tag{2.17}$$

- *Rectified Linear Unit (ReLU)*

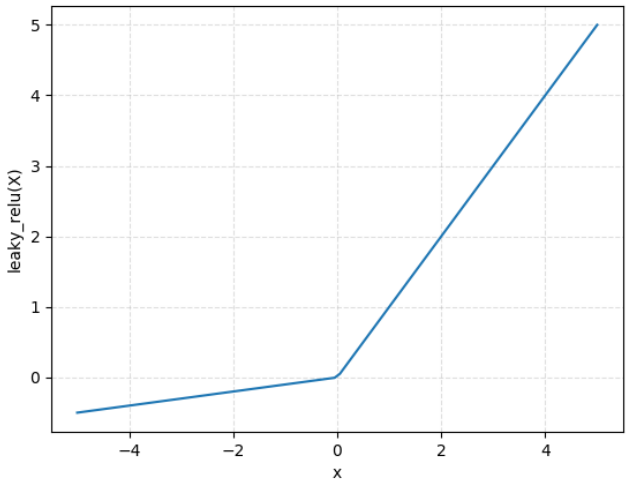


Σχήμα 2.10: Συνάρτηση ReLU

Η συνάρτηση ReLU παρουσιάζει αρκετά προτερήματα και η χρήση της στα σύγχρονα συστήματα μηχανικής μάθησης είναι συχνή. Αντιμετωπίζει με επιτυχία το πρόβλημα της εξαφάνισης κλίσης και η σύγκλισή της είναι ταχύτερη από τις δύο προηγούμενες μεθόδους. Η μαθηματική εξίσωση της ReLU 2.18 είναι πολύ απλή και αποδοτική, ο μόνος περιορισμός που εμφανίζει είναι η χρήση της μόνο στα ενδιάμεσα επίπεδα ενός νευρωνικού δικτύου.

$$f(x) = max(0, x) = \begin{cases} x, & x > 0 \\ 0, & otherwise \end{cases} \tag{2.18}$$

- *Leaky Rectified Linear Unit (Leaky ReLU)*

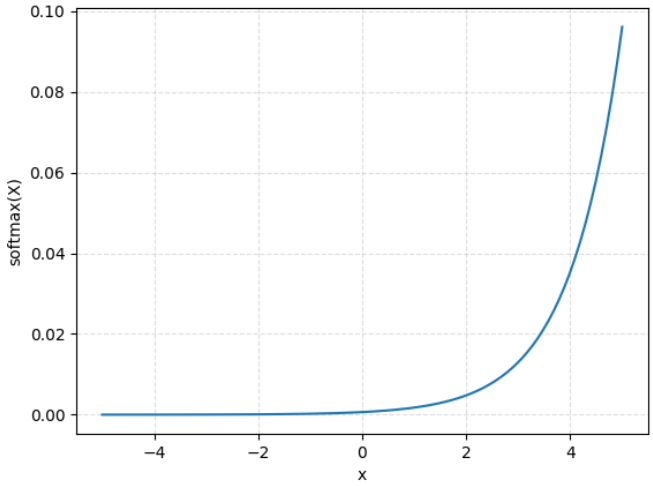


Σχήμα 2.11: Συνάρτηση Leaky ReLU

Η συνάρτηση Leaky ReLU είναι μια βελτιωμένη εκδοχή της ReLU που καταφέρνει να δώσει λύση στο πρόβλημα των Νεκρών Νευρώνων (Dead Neurons) που μπορεί να προκαλέσει η ReLU. Σύμφωνα με το πρόβλημα αυτό η ReLU μπορεί κατά την διάρκεια εκπαίδευσης να οδηγήσει στην ανανέωση κάποιου βάρους και να προκαλέσει απενεργοποίηση της εξόδου για όλα τα δεδομένα εισόδου. Όπως φαίνεται στο σχήμα 2.11, σε αντίθεση με την ReLU, διατηρεί μια πολύ μικρή κλίση για της αρνητικές τιμές αποφεύγοντας έτσι το πρόβλημα των νεκρών νευρώνων.

$$f(x) = \begin{cases} x, & x > 0 \\ \alpha x, & x \leq 0 \end{cases} \tag{2.19}$$

- *Συνάρτηση Softmax*



Σχήμα 2.12: Συνάρτηση Softmax

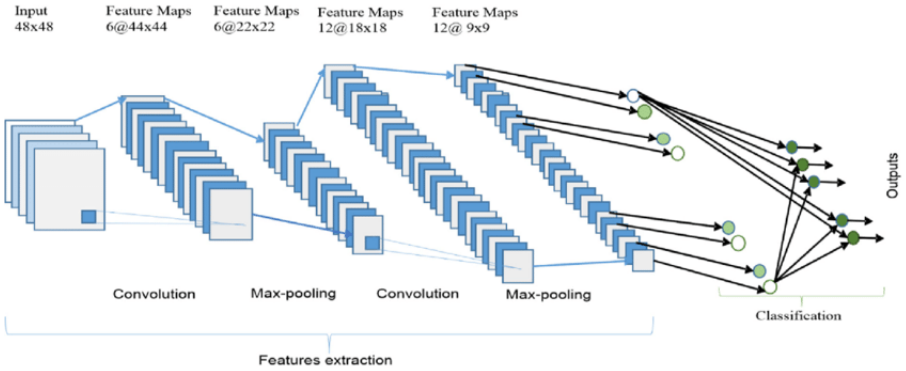
Η συνάρτηση softmax ξεχωρίζει από τις προηγούμενες κατηγορίες συναρτήσεων καθώς εκφράζει πιθανότητες στην έξοδο της. Χρησιμοποιείται συχνά για την κατηγοριοποίηση των δεδομένων εισόδου σε κλάσεις, για αυτό τον λόγο συναντάται συνήθως ως συνάρτηση ενεργοποίησης του τελευταίου επιπέδου σε ένα νευρωνικό δίκτυο. Πιο συγκεκριμένα, η softmax κανονικοποιεί τις εξόδους για κάθε κλάση και στη συνέχεια διαιρεί με το άθροισμα όλων το εξόδων, έτσι εκφράζει την πιθανότητα κάθε είσοδος να ανήκει στην αντίστοιχη κλάση. Η σχέση 2.20 δίνει την δυνατότητα οι έξοδοι να εκφράζονται ως πιθανότητες.

$$\sum_{k=1}^{K} y_k = \sum_{k=1}^{K} Softmax(z)_k = 1 \tag{2.20}$$

2.4 Συνελικτικά Νευρωνικά Δίκτυα (CNN)

Τα συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Networks - CNNs) αποτελούν μία κλάση των τεχνητών νευρωνικών δικτύων. Η χρήση τους είναι πολύ συχνή στην ανάπτυξη συστημάτων βαθιάς μηχανικής μάθησης και πιο συγκεκριμένα στην ανάπτυξη εφαρμογών που απαιτούν την ανάλυση εικόνας και βίντεο. Το όνομά ΄συνελικτικά νευρωνικά δίκτυα΄ προέρχεται από την μαθηματική πράξη της συνέλιξης. Η συνέλιξη είναι μια ειδική γραμμική πράξη και τα συνελικτικά νευρωνικά δίκτυα την χρησιμοποιούν ως μια γενική μέθοδο πολλαπλασιασμού πινάκων.

Ιστορικά, το 1968 η εργασία των D. Hubel και T. Wiesel [11] για τον διαχωρισμό των οπτικών κυττάρων του εγκεφάλου σε απλά και σύνθετα άνοιξε το δρόμο για την δημιουργία των συνελικτικών νευρωνικών δικτύων. Αργότερα, το 1980 ο Kunihiko Fukushima εμπνευσμένος από τους προηγούμενους θα είναι ο πρώτος που θα παρουσιάσει το τεχνητό νευρωνικό



Πηγή: researchgate.net/figure/The-overall-architecture-of-the-Convolutional-Neural-Network-CNN-includes-an-input_fig4_331540139

Σχήμα 2.13: Συνελικτικό Νευρωνικό Δίκτυο

δίκτυο neocognitron [17] που θα περιλαμβάνει συνελικτικά και υποδειγματοληπτικά επίπεδα. Για να γίνουν δημοφιλή τα CNNs στα συστήματα βαθιάς μηχανικής μάθησης χρειάστηκαν να περάσουν αρκετά χρόνια. Το 2012 ο Alex Krizhevsky με την αρχιτεκτονική που παρουσίασε ως AlexNet [3] κατάφερε να κερδίσει τον διαγωνισμό αναγνώρισης εικόνων ImageNet. Συστήματα που βασίζονται στην αρχιτεκτονική AlexNet αναπτύσσονται μέχρι και σήμερα.

2.4.1 Τρόπος λειτουργίας

Τα συνελικτικά νευρωνικά δίκτυα, όπως προαναφέρθηκε, έχουν σχεδιαστεί για την ανάλυση εικόνων. Η αρχιτεκτονική τους είναι ανάλογη με τα συνδετικά πρότυπα ενός ανθρώπινου εγκέφαλου και η οργάνωση τους είναι εμπνευσμένη από τον οπτικό φλοιό. Σε αντιστοίχιση με τον ανθρώπινο εγκέφαλο τα CNNs λαμβάνοντας μια εικόνα καταφέρνουν να ξεχωρίσουν και να αναγνωρίσουν τα σημεία ενδιαφέροντος. Τα σημεία αυτά σε μία εικόνα είναι δισδιάστατα σχήματα και τα CNNs καταφέρνουν να τα αναγνωρίσουν ακόμη και όταν οι παραμορφώσεις της εικόνας είναι υψηλές. Για να το καταφέρουν αυτό τα συνελικτικά νευρωνικά δίκτυα εκπαιδεύονται με έναν επιβλεπόμενο τρόπο ακολουθώντας συγκεκριμένα βήματα εκπαίδευσης. Τα βήματα που ακολουθούν είναι:

- *Εξαγωγή Χαρακτηριστικών (Feature Extraction)*

  Σαν πρώτο βήμα κάθε νευρώνας εξάγει τοπικά χαρακτηριστικά αξιοποιώντας τις εισόδους που παίρνει από το προηγούμενο επίπεδο. Για κάθε χαρακτηριστικό που εξάγεται διατηρείται η πληροφορία για τη σχετική θέση του και μειώνεται η σημαντικότητα της, έτσι ώστε να αποκτούν προτεραιότητα τα χαρακτηριστικά που δεν έχουν αναγνωριστεί ακόμα.

- *Αντιστοίχιση Χαρακτηριστικών (Feature Mapping)*

  Κάθε χαρακτηριστικό που υπολογίστηκε από την προηγούμενη διαδικασία αποθηκεύεται στη συνέχεια σε έναν χάρτη χαρακτηριστικών. Πολλοί τέτοιοι χάρτες εξάγονται από κάθε συνελικτικό επίπεδο σε ένα CNN. Κάθε χάρτης προέρχεται από την συνέλιξη ενός

φίλτρου και του διανύσματος που δέχεται ως είσοδο το επίπεδο. Η διαδικασία αυτή, του φιλτραρίσματος, συνήθως μειώνει τις ελεύθερες παραμέτρους και ακολουθείται από μια διαδικασία ενεργοποίησης. Ως συνάρτηση ενεργοποίησης στο βήμα αυτό χρησιμοποιείται συνήθως η ReLU.

- *Υποδειγματοληψία (Downsampling)*

  Μετά από κάθε συνελικτικό επίπεδο υπάρχει και ένα συγκεντρωτικό επίπεδο. Σκοπός του επιπέδου αυτού είναι να μειώσει τις διαστάσεις των χαρτών χαρακτηριστικών ώστε ελαττωθεί η υπολογιστική ισχύς που απαιτείται για την επεξεργασία των δεδομένων.

- *Αντιστοίχιση Προβλέψεων (Prediction Mapping)*

  Στο τέλος ενός συνελικτικού νευρωνικού δικτύου βρίσκεται μια σειρά από πλήρως συνδεδεμένα επίπεδα (Fully Connected Layers). Τα πλήρως συνδεδεμένα επίπεδα ακολουθούν τα επίπεδα που περιγράφηκαν προηγουμένως με σκοπό τη μετατροπή της πληροφορίας που ρέει στο δίκτυο στην επιθυμητή έξοδο και την εξαγωγή προβλέψεων.

Ο τρόπος λειτουργίας των συνελικτικών νευρωνικών δικτύων ακολουθεί τα πρότυπα των πολυεπίπεδων νευρωνικών δικτύων. Ένα προωθητικό πέρασμα (forward pass), που τροφοδοτεί τα δεδομένα εισόδου στο δίκτυο, ακολουθείται από ένα οπισθοδρομικό (backward pass) πέρασμα και η διαδικασία αυτή επαναλαμβάνεται αρκετές φορές. Στο προωθητικό πέρασμα η ροή της πληροφορίας, από την είσοδο στην έξοδο, περνάει από τα διαδοχικά επίπεδα επεξεργασίας. Κάθε επίπεδο δέχεται ως είσοδο την έξοδο του προηγούμενου επιπέδου, τη μετασχηματίζει και την προωθεί στο επόμενο επίπεδο. Όταν τελικά γίνει μια πρόβλεψη για την έξοδο του δικτύου, η τιμή αυτή θα διαδοθεί στο δίκτυο με το οπισθοδρομικό πέρασμα και με σκοπό την ανανέωση των παραμέτρων του δικτύου για την επίτευξη καλύτερης πρόβλεψης.

2.4.2 Επίπεδα επεξεργασίας

Από αρχιτεκτονικής σκοπιάς τα συνελικτικά νευρωνικά δίκτυα δομούνται από ένα σύνολο σειριακών επιπέδων επεξεργασίας. Ένα σύνολο από ομαδοποιημένους νευρώνες συνθέτει κάθε επίπεδο και καθορίζει την λειτουργία του. Στη συνέχεια παρουσιάζεται ο τρόπος κατασκευής και η λειτουργία των διάφορων επιπέδων επεξεργασίας σε ένα συνελικτικό νευρωνικό δίκτυο.

2.4.2.1 Επίπεδο Εισόδου

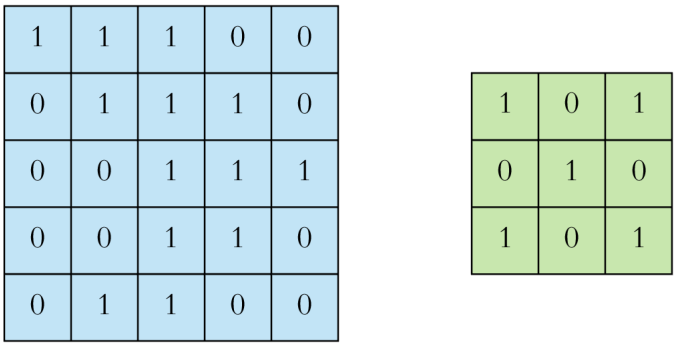
Το επίπεδο εισόδου (Input layer) είναι το πρώτο επίπεδο επεξεργασίας σε ένα τεχνητό νευρωνικό δίκτυο υπεύθυνο για την τροφοδότηση των δεδομένων εισόδου στο δίκτυο. Οι διαστάσεις των δεδομένων εισόδου καθορίζουν και τις διαστάσεις του επιπέδου εισόδου. Για παράδειγμα, σε ένα συνελικτικό νευρωνικό δίκτυο όπου τα δεδομένα εισόδου είναι ψηφιακές εικόνες με αναπαράσταση RGB διαστάσεων 1024x768 pixels, το επίπεδο εισόδου θα έχει μήκος 1024 νευρώνες, πλάτος 768 και ύψος 3, όσα είναι δηλαδή τα κανάλια σε μια RGB εικόνα (Red, Green, Blue).

2.4.2.2 Συνελικτικό Επίπεδο

Το συνελικτικό επίπεδο (Convolutional layer) είναι το κύριο δομικό στοιχείο ενός συνελικτικού νευρωνικού δικτύου αλλά και το πιο σύνθετο. Δέχεται σαν είσοδο την έξοδο του προηγούμενου επιπέδου, την μετασχηματίζει, εξάγει χαρακτηριστικά και δημιουργεί πίνακες (χάρτες) χαρακτηριστικών. Για να γίνει πιο εύκολα κατανοητός ο τρόπος λειτουργίας του επιπέδου αυτού στη συνέχεια γίνεται παρουσίαση των διαδικασιών και των εννοιών που το συνθέτουν:

- *Συνέλιξη (Convolution)*

  Η συνέλιξη είναι μια μαθηματική πράξη που συνδέει δύο διαφορετικά σύνολα πληροφορίας. Στην περίπτωση των συνελικτικών επιπέδων σε ένα CNN, η πράξη της συνέλιξης γίνεται μεταξύ των δεδομένων εισόδου και ενός φίλτρου (filter) ή αλλιώς πυρήνα (kernel). Με απλά λόγια, η πράξη της συνέλιξης είναι η ολίσθηση ενός παραθύρου, εν προκειμένω του πυρήνα, πάνω στα δεδομένα εισόδου. Το παράθυρο αυτό ολισθαίνοντας θα περάσει πάνω από όλα τα στοιχεία της εισόδου. Σε κάθε βήμα της ολίσθησης πραγματοποιείται ένα-προς-ένα πολλαπλασιασμός πινάκων μεταξύ του πυρήνα και των στοιχείων που καλύπτει το παράθυρο στον πίνακα των δεδομένων εισόδου. Στο σχήμα 2.14 δίνεται ένα παράδειγμα ενός πίνακα εισόδου και ενός πίνακα πυρήνα.

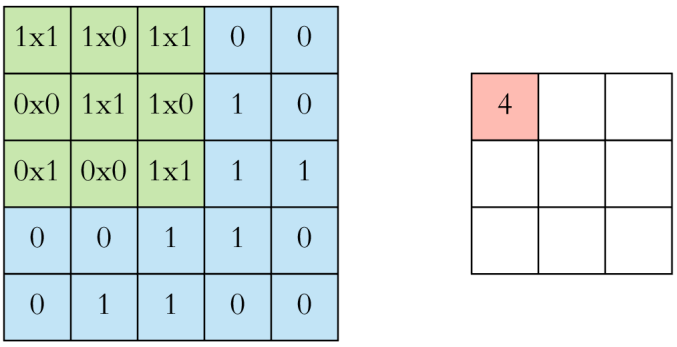


Πηγή: towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2

Σχήμα 2.14: Παράδειγμα εισόδου και πυρήνα

- *Πίνακες Χαρακτηριστικών (Feature Maps)*

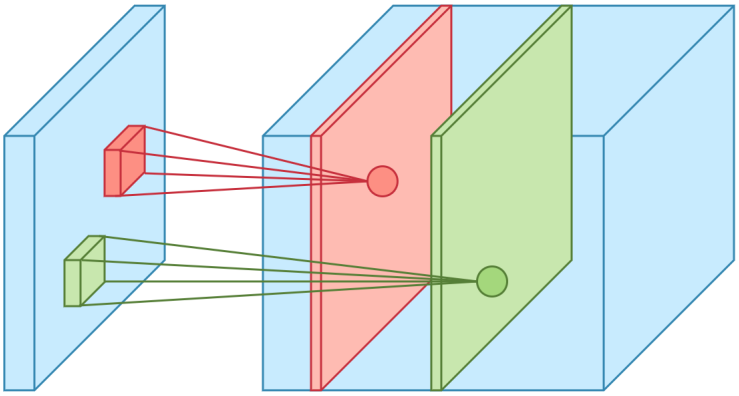
  Το αποτέλεσμα του πολλαπλασιασμού των πινάκων εισόδου και πυρήνα αποθηκεύεται σε έναν πίνακα χαρακτηριστικών, όπως φαίνεται στο σχήμα 2.15.



Πηγή: towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2

Σχήμα 2.15: Κατασκευή πίνακα χαρακτηριστικών

Ο αριθμός και οι διαστάσεις των φίλτρων ορίζεται στις παραμέτρους του επιπέδου. Για την ολίσθηση κάθε φίλτρου δημιουργείται ένας ξεχωριστός δισδιάστατος πίνακας χαρακτηριστικών. Κάθε τέτοιος πίνακας περιγράφει ξεχωριστά χαρακτηριστικά για την είσοδο. Αυτοί οι πίνακες στοιβάζονται (βλέπε σχήμα 2.16), έτσι στην έξοδο του επιπέδου θα σχηματιστεί ένας τρισδιάστατος πίνακας με ύψος όσος είναι ο αριθμός των πυρήνων που έχει καθοριστεί.



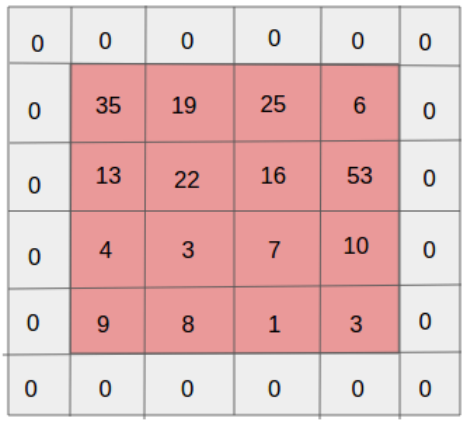
Πηγή: towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2

Σχήμα 2.16: Στοιβαγμένοι πίνακες χαρακτηριστικών

- *Υπερπαράμετροι (Hyperparameters)*

  Οι υπερπαράμετροι του επιπέδου συνέλιξης θα καθορίσουν το μέγεθος των πινάκων χαρακτηριστικών και την φύση της πληροφορίας που περιέχουν. Οι τέσσερις βασικές υπερπαράμετροι που πρέπει να καθοριστούν είναι το μέγεθος του πυρήνα (kernel size), ο αριθμός των πυρήνων (kernel count), το βήμα ολίσθησης (stride) και η επέκταση του περιθωρίου (padding). Όπως αναφέρθηκε ο αριθμός των πυρήνων θα καθορίσει

την τρίτη διάσταση στη έξοδο του συνελικτικού επιπέδου, όπως επίσης και το πλήθος των διαφορετικών χαρακτηριστικών. Το μέγεθος του πυρήνα καθορίζει τον όγκο της πληροφορίας που αναπαριστά κάθε στοιχείο στον πίνακα εξόδου. Τέλος, το βήμα ολίσθησης καθορίζει το μήκος και το πλάτος της εξόδου, αν το βήμα είναι πάνω από 1 αυτές οι διαστάσεις είναι μικρότερες από ότι είναι στην είσοδο του επιπέδου. Για να αποφύγουμε την μείωση των διαστάσεων μπορούμε να χρησιμοποιήσουμε την τελευταία υπερπαράμετρο για να επεκτείνουμε το περιθώριο της εισόδου όπως στο σχήμα 2.17.



Πηγή: towardsdatascience.com/introduction-to-convolutional-neural-networks-cnn-with-tensorflow-57e2f4837e18

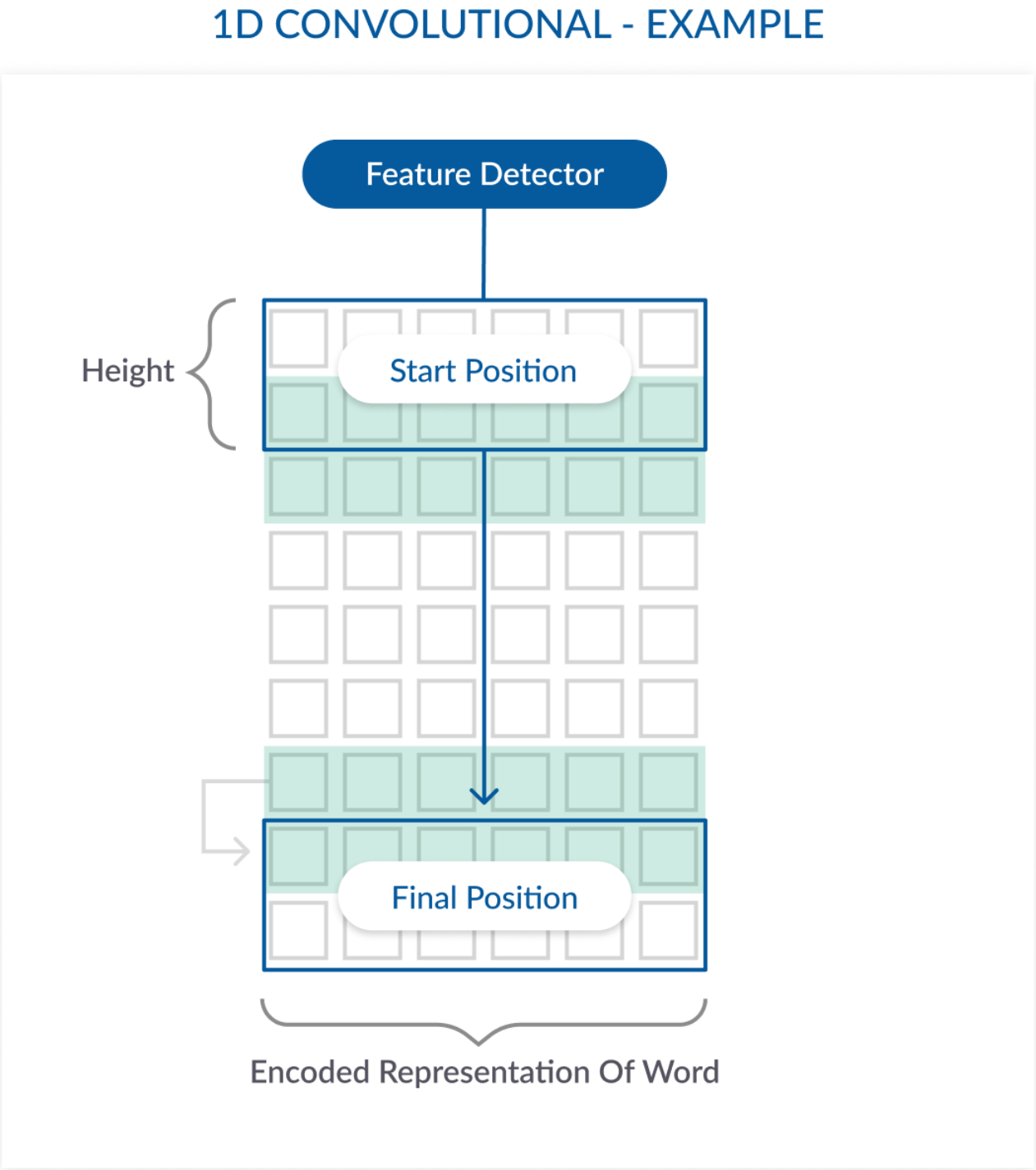
Σχήμα 2.17: Padding στα δεδομένα εισόδου

- *Μη-γραμμικότητα (Non-linearity)*

  Η εισαγωγή της μη-γραμμικότητας είναι αναγκαία ώστε ένα τεχνητό νευρωνικό δίκτυο να είναι ισχυρό. Σε ένα CNN η έννοια της μη-γραμμικότητας εφαρμόζεται στο συνελικτικό επίπεδο. Συγκεκριμένα, το αποτέλεσμα της συνέλιξης περνάει πρώτα από μία μη-γραμμική συνάρτηση ενεργοποίησης, πριν αποθηκευτεί. Στην πράξη, η συνάρτηση ενεργοποίησης που χρησιμοποιείται είναι η ReLU. Τελικά, αυτό που καταφέρνει η ReLU, σύμφωνα με τη σχέση 2.18, είναι να μηδενίσει όλες τις αρνητικές τιμές σε έναν πίνακα χαρακτηριστικών.

2.4.2.3 Συνελικτικό επίπεδο μίας διάστασης (1D))

Σε αντίθεση με ένα 2D CNN, όπου οι πυρήνες ή τα φίλτρα διασχίζουν και τις δύο διαστάσεις του χώρου μιας εικόνας, δηλαδή, από αριστερά προς τα δεξιά και από πάνω προς τα κάτω, οι πυρήνες σε 1D CNN προχωρούν μόνο σε μία διάσταση, που είναι η χρονική διάσταση σε περίπτωση δεδομένων χρονοσειρών και έτσι μπορούν να εξαγάγουν τοπικά χρονικά συναφή χαρακτηριστικά.

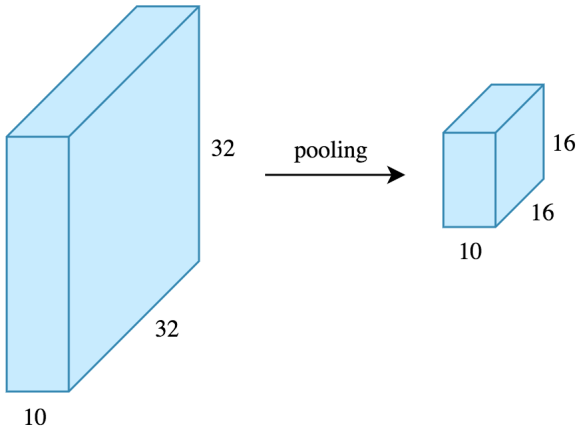


Πηγή: https://missinglink.ai/guides/keras/keras-conv1d-working-1d-convolutional-neural-networks-keras/

Σχήμα 2.18: Παράδειγμα συνελικτικού επιπέδου μιας διάστασης

2.4.2.4 Συγκεντρωτικό Επίπεδο

Ένα συγκεντρωτικό επίπεδο (Pooling layer) συνηθίζεται να ακολουθεί κάθε συνελικτικό επίπεδο με σκοπό την μείωση των διαστάσεων του χώρου που αναπαριστά τα δεδομένα. Ουσιαστικά, όπως φαίνεται και από το σχήμα 2.19, το συγκεντρωτικό επίπεδο δειγματοληπτεί τους πίνακες χαρακτηριστικών, μειώνοντας το μήκος και το πλάτος τους, ενώ το ύψος τους παραμένει ανέπαφο. Αυτό, επιτρέπει να ελαττωθεί ο αριθμός των παραμέτρων του συστήματος, το οποίο επιταχύνει την διαδικασία εκπαίδευσης και αντιμετωπίζει το πρόβλημα της υπερπροσαρμογής.



Πηγή: towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2

Σχήμα 2.19: Εφαρμογή υποδειγματοληψίας

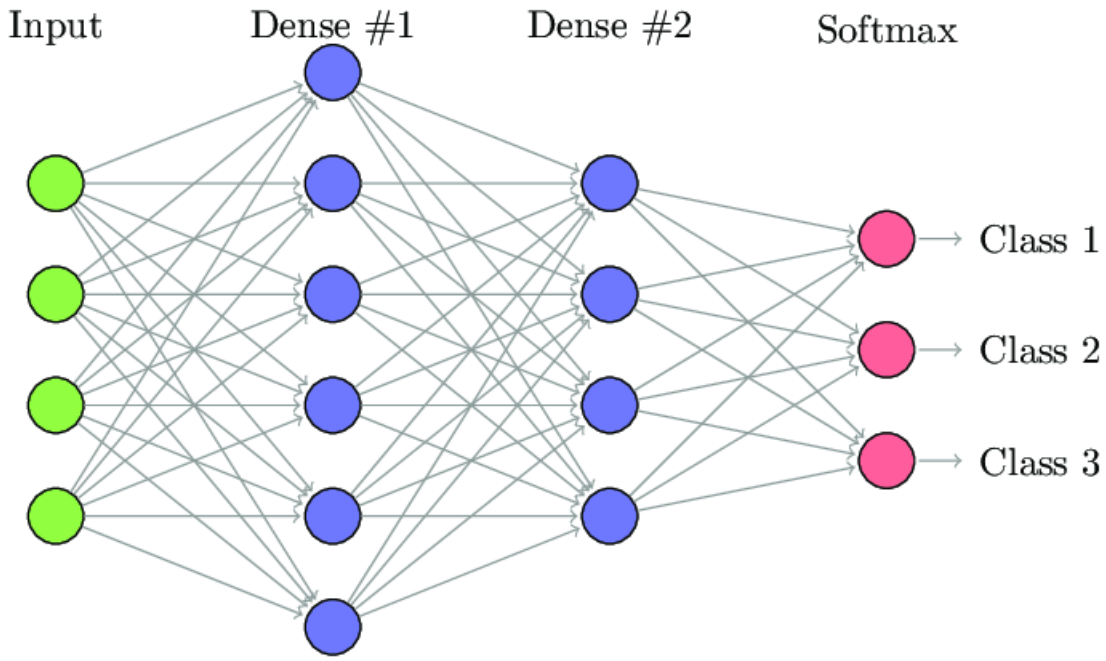
Ο τρόπος που γίνεται η υποδειγματοληψία στα επίπεδα αυτά θυμίζει πολύ την διαδικασία της συνέλιξης. Συγκεκριμένα, πάλι ένα παράθυρο ολισθαίνει πάνω από όλα τα δεδομένα εισόδου και υπολογίζει σε κάθε βήμα μία τιμή για τα στοιχεία που καλύπτει. Το μέγεθος του παραθύρου και το βήμα ολίσθησης καθορίζονται στις υπερπαραμέτρους του συγκεντρωτικού επιπέδου. Η τιμή που θα υπολογιστεί σε κάθε βήμα ολίσθησης καθορίζεται από την μέθοδο pooling που θα εφαρμοστεί. Μερικές από τις πιο συνήθεις μεθόδους είναι το max pooling που επιλέγει την μέγιστη τιμή στην περιοχή του παραθύρου, το average pooling που υπολογίζει την μέση τιμή των στοιχείων της περιοχής και το sum pooling που υπολογίζει το άθροισμα των στοιχείων της περιοχής.

2.4.2.5 Πλήρως Συνδεδεμένο Επίπεδο

Τα πλήρως συνδεδεμένα επίπεδα (Fully connected layers) εμφανίζονται σε ομάδες. Σε ένα CNN μια ομάδα από πληρώς συνδεδεμένα επίπεδα συναντάται στο τέλος του δικτύου μετά από μια αλληλουχία συνελικτικών και συγκεντρωτικών επιπέδων. Σε ένα πλήρως συνδεδεμένο επίπεδο κάθε νευρώνας του επιπέδου συνδέεται με όλους του νευρώνες του προηγούμενου επιπέδου, όπως στο σχήμα 2.20. Ένα τέτοιο επίπεδο αναμένει να πάρει ως είσοδο ένα διάνυσμα δεδομένων μίας διάστασης, τα συνελικτικά και συγκεντρωτικά επίπεδα όμως δίνουν διανύσματα

τριών διαστάσεων. Για τον λόγο αυτό σε ένα CNN πριν από το πλήρως συνδεδεμένο επίπεδο εφαρμόζεται ένας μετασχηματισμός των δεδομένων που ονομάζεται ισοπέδωση (flattening) και μετατρέπει τα τρισδιάστατα διανύσματα σε μονοδιάστατα, χωρίς να χάνεται καμία πληροφορία.

Σκοπός μιας ομάδας πλήρως συνδεδεμένων επιπέδων είναι να αξιοποιήσουν τα χαρακτηριστικά που έχουν προέλθει από τα συνελικτικά και συγκεντρωτικά επίπεδα, ώστε να παράξουν την επιθυμητή έξοδο. Για παράδειγμα, σε ένα πρόβλημα ταξινόμησης εικόνων τα πλήρως συνδεδεμένα επίπεδα του δικτύου αναλαμβάνουν να επιλέξουν σε ποια κλάση ανήκει κάθε εικόνα της εισόδου. Τέλος, η χρησιμότητα των επιπέδων αυτών δεν περιορίζεται μόνο στην ταξινόμηση της εισόδου αλλά και στην ικανότητα εκμάθησης μη γραμμικών συνδυασμών των χαρακτηριστικών της εισόδου.



Πηγή: researchgate.net/figure/Example-of-fully-connected-neural-network_fig2_331525817

Σχήμα 2.20: Πλήρως συνδεδεμένα επίπεδα

2.4.2.6 Ομαλοποίηση (Regularization)

Ένα κεντρικό πρόβλημα στη μηχανική μάθηση είναι πώς να φτιάξουμε έναν αλγόριθμο που θα το αποδίδει καλά όχι μόνο στα δεδομένα εκπαίδευσης, αλλά και στα δεδομένα ελέγχου. Πολλές στρατηγικές που χρησιμοποιούνται στη μηχανική μάθηση έχουν σχεδιαστεί ρητά για να μειώσουν το σφάλμα δοκιμής, πιθανώς εις βάρος του αυξημένου σφάλματος εκπαίδευσης. Αυτές οι στρατηγικές είναι γνωστές συλλογικά ως **Ομαλοποίηση (Regularization)**.

Υπάρχουν διάφορες τεχνικές ομαλοποίησης. Δύο κύριοι τύποι ομαλοποίησης είναι οι L1 και L2.

Ένα μοντέλο γραμμικής παλινδρόμησης που εφαρμόζει τον κανόνα L1 για ομαλοποίηση ονομάζεται παλινδρόμηση λάσο(lasso regression) και ένα που εφαρμόζει (τετράγωνο) κανόνα L2 για ομαλοποίηση ονομάζεται παλινδρόμηση κορυφογραμμής (ridge regression). Για να εφαρμόστεί κάποιο από τα δύο, πρέπει να τροποποιηθεί η συνάρτηση κόστους. Στα σχήματα ;; και 2.23 φαίνονται αυτές οι συναρτήσεις.

$$Loss = Error(y, \hat{y})$$

Σχήμα 2.21: Συνάρτηση κόστους χωρίς ομαλοποίηση

$$Loss = Error(y, \hat{y}) + \lambda \sum_{i=1}^{N} |w_i|$$

Σχήμα 2.22: Συνάρτηση κόστους με ομαλοποίηση L1

$$Loss = Error(y, \hat{y}) + \lambda \sum_{i=1}^{N} w_i^2$$

Πηγή: https://towardsdatascience.com/intuitions-on-l1-and-l2-regularisation-235f2db4c261

Σχήμα 2.23: Συνάρτηση κόστους με ομαλοποίηση L2

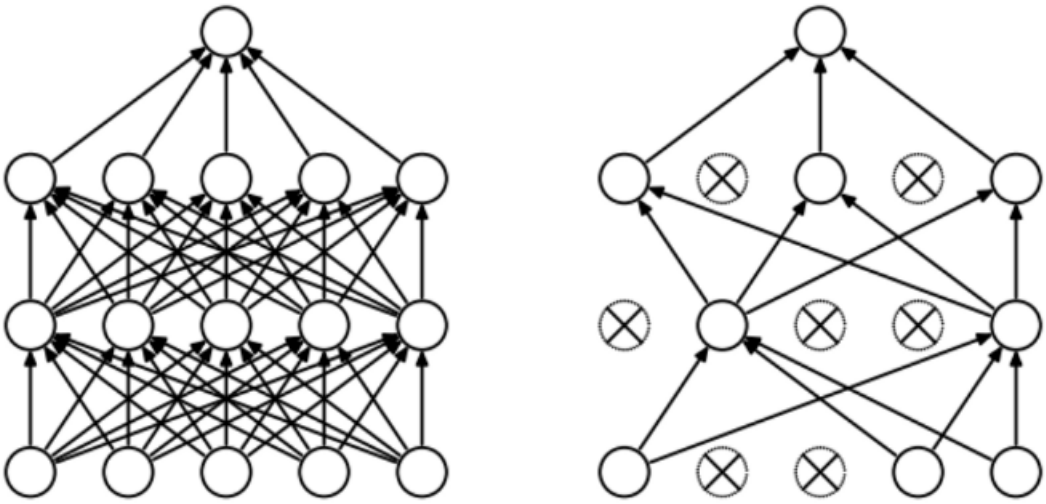
2.4.2.7 Στρώμα εγκατάλειψης (Dropout Layer)

Μία ακόμη τεχνική ομαλοποίησης είναι και αυτή του dropout. Ο όρος dropout (ελλ. ίσως αφαίρεση, εγκατάλειψη, διάλυση, αρραίωση) αναφέρεται σε εγκατάλειψη μονάδων (τόσο κρυφών όσο και ορατών) σε ένα νευρωνικό δίκτυο. Με απλά λόγια, το dropout αναφέρεται σε μονάδες που αγνοούνται (δηλ. Νευρώνες) κατά τη φάση της εκπαίδευσης ορισμένου συνόλου νευρώνων, οι οποίες επιλέγεται τυχαία. Με το "αγνοούνται", εννοούμε ότι αυτές οι μονάδες δεν λαμβάνονται υπόψη κατά τη διάρκεια μιας συγκεκριμένης διάσχισης του νευρωνικού είτε προς τα εμπρός ή προς τα πίσω.

Πιο συγκεκριμένα, σε κάθε στάδιο εκπαίδευσης, μεμονωμένοι κόμβοι είτε αφαιρούνται έξω από το δίκτυο με πιθανότητα 1-p είτε διατηρούνται με πιθανότητα p, έτσι ώστε να μείνει ένα μειωμένο δίκτυο. Οι εισερχόμενες και εξερχόμενες ακμές σε έναν αποκλεισμένο κόμβο αφαιρούνται επίσης.

Πολλές φορές χρειαζόμαστε την αφαίρεση για να αποφευχθεί η υπερπροσαρμογή (overfitting).

Στο σχήμα 2.24 μπορούμε να δούμε ένα παράδειγμα εφαρμογής της εγκατάλειψης σε ένα νευρωνικό δίκτυο με 2 κρυφά στρώματα.



Πηγή: Srivastava (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting [45]

Σχήμα 2.24: Παράδειγμα εφαρμογής dropout. Αριστερά: Ένα τυπικό νευρικό δίκτυο με 2 κρυφά στρώματα. Δεξιά: Ένα παράδειγμα αραιωμένου δικτυού που παράγεται με την εφαρμογή της αφαίρεσης στο δίκτυο στα αριστερά.Οι μονάδες με το γράμμα X αφαιρούνται

2.5 Μετρικές Αξιολόγησης

Η προεπεξεργασία των δεδομένων και εκπαίδευση είναι αντικείμενα υψίστης σημασίας για την ανάπτυξη ενός μοντέλου μηχανικής μάθησης, εξίσου σημαντική είναι όμως και η παρακολούθηση της επίδοσης του μοντέλου. Δηλαδή, πόσο καλά μπορεί να γενικεύει το μοντέλο για άγνωστα δεδομένα. Οι μετρικές αξιολόγησης εξυπηρετούν αυτόν το σκοπό, συγκεκριμένα μετρούν κάποιο μέγεθος του εκπαιδευμένου μοντέλου ως προς κάποιο χαρακτηριστικό. Χωρίς την χρήση μετρικών αξιολόγησης η βελτίωση της προβλεπτικής ικανότητας του μοντέλου ή η σύγκρισή του με άλλα μοντέλα δεν θα ήταν εφικτή. Ακόμα, αξίζει να αναφερθεί ότι η χρήση των μετρικών αξιολόγησης δεν είναι καθολική αλλά η φύση του κάθε προβλήματος ενθαρρύνει την χρήση διαφορετικών μετρικών αξιολόγησης. Στη συνέχεια παρουσιάζονται μερικές από τις σημαντικότερες μετρικές αξιολόγησης. Πρώτα, όμως, γίνεται αναφορά στις κλάσεις που μπορούν να ανήκουν οι προβλέψεις του συστήματος:

- *True Positive (TP)*:

  Το σύνολο τις εξόδου για το οποίο η πρόβλεψη είναι σωστή και η προβλεπόμενη κλάση είναι θετική.

- *True Negative (TN)*:

  Το σύνολο τις εξόδου για το οποίο η πρόβλεψη είναι σωστή και η προβλεπόμενη κλάση είναι αρνητική.

- *False Positive (FP)*:

  Το σύνολο τις εξόδου για το οποίο η πρόβλεψη είναι λανθασμένη και η προβλεπόμενη κλάση είναι θετική.

- *False Negative (FN)*:

  Το σύνολο τις εξόδου για το οποίο η πρόβλεψη είναι λανθασμένη και η προβλεπόμενη κλάση είναι αρνητική.

2.5.1 Μετρικές σε προβλήματα ταξινόμησης δύο κλάσεων (Metrics for binary-classification problems)

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Σχήμα 2.25: Κλάσεις προβλέψεων

1. *Accuracy*

   Η μετρική αξιολόγησης Accuracy ή ακρίβεια όπως φαίνεται και από τη σχέση 2.21, περιγράφει τον λόγο των σωστά ταξινομημένων δειγμάτων προς το σύνολο όλων των δειγμάτων.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2.21}$$

   Η μετρική Accuracy είναι η πιο σημαντική μετρική και δίνει μια άμεση και απλή αξιολόγηση του μοντέλου. Η χρήση της συνίσταται για δεδομένα που είναι καλά ισορροπημένα.

2. *Precision*

   Η μετρική αξιολόγησης Precision (βλέπε σχέση 2.22) που εκφράζει τον λόγο των σωστά ταξινομημένων θετικών προβλέψεων προς το σύνολο των προβλέψεων που έχουν ταξινομηθεί ως θετικές.

$$Precision = \frac{TP}{TP + FP} \tag{2.22}$$

   Η μετρική Precision χρησιμοποιείται σε προβλήματα που η εγκυρότητα της πρόβλεψης είναι μεγάλης σημασίας.

3. *Recall*

   Η μετρική αξιολόγησης Recall (βλέπε σχέση 2.23) που εκφράζει το λόγο των σωστά ταξινομημένων θετικών προβλέψεων προς το σύνολο των θετικών προβλέψεων.

$$Recall = \frac{TP}{TP + FN} \tag{2.23}$$

   Η μετρική Recall χρησιμοποιείται σε προβλήματα που έχουν ως σκοπό την μεγιστοποίηση των θετικών προβλέψεων.

4. *F1 Score*

   Η μετρική αξιολόγησης F1 Score (βλέπε σχέση 2.24) που εκφράζει τον αρμονικό μέσο όρο των μετρικών Precision και Recall.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{2.24}$$

   Η χρήση της μετρικής F1 Score γίνεται όταν το πρόβλημα απαιτεί καλό Precision και Recall.

5. *Crossentropy*

   Η μετρική Crossentropy ή Log Loss λαμβάνει υπόψιν την αβεβαιότητα της πρόβλεψης βασιζόμενη στο πόσο διαφέρει από την πραγματική τιμή. Χρησιμοποιείται σε προβλήματα δυαδικής ταξινόμησης και υπολογίζεται από τον τύπο 2.25.

$$Crossentropy = -(y \log(p) + (1 - y) \log(1 - p)) \tag{2.25}$$

   Όπου p είναι η πιθανότητα η πρόβλεψη να είναι 1 και y είναι η πρόβλεψη του μοντέλου. Η μετρική αυτή χρησιμοποιείται όταν η έξοδος του μοντέλου είναι πιθανοτικές προβλέψεις.

6. *Categorical Crossentropy*

   Αυτή η μετρική αξιολόγησης είναι ίδια με την μετρική Crossentropy η μόνη διαφορά είναι ότι χρησιμοποιείται σε προβλήματα που οι κλάσεις ταξινόμησης είναι παραπάνω από δύο.

$$CategoricalCrossentropy = \frac{-1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} y_{ij} * \log(p_{ij}) \tag{2.26}$$

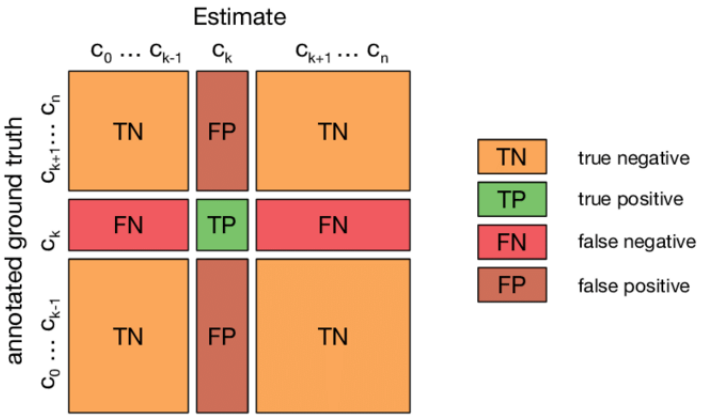
   Όπου το y_{ij} είναι 1 αν το δείγμα i ανήκει στην κλάση j, αλλιώς είναι 0 και p_{ij} είναι η πιθανότητα το μοντέλο να προβλέψει ότι το δείγμα i ανήκει στην κλάση j.

2.5.2 Μετρικές σε προβλήματα ταξινόμησης πολλαπλών κλάσεων (Metrics for multi-class problems)

Σε ένα τυπικό πρόβλημα ταξινόμησης πολλαπλών τάξεων, πρέπει να κατηγοριοποιήσουμε κάθε δείγμα σε 1 από N διαφορετικές κατηγορίες.

Το 2.26 απεικονίζει έναν **Πίνακα Σύγχυσης (Confusion Matrix)** για την κατάσταση πολλαπλών τάξεων με κλάσεις n. Ο πίνακας σύγχυσης δίνει το ποσό των αποτυχημένων ταξινομήσεων για κάθε κλάση. Οι εκτιμήσεις σημείων συλλέγονται στον πίνακα σύγχυσης $C(c_{ij})$, όπου c_{ij} είναι ο αριθμός των χρονικών βημάτων όπου η κλάση ήταν στην πραγματικότητα i αλλά η κλάση j έχει υπολογιστεί. Σε γενικές γραμμές, ο πίνακας σύγχυσης παρέχει τέσσερις τύπους αποτελεσμάτων ταξινόμησης (καθένας από αυτούς κωδικοποιείται με διαφορετικό χρώμα στο σχήμα 2.26) σε σχέση με έναν στόχο ταξινόμησης k: [28]

- true positives (tp) Έγινε πρόβλεψη της κλάσσης ενώ πραγματικά είναι.
- true negatives (tn) Δεν έγινε πρόβλεψη της κλάσσης και δεν είναι.
- false positives (fp) Έγινε πρόβλεψη της κλάσσης ενώ πραγματικά δεν είναι.
- false negatives (fn) Δεν έγινε πρόβλεψη της κλάσσης ενώ πραγματικά είναι.



Πηγή: Activity, Context, and Plan Recognition with Computational Causal Behaviour Models - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Confusion-matrix-for-multi-class-classification-The-confusion-matrix-of-a_fig7_314116591

Σχήμα 2.26: Πίνακας σύγχυσης για ταξινόμηση πολλαπλών κατηγοριών. Ο πίνακας σύγχυσης μιας ταξινόμησης με n τάξεις. Κατά την εξέταση της κλάσης $k(0 \leq k \leq n)$, μπορούν να ληφθούν τα τέσσερα διαφορετικά αποτελέσματα ταξινόμησης: TP (πράσινα), TN (πορτοκαλί), FP (καφέ) και FN (κόκκινα).

Τα **precision** και **recall** μπορούν να υπολογιστούν παρόμοια με την ταξινόμηση δύο κατηγοριών, με την διαφορά ότι έχουμε πολλαπλά FN.

2.5.2.1 F1-score για ταξινόμηση πολλαπλών κλάσεων

Το F1-score κάθε κατηγορίας υπολογίζεται αντίστοιχα με με την ταξινόμηση δύο κατηγοριών όπως στην σχέση 2.24

Θέλουμε όμως και μετρικές για το συνολικό F1-score του ταξινομητή. Να υπολογίσουμε δηλαδή τον μέσο όρο για το F1-score. Υπάρχουν 3 διαφορετικές τεχνικές

1. **Micro Average ή Accuracy**

   Στο Micro Average η συνάρτηση για τον υπολογισμό του f1 λαμβάνει υπόψην τα συνολικά αληθινά θετικά, ψευδώς αρνητικά και ψευδώς θετικά (ανεξάρτητα από την πρόβλεψη για κάθε ετικέτα στο σύνολο δεδομένων). Υπολογίζει δηλαδή το ποσοστό των σωστών προβλέψεων ανεξαρτήτως κλάσσης (Accuracy).

2. **Macro Average ή F1_m**

   Στο Macro Average η συνάρτηση για τον υπολογισμό του f1 λαμβάνει υπόψιν κάθε ετικέτα και επιστρέφει τον μέσο όρο του f1-score κάθε ετικέτας. Δεν λαμβάνεται υπόψη η αναλογία για κάθε ετικέτα στο σύνολο δεδομένων.

3. **Weighted Average**

   Στο Weighted Average η συνάρτηση για τον υπολογισμό του f1 λαμβάνει υπόψιν κάθε ετικέτα και επιστρέφει τον σταθμισμένο μέσο όρο του f1-score κάθε ετικέτας, λαμβάνει δηλαδή υπόψιν την αναλογία για κάθε ετικέτα στο σύνολο δεδομένων.

Έτσι μπορεί να επιλεχθεί μία ή περισσότερες από αυτές τις μετρικές για να υπολογίζουμε εάν ένας ταξινομητής είναι καλύτερος από έναν άλλο. Βέβαια, θέλει προσοχή σε αυτήν την σύγκριση.

Αν και είναι πράγματι βολικό για μια γρήγορη, υψηλού επιπέδου σύγκριση, το κύριο ελάττωμά τους είναι ότι δίνουν ίσο βάρος στο precision και στο recall.Η ταξινόμηση ενός άρρωστου ως υγιούς έχει διαφορετικό κόστος από την ταξινόμηση ενός υγιούς ατόμου ως άρρωστου και αυτό πρέπει να αντικατοπτρίζεται στον τρόπο με τον οποίο χρησιμοποιούνται τα βάρη και το κόστος για την επιλογή του καλύτερου ταξινομητή για ένα συγκεκριμένο πρόβλημα. Αυτό ισχύει για τους δυαδικούς ταξινομητές και το πρόβλημα επιδεινώνεται κατά τον υπολογισμό βαθμολογιών F1 πολλαπλών κατηγοριών όπως βαθμολογίες macro, micro, weighted average. Στην περίπτωση πολλαπλών κατηγοριών, διαφορετικά σφάλματα πρόβλεψης έχουν διαφορετικές επιπτώσεις. Η πρόβλεψη του X ως Y είναι πιθανό να έχει διαφορετικό κόστος από την πρόβλεψη του Z ως W, ούτω καθεξής. Οι τυπικές βαθμολογίες F1 δεν λαμβάνουν υπόψη καμία γνώση τομέα.

Εάν όμως θεωρούμε πως η πρόβλεψη του ως έχει ίδιο κόστος με κάποιο άλλο ζευγάρι προβλέψεων, τότε μπορούμε να χρησιμοποιήσουμε αυτές τις μετρικές.

Ακόμα, σε περίπτωση που έχουμε *μη ισορροπημένο σετ δεδομένων*, θεωρείται καλύτερη πρακτική να υπολογίζουμε το macro average ή αλλιώς f1-m. Έτσι, μπορούμε να καταλάβουμε εάν όντως γίνεται σωστή ταξινόμηση πολλών κατηγοριών. Ενώ το micro average ή αλλιώς accuracy μπορεί να δείχνει διαστρευλωμένη την ταξινόμηση, σε περίπτωση που ο ταξινομητής πετυχαίνει υψηλό ποσοστό επιτυχίας, επειδή οι κλάσσεις με τα περισσότερα δεδομένα μπορεί να έχει πολύ υψηλό ποσοστό επιτυχίας, ενώ οι κλάσσεις με τα λιγότερα δεδομένα πολύ χαμηλό.

2.6 Επιλογή νευρωνικού δικτύου για κατηγοριοποίηση είδους μουσικής

Για την κατηγοριοποίηση μουσικής θα χρησιμοποιήσουμε συνελικτικά νευρωνικά δίκτυα (CNN) και συγκερκιμένα 1D - CNN. Ως είσοδο θα έχουμε ένα σετ δεδομένων που θα αποτελείται από δισδιάστατους πίνακες. Θα μιλήσουμε για τα σετ δεδομένων σε επόμενο κεφάλαιο. Θα χρησιμοποιήσουμε όλα τα επίπεδα επεξεργασίας που αναφέραμε στο 2.4.2 ενώ για την εκπαίδευση του νευρωνικού θα χρησιμοποιηθούν οι αλγόριθμοι backpropagation (2.3.6.2) και κατάβασης κλίσης (2.3.7.1). Ακόμα, θα χρησιμοποιηθεί ο αλγόριθμος βελτιστοποίησης Adam. Ακόμα χρησιμοποιείται η Categorical Crossentropy(2.26) για το σφάλμα (Loss) της συνάρτησης κόστους του νευρωνικού.

Τελικά, μόλις ολοκληρωθεί η εκπαίδευση των μοντέλων το τελικό βήμα αφορά την αξιολόγηση των εξόδων που αποδίδουν. Για την καταγραφή των αποτελεσμάτων θα χρησιμοποιηθούν ορισμένες από τις μετρικές αξιολόγησης, όπως περιγράφονται στο 2.5.2, επί των δεδομένων ελέγχου. Η πειραματική διαδικασία θα αξιολογηθεί κυρίως βάση των μετρικών macro average ή f1-m και της micro average ή Accuracy, όπως περιγράφτηκαν στο 2.5.2.1.

Θα μιλήσουμε για την κατασκευή του νευρωνικού δικτύου που θα χρησιμοποιήσουμε στο κεφάλαιο 5.

Κεφάλαιο 3

Θεωρία Μουσικής

3.1 Εισαγωγή στην θεωριά είδους μουσικής

Οι κατηγορίες από τα είδη μουσικής επιτρέπουν στους ανθρώπους να ομαδοποιούν μουσική σύμφωνα με τις αντιληπτές ομοιότητες μεταξύ τους. Θεωρείται από γνωστικούς ψυχολόγους ότι η κατηγοριοποίηση γενικά μας επιτρέπει να αποδεχόμαστε τις πληροφορίες, ομαδοποιώντας τις με ουσιαστικούς τρόπους που μπορούν να αυξήσουν την κατανόησή μας για αυτά. Αν και το μουσικό είδος αναφέρεται μερικές φορές ως πρωταρχικό εργαλείο μάρκετινγκ ή τεχνητό σύστημα σήμανσης που χρησιμοποιείται από ακαδημαϊκούς, δεν πρέπει να ξεχνάμε τη βαθύτερη σημασία του.

Υπάρχουν ορισμένα σημαντικά ζητήματα που πρέπει να λάβουμε υπόψιν σχετικά με το είδος. Πώς δημιουργούνται τα είδη, πώς συμφωνούνται και διαδίδονται, πώς ορίζονται, πώς αντιλαμβάνονται και αναγνωρίζονται, πώς αλλάζουν, πώς αλληλοσυνδέονται και πώς τα χρησιμοποιούμε είναι όλοι σημαντικοί τομείς έρευνας.

Η απάντηση σε αυτές τις ερωτήσεις δεν είναι εύκολη υπόθεση. Μπορεί να είναι δύσκολο να βρεθούν σαφείς, συνεπείς και αντικειμενικοί ορισμοί των ειδών και τα είδη σπάνια οργανώνονται με συνεπή ή ορθολογικό τρόπο. Οι διαφορές μεταξύ των ειδών είναι ασαφείς κατά καιρούς, οι κανόνες που διακρίνουν τα είδη είναι συχνά ασαφείς ή ασυνεπείς, οι κρίσεις ταξινόμησης είναι υποκειμενικές και τα είδη μπορούν να αλλάξουν με την πάροδο του χρόνου. Οι κατηγορίες που προκύπτουν είναι αποτέλεσμα πολύπλοκων αλληλεπιδράσεων πολιτιστικών παραγόντων, στρατηγικών μάρκετινγκ, ιστορικών συμβάσεων, επιλογών από μουσικούς βιβλιοθηκονόμους, κριτικούς και εμπόρους λιανικής και τις αλληλεπιδράσεις ομάδων μουσικών και συνθετών. [33]

Σε αυτό το σημείο, θα αναφέρουμε γενικά την έννοια του διαφορετικής κατηγορίας ή του διαφορετικού είδους, και πως μπορούν αυτά να διαχωριστούν συστηματικά.

3.1.1 Κλασική ταξινόμηση είδους

Η παλαιότερη γνωστή θεωρία ταξινόμησης, που προέρχεται από τον Αριστοτέλη και ονομάζεται **κλασική θεωρία**. Η βασική ιδέα της κλασικής θεωρίας περιέχεται στην υπόθεσή πως τα χαρακτηριστικά που αντιπροσωπεύουν μια έννοια είναι (1) απαραίτητα και (2) από κοινού αρκούν για να ορίσουν αυτήν την έννοια. Για να είναι ένα χαρακτηριστικό απαραίτη-

το, πρέπει κάθε περίπτωση (instance) της οντότητας να το έχει. Για να είναι επαρκές ένα ένα σύνολο χαρακτηριστικών, πρέπει κάθε οντότητα που έχει αυτό το σύνολο να είναι ένα παράδειγμα της έννοιας.

Υπάρχουν πολλά προβλήματα με την κλασική θεωρία, όπως φημίζεται από τον Wittgenstein (1953) [49] και επιβεβαιώθηκαν τόσο πειραματικά όσο και θεωρητικά από πολλούς άλλους. Για παράδειγμα, διαφορετικές οντότητες που ανήκουν στην ίδια κατηγορία μπορούν να έχουν κάποια βασικά διαφορετικά χαρακτηριστικά.

Ένα επιπλέον πρόβλημα είναι ότι οι ταξινομήσεις μπορεί συχνά να εξαρτώνται από το περιβάλλον, να εξαρτώνται από την τρέχουσα κατάσταση του νου και το υπόβαθρο ενός ανθρώπου. Η κλασική θεωρία αναλαμβάνει λανθασμένα τους απόλυτους κανόνες σε όλες τις περιπτώσεις.

Ένα άλλο πρόβλημα είναι ότι πολλά πειράματα έχουν δείξει ότι οι άνθρωποι συχνά θεωρούν ότι ορισμένες οντότητες είναι πιο αντιπροσωπευτικά παραδείγματα μιας κατηγορίας από άλλες. Αυτό ονομάζεται κεντρικότητα (centrality) και σχετίζεται με την έννοια ότι οι κατηγορίες μπορούν να έχουν βαθμούς συμμετοχής καθώς και ασαφή όρια. [44, 34] Για παράδειγμα, στη θεωρία της μουσικής ξεχωρίζονται οι περίοδοι της κλασικής μουσικής και της ρομαντικής. Ο Ludwig van Beethoven, συνθέτης έζησε κατά τη μεταβατική περίοδο μεταξύ των κλασικής-μπαρόκ και ρομαντικής εποχής.[26] Έτσι κάποια κομμάτια του μπορούν να θεωρηθούν κλασικά, άλλα ρομαντικά και άλλα μία μίξη κλασικής και ρομαντικής.

Έχουν γίνει πολλές προσπάθειες τροποποίησης της κλασικής θεωρίας, προκειμένου να ληφθούν υπόψη τέτοια προβλήματα.

3.1.2 Ταξινόμηση είδους με πρότυπο μοντέλο(Exemplar-based classification)

Μία από τις πιο σημαντικές εναλλακτικές θεωρίες που προτάθηκαν ήταν η θεωρία ότι οι κατηγορίες, αντί να ορίζονται από κανόνες, ορίζονται με βάση πρότυπα παραδείγματα. Έτσι οι ταξινομήσεις γίνονται συγκρίνοντας ομοιότητες με τα πρότυπα αυτά παραδείγματα (Poser Keele 1968, 1970, Reed 1972, Rosch 1973a, 1973b, Rosch, Simpson and Miller 1976). Μια οντότητα μπορεί επομένως να θεωρηθεί, ότι ανήκει στην ίδια κατηγορία με εκείνη του πρότυπου με το οποίο είναι στην αντίληψή μας το πιο όμοιο.

Το πρότυπο μοντέλο ξεπερνά πολλές από τις ασυνέπειες της κλασικής θεωρίας. Για παράδειγμα, δεν είναι πλέον απαραίτητο για κάθε οντότητα που ανήκει σε μια κατηγορία να μοιράζεται ορισμένες ιδιότητες με όλες τις άλλες οντότητες που ανήκουν στην ίδια κατηγορία. Τα όρια μεταξύ κατηγοριών μπορεί να είναι πιο ασαφή και πιο ευέλικτα. Η πρότυπη θεωρία μπορεί επίσης να συμβάλει στην αβεβαιότητα σε ποια κατηγορία ανήκουν.

Τα μοντέλα βασισμένα σε πρότυπα παραδείγματα τροποποιήθηκαν για να επιτρέψουν τον καθορισμό κατηγοριών με πολλαπλά παραδείγματα ανά πρότυπο και όχι μόνο με ένα (Rosch 1975; Rosch 1978). Σε κάποιες περιπτώσεις εξακολουθεί να υπάρχει ένα παράδειγμα που είναι πιο ισχυρό από άλλα παραδείγματα, ενώ άλλες εκδόσεις επιτρέπουν στα πολλαπλά παραδείγματα να είναι εξίσου σημαντικά. [34]

3.1.3 Γενική θεωρία αναγνώρισης (General recognition theory)

Η γενική θεωρία αναγνώρισης (GRT) (Ashby 1989, Ashby 1992)[[**Ασσβ**]] είναι μια τροποποιημένη έκδοση μοντέλων που βασίζονται σε παραδείγματα που δεν απαιτούν τον μεγάλο αριθμό συγκρίσεων ομοιότητας που απαιτούνται από τα περισσότερα πρότυπα μοντέλα. Η βασική ιδέα πίσω από τη ΓΘΑ είναι ότι οι άνθρωποι χωρίζουν τον χώρο αντίληψης σε περιοχές και συνδέουν μια ετικέτα κατηγορίας σε κάθε περιοχή. Κάθε περιοχή χωρίζεται από ένα όριο απόφασης που δίνεται από κάποια συνάρτηση, και εκχωρείται σε κάθε στιγμιότυπο η κατηγορία της περιοχής στην οποία εμπίπτει. Κάθε ταξινόμηση συνεπώς συνεπάγεται πρώτα τη αντιστοίχηση του πρότυπου μοντέλου στην κατάλληλη περιοχή στον χώρο αντίληψής μας και στη συνέχεια επισήμανση της κατηγορίας που αντιστοιχεί σε αυτήν την περιοχή. Η σύνδεση κατηγοριών με αφηρημένες περιοχές του αντιληπτικού χώρου καθιστά περιττό τον υπολογισμό της ομοιότητας με όλα τα παραδείγματα κάθε φορά που παρουσιάζεται μια οντότητα διερευνητή.

Στο GRT, οι περιοχές απόφασης στον χώρο αντίληψης σχηματίζονται με βάση την ένωση, τυπικά, περιοχών πολυμεταβλητής κανονικής πυκνότητας πιθανότητας που περιβάλλουν κάθε πρότυπο μοντέλο. [34]

3.1.4 Ταξινόμηση είδους μουσικής

Η έρευνα της ψυχολόγου Eleanor Rosch έδειξε ότι οι άνθρωποι τείνουν να θεωρούν τις κατηγορίες ότι έχουν κάποια τυπικά ή πρότυπα μέλη και άλλα λιγότερο τυπικά μέλη, ταξινομούν τα είδη μουσικής με την τεχνική του πρότυπου μοντέλου. (3.1.2) Αυτό φαίνεται σίγουρα σύμφωνο με τον τρόπο με τον οποίο ορισμένες ηχογραφήσεις φαίνονται χαρακτηριστικές ενός μουσικού είδους, ενώ άλλες φαίνονται λιγότερο, ενώ αφήνουν μικρή αμφιβολία ως προς τη συμμετοχή τους στο συγκεκριμένο είδος.

Η Marie-Laure Ryan θεωρεί πως:

Μπορούμε να σκεφτούμε τα είδη ως συλλόγους που επιβάλλουν έναν ορισμένο αριθμό κανόνων για ένταξη, αλλά ανέχεται ως εν μέρει μέλη εκείνα τα άτομα που μπορούν να πληρούν μόνο μερικές από τις προϋποθέσεις και που δεν φαίνεται να ανήκουν σε καμία άλλη λέσχη. Καθώς αυτά τα εν μέρει μέλη γίνονται όλο και περισσότερα, οι όροι εισδοχής μπορούν να τροποποιηθούν, έτσι ώστε και αυτοί, να γίνουν πλήρη μέλη. Μόλις γίνει δεκτός στο σύλλογο, ωστόσο, ένα μέλος παραμένει μέλος, ακόμα κι αν δεν μπορεί να ικανοποιήσει τους νέους κανόνες εισδοχής. (Ryan 1981, 118)

Η έρευνα έχει δείξει ότι τα άτομα είναι πιο εξοικειωμένα με υποκατηγορίες μέσα σε ένα μουσικό είδος που τους αρέσει παρά σε είδη που δεν τους αρέσουν (Hargreaves & North 1999). Ακόμα και οι καλά εκπεδευμένοι μουσικοί μπορεί να έχουν πολύ λιγότερη γνώση και κατανόηση των ειδών στα οποία δεν είναι επιδέξιοι από ότι ακόμη και οι απλοί ακροατές που ενδιαφέρονται για αυτά τα είδη. Επιπλέον, τα χαρακτηριστικά γνωρίσματα που χρησιμοποιούν διαφορετικά άτομα για την αναγνώριση των ειδών μπορεί να διαφέρουν ανάλογα με τα είδη που είναι εξοικειωμένα. Όλα αυτά υπονοούν ότι υπάρχουν λίγοι άνθρωποι, αν υπάρχουν, στους οποίους μπορούμε να βασιστούμε ώστε αυτοί να ταξινομήσουν τα αυθαίρετα είδη αξιόπιστα και χρησιμοποιώντας σταθερούς μηχανισμούς. Ένα σύστημα υπολογιστή που μπορεί να εξοι-

κειωθεί με ένα πολύ μεγαλύτερο εύρος μουσικής από ότι οι περισσότεροι άνθρωποι θα ήταν πρόθυμοι ή θα μπορούσαν να είναι σε θέση να ταξινομήσουν, παρόλο που συγκεκριμένα άτομα μπορεί να εξακολουθούν να αποδίδουν καλύτερα στο τομέα που είναι εξειδικευμένοι. [33]

Στα πλαίσια του πειράματος, χρησιμοποιήσαμε κατηγορίες που ήδη έχουν γίνει από διάφορες τεχνικές και από διάφορους ανθρώπους. Περισσότερα για αυτό θα μιλήσουμε σε επόμενο κεφάλαιο.

3.2 Χαρακτηριστικά γνωρίσματα (Features) με τα οποία γίνεται κατηγοριοποίηση

Υπάρχουν τρεις κύριοι τύποι μουσικών πληροφοριών που παραδοσιακά χρησιμεύουν ως περιπτώσεις στην αυτόματη ταξινόμηση μουσικής:

- **Ακουστικά δεδομένα**

  Ψηφιακές αναπαραστάσεις φυσικών ηχητικών σημάτων. Αυτά συνήθως αποθηκεύονται σε μορφές όπως MP3, WAV και FLAC. Από τα ακουστικά δεδομένα εξάγουμε κυρίως **χαρακτηρισιτκά με περιεχόμενο χαμηλού επιπέδου** όπως φασματικές πληροφορίες ή πληροφορίες χρονικού τομέα που εξάγονται απευθείας από ηχητικά σήματα. Τα περισσότερα χαρακτηριστικά αυτού του τύπου δεν παρέχουν πληροφορίες που είναι διαισθητικά μουσικές, αλλά μπορούν να έχουν σημαντική διακριτική ισχύ όταν υποβάλλονται σε επεξεργασία από υπολογιστές και μερικές φορές μπορούν επίσης να έχουν ψυχο-ακουστική σημασία. Συντελεστές Cepstral Mel-Frequency, Zero Crossings και Signal RMS είναι παραδείγματα τέτοιων χαρακτηριστικών.

- **Συμβολικές αναπαραστάσεις μουσικής**

  Αναπαράσταση ήχου με βάση αφηρημένα σύμβολα που έχουν νόημα μουσικά, όπως η μουσική σημειογραφία που χρησιμοποιείται στα μουσικά θέματα. Οι μορφές αρχείων όπως MIDI, OSC, Music XML και Humdrum αποθηκεύουν συμβολικά δεδομένα.

  Από τις συμβολικές αναπαραστάσεις δεδομένων εξάγουμε κυρίως **χαρακτηριστικά με περιεχόμενο υψηλού επιπέδου**, δηλαδή πληροφορίες που διατυπώνονται με τέτοιο τρόπο ώστε να έχουν νόημα για μουσικά σχετικούς ανθρώπους. Οι μετρήσεις της ποσότητας της χρωματικής κίνησης, της ποσότητας rubato (εύλικτος ρυθμός), των οργάνων που υπάρχουν και των πληροφοριών που σχετίζονται με τους στίχους τραγουδιών είναι παραδείγματα τέτοιων χαρακτηριστικών.

- **Πολιτιστικά δεδομένα**

  Πληροφορίες που σχετίζονται με τη μουσική του ενδιαφέροντος, αλλά δεν αποτελούν άμεση αναπαράσταση, αφηρημένη ή άλλη, του πραγματικού ήχου που περιλαμβάνει τη μουσική. Το Διαδίκτυο παρέχει την πιο συχνά χρησιμοποιούμενη πηγή πολιτιστικών δεδομένων, καθώς διαθέτει πόρους όπως επεξεργασμένα αποθετήρια μεταδεδομένων, μη

επεξεργασμένες ετικέτες ακροατή, λίστες αναπαραγωγής και ιστοσελίδες με δυνατότητα αναζήτησης γενικά. Άλλες πιθανές πηγές πολιτιστικών δεδομένων περιλαμβάνουν εξειδικευμένα κείμενα όπως κριτικές άλμπουμ, στατιστικά στοιχεία όπως στατιστικά πωλήσεων, έρευνες, τα αποτελέσματα ψυχολογικών πειραμάτων και εικόνες της τέχνης του άλμπουμ. [34]

3.3 Άντληση χαρακτηριστικών γνωρισμάτων μέσω συνελικτικού νευρωνικού δικτύου.

Η άντληση πληροφορίας από μουσική (Music Information Retrieval - MIR) είναι ένα πεδίο μελέτης με συνεχή ανάπτυξη τα τελευταία χρόνια, οδηγούμενο από την ανάγκη αυτόματης επεξεργασίας μεγάλου όγκου μουσικών τραγουδιών. Πολλές είναι οι επιστήμες και τα επιστημονικά πεδία από τα όποια η MIR αντλεί γνώση, όπως η ψυχολογία, η μουσικολογία, η επεξεργασία σήματος, η μηχανική μάθηση και πολλά άλλα. Μερικά αντικείμενα μελέτης στην MIR είναι η αναγνώριση του είδους μουσικής στο οποία ανήκει ένα μουσικό κομμάτι, ο διαχωρισμός φωνής και ενόργανης μουσικής, η παρακολούθηση ρυθμού και τονικού ύψους, η αναγνώριση μουσικών οργάνων αλλά και η αναγνώριση συναισθήματος ενός μουσικού κομματιού. [34]

Τα συνελικτικά νευρωνικά δίκτυα CNN έχουν το βασικό πλεονέκτημα ότι ανιχνεύει αυτόματα τα σημαντικά χαρακτηριστικά (feature extraction) χωρίς καμία ανθρώπινη επίβλεψη. Για παράδειγμα, με δεδομένες πολλές εικόνες γάτας και σκύλου, μαθαίνει ξεχωριστά χαρακτηριστικά για κάθε τάξη από μόνη της. Έτσι ένα CNN μπορεί να μάθει τα διαφορετικά χαρακτηριστικά δεδομένων κομμάτια που ανήκουν σε ξεχωριστές κλάσεις.

Κεφάλαιο 4

Δεδομένα για ταξινόμηση

4.1 (MIDI) δεδομένα

Το **MIDI** (Musical Instrument Digital Interface, ελλ. Ψηφιακή Διασύνδεση Μουσικών Οργάνων) είναι ένα σύστημα κωδικοποίησης που χρησιμοποιείται για την αντιπροσώπευση, μεταφορά και αποθήκευση μουσικών πληροφοριών. Ουσιαστικά πρόκειται για πληροφορία μουσικής σημειογραφίας ή πληροφορία παρτιτούρας που μπορεί να αποθηκευτεί ηλεκτρονικά. Αντί να περιέχουν πραγματικά δείγματα ήχου όπως κάνουν οι μέθοδοι κωδικοποίησης ήχου, τα αρχεία MIDI αποθηκεύουν οδηγίες που μπορούν να σταλούν σε συνθεσάιζερ. Η ποιότητα του ήχου που παράγεται κατά την αναπαραγωγή ενός αρχείου MIDI εξαρτάται σε μεγάλο βαθμό από το synthesizer (ή από τον υπολογιστή) στον οποίο αποστέλλονται οι οδηγίες MIDI. Στην πραγματικότητα, οι ηχογραφήσεις MIDI δίνουν σε όλους τις ίδιες πληροφορίες που θα βρει κανείς σε ένα μουσικό θέμα (παρτιτούρα), μπορούμε δηλαδή να το θεωρούμε ως ψηφιοποιημένη παρτιτούρα.

Το MIDI αποτελείται από μια απλή διεπαφή υλικού (hardware interface) και από ένα πιο περίπλοκο πρωτόκολλο μετάδοσης (transmition protocol). Εμείς θα ασχοληθούμε μόνο με το πρωτόκολλο MIDI.

4.2 Γιατί συμβολικά (MIDI) δεδομένα

Όπως έχουμε αναφέρει, από τις συμβολικές αναπαραστάσεις δεδομένων εξάγουμε κυρίως **χαρακτηριστικά με περιεχόμενο υψηλού επιπέδου**, δηλαδή πληροφορίες σχετικά με την μουσική. Στην πραγματικότητα, οι εγγραφές MIDI δίνουν τις ίδιες πληροφορίες που θα μπορούσε κανείς να βρει σε μία μουσική παρτιτούρα.

Οι εγγραφές MIDI δηλαδή έχουν ορισμένα σημαντικά **πλεονεκτήματα** σε σχέση με τις ηχογραφήσεις. Είναι πολύ πιο συμπαγής, κάτι που με τη σειρά του τα καθιστά ευκολότερα στην αποθήκευση και πολύ πιο γρήγορα στην επεξεργασία και την ανάλυση. Οι εγγραφές MIDI είναι επίσης ευκολότερες στην επεξεργασία, καθώς αποθηκεύουν απλές οδηγίες που είναι εύκολο να προβληθούν και να αλλάξουν. Αυτό έρχεται σε αντίθεση με τις ηχογραφήσεις, όπου προς το παρόν δεν είναι δυνατό (με εξαίρεση την απλή μονοφωνική μουσική) να εξάγουμε

σωστά τις πραγματικές νότες που παίζονται.

Όμως, το πρότυπο MIDI είναι γνωστό ότι έχει μια σειρά από αδυναμίες και μειονεκτήματα που καλό είναι να έχουμε στο μυαλό μας. Υπάρχει ένα σχετικά χαμηλό θεωρητικό όριο σχετικά με την ποσότητα πληροφοριών ελέγχου που μπορεί να ενσωματώσει το MIDI. Επιπλέον, μπορεί να είναι δύσκολο και χρονοβόρο να καταγράψουμε σωστά σύνθετες οδηγίες σύνθεσης. Στην πραγματικότητα, μια βασική εγγραφή MIDI μιας ανθρώπινης απόδοσης θα ακούγεται σχεδόν πάντα χειρότερα από μια ηχογράφηση, εν μέρει επειδή είναι αδύνατο να καταγραφεί σωστά το πλήρες εύρος των παραμέτρων ελέγχου πολλών οργάνων και εν μέρει λόγω των περιορισμών στους συνθεσάιζερ.

Το MIDI είναι επομένως πολύ πιο βολικό από τον ήχο εάν κάποιος επιθυμεί να εξαγάγει ακριβείς **μουσικές πληροφορίες υψηλού επιπέδου**. Για τους σκοπούς της ανάλυσης του είδους, αυτό είναι ένα σημαντικό πλεονέκτημα, καθώς είναι επιθυμητό να αναζητούμε μοτίβα που σχετίζονται με νότες, ρυθμούς, χορδές και όργανα, τα οποία είναι εύκολο να εξαγάγετε από το ΜΙΔΙ, αλλά προς το παρόν είναι δύσκολο και αδύνατο να εξάγουμε από τον ήχο.

4.3 Μορφή ενός αρχείου MIDI

Τα αρχεία MIDI περιέχουν μία ή περισσότερες ροές MIDI, με πληροφορίες για το κομμάτι. Μπορούν να υποστηριχθούν όλα τα τραγούδια, όλες η ακολουθίες και οι δομές κομματιών, οι πληροφορίες για το ρυθμό, το τέμπο(tempo),το χρόνο. Ακόμα τα ονόματα κομματιών και άλλες περιγραφικές πληροφορίες μπορούν να αποθηκευτούν με τα δεδομένα MIDI. Αυτή η μορφή υποστηρίζει πολλαπλά κομμάτια και πολλαπλές ακολουθίες, έτσι ώστε εάν ο χρήστης ενός προγράμματος που υποστηρίζει πολλαπλά κομμάτια σκοπεύει να μετακινήσει ένα αρχείο σε άλλο, αυτή η μορφή μπορεί να το επιτρέψει.

Τα αρχεία MIDI περιλαμβάνουν πληροφορίες επικεφαλίδας (header chunk) και μια λίστα κομματιών(tracks).

Τα κομμάτια επικεφαλίδας ορίζουν κάποιες συγκεκριμένες πληροφορίες σχετικά με τα δεδομένα του αρχείου. Μας δίνεται σαν πληροφορία πόσα τικ (ticks) αντιστοιχούν σε ένα τέταρτο (Ticks Per Quarter Note). Ποιός είναι ο ρυθμός (MIDI time signature), το κλειδί (key signature) στο οποίο ανήκει το κομμάτι κ.α.

Κάθε κομμάτι track περιλαμβάνει μία σειρά από συμβάντα (events). Κάθε συμβάν αποτελείται από δύο στοιχεία: ένα χρόνο (time) MIDI και ένα μήνυμα (message) MIDI. Αυτά τα ζεύγη χρόνου/μηνυμάτων (time/message) ακολουθούν το ένα το άλλο σε ένα αρχείο MIDI.

4.3.1 Χαρακτηριστικά μηνύματος

Τα μηνύματα MIDI αποστέλλονται ως ακολουθία ενός ή περισσότερων byte. Το πρώτο byte είναι ένα byte εντολών (STATUS), ακολουθούμενο συχνά από byte δεδομένων (DATA) με πρόσθετες παραμέτρους.

Οι κύριες εντολές είναι τα πάτημα νότας (Note-on) και πότε αυτή αφήνεται (Note-off) που επιτρέπουν την έναρξη / διακοπή της αναπαραγωγής μιας μουσικής νότας. Το μήνυμα Note-

on αποστέλλεται όταν ο εκτελεστής πατά ένα πλήκτρο των πλήκτρων μουσικής. Περιέχει παραμέτρους για τον καθορισμό του τονικού ύψους της νότας καθώς και της ταχύτητας (δηλ. ένταση της νότας όταν χτυπιέται). Όταν ένα συνθεσάιζερ λαμβάνει αυτό το μήνυμα, αρχίζει να παίζει τη νότα με το σωστό επίπεδο βήματος και δύναμης. Όταν ληφθεί το μήνυμα Note-off, η αντίστοιχη νότα απενεργοποιήται από το συνθεσάιζερ.

Υπάρχουν τουλάχιστον δύο καταστάσεις λειτουργίας, δηλαδή μονοφωνικό και πολυφωνικό.

- Μονοφωνικό

  Η έναρξη μιας νέας εντολής "Note-on" υποδηλώνει τον τερματισμό της προηγούμενης νότας.

- Πολυφωνικό

  Μπορεί να ακούγονται πολλές νότες ταυτόχρονα, έως ότου οι νότες σταματήσουν να ακούγονται με την σταδιακή πτώση έντασής τους (decay), ή όταν ληφθούν ρητές εντολές "Note-off".

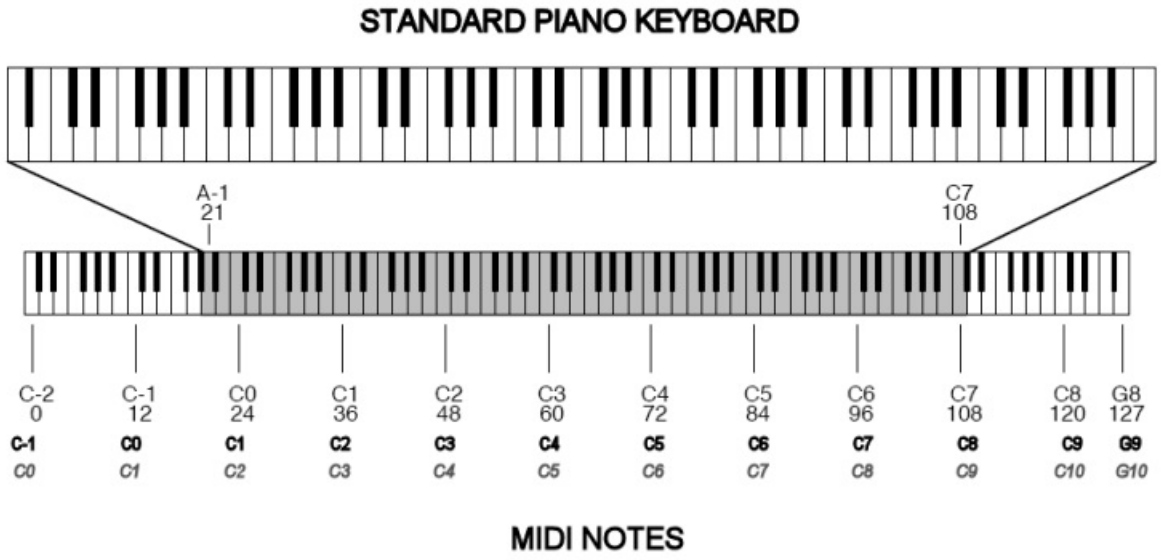
Το πρωτόκολλο MIDI επιτρέπει την αποστολή μηνυμάτων πάνω από **16 ανεξάρτητα κανάλια MIDI (channels)**, επιτρέποντας 16 όργανα.

4.3.2 Τονικό ύψος (pitch)

Το **τονικό ύψος (pitch)** μίας νότας έχει τιμές από 0 έως 127. Υπάρχουν τρεις συμβάσεις ονομασίας για τις ίδιες τις νότες των τιμών αυτών. Η πρώτη, και ίσως πιο συνηθισμένη βασίζεται στα πλήκτρα του πιάνου και αριθμεί την κλίμακα MIDI ως "C-2" έως "G8".Στο σχήμα 4.1 φαίνεται η αντιστοιχία νότας και τιμής στο MIDI.

Νούμερο οκτάβας	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
-2	0	1	2	3	4	5	6	7	8	9	10	11
-1	12	13	14	15	16	17	18	19	20	21	22	23
0	24	25	26	27	28	29	30	31	32	33	34	35
1	36	37	38	39	40	41	42	43	44	45	46	47
2	48	49	50	51	52	53	54	55	56	57	58	59
3	60	61	62	63	64	65	66	67	68	69	70	71
4	72	73	74	75	76	77	78	79	80	81	82	83
5	84	85	86	87	88	89	90	91	92	93	94	95
6	96	97	98	99	100	101	102	103	104	105	106	107
7	108	109	110	111	112	113	114	115	116	117	118	119
8	120	121	122	123	124	125	126	127	-	-	-	-

Πίνακας 4.1: Αντιστοίχιση νοτών με αναπαράσταση σε MIDI για διαφορετικές οκτάβες



Πηγή: http://www.midisolutions.com/chapter3.htm
Σχήμα 4.1: Αντιστοίχιση νοτών με το πλήκρα ενός πιάνου

4.3.3 Ένταση ήχου νότας (velocity)

Η **ένταση (velocity)** μιας νότας έχει τιμές από 0 έως 127, καλύπτοντας το εύρος από μια πρακτικά υπερβολικά χαμηλή ένταση νότας έως το μέγιστο επίπεδο έντασης μιας νότας. Αντιστοιχεί βασικά στην κλίμακα των αποχρώσεων που βρίσκονται στη μουσική σημειογραφία,(πιάνο (p), πιανίσιμο (pp), φόρτε (f), φορτίσιμο (ff).) Η αντιστοίχιση φαίνεται στο σχήμα 4.2.

Πηγή: https://www.researchgate.net/figure/Parameter-data-byte-associated-with-the-note-on-command_fig3_316955785/
Σχήμα 4.2: Σχέση έντασης νότας με μουσική σημειογραφία

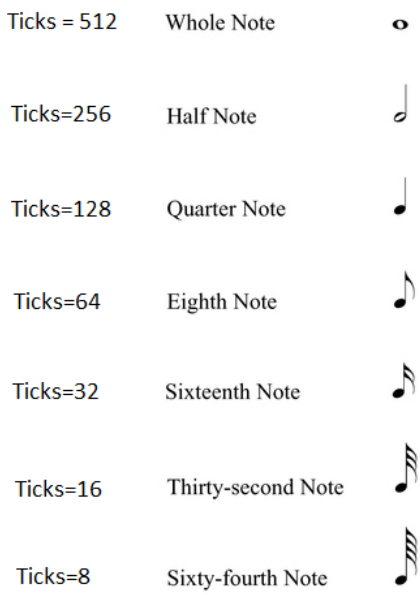
4.3.4 Μέτρηση του χρόνου (time)

Για να γίνει η **μέτρηση του χρόνου (time)** πρέπει να περιμένουμε και κάποιο επόμενο μήνυμα από την ροή των δεδομένων MIDI. Μετριέται δηλαδή σε MIDI ticks από το προηγούμενο μήνυμα που λάβαμε. Αυτή η μέθοδος καθορισμού του χρόνου ονομάζεται χρόνος δέλτα και φαίνεται στον πίνακα 4.2

delta time (ΔT)	elapsed time until event
$t_1 = \Delta T_1 = T_1 - 0$	$T_1 = t_1$
$t_2 = \Delta T_2 = T_2 - T_1$	$T_2 = t_1 + t_2$
$t_3 = \Delta T_3 = T_3 - T_2$	$T_3 = t_1 + t_2 + t_3$
...	...
$t_n = \Delta T_n = T_n - T_{n-1}$	$T_n = t_1 + t_2 + ... + t_n$
...	...

Πίνακας 4.2: Εξήγηση των χρόνων δέλτα σε μια σειρά συμβάντων

Έτσι μπορούμε να βρούμε τον χρόνο που η νότα ήταν πατημένη βρίσκοντας την διαφορά του χρόνου μεταξύ δύο μηνυμάτων τις νότας. Συγκρίνοντάς το με το τικ (χτύπημα) ανά τέταρτο (ticks per quarter) μπορούμε να βρούμε και την αξία της νότας στη μουσική. Για παράδειγμα εάν το 128 ticks αντιστοιχούν σε ένα τέταρτο τότε για το μισό θα αντιστοιχούν 256, ενώ για το όγδοο 64. Στο σχήμα 4.3 μπορούμε να δούμε ένα παράδειγμα αντιστοίχισης πλήθος το ticks με τις αξίες των νοτών.

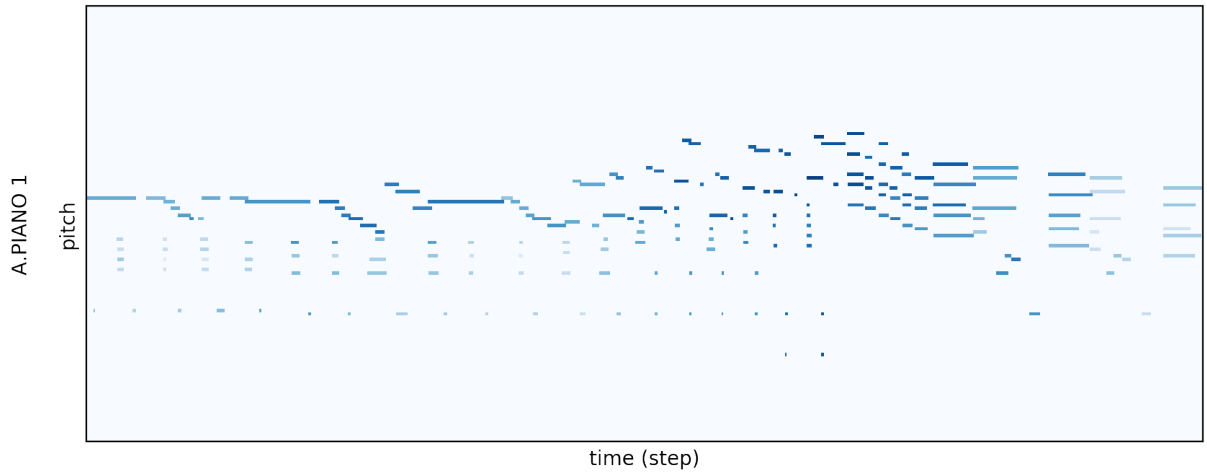


Σχήμα 4.3: Αριθμός χτύπων σε σχέση με την διάρκεια μιας νότας.

4.4 Μετατροπή σε pianoroll

Παρόλο που το πρωτόκολλο MIDI έχει όλα τα προαναφερθέντα θετικά, εμείς θα χρησιμοποιήσουμε pianoroll ως το dataset για το νευρωνικό μας δίκτυο. Πρώτα όμως ας εξηγήσουμε τί είναι το pianoroll και ποια η αντιστοίχιση των δύο μορφών MIDI και pianoroll.

Το Pianoroll (ελλ. ρολό πιάνου) είναι μια μορφή αποθήκευσης μουσικής που αντιπροσωπεύει ένα μουσικό κομμάτι από έναν πίνακα σκορ (score-like matrix). Οι κατακόρυφοι και οριζόντιοι άξονες αντιπροσωπεύουν το τονικό ύψος (pitch) και τη διάρκεια (duration) των νοτών, αντίστοιχα. Οι τιμές αντιπροσωπεύουν τις εντάσεις (velocities) των νοτών.



Πηγή: https://salu133445.github.io/lakh-pianoroll-dataset/representation

Σχήμα 4.4: Παράδειγμα pianoroll, όπου οι κατακόρυφος και οριζόντιος άξονας αντιπροσωπεύει τονικό ύψος(pitch) και χρόνο(time), αντίστοιχα.

Ο άξονας του χρόνου μπορεί να είναι ως απόλυτος χρονος ή σε συμβολικός χρόνος. Για απόλυτο χρονισμό, χρησιμοποιείται ο πραγματικός χρόνος εμφάνισης της νότας. Για συμβολικό χρονισμό, οι πληροφορίες για τον ρυθμό tempo αφαιρούνται και έτσι κάθε χτύπος έχει το ίδιο μήκος, ανεξάρτητα της ταχύτητας του κομματιού. [12, 38]

Πολλές μορφές pianoroll, όπως αυτές που εφαρμόζονται στο pretty midi [39], χρησιμοποιούν τον απόλυτο χρονισμός, όπου ο πραγματικός χρόνος (σε δευτερόλεπτα) των νοτών που είναι πατημένες χρησιμοποιούνται για τον χρονικό άξονα. Όμως, αυτό εισάγει χαρακτηριστικά αποδόσεων στα pianorolls.

Αντίθετα στο **pypianoroll** [20] που θα χρησιμοποιήσουμε για τα πειράματα, χρησιμοποιούμε συμβολικό χρονισμό και ο ρυθμός ανάλυσης ορίζεται ως 24 ανά χτύπο για να καλύψουμε κοινά χρονικά μοτίβα, όπως τρίηχα και τριακοστά δεύτερα. Το τονικό ύψος έχει 128 δυνατότητες, καλύπτοντας από το C-2 έως το G8. Για παράδειγμα, ένα μέτρο σε 4/4 χρόνο με μόνο ένα κομμάτι μπορεί να αναπαρασταθεί ως πίνακας *96 x 128*. Ενώ ο ρυθμός (tempo) αποθηκεύεται ως έξτρα πίνακας, αφού αφαιρείται η πληροφορία για το ρυθμό από τους πίνακες.

Σημειώνουμε πως κατά τη μετατροπή από αρχεία MIDI σε pianorolls, προστίθεται μια επιπλέον παύση ελάχιστου μήκους (ενός βήματος χρόνου) μεταξύ δύο διαδοχικών (χωρίς παύση) νοτών του ίδιου βήματος για να τις διακρίνετε από μία μόνο νότα. Δεν θεωρούμε όμως πως αυτή η αλλαγή επηρεάζει την μουσική πληροφορία του κομματιού.

Στο Pypianoroll, χρησιμοποιείται η κλάσση Multitrack (ελλ. πολυμερές ή πολυκαναλικό κομμάτι), ως βασική κλάσση για τα πολυμερή pianorolls.Όπως φαίνεται και στους πίνακες 4.3

και 4.4, ένα αντικείμενο Multitrack περιλαμβάνει μία λίστα από αντικείμενα Tracks(κομμάτια), τα οποία περιέχουν πίνακες (matrices) pianoroll, ένα νόυμερο programm, μία λογική τιμή is_drum αληθές ή ψευδές για το εάν το κομμάτι ανήκει στα κρουστά(drums) και το όνομα του κομματιού. Ένα αντικείμενο Multitrack περιέχει επίσης το beat resolution, δηλαδή τον αριθμό των βημάτων που χρησιμοποιούνται για την αναπαράσταση του ρυθμού, περιέχει έναν πίνακα με τον ρυθμό (tempo array), έναν πίνακα με τις θέσεις των πρώτων χτύπων κάθε μέτρου downbeat και το όνομά του.

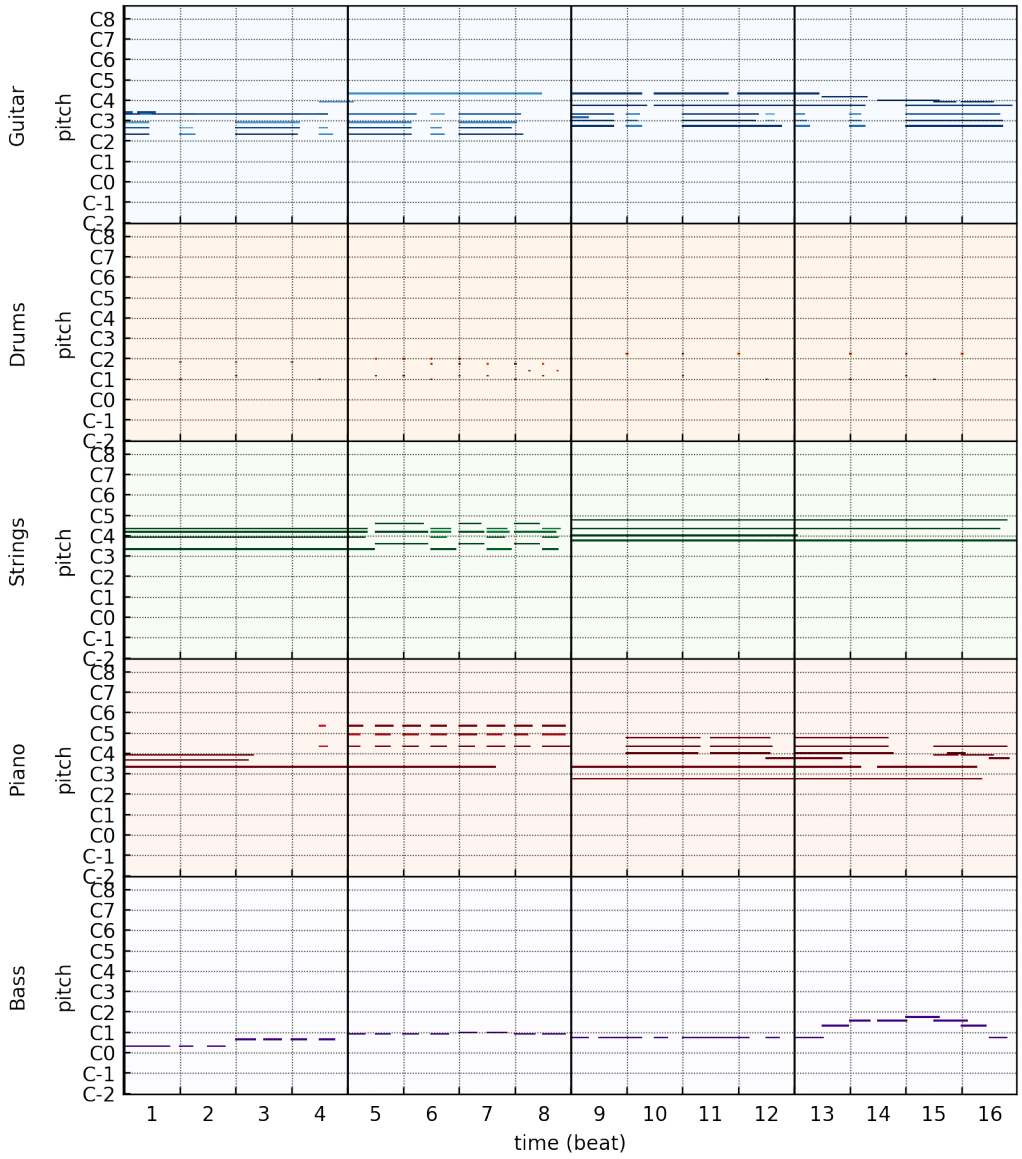
Attribute	Description
tracks	List of Track objects
beat	resolution Resolution of a beat (in time step)
tempo	Array that records the tempo value (in bpm) at each time step
downbeat	Array that indicates the locations of downbeats (the first beat of a bar)
name	Name of the multitrack

Πίνακας 4.3: Χαρακτηριστικά ενός αντικειμένου πολυκαναλιού (Attributes of a Multitrack object)

Attribute	Description
pianoroll	Pianoroll matrix
program	Program number according to General MIDI Level 1 specification
is_drum	Whether it is a percussion track
name	Name of the track

Πίνακας 4.4: Χαρακτηριστικά ενός αντικειμένου καναλιού(Attributes of a track object.)

Pianoroll με πολλαπλά μέρη ενός κομματιού (Multitrack) Αντιπροσωπεύουμε ένα μουσικό κομμάτι multitrack με ένα pianoroll πολλαπλών μερών του κομματιού(tracks), το οποίο είναι ένα σύνολο από pianoroll, όπου κάθε pianoroll αντιπροσωπεύει ένα συγκεκριμένο μέρος στο αρχικό κομμάτι της μουσικής. Δηλαδή, ένα μουσικό κομμάτι με M μέρη (M-track) θα μετατραπεί σε ένα σύνολο M pianorolls. Για παράδειγμα, σε 4/4 χρόνο με M κομμάτια μπορεί να αναπαρασταθεί ως πίνακα διαστάσεων *time steps x 128 x M*.



Πηγή: https://salu133445.github.io/lakh-pianoroll-dataset/representation

Σχήμα 4.5: Παράδειγμα multitrack pianoroll

Κεφάλαιο 5

Ανάλυση και προεπεξεργασία δεδομένων

Στο κεφάλαιο αυτό θα αναλυθούν τα δεδομένα που θα χρησιμοποιήσουμε καθώς και η προεπεξεργασία που χρειάζεται για να έρθουν σε κατάλληλη μορφή.

5.1 Ανάλυση δεδομένων

Θα χρησιμοποιήσουμε pianoroll ως δεδομένα, όπως αναφέραμε και στο 4.4. Συγκεκριμένα θα χρησιμοποιήσουμε το **The Lakh Pianoroll Dataset - LPD**[12, 38] που είναι μια συλλογή από 174,154 πολυμερών (multitrack) pianorolls που προέρχονται από το **Lakh MIDI Dataset (LMD)**.[38, 7]

Υπάρχουν διάφορα υποσύνολα του του LPD και συγκεκριμένα τα:

- lpd-full

  Το lpd-full περιέχει 174,154 πιανορολλς πολυμερών κομματιών που προέρχονται από το σύνολο δεδομένων Lakh MIDI (LMD).

- lpd-matched

  Το lpd-matched περιέχει 115,160 pianoroll πολυμερών κομματιών που προέρχονται από την αντίστοιχη έκδοση του LMD. Αυτά τα αρχεία αντιστοιχίζονται σε καταχωρίσεις στο σύνολο δεδομένων Million Song (MSD).

- lpd-cleaned

  Το lpd-cleaned περιέχει 21.425 πιανορολλς πολυμερών κομματιών που συλλέγονται από το υποσύνολο lpd-matched με τους ακόλουθους κανόνες:

  – Αφαιρούμε όσα έχουν παραπάνω από μία μετατροπή στο ρυθμό (time signature change)

  – Αφαιρούμε όσα έχουν ρυθμό διαφορετικό από 4/4

– Διατηρήστε μόνο ένα αρχείο που έχει την υψηλότερο βαθμό εμπιστοσύνης για την αντιστοίχιση για κάθε τραγούδι. Η εμπιστοσύνη αυτή αναφέρεται στο κατα πόσο το αρχείο MIDI ταιριάζει με οποιαδήποτε καταχώριση στο MSD

Εμείς θα χρησιμοποιήσουμε το lpd-cleansed για τα πειράματά μας, όντας το πιο απλό και ξεκάθαρο.

Ετικέτες

Οι ετικέτες για το σύνολο δεδομένων Lakh Pianoroll (LPD) προέρχονται από τρεις διαφορετικές πηγές: το σύνολο δεδομένων Last.fm[7], τα σημεία αναφοράς(benchmarks) του Million Song Dataset (MSD) [7] και τους σχολιασμούς του Tagtraum genre[42]. Από το MSD χρησιμοποιούμε τα σετ δεδομένων TopMAGD, MASD και ένα άλλο υποσύνολο του MASD, το MASD-labels-cleansed.

MASD-cleansed	
Genre Name	Number of Tracks
Country Traditional	438
Dance	462
Metal Alternative	263
Pop Contemporary	844
Pop Indie	290
Pop Latin	205
Rock Alternative	167
Rock College	197
Rock Contemporary	561
Rock Hard	332

Πίνακας 5.1: Κατηγορίες και πλήθος κομματιών του MASD-cleansed

MSD Allmusic Style Dataset (MASD)	
Genre Name	Number of Tracks
Big Band	61
Blues Contemporary	26
Country Traditional	438
Dance	462
Electronica	112
Experimental	149
Folk International	131
Gospel	104
Grunge Emo	96
Hip Hop Rap	164
Jazz Classic	64
Metal Alternative	263
Metal Death	82
Metal Heavy	86
Pop Contemporary	844
Pop Indie	290
Pop Latin	205
Punk	35
Reggae	35
RnB Soul	125
Rock Alternative	167
Rock College	197
Rock Contemporary	561
Rock Hard	332
Rock Neo Psychedelia	127

Πίνακας 5.2: Κατηγορίες και πλήθος κομματιών του MASD

MSD Allmusic Genre Dataset (TOP-MAGD)	
Genre Name	Number of Tracks
Blues	22
Country	512
Electronic	889
Folk	45
International	206
Jazz	157
Latin	360
NewAge	67
Pop_Rock	4345
Rap	164
Reggae	45
RnB	397
Vocal	114

Πίνακας 5.3: Κατηγορίες και πλήθος κομματιών του TOP-MAGD

Tagtraum	
Genre Name	Number of Tracks
Blues	19
Country	448
Electronic	352
Folk	33
Jazz	127
Latin	70
Metal	112
NewAge	29
Pop	1086
Punk	12)
Rap	89
Reggae	53
RnB	264
Rock	1668
World	10

Πίνακας 5.4: Κατηγορίες και πλήθος κομματιών του Tagtraum

Lastfm			
Genre Name	Number of Tracks	Genre Name	Number of Tracks
00s	1453	fun	1253
60s	1140	funk	742
70s	1579	guitar	912
80s	2328	happy	1230
90s	1984	hardcore	123
acoustic	768	hip-hop	477
alternative	2076	house	681
amazing	763	indie	1231
ambient	413	instrumental	677
american	1814	jazz	673
awesome	1733	lounge	411
beautiful	2205	love	3227
blues	620	loved	1024
british	1448	melancholy	871
catchy	1496	mellow	1688
chill	1257	metal	585
chillout	1314	oldies	2507
classic	1668	party	1836
cool	1322	piano	669
country	911	pop	5808
cover	730	psychedelic	305
dance	2904	punk	461
downtempo	316	rap	289
electro	488	reggae	198
electronic	1768	relax	792
electronica	878	rnb	1118
experimental	213	rock	4693
favorite	1475	sad	1061
favorites	3339	sexy	1252
favourite	1443	soul	1537
favourites	1624	soundtrack	1140
female	1051	techno	539
folk	669	trance	643

Πίνακας 5.5: Κατηγορίες και πλήθος κομματιών του Lastfm

5.2 Επεξεργασία δεδομένων

Στο LPD,χρησιμοποιείται συμβολικός χρονισμός και ο ρυθμός ανάλυσης beat resolution ορίζεται ως 24 ανά χτύπο για να καλύψουμε κοινά χρονικά μοτίβα, όπως τρίηχα και τριακοστά δεύτερα. Το τονικό ύψος της φωνής έχει 128 δυνατότητες, καλύπτοντας από το C-2 έως το G8. Για παράδειγμα, ένα μέτρο σε 4/4 χρόνο με μόνο ένα κομμάτι (track) μπορεί να αναπαρασταθεί ως πίνακας *96 x 128*.

Έτσι στο pianoroll όλα τα κομμάτια μπορούν να αναπαρασταθούν ως πίνακες time_steps x 128, όπου time_steps είναι ένας αριθμός που αναπαραστά κατάλληλα τον συμβολικό χρονισμό.

5.2.1 Υποδειγματοληψία - Μείωση των χρονικών βημάτων ανά χτύπο (Downsampling - Reducing beat resolution)

Μειώνουμε τα χρονικά βήματα ανά χτύπο από 24 σε 12. Αυτό το υλοποιούμε εξάγωντας κάθε δεύτερο στοιχείο του πίνακα. Αυτή η διαδικασία λέγεται υποδειγματοληψία (downsampling) Αυτό έχει το σημαντικό πλεονέκτημα ότι μειώνεται στα δύο η διάσταση του πίνακα. Έτσι από time_steps x 128 θα γίνει ($new\_time\_steps = time\_steps/2$) x 128.

Έχει όμως το μειονέκτημα ότι δεν μπορούμε να διακρίνουμε νότες με πολύ μικρή διάρκεια.

Πιο συγκεκριμένα, αρχικά είχαμε 24 χρονικά βήματα ανά χτύπο. Ο χτύπος αντιπροσωπεύει ένα τέταρτο σύμφωνα με τους κανόνες του lpd-cleansed του Lakh Pianoroll Dataset 5.1. Έτσι πριν την υποδειγματοληψία, μπορούσαμε να ξεχωρίσουμε νότες μέχρι και τρίηχο εξηκοστό τέταρτο που είναι αντίστοιχο με στακάτο εξηκοστό τέταρτο (1/64) ή εάν το αποκαλούσαμε τελείως μαθηματικά και όχι μουσικά ένα ενενηκοστό έκτο (1/96)).

Μετά την μετατροπή έχουμε 12 χρονικά βήματα ανά χτύπο. Άρα πλέον μπορούμε να ξεχωρίσουμε νότες μέχρι και τρίηχο τριακοστό δεύτερο που είναι αντίστοιχο με στακάτο τριακοστό δεύτερο (1/32)), που μαθηματικά θεωρείται ως ένα τεσσαρακοστό όγδοο (1/48)).

Κάνοντας αυτήν την υποδειγματοληψία έχουμε 48 χρονικά βήματα ανά 4/4, δηλαδή ανά μουσικό μέτρο.

Εν τέλει, μειώνοντας τη διάσταση κατά δύο, θεωρούμε πως δεν χάνουμε σπουδαία πληροφορία, το οποίο επιβεβαιώσαμε και με σύντομα πειράματα. Κερδίζουμε όμως μείζονα χρόνο για την εκπαίδευση του νευρωνικού μας δικτύου.

5.2.2 Μείωση διαστάσεων

Αφαιρέσαμε από την διάσταση του τονικού ύψους (pitch) τις πολύ υψηλής και πολύ χαμηλής συχνότητας νότες. Έτσι η διάσταση του πίνακα έγινε από time_steps x 128 σε

Διάσταση πίνακα pianoroll = time_steps x 88

5.2.3 Δημιουργία ενός πίνακα(matrix) για κάθε κομμάτι

Πιο συγκεκριμένα εάν ένα κομμάτι έχει πολλά tracks (π.χ. κάθε track είναι διαφορετικό όργανο), τότε μπορούν να αναπαρασταθούν με πολλούς πίνακες, έναν πίνακα άνα όργανο. Επειδή όμως αυτό είναι μη σταθερό μέγεθος (ένα κομμάτι μπορεί να έχει 1 όργανο, άλλα

περισσότερα) επιλέξαμε να κρατάμε έναν πίνακα για κάθε κομμάτι. Αυτός ο πίνακας δημιουργήθηκε κρατώντας το άθροισμα των πινάκων του κάθε οργάνου και διαιρώντας δια το πλήθος των οργάνων. Βρίσκουμε δηλαδή τον μέσο όρο των τιμών όλων των πινάκων.

$$M_{88,48} = \left(\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,48} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,48} \\ \vdots & \vdots & \ddots & \vdots \\ a_{88,1} & a_{88,2} & \cdots & a_{88,48} \end{bmatrix} + \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,48} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,48} \\ \vdots & \vdots & \ddots & \vdots \\ b_{88,1} & b_{88,2} & \cdots & b_{88,48} \end{bmatrix} \right)/2$$

Σχήμα 5.1: Παράδειγμα πίνακα για ένα κομμάτι 4/4 ενός μέτρου(48 χρονικά βήματα ανά χτύπο) με δύο όργανα. Υπολογίζουμε το άθροισμα των δύο πινάκων και μετά τον μέσο όρο των πινάκων

Πρακτικά μπορούμε να σκεφτούμε ένα κομμάτι ότι αντιπροσωπεύεται από έναν πίνακα που δείχνει τον χρόνο και την νότα και την ένταση που αυτή παίζεται, για όλα τα μουσικά όργανα.

5.2.4 Κανονικοποίηση πίνακα

Ύστερα κανονικοποιούμε τον πίνακα σε τιμές [0,1]. Αυτό το κάνουμε διαιρώντας όλες τις τιμές του πίνακα, με την μέγιστη τιμή του πίνακα.

5.2.5 Αλλαγή σε δύο πινάκες ανά κομμάτι

Αρχικά προσθέσαμε όλα τα όργανα και κρατούσαμε έναν πίνακα. Όμως μετά από πειραματικά αποτελέσματα φάνηκε πως αυτό δεν ήταν καλή πρακτική.

Έτσι κρατάμε **δύο πίνακες** για κάθε κομμάτι, έναν για όλα τα μουσικά όργανα που δεν είναι κρουστά (drums) και έναν για όλα τα όργανα που είναι κρουστά (drums). Ας αποκαλέσουμε τα πρώτα όργανα ως τα **Μελωδικά όργανα** ενώ τα δεύτερα **Κρουστά όργανα**. Αυτό το επιλέξαμε ώστε να θεωρούμε ότι ο πρώτος πίνακας περιέχει κυρίως την πληροφορία για την **αρμονία** και την **μελωδία** και ο δεύτερος πίνακας για τον **ρυθμό**.

Ο διαχωρισμός αυτός έγινε εύκολα αφού όπως αναφέραμε και στο 4.4 υπάρχει σε κάθε Track η πληροφορία εάν είναι κρουστό (is_drum).

Σε περίπτωση που κάποιο κομμάτι δεν έχει μία από τις δύο κατηγορίες (μελωδικά ή κρουστά όργανα) τότε ο πίνακας της συγκεκριμένης κατηγορίας θα είναι ο μηδενικός πίνακας.

5.2.6 Κόψιμο-διαχωρισμός σε μικρότερα μέρη (Slicing)

Όπως είδαμε παραπάνω, ένα pianoroll matrix έχει δύο διαστάσεις, αυτή των χρονικών βημάτων time steps και αυτή του τονικού ύψους pitch. Μετασχηματίσαμε εύρος του τονικού ύψος από 128 σε 88. Κάθε κομμάτι όμως έχει διαφορετική διάρκεια σε δευτερόλεπτα και αντίστοιχα και σε χρονικά βήματα.

Όμως για την εκπαίδευση του νευρωνικού μας δικτύου θα χρειαστούμε πίνακες ίδιου μεγέθους. Έτσι, κάθε κομμάτι θα χωριστεί σε μικρότερα μέρη. Θα τα ονομάσουμε αυτά **κομμένα τραγούδια (sliced songs)**.

Έγιναν πολλές διαφορετικές διαμερίσεις με χρονικά βήματα

- Διαμερίσεις που αντιστοιχούν σε συγκεκριμένα μέτρα όπως:

  96, 192, 384, 576, 768, 1152, 1536 που αντιστοιχούν σε 2, 4, 8, 12, 16, 20, 24, 32 μουσικά μέτρα αντίστοιχα.

- Διαμερίσεις που δεν αντιστοιχούν σε συγκεκριμένα μέτρα όπως:

  128, 256, 512, 1024

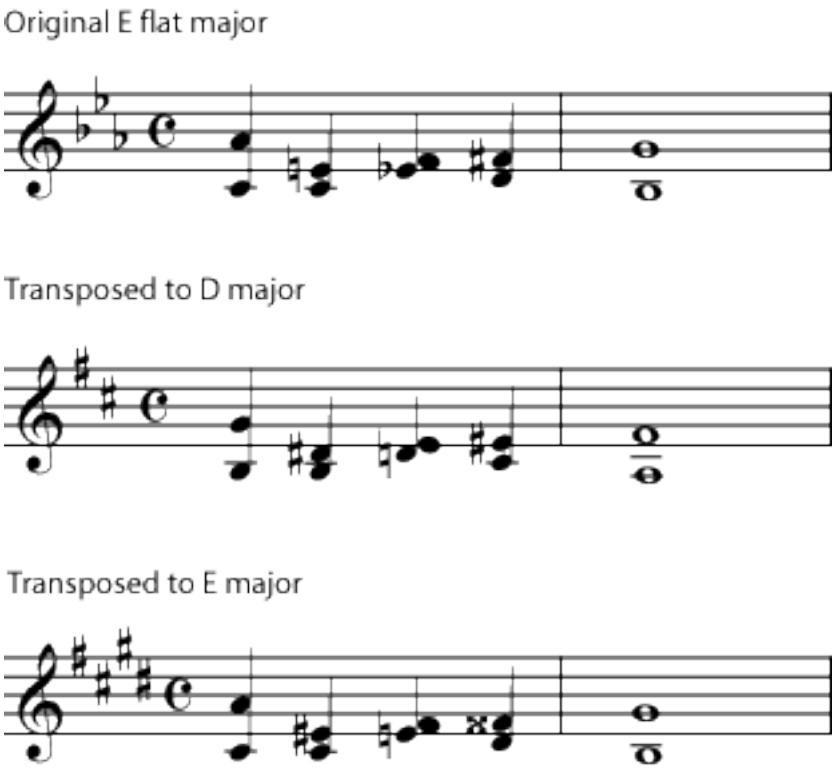
5.2.7 Αλλαγή κλειδιού (Changing Keys) - Αλλαγή Τονικότητας - Μετατροπία (Transposition)

Όλες τις διαμερίσεις του κάθε κομματιού τις αποθηκεύουμε ως ξεχωριστά αρχεία. Κάνουμε όμως ένα ακόμα βήμα πριν χρησιμοποιηθούν από το νευρωνικό μας δίκτυο.

Κάνουμε μεταφορά όλων των νοτών κατά ένα τυχαίο πραγματικό μέγεθος που μπορεί να είναι από 12 ημιτόνια χαμηλότερα έως 12 ημιτόνια υψηλότερα. Αυτό θεωρήσαμε πως είναι σημαντικό για την γενίκευση του νευρωνικού δικτύου, ώστε να μην εξάγεται ως χαρακτηριστικό μία τονικότητα ενός κομματιού.

Αυτό σε μουσικούς όρους λέγεται αλλαγή τονικότητας ή **μετατροπία (Transposition)**, ενώ για το πείραμά μας σημαίνει μεταφοράshift όλων των στοιχείων του πίνακα pianoroll στον άξονα του τονικού ύψους κατά κάποιες θέσεις. Αυτό σημαίνει ότι είναι πιθανό κάποιες κάποιες τιμές του πίνακα που βρίσκονται στην άκρη του, μπορεί να χαθούν.

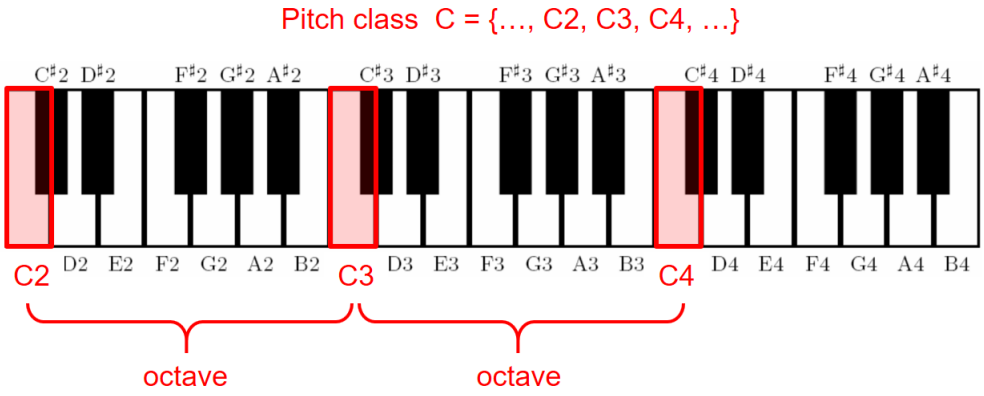
Στο σχήμα 6.14 μπορούμε να δούμε αλλαγή της τονικότητας του κομματιού από Μι ύφεση μείζονα (E flat major), σε Ρε μείζονα (D major) και σε Μι μείζονα (E major). Ουσιαστικά πρόκειται για μεταφορά όλου του κομματιού κατά ένα ημιτόνιο χαμηλότερα στην πρώτη περίπτωση και κατά ένα ημιτόνιο υψηλότερα στην δεύτερη.



Πηγή: https://www.earmaster.com/music-theory-online/ch06/chapter-6-4.html

Σχήμα 5.2: Παράδειγμα αλλαγής κλειδιού

Αποφασίσαμε το εύρος της μετατροπίας να είναι από -12 έως 12 ημιτόνια, ώστε να περιλαμβάνει μετατροπία έως μία οκτάβα υψηλότερα ή χαμηλότερα. Για παράδειγμα εάν είμαστε στην τονικότητα Ντο - C3 (C3 όπως έχει καθοριστέι από το midi 4.3.2) τότε μπορεί να γίνει τυχαία μετατροπία σε εύρος Ντο - C2 και Ντο - C4, όπως φαίνεται στο σχήμα 5.3



Πηγή: https://www.audiolabs-erlangen.de/resources/MIR/FMP/C1/C1S1_MusicalNotesPitches.html

Σχήμα 5.3: Παράδειγμα αλλαγής κλειδιού - 2

5.2.8 Προμήθεια ισορροπημένων δεδομένων εισόδου

Χρησιμοποιούμε γεννήτρια δεδομένων (data generator) για την προμήθεια των δεδομένων στο νευρωνικό μας δίκτυο. Όπως είδαμε στο 5.1, τα σετ δεδομένων είναι μη ισορροπημένα. Η ανισορροπία δεδομένων σημαίνει πως υπάρχει μια άνιση κατανομή κλάσεων σε ένα σύνολο δεδομένων. Για παράδειγμα το σετ δεδομένων TOP-MAGD περιέχει 13 κατηγορίες και 7323 κομμάτια από τα οποία τα 4345 ανήκουν στην κατηγορία Pop Rock. Αυτό σημαίνει πως η κατηγορία αυτή περιέχει το 59 % των δεδομένων.

Αυτό είναι πρόβλημα, διότι το νευρωνικό δίκτυο θα μάθει να αναγνωρίζει περισσότερο αυτήν την συγκεκριμένη κατηγορία, και δεν θα μπορεί να εκπαιδευτεί για τις πιο σπάνιες κλάσσεις.

Μία λύση σε αυτό το πρόβλημα είναι η προσθήκη ενός αλγορίθμου που να διαβάζει δεδομένα, τα οποία ανάλογα με το πλήθος των στοιχείων της κλάσσης στην οποία ανήκουν, θα έχουν αντιστρόφως ανάλογη πιθανότητα επιλογής τους.

Πιο συγκεκριμένα, χρησιμοποιούμε έναν πίνακα που περιέχει πιθανότητες επιλογής κάθε στοιχείου. Εάν για παράδειγμα είχαμε 100 στοιχεία δύο κατηγοριών, με τα 80 στοιχεία να ανήκουν στην A και τα 20 στοιχεία να ανήκουν στην B, κάθε στοιχείο του B θα έχει $80/20 = 4$ φορές μεγαλύτερη πιθανότητα να επιλεχθεί από το νευρωνικό δίκτυο από ότι ένα στοιχείο του A. Αφού όμως το A έχει 4 φορές περισσότερα στοιχεία, τότε η επιλογή των δεδομένων εισόδου θα είναι περίπου ισορροπημένη.

Έτσι, εάν είχαμε τα στοιχεία α1 έως α80 και β1 έως β20 , τότε επιλέγοντας σύμφωνα με τον αλόγριθμό μας η πιθανότητα για να επιλεχθεί ένα συγκεκριμένο στοιχείο α, θα ήταν $p_{ai} = 0.00625$ και η πιθανότητα για να επιλεχθεί ένα συγκεκριμένο στοιχείο β θα ήταν $p_{bj} = 0.025$. Άρα θα είχαμε πιθανότητα $p_a = 0.00625 * 80 = 0.5$ να επιλεχθεί ένα στοιχείο του A, και $p_b = 0.025 * 20 = 0.5$.

5.2.9 Συνδυασμός - συνένωση δεδομένων ελέγχου (combination)

Όπως είπαμε στο 5.2.6 κάθε μουσικό κομμάτι θα χωριστεί σε μικρότερα μέρη (sliced songs).

Έτσι το νευρωνικό δίκτυο θα εκπαιδευτεί και θα ταξινομεί αυτά τα sliced songs. Εμείς όμως θέλουμε να κάνουμε πρόβλεψη για όλο το κομμάτι (whole song). Για αυτό το λόγο, θα συνενώσουμε τα αποτελέσματα και θα βρούμε τον μέσο όρο όλων των πιθανοτήτων των προβλέψεων.

Για παράδειγμα εάν είχαμε ένα ολόκληρο κομμάτι και το χωρίζαμε στα δύο και είχαμε τρεις κλάσεις, τότε θα κάναμε την πρόβλεψη και στα δύο, θα μας επιστρέψει ένας πίνακας πιθανοτήτων με την πιθανότητα κάθε κομμένο κομμάτι να ανήκει σε μία από τις τρεις κατηγορίες. Έστω ότι το πρώτο κομμένο κομμάτι επέστρεφε πίνακα με πιθανότητες $p_a = [0.2, 0.2, 0.6]$ και το δεύτερο κομμένο κομμάτι $p_b = [0.3, 0.4, 0.3]$, τότε θα βρίσκαμε τον μέσο όρο των δύο πιθανοτήτων $p = (p_a + p_b)/2 = [0.25, 0.3, 0.45]$. Έτσι θα προβλέπαμε ότι το ολόκληρο κομμάτι ανήκει στην τρίτη κατηγορία.

Κεφάλαιο 6

Πειραματική Διαδικασία

Θα καθορίσουμε τις αρχιτεκτονικές και τον τρόπο με τον οποίο θα τροφοδοτηθούν τα δεδομένα στα μοντέλα. Θα οριστούν ποια δεδομένα θα χρησιμοποιηθούν και πως θα γίνει η εκπαίδευση τους.

Τα δεδομένα σε ένα πρόβλημα επιβλεπόμενης μάθησης, όπως το πρόβλημα που προσπαθεί να αντιμετωπίσει η παρούσα διπλωματική, εμφανίζονται σε ζευγάρια (X, Υ), όπου το X αναπαριστά τα δεδομένα εισόδου ενώ το Υ την τιμή στόχο (label/target) που προσπαθεί να προβλέψει το μοντέλο. Τα εκάστοτε δεδομένα εισόδου για την σωστή εκπαίδευση και επαλήθευση των μοντέλων χωρίζονται σε τρεις κατηγορίες, τα δεδομένα εκπαίδευσης (training data),τα δεδομένα επαλήθευσης (validation data) και τα δεδομένα ελέγχου (testing data). Τα δεδομένα εκπαίδευσης αποτελούν το μεγαλύτερο μέρος των δεδομένων της εισόδου και όπως είναι προφανές από την ονομασία τους είναι τα δεδομένα που χρησιμοποιούνται για την εκπαίδευση ενός μοντέλου. Τα δεδομένα επαλήθευσης συνήθως συνθέτουν το μικρότερο τμήμα των δεδομένων εισόδου και χρησιμοποιούνται για την επαλήθευση των αποτελεσμάτων σε κάθε στάδιο της εκπαίδευσης. Τέλος, τα δεδομένα ελέγχου χρησιμοποιούνται μετά την ολοκλήρωση της εκπαίδευσης του μοντέλου για την καταγραφή της επίδοσης του.

Δύο ακόμη σημαντικοί παράγοντες για την αποτελεσματική εκπαίδευση ενός μοντέλου είναι οι επιλογή των κατάλληλων τιμών για τον αριθμό των εποχών (epochs) και για το μέγεθος τεμαχίου (batch size). Ο αριθμός των εποχών αναφέρεται στον αριθμό των φορών που θα επαναληφθεί η διαδικασία εκπαίδευσης πάνω στο σύνολο των δεδομένων εκπαίδευσης, μέχρι να επιλεγούν οι τελικές τιμές των παραμέτρων του δικτύου. Το μέγεθος τεμαχίου αναφέρεται στον αριθμό των δεδομένων εισόδου που θα χρησιμοποιηθούν ως τεμάχιο για τον υπολογισμό της κλίσης και τη χρήση του για την ανανέωση των παραμέτρων του δικτύου, όπως περιγράφουν οι αλγόριθμοι Backpropagation και Κατάβασης Κλίσης.

Τελικά, μόλις ολοκληρωθεί η εκπαίδευση των μοντέλων το τελικό βήμα αφορά την αξιολόγηση των εξόδων που αποδίδουν. Για την καταγραφή των αποτελεσμάτων θα χρησιμοποιηθούν ορισμένες από τις μετρικές αξιολόγησης, όπως περιγράφονται στην ενότητα 2.5, επί των δεδομένων ελέγχου. Η πειραματική διαδικασία θα αξιολογηθεί κυρίως βάση των μετρικών Macro Average ή f1-m και λιγότερο ως προς το Accuracy ή Micro Average και το Weighted average, που περιγράφει των αριθμό των σωστών προβλέψεων προς το σύνολο όλων των

προβλέψεων, και Loss, που περιγράφει το σφάλμα της συνάρτησης κόστους εν προκειμένω της Categorical Crossentropy. Για την καλύτερη εποπτεία της επίδοσης στη συνέχεια θα παρουσιαστούν οι τιμές των μετρικών αξιολόγησης.

6.1 Κατασκευή CNN

Αφού κάναμε την προεπεξεργασία των δεδομένων, μπορούμε πλέον να κατασκευάσουμε το συνελικτικό νευρωνικό δίκτυο για να το εκπαιδεύσουμε και να κάνουμε σωστά προβλέψεις. Θα κατασκευάσουμε συνελικτικό νευρωνικό δίκτυο με τα ακόλουθα χαρακτηριστικά:

Ο ένας ή και οι δύο πίνακες των οργάνων είναι η είσοδος του νευρωνικού. Είτε χρησιμοποιούμε μόνο τον πίνακα των μελωδικών οργάνων ή και τον πίνακα των κρουστών οργάνων.

Χρησιμοποιούμε συνεχόμενα συνελικτικά στρώματα μιάς διάστασης (1D) με τις παραμέτρους filters, kernel_size, relu activation, επίπεδο dropout, max-pooling, και πλήρως συνδεδεμένο επίπεδα (fully connected layers) που θεωρείται το Flatten μαζί με το Dense. Τέλος χρησιμοποιούμε ένα Dense με παράμετρο τις κλάσσεις του σετ δεδομένων και μέσω της συνάρτησης ενεργοποίησης softmax φτάνουμε στην έξοδο η οποία θα είναι ένας πίνακας με τις πιθανότητες για κάθε μία από τις n κλάσσεις που θα έχει.

Ο αριθμός (epochs) ήταν μεταβλητός και έγινε εμπειρικά μέχρι να εκπαιδευτει το νευρωνικό ενώ το μέγεθος τεμαχίου (batch size) ήταν 128.

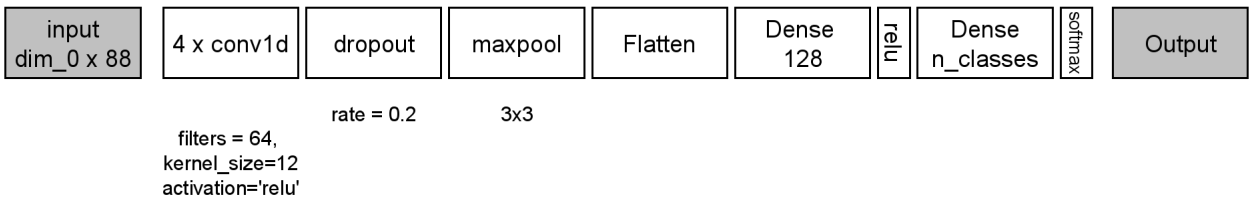
6.1.1 Επιλογή βιβλιοθήκης Tensorflow της python

Για την πειραματική διαδικασία χρησιμοποιήσαμε πολλές βιβλιοθήκες τις γλώσσας προγραμματισμού python. Για την κατασκευή, εκπαίδευση, έλεγχο των νευρωνικών δικτύων χρησιμοποιήσαμε την βιβλιοθήκη **Tensorflow** [1] και συγκεκριμένα την έκδοση r1.15. Το TensorFlow είναι μια βιβλιοθήκη ανοιχτού κώδικα για αριθμητικούς υπολογισμούς και μηχανική εκμάθηση μεγάλης κλίμακας.

6.1.2 Αρχιτεκτονικές νευρωνικών δικτύων

Κάναμε διάφορες δοκιμές με πολλές μεταβλητές τιμές των υπερπαραμέτρων και καταλείξαμε κυριώς στα 4 ακόλουθα νευρωνικά δίκτυα:

6.1.2.1 Νευρωνικό δίκτυο i

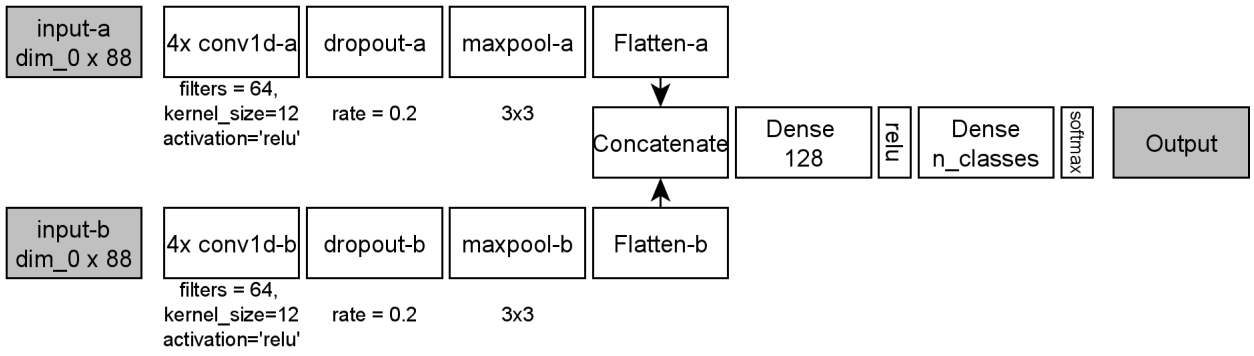


Σχήμα 6.1: Αρχιτεκτονική i νευρωνικού δικτύου. Ο πίνακας των μελωδικών οργάνων είναι η είσοδος. Χρησιμοποιούμε 4 συνεχόμενα συνελικτικά στρώματα με τις παραμέτρους filters=64, kernel_size=12, activation="relu", ύστερα από επίπεδο dropout, max-pooling 3x3, μετά από ένα πλήρως συνδεδεμένο επίπεδο (fully connected layer) που θεωρείται το το Flatten μαζί με το Dense και μετά από Dense με παράμετρο τις κλάσσεις του σετ δεδομένων και μέσω της συνάρτησης ενεργοποίησης softmax φτάνουμε στην έξοδο.

6.1.2.2 Νευρωνικό δίκτυο ii

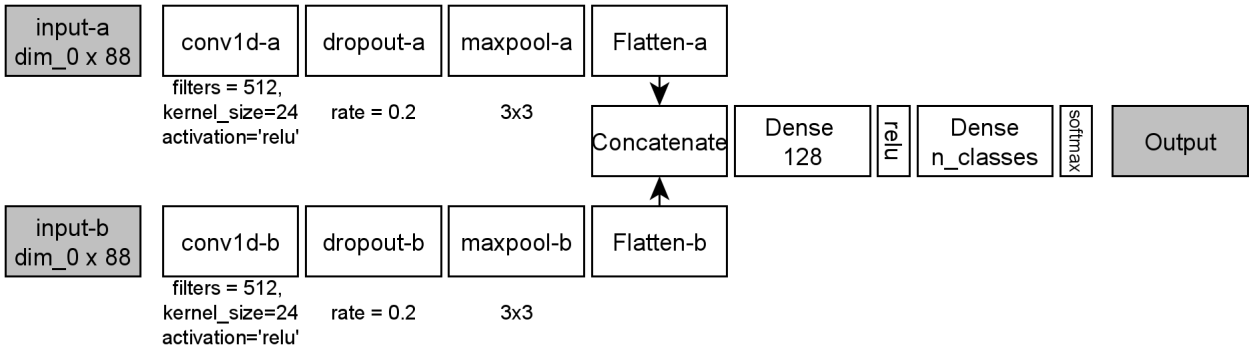
Σχήμα 6.2: Αρχιτεκτονική ii νευρωνικού δικτύου. Παρόμοια αρχιτεκτινική με το νευρωνικό i με τη διαφορά ότι χρησιμοποιούμε ένα συνελικτικό στρώμα με άλλες παραμέτρους.

6.1.2.3 Νευρωνικό δίκτυο iii



Σχήμα 6.3: Αρχιτεκτονική iii νευρωνικού δικτύου. Αρχιτεκτονική παρόμοια με το νευρωνικό i, αλλά πλεον έχουμε δύο εισόδους.Ο πίνακας των μελωδικών οργάνων είναι η είσοδος a και ο πίνακας των κρουστών είναι ο πίνακας b. Τα flatten συνδέονται μαζί στο στάδιο (concatenate).

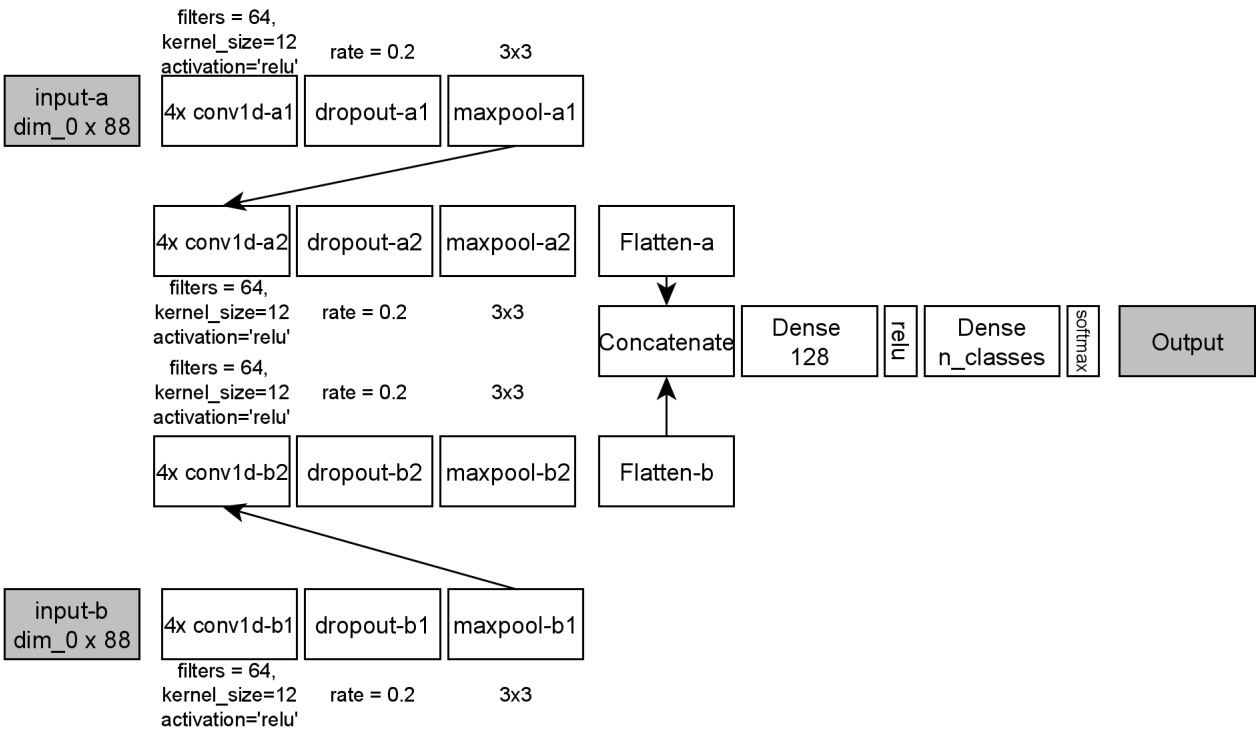
6.1.2.4 Νευρωνικό δίκτυο iv



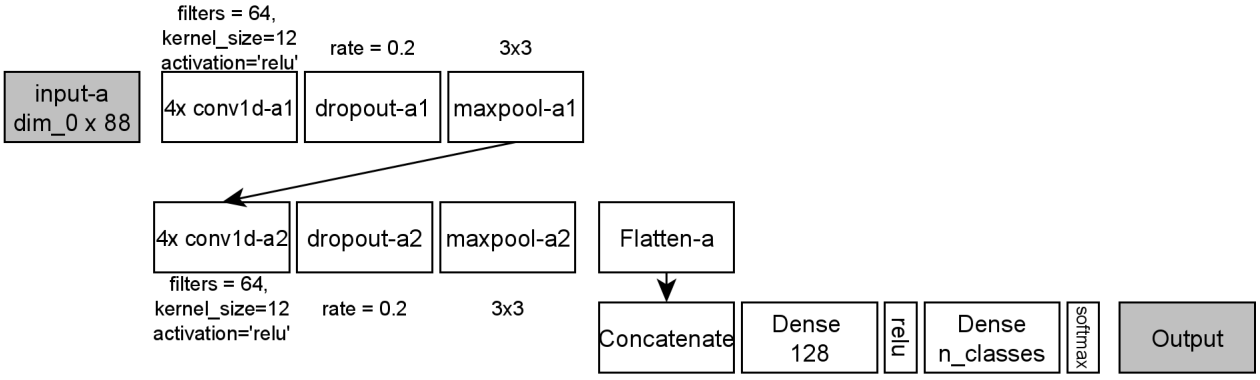
Σχήμα 6.4: Αρχιτεκτονική iv νευρωνικού δικτύου. Παρόμοια αρχιτεκτονική με το iii με τη διαφορά ότι χρησιμοποιούμε ένα ένα συνελικτικό στρώμα για κάθε είσοδο, όπως στην αρχιτεκτονική ii.

6.1.2.5 Άλλα νευρωνικά δίκτυα

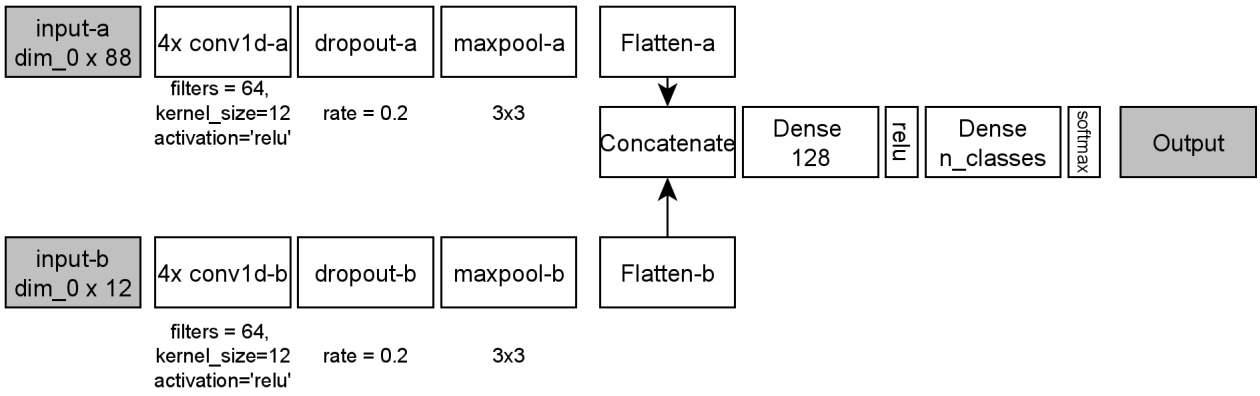
Κάναμε κάποια λίγα πειράματα και στις ακόλουθες τρεις αρχιτεκτονικές νευρωνικού δικτύου. Αλλά, δεν είχαμε σημαντικές διαφοροποιήσεις ή είχαμε μικρές μειώσεις στην επιτυχία πρόβλεψης των αποτελέσματά μας, οπότε κρατήσαμε μόνο τα πρώτα 4 νευρωνικά δίκτυα για τα περισσότερα πειράματα.



Σχήμα 6.5: Αρχιτεκτονική πειραματικού νευρωνικού δικτύου πείραμα - a. Χρησιμοποιούνται δύο φορές στη σειρά η ακολουθία 4x συνελικτικό επίπεδο - dropout - maxpool. Δύο είσοδοι.



Σχήμα 6.6: Αρχιτεκτονική πειραματικού νευρωνικού δικτύου πείραμα - b. Χρησιμοποιούνται δύο φορές στη σειρά η ακολουθία 4x συνελικτικό επίπεδο - dropout - maxpool. Μία είσοδος.

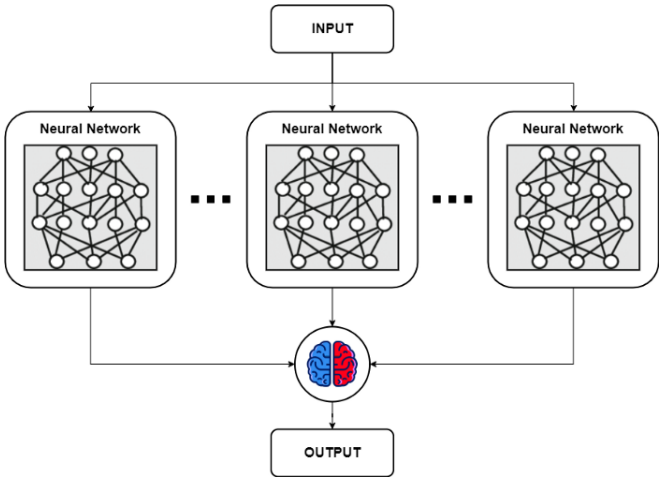


Σχήμα 6.7: Αρχιτεκτονική πειραματικού νευρωνικού δικτύου πείραμα - c. Παρόμοιο με την αρχιτεκτονική του νευρωνικού δικτύου iv, με τη διαφορά πως έχουμε μόνο μια οκτάβα (12 νότες - 12 θέσεις στον πίνακα) για τα κρουστά όργανα. Ουσιαστικά συμπιέσαμε τον πίνακα από 88 διαστάσεις σε 12 αντιστοιχίζοντας όλα τις ίδιες ονομαστικές νοτές μαζί. Π.χ. θεωρούμε όλες τις νότες ντο ταυτόχρονα αντιπροσωπευμένες από μία θέση στον πίνακα. Αυτό το κάναμε για μείωση τον διαστάσεων, θεωρώντας πως δεν χάνουμε σπουδαία πληροφορία.

6.2 Αποτελέσματα

Όπως αναφέραμε και παραπάνω 5.2.6 χρησιμοποιήσαμε πολλές διαφορετικές διαμερίσεις πρώτης διάστασης για να τρέξουμε τα πειράματά μας.

Θα συνδυάσουμε τα αποτελέσματα από τις διαφορετικές διαμερίσεις των νευρωνικών μας δικτύων για ένα συγκεκριμένο σετ δεδομένων για να υπολογίσουμε ένα ακόμη καλύτερο αποτέλεσμα στην πρόβλεψή μας. Επίσης, θα ενώσουμε διαφορετικές αρχιτεκτονικές νευρωνικών δικτύων. Η διαδικασία αυτή λέγεται ensembling.



Πηγή: towardsdatascience.com/neural-networks-ensemble-33f33bea7df3

Σχήμα 6.8: Ensemble νευρωνικών

6.2.1 Αποτελέσματα πειραμάτων για MASD-labels-cleansed

Για το σετ δεδομένων MASD-labels-cleansed (5.1), χρησιμοποιήσαμε τις αρχιτεκτονικές των νευρωνικών δικτύων i, ii και iii Ακόμα, χρησιμοποιήσαμε συνδυασμούς (ensembling) των νευρωνικών για ακόμη καλύτερα αποτελέσματα.

Στο συγκεκριμένο σύνολο δεδομένων βλέπουμε γενικά οι πιο πολυπληθείς κατηγορίες ταξινομούνται καλύτερα από ότι αυτές με λιγότερο πλήθος. Επίσης βλέπουμε πως είναι πιθανότερο ο ταξινομητής να μπερδέψει κατηγορίες που είναι κοντινές, για παράδειγμα τις κατηγορίες Pop Contemporary και Pop Indie.

- **Νευρωνικό i**

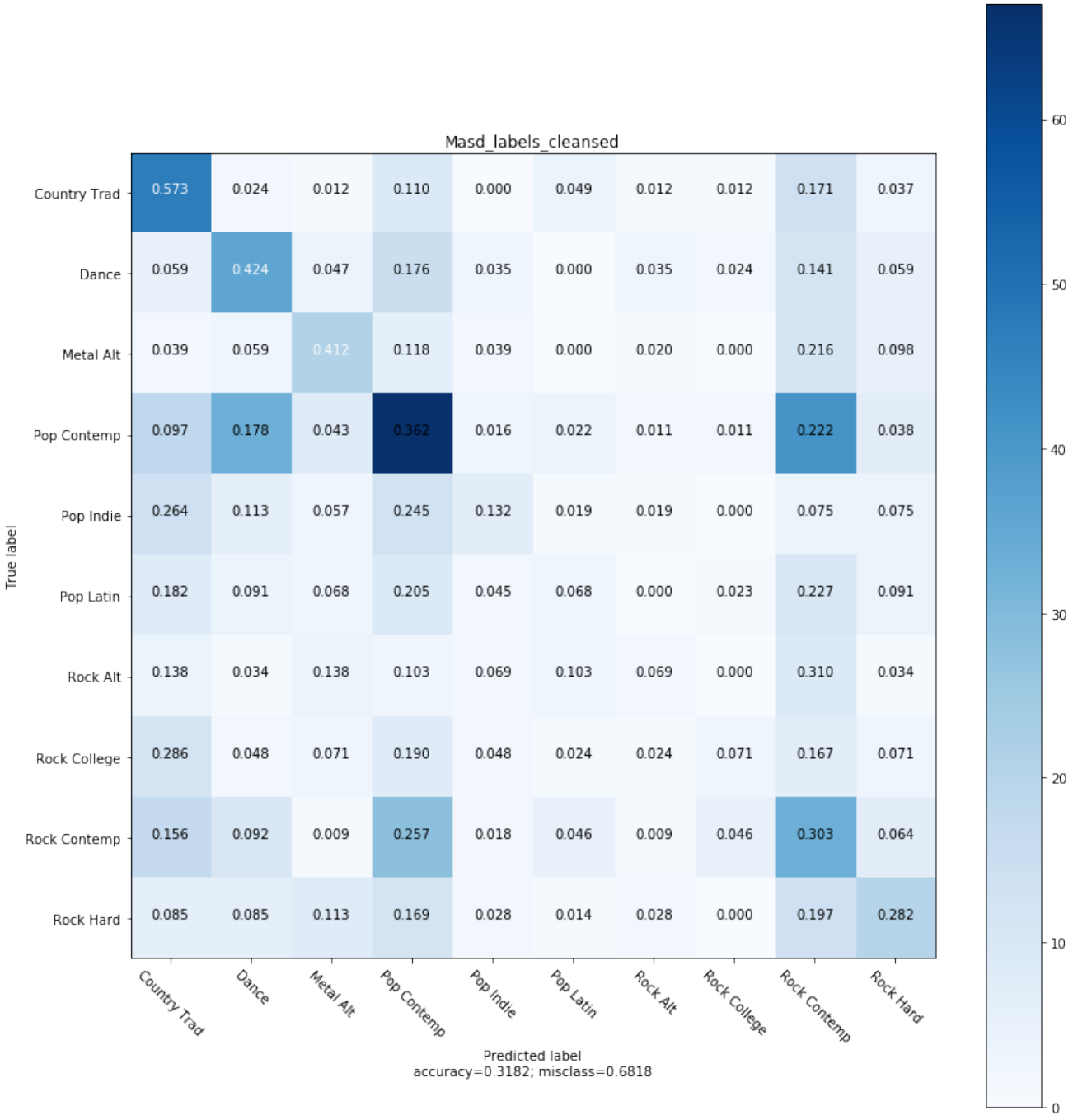
Πίνακας 6.1: νευρωνικό i - MASD-labels-cleansed

	Ποσοστό επιτυχίας για διαφορετική διαμέριση				
Μετρική	128	256	384	512	ensemble
accuracy	0.3146	0.2844	0.2644	0.2905	**0.3182**
f1_m	0.2575	0.2431	0.2274	0.2514	**0.2618**
weighted avg	0.3005	0.2513	0.2491	0.2838	**0.3037**

Πίνακας 6.2: Αναφορά κατηγοριοποίησης νευρωνικού i

Class	Precision	Recall	F-score	Support
Country Traditional	0.3534	0.5732	0.4372	82
Dance	0.3495	0.4235	0.383	85
Metal Alternative	0.375	0.4118	0.3925	51
Pop Contemporary	0.3941	0.3622	0.3775	185
Pop Indie	0.2800	0.1321	0.1795	53
Pop Latin	0.1364	0.0682	0.0909	44
Rock Alternative	0.1429	0.069	0.093	29
Rock College	0.2143	0.0714	0.1071	42
Rock Contemporary	0.2129	0.3028	0.2500	109
Rock Hard	0.3390	0.2817	0.3077	71
accuracy			0.3182	751
macro avg	0.2797	0.2696	0.2618	751
avg	0.3089	0.3182	0.3037	751

Σχήμα 6.9: Πίνακας σύγχυσης νευρωνικού i



- **Νευρωνικό ii**

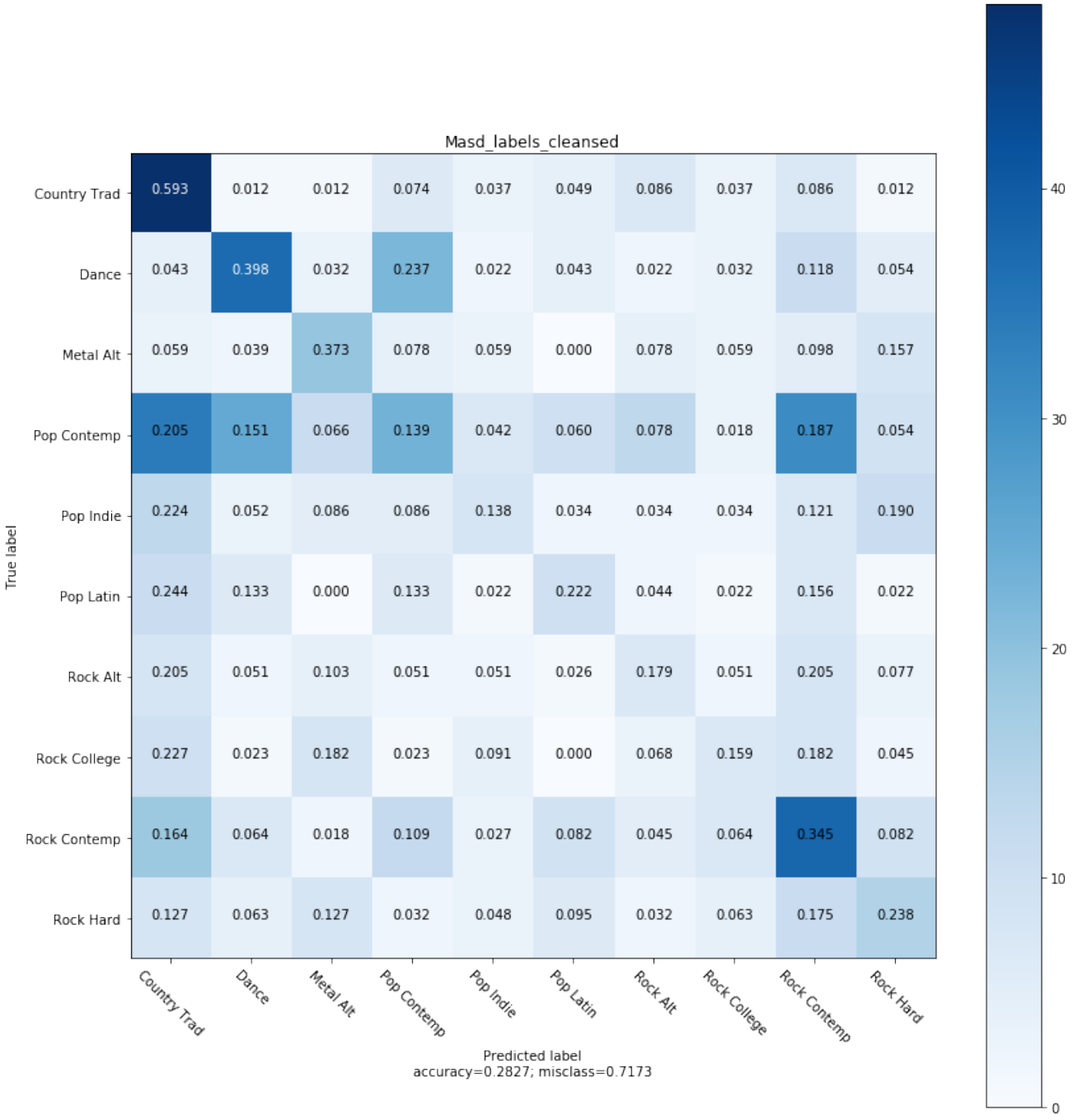
Πίνακας 6.3: νευρωνικό ii - MASD-labels-cleansed

	Ποσοστό επιτυχίας για διαφορετική διαμέριση				
Μετρική	512	576	768	960	ensemble
accuracy	0.2533	0.2610	0.2389	0.2387	**0.2827**
f1_m	0.2348	0.2451	0.2306	0.2302	**0.2615**
weighted avg	0.2501	0.2505	0.2361	0.2310	**0.2691**

Πίνακας 6.4: Αναφορά κατηγοριοποίησης νευρωνικού ii

Class	Precision	Recall	F-score	Support
Country Traditional	0.3057	0.5926	0.4034	81
Dance	0.4205	0.3978	0.4088	93
Metal Alternative	0.3115	0.3725	0.3393	51
Pop Contemporary	0.2771	0.1386	0.1847	166
Pop Indie	0.2222	0.1379	0.1702	58
Pop Latin	0.2174	0.2222	0.2198	45
Rock Alternative	0.1489	0.1795	0.1628	39
Rock College	0.2	0.1591	0.1772	44
Rock Contemporary	0.2857	0.3455	0.3128	110
Rock Hard	0.2344	0.2381	0.2362	63
accuracy			0.2827	750
macro avg	0.2623	0.2784	0.2615	750
avg	0.279	0.2827	0.2691	750

Σχήμα 6.10: Πίνακας σύγχυσης νευρωνικού ii



- **Νευρωνικό iii**

Πίνακας 6.5: νευρωνικό iii - MASD-labels-cleansed

	Ποσοστό επιτυχίας για διαφορετική διαμέριση						
Μετρική	192	384	576	768	960	1152	Ensemble
accuracy	0.2713	0.2726	0.2463	0.3023	0.2925	0.2767	**0.3098**
f1_m	0.2442	0.2444	0.2179	0.2638	0.2528	0.2356	**0.2663**
weighted avg	0.2661	0.2674	0.2426	0.2939	0.2836	0.2662	**0.2965**

Πίνακας 6.6: Αναφορά κατηγοριοποίησης νευρωνικού iii

Class	Precision	Recall	F-score	Support
Country Traditional	0.3171	0.321	0.319	81
Dance	0.4318	0.4043	0.4176	94
Metal Alternative	0.3953	0.3269	0.3579	52
Pop Contemporary	0.2943	0.5	0.3705	166
Pop Indie	0.2727	0.1552	0.1978	58
Pop Latin	0.1818	0.0889	0.1194	45
Rock Alternative	0.2	0.0769	0.1111	39
Rock College	0.3636	0.1818	0.2424	44
Rock Contemporary	0.252	0.2909	0.27	110
Rock Hard	0.3421	0.2063	0.2574	63
accuracy			0.3098	752
macro avg	0.3051	0.2552	0.2663	752
avg	0.3095	0.3098	0.2965	752

Σχήμα 6.11: Πίνακας σύγχυσης νευρωνικού iii

- **Ένωση Νευρωνικών i - ii - iii**

Μπορούμε επίσης να ενώσουμε (ensemble) τα διαφορετικά νευρωνικά i - ii - iii. Μπορούμε να επιλέξουμε όλα τα διαφορετικά νευρωνικά με όλες τις διαφορετικές διαστάσεις και τα ενώνουμε Στο 6.7 βλέπουμε την αναφορά κατηγοριοποίησης και στο σχήμα 6.8 τον πίνακα σύγχυσης.

Πίνακας 6.7: Συνδυασμοί νευρωνικών για MASD-labels-cleansed

	Ποσοστό επιτυχίας για συνδυασμούς νευρωνικών			
Μετρική	Ένωση i - ii	Ένωση i - iii	Ένωση ii - iii	Ένωση i - ii - iii
accuracy	0.3182	0.3098	0.3324	**0.3404**
f1_m	0.2618	0.2663	0.2903	**0.2999**
weighted avg	0.3037	0.2965	0.3207	**0.3284**

Πίνακας 6.8: Αναφορά κατηγοριοποίησης συνδυασμού νευρωνικών i -ii - iii

Class	Precision	Recall	F-score	Support
Country Traditional	0.3431	0.4321	0.3825	81
Dance	0.4362	0.4362	0.4362	94
Metal Alternative	0.3846	0.3846	0.3846	52
Pop Contemporary	0.3226	0.4819	0.3865	166
Pop Indie	0.2903	0.1552	0.2022	58
Pop Latin	0.2857	0.1333	0.1818	45
Rock Alternative	0.1765	0.0769	0.1071	39
Rock College	0.4	0.1818	0.25	44
Rock Contemporary	0.2742	0.3091	0.2906	110
Rock Hard	0.4651	0.3175	0.3774	63
accuracy			0.3404	752
macro avg	0.3378	0.2909	0.2999	752
avg	0.3404	0.3404	0.3284	752

Σχήμα 6.12: Πίνακας σύγχυσης ένωσης i - ii - iii

6.2.2 Αποτελέσματα πειραμάτων για MASD

Για το σετ δεδομένων MASD (5.1), χρησιμοποιήσαμε τις αρχιτεκτονικές των νευρωνικών δικτύων i, ii και iii Ακόμα, χρησιμοποιήσαμε συνδυασμούς (ensembling) των νευρωνικών για ακόμη καλύτερα αποτελέσματα.

Στο συγκεκριμένο σύνολο δεδομένων βλέπουμε γενικά οι πιο πολυπληθείς κατηγορίες ταξινομούνται καλύτερα από ότι αυτές με λιγότερο πλήθος. Σε κάποιες κατηγορίες με μικρό πλήθος βλέπουμε πως ο ταξινομητής μας δεν κάνει καθόλου καλή ταξινόμηση. Βλέπουμε πως συνήθως το μεγαλύτερο πλήθος αυτών των στοιχείων ταξινομούνται σε κοντινές κατηγορίες ή τις πιο δημοφιλείς κατηγορίες. Για παράδειγμα, βλέπουμε στον πίνακα σύγχυσης του νευρωνικού i (6.13) πως ένα μεγάλο μέρος κατηγορία Rock Alt κατηγοριοποιείται ως Rock Contemporary, που είναι αρκετά κοντινή κατηγορία αλλά και ένα άλλο μεγάλο μέρος της ως Pop Contemporary, που είναι λιγότερο σχετικό αλλά πολυπληθής κατηγορία. Μπορεί να θεωρούμε ότι το φαινόμενο της πρώτης περίπτωσης να μην είναι τόσο σημαντικό όσο ή λάθος πρόβλεψη στην δεύτερη περίπτωση.

Βλέπουμε όμως πως ο ταξινομητής εν τέλει μπορεί να πετύχει accuracy ως 0.246 και f1_m έως 0.18. Δεν μπορεί δηλαδή να πετύχει υψηλό ποσοστό για το MASD με τις 25 διαφορετικές κατηγορίες.

- **Νευρωνικό i**

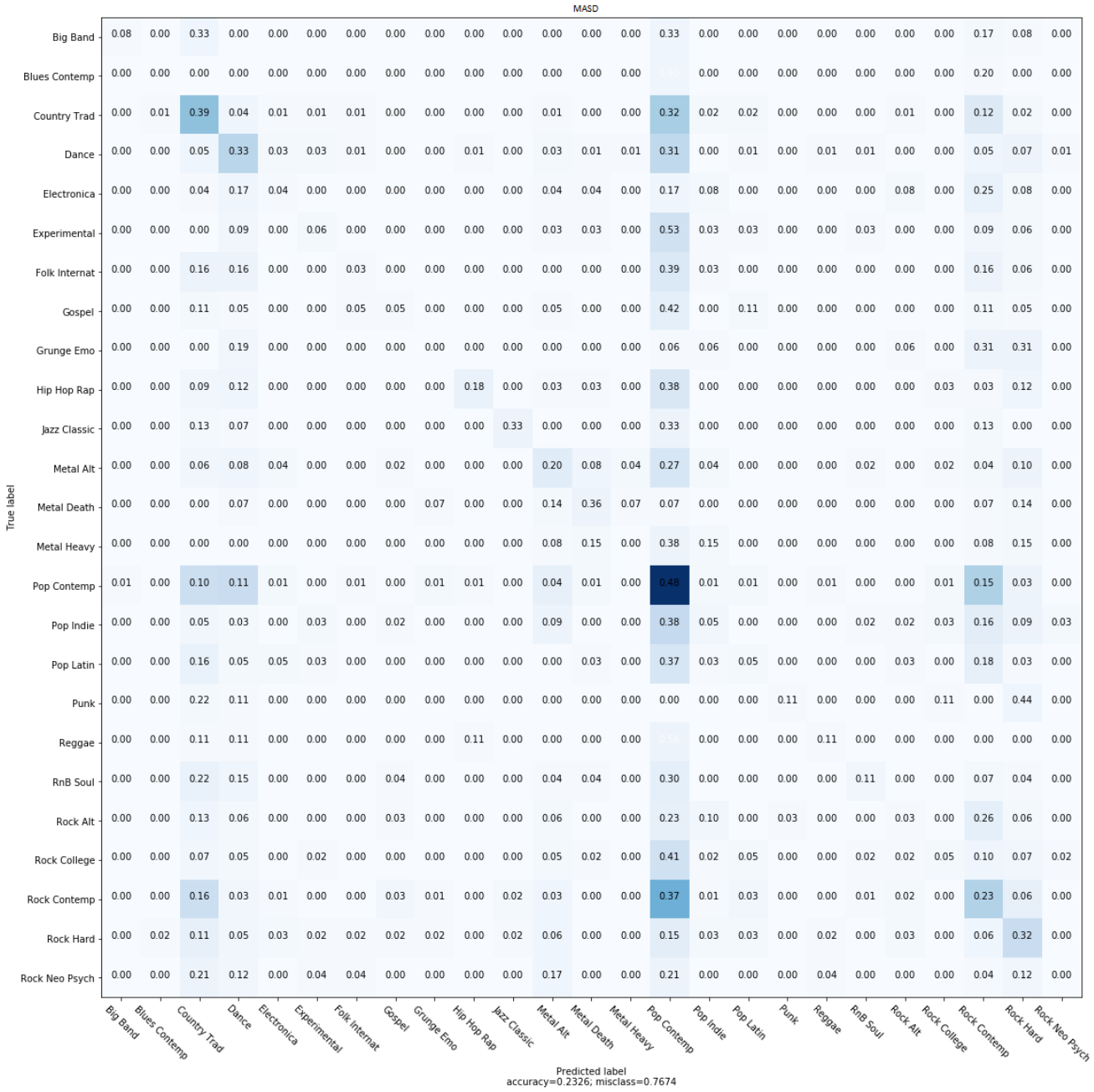
Πίνακας 6.9: νευρωνικό i - MASD

	Ποσοστό επιτυχίας για διαφορετική διαμέριση					
Μετρική	256	512	768	1024	1536	Ensembled
accuracy	0.1424	0.2068	0.1877	0.2000	0.1796	**0.2326**
macro avg	0.1299	0.1417	0.1301	0.1298	0.1256	**0.1520**
weighted avg	0.1256	0.1901	0.1729	0.1811	0.1697	**0.2036**

Πίνακας 6.10: Αναφορά κατηγοριοποίησης MASD - νευρωνικό i

Class	Precision	Recall	F-score	Support
Big Band	0.5	0.0833	0.1429	12
Blues Contemporary	0.0	0.0	0.0	5
Country Traditional	0.2667	0.3871	0.3158	93
Dance	0.2551	0.3333	0.289	75
Electronica	0.0769	0.0417	0.0541	24
Experimental	0.1818	0.0625	0.093	32
Folk International	0.1429	0.0323	0.0526	31
Gospel	0.1111	0.0526	0.0714	19
Grunge Emo	0.0	0.0	0.0	16
Hip Hop Rap	0.6667	0.1765	0.2791	34
Jazz Classic	0.625	0.3333	0.4348	15
Metal Alternative	0.2041	0.1961	0.2	51
Metal Death	0.25	0.3571	0.2941	14
Metal Heavy	0.0	0.0	0.0	13
Pop Contemporary	0.243	0.478	0.3222	182
Pop Indie	0.125	0.0517	0.0732	58
Pop Latin	0.1176	0.0526	0.0727	38
Punk	0.5	0.1111	0.1818	9
Reggae	0.2	0.1111	0.1429	9
RnB Soul	0.3333	0.1111	0.1667	27
Rock Alternative	0.0833	0.0323	0.0465	31
Rock College	0.2222	0.0488	0.08	41
Rock Contemporary	0.1985	0.2308	0.2134	117
Rock Hard	0.2381	0.3226	0.274	62
Rock Neo Psychedelia	0.0	0.0	0.0	24
accuracy			0.2326	1032
macro avg	0.2217	0.1442	0.152	1032
avg	0.2239	0.2326	0.2036	1032

Σχήμα 6.13: Πίνακας σύγχυσης νευρωνικού i



- **Νευρωνικό iii**

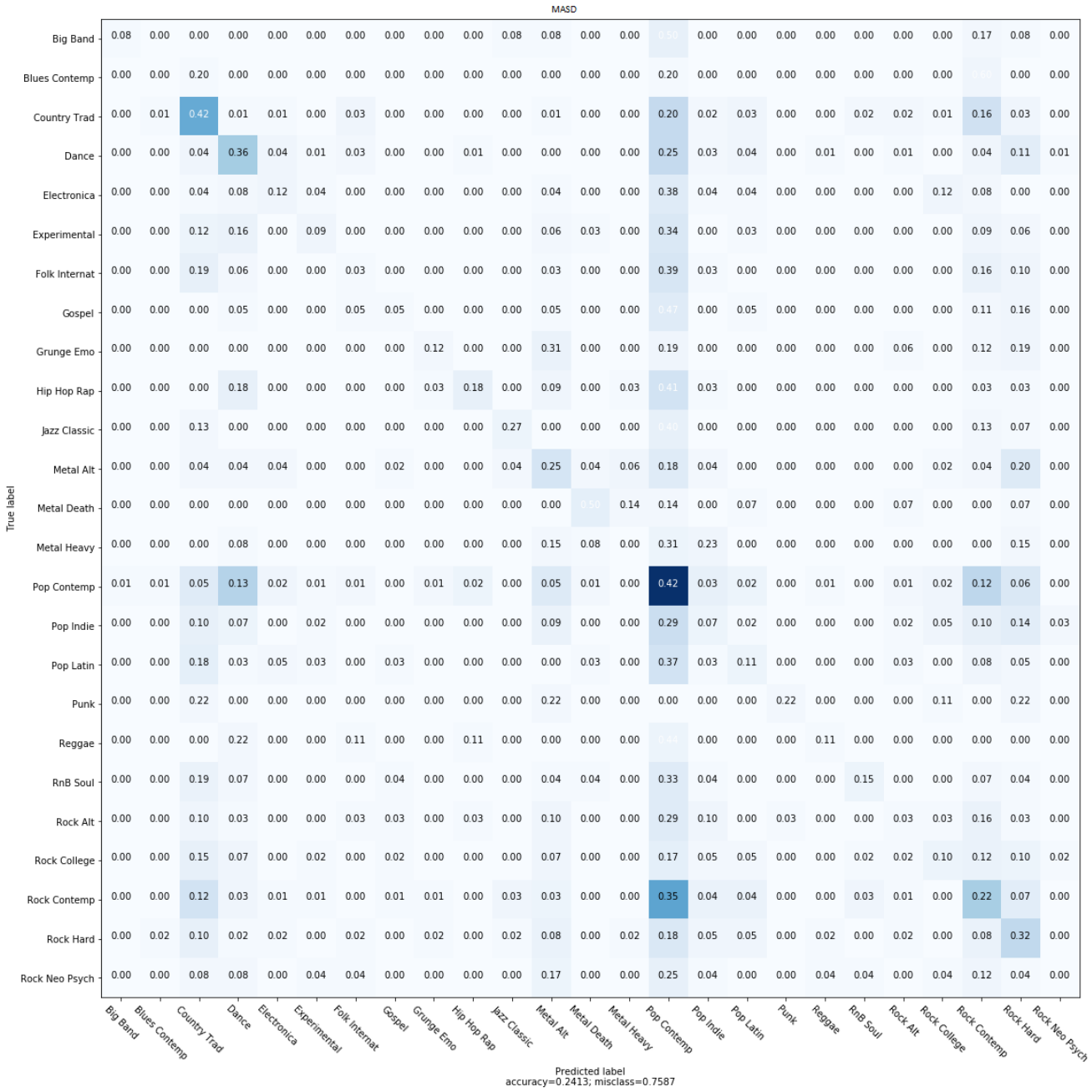
Πίνακας 6.11: νευρωνικό iii - MASD

	Ποσοστό επιτυχίας για διαφορετική διαμέριση										
Μετρική	192	384	576	768	960	1152	1536	256	512	1024	ensembled
accuracy	0.1950	0.1814	0.1823	0.2043	0.2016	0.2186	0.1775	0.2141	0.2175	0.2179	**0.2413**
macro avg	0.1626	0.1543	0.1417	0.1580	0.1604	0.1540	0.1320	0.1625	0.1602	0.1697	**0.1817**
weighted avg	0.1950	0.1814	0.1823	0.2043	0.2016	0.2186	0.1775	0.2141	0.2175	0.2179	**0.2219**

Πίνακας 6.12: Αναφορά κατηγοριοποίησης MASD - νευρωνικό iii

Class	Precision	Recall	F-score	Support
Big Band	0.5	0.0833	0.1429	12
Blues Contemporary	0.0	0.0	0.0	5
Country Traditional	0.3305	0.4194	0.3697	93
Dance	0.3034	0.36	0.3293	75
Electronica	0.1875	0.125	0.15	24
Experimental	0.2727	0.0938	0.1395	32
Folk International	0.0769	0.0323	0.0455	31
Gospel	0.1429	0.0526	0.0769	19
Grunge Emo	0.2857	0.125	0.1739	16
Hip Hop Rap	0.4615	0.1765	0.2553	34
Jazz Classic	0.3636	0.2667	0.3077	15
Metal Alternative	0.194	0.2549	0.2203	51
Metal Death	0.4667	0.5	0.4828	14
Metal Heavy	0.0	0.0	0.0	13
Pop Contemporary	0.239	0.4176	0.304	182
Pop Indie	0.1053	0.069	0.0833	58
Pop Latin	0.1379	0.1053	0.1194	38
Punk	0.6667	0.2222	0.3333	9
Reggae	0.2	0.1111	0.1429	9
RnB Soul	0.3636	0.1481	0.2105	27
Rock Alternative	0.0769	0.0323	0.0455	31
Rock College	0.2222	0.0976	0.1356	41
Rock Contemporary	0.2203	0.2222	0.2213	117
Rock Hard	0.2083	0.3226	0.2532	62
Rock Neo Psychedelia	0.0	0.0	0.0	24
accuracy			0.2413	1032
macro avg	0.241	0.1695	0.1817	1032
avg	0.2351	0.2413	0.2219	1032

Σχήμα 6.14: Πίνακας σύγχυσης νευρωνικού iii



- **Νευρωνικό iv**

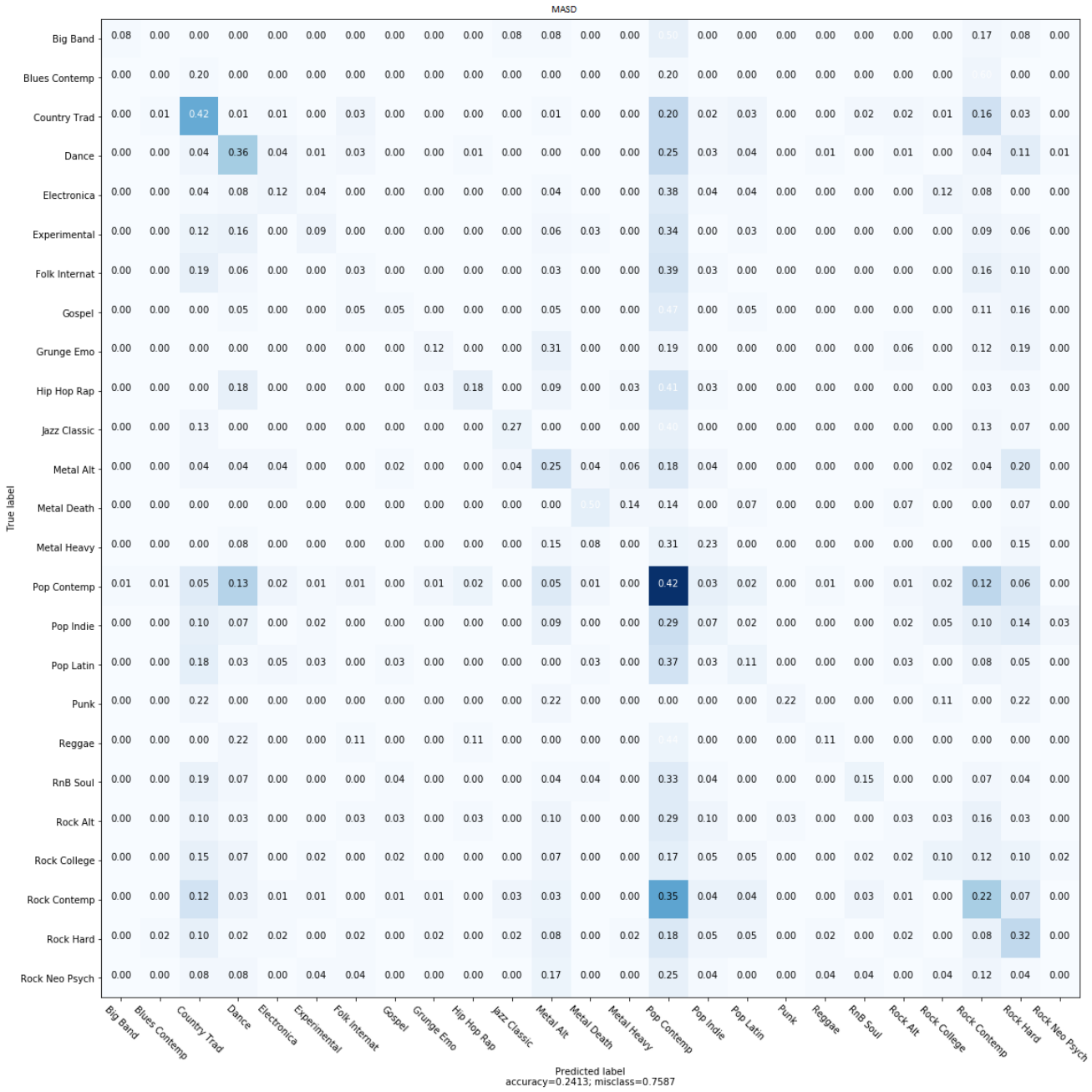
Πίνακας 6.13: νευρωνικό iv - MASD

	Ποσοστό επιτυχίας για διαφορετική διαμέριση			
Μετρικές	384	768	960	Ensembled
accuracy	0.2163	0.1814	0.1986	**0.2134**
macro avg	0.1668	0.1543	0.1628	**0.1720**
weighted avg	0.2077	0.1764	0.1920	**0.2013**

Πίνακας 6.14: Αναφορά κατηγοριοποίησης MASD - νευρωνικό iv

Class	Precision	Recall	F-score	Support
Big Band	0.3333	0.0833	0.1333	12
Blues Contemporary	0.0	0.0	0.0	5
Country Traditional	0.2832	0.3441	0.3107	93
Dance	0.2874	0.3333	0.3086	75
Electronica	0.125	0.125	0.125	24
Experimental	0.1053	0.0625	0.0784	32
Folk International	0.0833	0.0333	0.0476	30
Gospel	0.1429	0.0526	0.0769	19
Grunge Emo	0.1429	0.0625	0.087	16
Hip Hop Rap	0.4167	0.1471	0.2174	34
Jazz Classic	0.5455	0.4	0.4615	15
Metal Alternative	0.1714	0.2353	0.1983	51
Metal Death	0.4667	0.5	0.4828	14
Metal Heavy	0.0	0.0	0.0	13
Pop Contemporary	0.2243	0.3352	0.2687	182
Pop Indie	0.1321	0.1207	0.1261	58
Pop Latin	0.1026	0.1053	0.1039	38
Punk	0.5	0.2222	0.3077	9
Reggae	0.2	0.1111	0.1429	9
RnB Soul	0.2381	0.1852	0.2083	27
Rock Alternative	0.1053	0.0645	0.08	31
Rock College	0.1481	0.0976	0.1176	41
Rock Contemporary	0.2021	0.1624	0.1801	117
Rock Hard	0.1939	0.3065	0.2375	62
Rock Neo Psychedelia	0.0	0.0	0.0	24
accuracy			0.2134	1031
macro avg	0.206	0.1636	0.172	1031
avg	0.2071	0.2134	0.2013	1031

Σχήμα 6.15: Πίνακας σύγχυσης νευρωνικού iv



- **Ένωση Νευρωνικών i - iii - iv**

Μπορούμε επίσης να ενώσουμε (ensemble) τα διαφορετικά νευρωνικά i - iii - iv. Μπορούμε να επιλέξουμε όλα τα διαφορετικά νευρωνικά με όλες τις διαφορετικές διαστάσεις και τα ενώνουμε Στο 6.15 βλέπουμε την αναφορά κατηγοριοποίησης και στο σχήμα 6.16 τον πίνακα σύγχυσης.

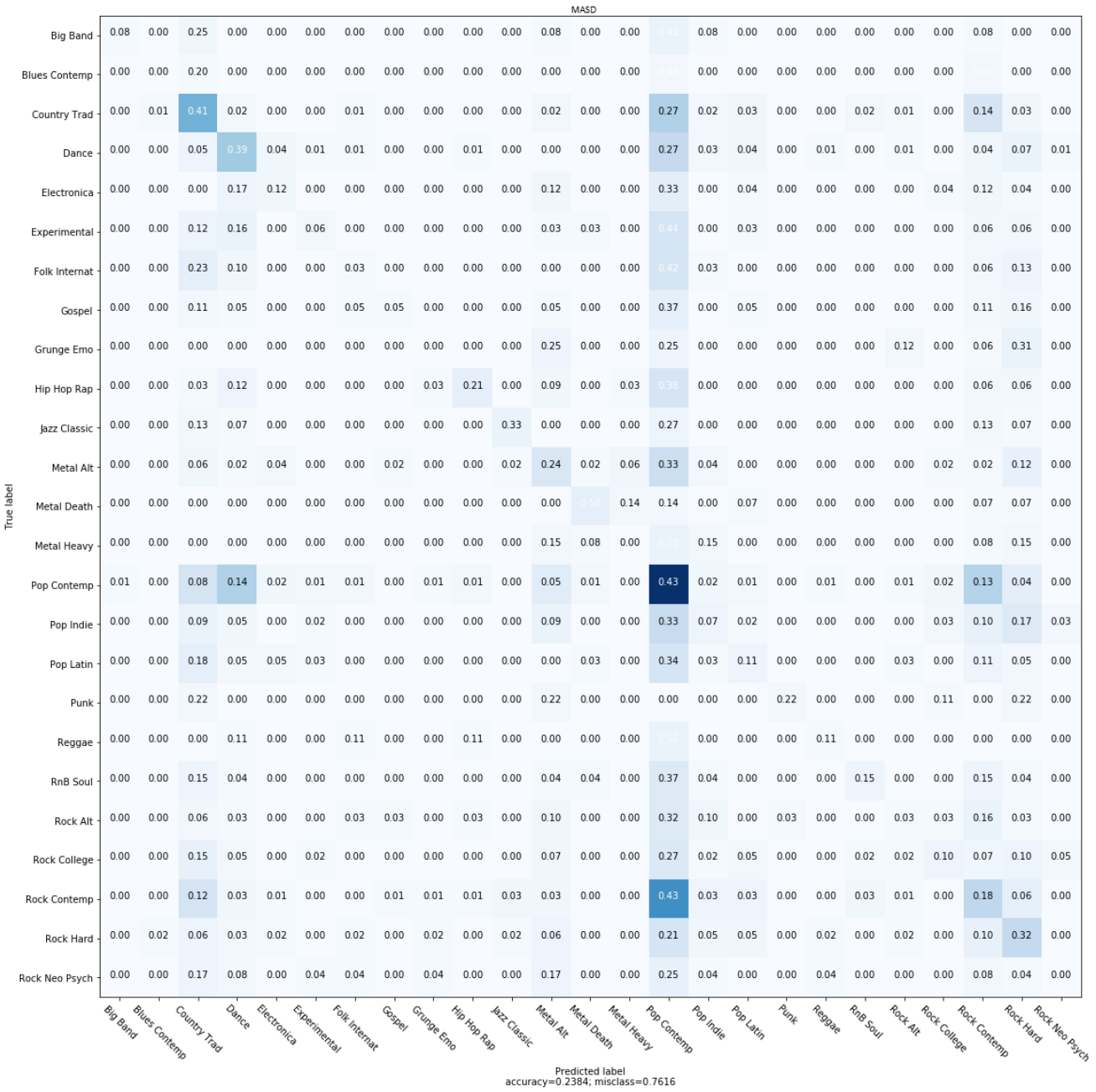
Πίνακας 6.15: Συνδυασμοί νευρωνικών για MASD

	Ποσοστό επιτυχίας για συνδυασμούς νευρωνικών			
Μετρική	Ένωση i - iii	Ένωση i - iv	Ένωση iii - iv	Ένωση i - iii - iv
accuracy	**0.2461**	0.2452	0.2384	0.2384
f1_m	0.1799	0.1708	**0.1821**	0.1790
weighted avg	**0.2204**	0.2183	0.2195	0.2158

Πίνακας 6.16: Αναφορά κατηγοριοποίησης συνδυασμού νευρωνικών

Class	Precision	Recall	F-score	Support
Big Band	0.5	0.0833	0.1429	12
Blues Contemporary	0.0	0.0	0.0	5
Country Traditional	0.2992	0.4086	0.3455	93
Dance	0.3152	0.3867	0.3473	75
Electronica	0.2	0.125	0.1538	24
Experimental	0.2222	0.0625	0.0976	32
Folk International	0.1	0.0323	0.0488	31
Gospel	0.25	0.0526	0.087	19
Grunge Emo	0.0	0.0	0.0	16
Hip Hop Rap	0.5385	0.2059	0.2979	34
Jazz Classic	0.5	0.3333	0.4	15
Metal Alternative	0.1905	0.2353	0.2105	51
Metal Death	0.5	0.5	0.5	14
Metal Heavy	0.0	0.0	0.0	13
Pop Contemporary	0.2225	0.4341	0.2942	182
Pop Indie	0.129	0.069	0.0899	58
Pop Latin	0.1538	0.1053	0.125	38
Punk	0.6667	0.2222	0.3333	9
Reggae	0.2	0.1111	0.1429	9
RnB Soul	0.4	0.1481	0.2162	27
Rock Alternative	0.1	0.0323	0.0488	31
Rock College	0.3077	0.0976	0.1481	41
Rock Contemporary	0.1892	0.1795	0.1842	117
Rock Hard	0.2198	0.3226	0.2614	62
Rock Neo Psychedelia	0.0	0.0	0.0	24
accuracy			0.2384	1032
macro avg	0.2482	0.1659	0.179	1032
avg	0.2361	0.2384	0.2158	1032

Σχήμα 6.16: Πίνακας σύγχυσης ένωσης i - iii - iv



6.2.3 Αποτελέσματα πειραμάτων για Top-MAGD

Για το σετ δεδομένων Top-MAGD (5.1), χρησιμοποιήσαμε τις αρχιτεκτονικές των νευρωνικών δικτύων i, iv Ακόμα, χρησιμοποιήσαμε συνδυασμούς (ensembling) των νευρωνικών για ακόμη καλύτερα αποτελέσματα.

Στο συγκεκριμένο σύνολο δεδομένων βλέπουμε γενικά οι πιο πολυπληθείς κατηγορίες ταξινομούνται πολύ καλά ενώ αυτές με λιγότερο πλήθος καθόλου καλά.

Αυτό συμβαίνει κυρίως επειδή το συγκεκριμένο σύνολο δεδομένων είναι πολύ ανισόρροπο.

Βλέπουμε όμως πως ο ταξινομητής εν τέλει μπορεί να πετύχει accuracy ως 0.595 και f1_m έως 0.195. Δεν μπορεί δηλαδή να πετύχει υψηλό ποσοστό f1.

- **Νευρωνικό i**

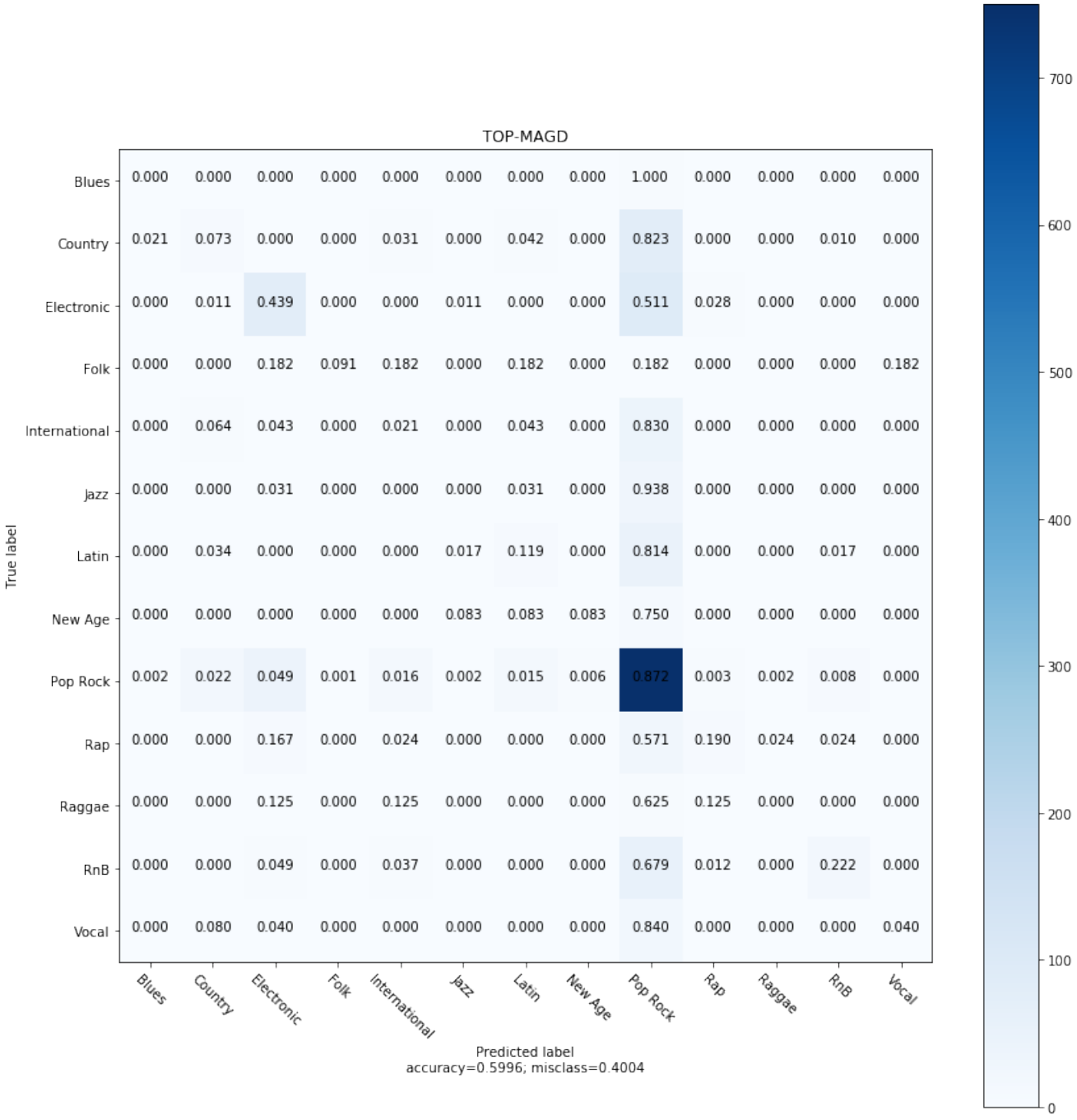
Πίνακας 6.17: νευρωνικό i

	Ποσοστό επιτυχίας για διαφορετική διαμέριση			
Μετρική	512	1024	768	Ensembled
accuracy	0.5879	0.5843	0.3482	**0.5996**
macro	0.1927	0.1752	0.1612	**0.1896**
weighted avg	0.5330	0.5200	0.3656	**0.5442**

Πίνακας 6.18: Αναφορά κατηγοριοποίησης TOP-MAGD - νευρωνικό i

Class	Precision	Recall	F-score	Support
Blues	0.0	0.0	0.0	3
Country	0.2	0.0729	0.1069	96
Electronic	0.5683	0.4389	0.4953	180
Folk	0.5	0.0909	0.1538	11
International	0.04	0.0213	0.0278	47
Jazz	0.0	0.0	0.0	32
Latin	0.2333	0.1186	0.1573	59
New Age	0.1667	0.0833	0.1111	12
Pop Rock	0.6482	0.8721	0.7437	860
Rap	0.4444	0.1905	0.2667	42
Raggae	0.0	0.0	0.0	8
RnB	0.6429	0.2222	0.3303	81
Vocal	0.3333	0.04	0.0714	25
accuracy			0.5996	1456
macro avg	0.2906	0.1654	0.1896	1456
avg	0.5365	0.5996	0.5442	1456

Σχήμα 6.17: Πίνακας σύγχυσης νευρωνικού i



- **Νευρωνικό iv**

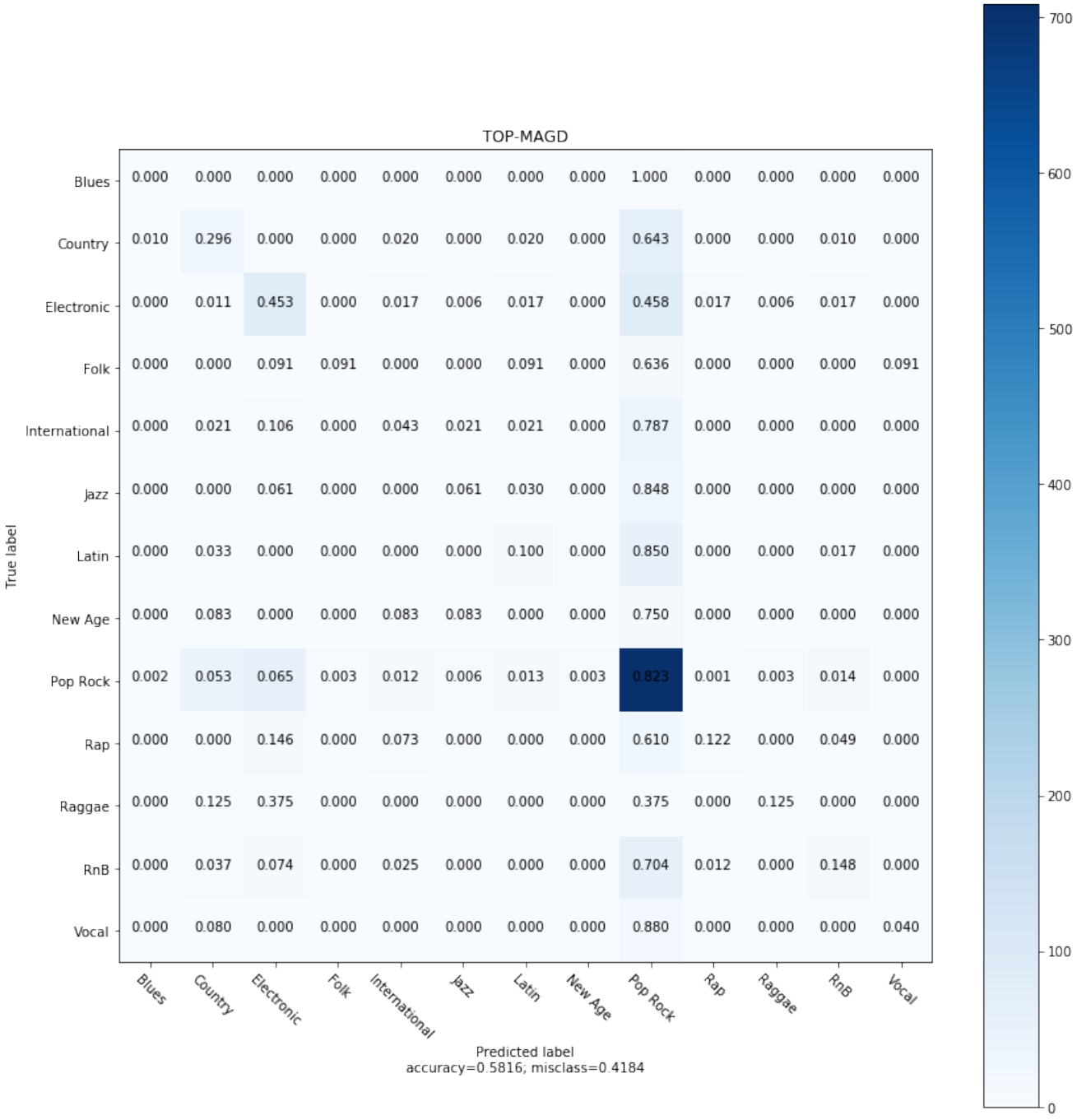
Πίνακας 6.19: νευρωνικό iv

	Ποσοστό επιτυχίας για διαφορετική διαμέριση								
Μετρική	256	512	1152	192	384	576	768	960	Ensembled
accuracy	0.5556	0.5735	0.4961	0.4870	0.4499	0.4831	0.4856	0.4856	**0.5816**
macro avg	**0.2066**	0.1971	0.1872	0.1754	0.1771	0.1916	0.1893	0.1893	0.1984
weighted avg	0.5314	0.5322	0.4866	0.4748	0.4598	0.4808	0.4825	0.4825	**0.5373**

Πίνακας 6.20: Αναφορά κατηγοριοποίησης TOP-MAGD - νευρωνικό iv

Class	Precision	Recall	F-score	Support
Blues	0.0	0.0	0.0	3
Country	0.3333	0.2959	0.3135	98
Electronic	0.5062	0.4525	0.4779	179
Folk	0.25	0.0909	0.1333	11
International	0.087	0.0426	0.0571	47
Jazz	0.2	0.0606	0.093	33
Latin	0.24	0.1	0.1412	60
New Age	0.0	0.0	0.0	12
Pop Rock	0.6466	0.8233	0.7243	860
Rap	0.5	0.122	0.1961	41
Raggae	0.2	0.125	0.1538	8
RnB	0.3871	0.1481	0.2143	81
Vocal	0.5	0.04	0.0741	25
accuracy			0.5816	1458
macro avg	0.2962	0.177	0.1984	1458
avg	0.5303	0.5816	0.5373	1458

Σχήμα 6.18: Πίνακας σύγχυσης νευρωνικού iv



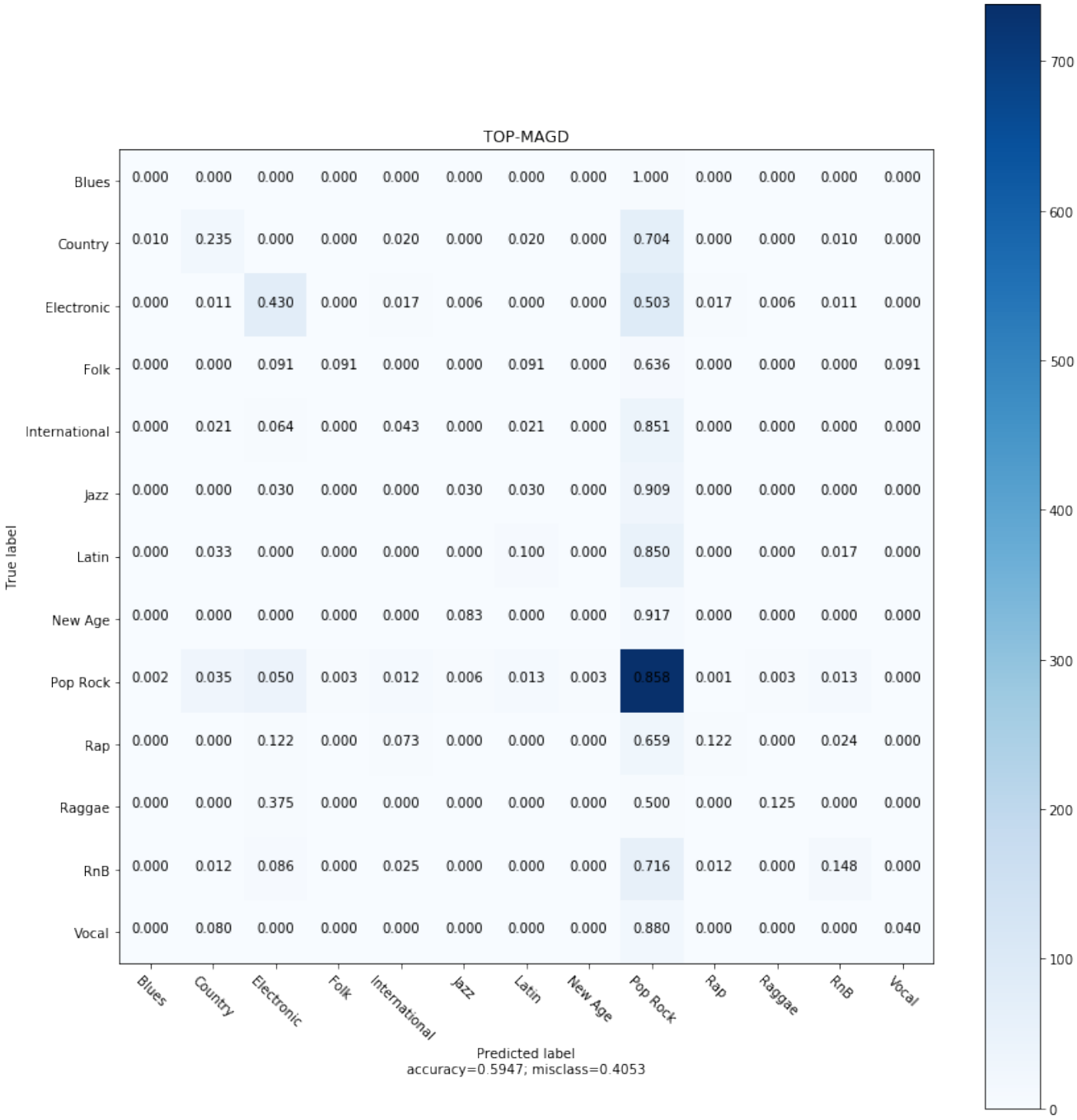
- **Ένωση Νευρωνικών i - iv**

Μπορούμε επίσης να ενώσουμε (ensemble) τα διαφορετικά νευρωνικά i - iv. Μπορούμε να επιλέξουμε όλα τα διαφορετικά νευρωνικά με όλες τις διαφορετικές διαστάσεις και τα ενώνουμε Στο 6.21 βλέπουμε την αναφορά κατηγοριοποίησης και στο σχήμα 6.21 τον πίνακα σύγχυσης.

Πίνακας 6.21: Αναφορά κατηγοριοποίησης συνδυασμού νευρωνικών

Class	Precision	Recall	F-score	Support
Blues	0.0	0.0	0.0	3
Country	0.377	0.2347	0.2893	98
Electronic	0.55	0.4302	0.4828	179
Folk	0.25	0.0909	0.1333	11
International	0.0909	0.0426	0.058	47
Jazz	0.125	0.0303	0.0488	33
Latin	0.2727	0.1	0.1463	60
New Age	0.0	0.0	0.0	12
Pop Rock	0.6417	0.8581	0.7343	860
Rap	0.5	0.122	0.1961	41
Raggae	0.2	0.125	0.1538	8
RnB	0.4286	0.1481	0.2202	81
Vocal	0.5	0.04	0.0741	25
accuracy			0.5947	1458
macro avg	0.3028	0.1709	0.1952	1458
avg	0.5378	0.5947	0.5417	1458

Σχήμα 6.19: Πίνακας σύγχυσης ένωσης i - iv



6.2.4 Αποτελέσματα πειραμάτων για Tagtraum

Για το σετ δεδομένων Tagtraum (5.1), χρησιμοποιήσαμε τις αρχιτεκτονικές των νευρωνικών δικτύων i, iii Ακόμα, χρησιμοποιήσαμε συνδυασμούς (ensembling) των νευρωνικών για ακόμη καλύτερα αποτελέσματα.

Στο συγκεκριμένο σύνολο δεδομένων βλέπουμε γενικά οι πιο πολυπληθείς κατηγορίες ταξινομούνται καλά ενώ αυτές με λιγότερο πλήθος καθόλου καλά.

Το μεγαλύτερο πρόβλημα σε αυτό το σετ δεδομένων είναι ότι κάποιες κατηγορίες έχουν πολύ μικρό πλήθος. Έτσι δεν μπορεί το νευρωνικό δίκτυο να τις κατηγοροποιήσει κατάλληλα.

Βλέπουμε όμως πως ο ταξινομητής εν τέλει μπορεί να πετύχει accuracy ως 0.505 και f1_m έως 0.2381. Δεν μπορεί δηλαδή να πετύχει υψηλό ποσοστό f1, διότι οι κατηγορίες με το μικρό πλήθος και την κακή επίδοση στην πρόβλεψη μειώνουν πολύ την τιμή αυτή.

- **Νευρωνικό i**

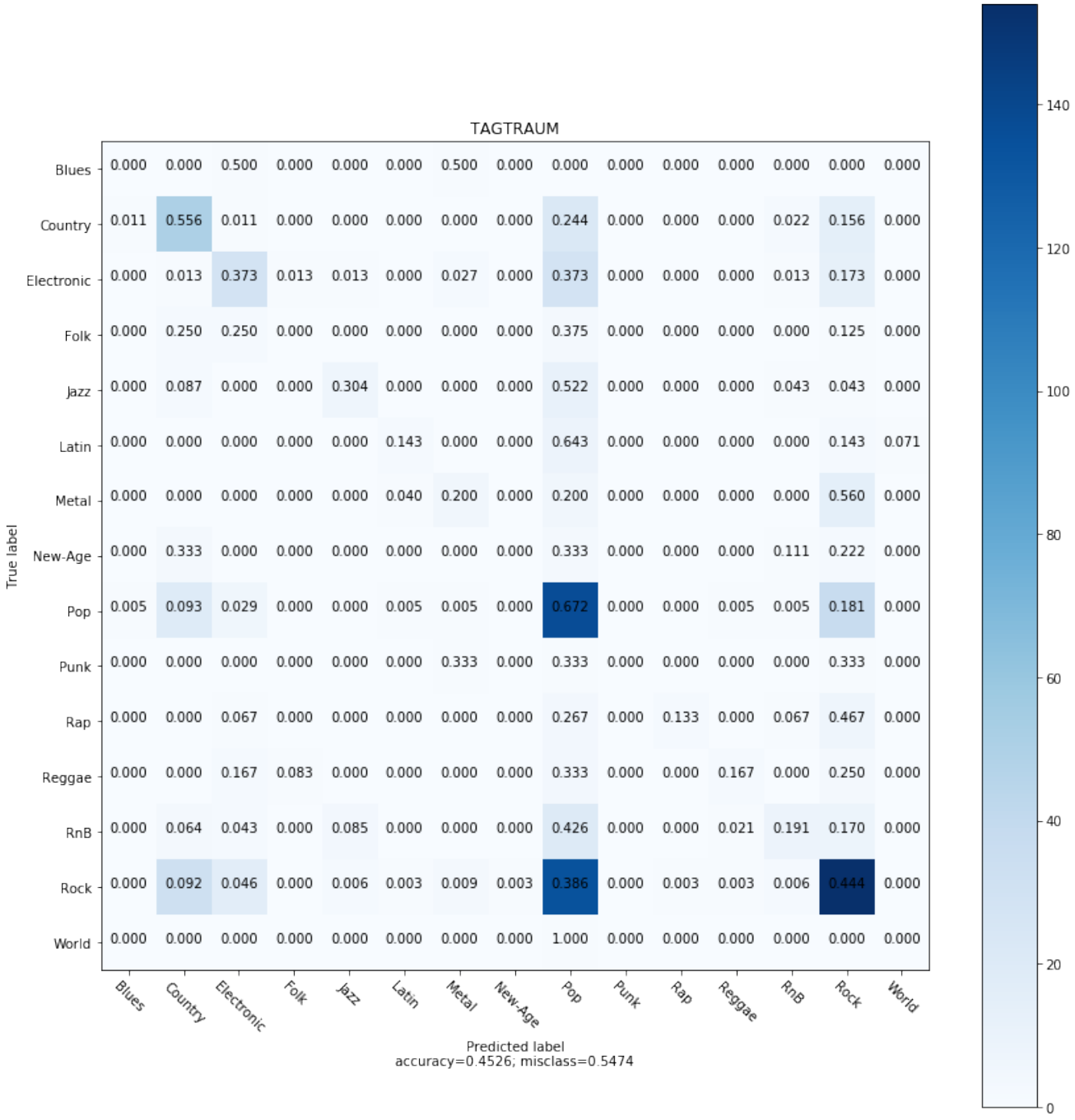
Πίνακας 6.22: νευρωνικό i

	Ποσοστό επιτυχίας για διαφορετική διαμέριση				
Μετρική	512	1024	256	384	Ensembled
accuracy	0.3952	0.3697	0.4274	0.3959	**0.4526**
macro avg	0.2346	0.1709	0.2253	**0.2405**	0.2317
weighted avg	0.3803	0.3600	0.4233	0.3899	**0.4405**

Πίνακας 6.23: Αναφορά κατηγοριοποίησης TAGTRAUM - νευρωνικό i

Class	Precision	Recall	F-score	Support
Blues	0.0	0.0	0.0	2
Country	0.4464	0.5556	0.495	90
Electronic	0.4746	0.3733	0.4179	75
Folk	0.0	0.0	0.0	8
Jazz	0.5	0.3043	0.3784	23
Latin	0.4	0.1429	0.2105	14
Metal	0.3846	0.2	0.2632	25
New-Age	0.0	0.0	0.0	9
Pop	0.3577	0.6716	0.4668	204
Punk	0.0	0.0	0.0	3
Rap	0.6667	0.1333	0.2222	15
Reggae	0.4	0.1667	0.2353	12
RnB	0.5	0.1915	0.2769	47
Rock	0.5992	0.4438	0.5099	347
World	0.0	0.0	0.0	1
accuracy			0.4526	875
macro avg	0.3153	0.2122	0.2317	875
avg	0.4819	0.4526	0.4405	875

Σχήμα 6.20: Πίνακας σύγχυσης νευρωνικού i



- **Νευρωνικό iii**

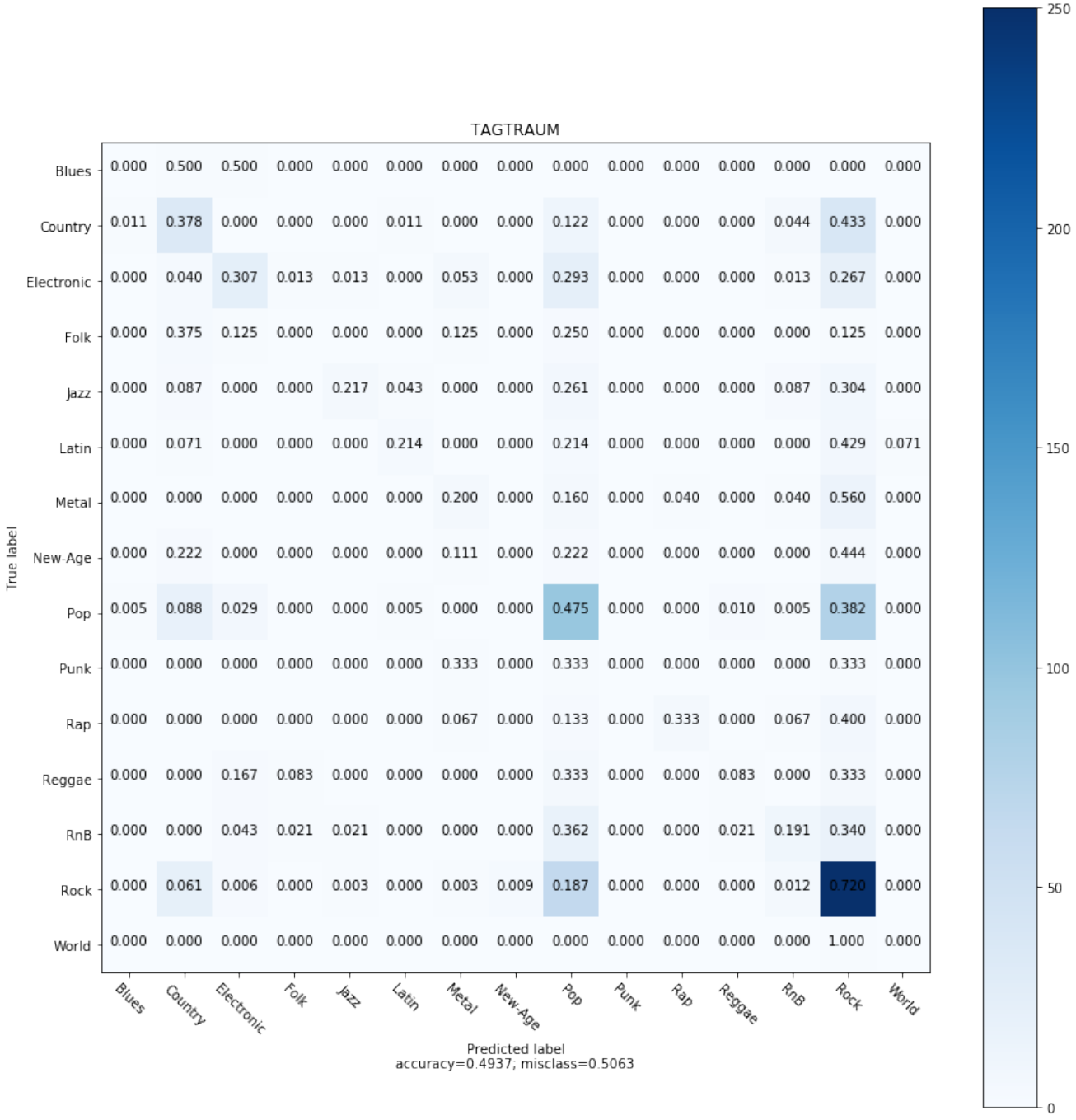
Πίνακας 6.24: νευρωνικό iii

	Ποσοστό επιτυχίας για διαφορετική διαμέριση						
Μετρική	1152	192	384	512	576	768	Ensembled
accuracy	0.4252	0.4153	0.4382	0.4582	0.4183	0.4153	**0.4937**
macro avg	0.1895	0.2111	0.2155	0.2056	0.2134	0.2111	**0.2405**
weighted avg	0.4099	0.4138	0.4277	0.4174	0.4179	0.4138	**0.4720**

Πίνακας 6.25: Αναφορά κατηγοριοποίησης TAGTRAUM - νευρωνικό iii

Class	Precision	Recall	F-score	Support
Blues	0.0	0.0	0.0	2
Country	0.4	0.3778	0.3886	90
Electronic	0.6216	0.3067	0.4107	75
Folk	0.0	0.0	0.0	8
Jazz	0.625	0.2174	0.3226	23
Latin	0.5	0.2143	0.3	14
Metal	0.3571	0.2	0.2564	25
New-Age	0.0	0.0	0.0	9
Pop	0.411	0.4755	0.4409	204
Punk	0.0	0.0	0.0	3
Rap	0.8333	0.3333	0.4762	15
Reggae	0.25	0.0833	0.125	12
RnB	0.3913	0.1915	0.2571	47
Rock	0.5593	0.7205	0.6297	347
World	0.0	0.0	0.0	1
accuracy			0.4937	875
macro avg	0.3299	0.208	0.2405	875
avg	0.4854	0.4937	0.472	875

Σχήμα 6.21: Πίνακας σύγχυσης νευρωνικού iii



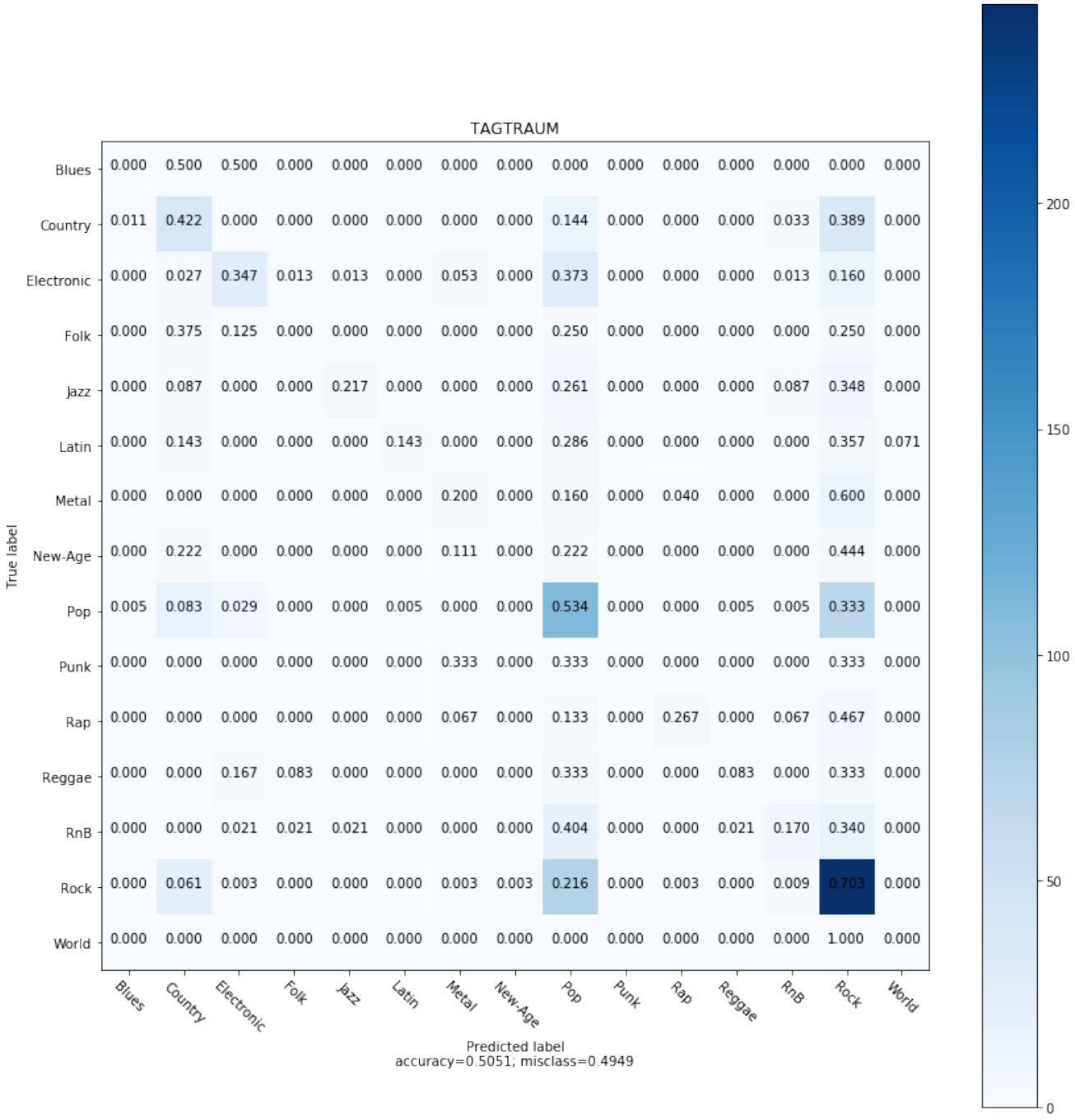
- **Ένωση Νευρωνικών i - iii**

Μπορούμε επίσης να ενώσουμε (ensemble) τα διαφορετικά νευρωνικά i - iii. Μπορούμε να επιλέξουμε όλα τα διαφορετικά νευρωνικά με όλες τις διαφορετικές διαστάσεις και τα ενώνουμε Στο 6.26 βλέπουμε την αναφορά κατηγοριοποίησης και στο σχήμα 6.23 τον πίνακα σύγχυσης.

Πίνακας 6.26: Αναφορά κατηγοριοποίησης συνδυασμού νευρωνικών

Class	Precision	Recall	F-score	Support
Blues	0.0	0.0	0.0	2
Country	0.4318	0.4222	0.427	90
Electronic	0.6842	0.3467	0.4602	75
Folk	0.0	0.0	0.0	8
Jazz	0.7143	0.2174	0.3333	23
Latin	0.6667	0.1429	0.2353	14
Metal	0.3846	0.2	0.2632	25
New Age	0.0	0.0	0.0	9
Pop	0.4052	0.5343	0.4609	204
Punk	0.0	0.0	0.0	3
Rap	0.6667	0.2667	0.381	15
Reggae	0.3333	0.0833	0.1333	12
RnB	0.4211	0.1702	0.2424	47
Rock	0.5782	0.7032	0.6346	347
Worl	0.0	0.0	0.0	1
accuracy			0.5051	875
macro avg	0.3524	0.2058	0.2381	875
avg	0.5059	0.5051	0.4839	875

Σχήμα 6.22: Πίνακας σύγχυσης ένωσης i - iii



6.2.5 Αποτελέσματα πειραμάτων για Lastfm

Για το σύνολο δεδομένων Lastfm (5.1), χρησιμοποιήσαμε τις αρχιτεκτονική του νευρωνικού i δικτύου.

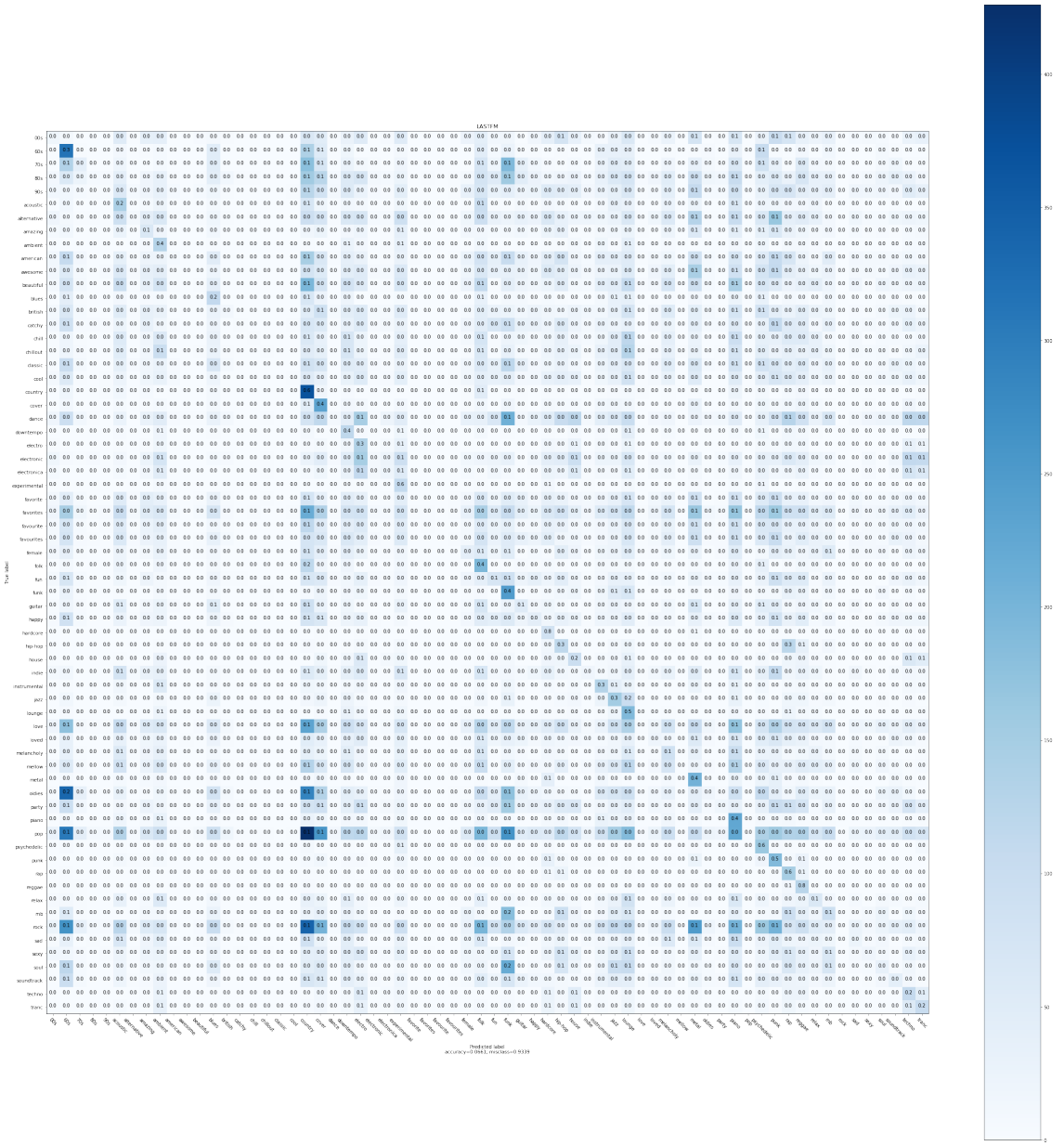
Όμως επειδή αυτό το σετ δεδομένων έχει πάρα πολλά δεδομένα και πάρα πολλές κλάσεις (66), δεν ήταν δυνατή η καλή εκπαίδευση αυτού του νευρωνικού δικτύου, επειδή χρειαζόταν πολύ χρόνο ή κάποιον υπερυπολογιστή.

Ενδεικτικά δείχνουμε συνοπτικά τα αποτελέσματα εκπαίδευσης παρόλο που βλέπουμε να μην έχει γίνει καλή εκπαίδευση. Παρατηρούμε ωστόσο, πως έχει ξεκινήσει η διαδικασία κατηγοριοποίησης, αφού δεν βλέπουμε τυχαία κατανεμημένα στοιχεία μόνο. Μερικά στοιχεία είναι πάνω στην διαγώνιο, δηλαδή έχει γίνει σωστή πρόβλεψη για αυτά.

Πίνακας 6.27: Αναφορά κατηγοριοποίησης νευρωνικού i

Class	Precision	Recall	F-score	Support
00s	0.0413	0.0115	0.018	1301
60s	0.0894	0.3316	0.1408	950
70s	0.1076	0.0284	0.0449	1338
80s	0.1111	0.0031	0.0061	1922
90s	0.0652	0.0018	0.0035	1671
acoustic	0.0648	0.2009	0.098	697
alternative	0.0686	0.0038	0.0072	1832
amazing	0.0581	0.0847	0.069	708
ambient	0.0641	0.3575	0.1087	358
american	0.0575	0.003	0.0057	1659
awesome	0.0625	0.0006	0.0012	1619
beautiful	0.0476	0.0015	0.003	1952
blues	0.0738	0.2072	0.1088	531
british	0.0859	0.0084	0.0153	1311
catchy	0.0625	0.0014	0.0028	1401
chill	0.0513	0.0017	0.0033	1171
chillout	0.0667	0.0017	0.0032	1211
classic	0.0	0.0	0.0	1551
cool	0.0541	0.0016	0.0032	1228
country	0.0725	0.6474	0.1304	570
cover	0.0711	0.4122	0.1212	558
dance	0.1023	0.0037	0.0072	2417
downtempo	0.0515	0.3966	0.0912	290
electro	0.0569	0.3358	0.0973	399
electronic	0.0614	0.0049	0.009	1440
electronica	0.0909	0.0248	0.039	765
experimental	0.0514	0.6056	0.0948	180
favorite	0.0357	0.0007	0.0015	1347
favorites	0.0833	0.001	0.002	3038
favourite	0.1429	0.0007	0.0015	1339
favourites	0.05	0.0013	0.0026	1523
female	0.0592	0.04	0.0478	949
folk	0.0619	0.3508	0.1052	553
fun	0.0618	0.0572	0.0594	1171
funk	0.0705	0.3722	0.1185	661
guitar	0.0599	0.0623	0.0611	818
happy	0.0587	0.0348	0.0437	1151
hardcore	0.0559	0.7619	0.1042	105
hip-hop	0.0608	0.2937	0.1007	395
house	0.0786	0.1939	0.1119	495
indie	0.07	0.0169	0.0273	1064
instrumental	0.1026	0.2764	0.1496	474
jazz	0.0731	0.2828	0.1162	541
lounge	0.0529	0.5162	0.096	370
love	0.05	0.001	0.002	2898
loved	0.0499	0.0195	0.028	975
melancholy	0.0563	0.1051	0.0733	818
mellow	0.0345	0.0006	0.0013	1555
metal	0.0727	0.4469	0.125	461
oldies	0.0752	0.008	0.0144	2138
party	0.1039	0.0097	0.0177	1658
piano	0.0595	0.4109	0.104	550
pop	0.0909	0.0004	0.0008	4705
psychedelic	0.0625	0.6059	0.1134	269
punk	0.0601	0.4713	0.1065	401
rap	0.0634	0.6218	0.1151	238
reggae	0.0574	0.7561	0.1066	164
relax	0.0587	0.0825	0.0686	727
rnb	0.07	0.094	0.0802	979
rock	0.0	0.0	0.0	3868
sad	0.0623	0.0237	0.0344	969
sexy	0.0657	0.0156	0.0252	1153
soul	0.0679	0.025	0.0365	1322
soundtrack	0.061	0.038	0.0468	1026
techno	0.0723	0.2443	0.1116	438
trance	0.0923	0.2388	0.1332	423
accuracy			0.0661	74759
macro avg	0.066	0.1691	0.0565	74759
avg	0.0657	0.0661	0.0301	74759

Σχήμα 6.23: Πίνακας σύγχυσης LASTFM



Κεφάλαιο 7

Συμπεράσματα και Προτάσεις

Στα προηγούμενα κεφάλαια έγινε μια προσπάθεια προσέγγισης του προβλήματος της αναγνώρισης είδους μουσικής με ανάλυση κομματιών μουσικής από συμβολικά δεδομένα MIDI, εφαρμόζοντας τις αρχές και τη λογική της επιβλεπόμενης μάθησης. Παρουσιάστηκαν υλοποιήσεις συστημάτων βαθιάς μηχανικής μάθησης. Η διαδικασία ολοκληρώθηκε με την εκπαίδευση και την αξιολόγηση των συστημάτων που υλοποιήθηκαν για κάθε μια από τις διαφορετικές προσεγγίσεις. Στο παρόν κεφάλαιο συνοψίζεται το έργο που επιτελέσθηκε στη προκείμενη διπλωματική εργασία και καταγράφονται τα αποτελέσματα που εξήχθησαν, δίνοντας στη συνέχεια προτάσεις για μελλοντική έρευνα και ανάπτυξη.

7.1 Συμπεράσματα

Για την ανάπτυξη ενός συστήματος Βαθιάς Μηχανικής Μάθησης που θα πετυχαίνει μεγάλες επιδόσεις στο πρόβλημα της αναγνώρισης είδους μουσικής με ανάλυση κομματιών από συμβολικά δεδομένα, πραγματοποιήθηκε σε πρώτο στάδιο εκτενής μελέτη παρεμφερών εργασιών και ερευνών με σκοπό την επιλογή των κατευθύνσεων δράσης. Έγινε μελέτη και εργασιών για αναγνώριση είδους μουσικής με ανάλυση κομματιών από ηχητικά δεδομένα είτε από στίχους κομματιών. Επιλέξαμε όμως να αναπτύξουμε σύστημα αναγνώρισης για συμβολικά δεδομένα γιατί αυτή είναι υψηλού επιπέδου πληροφορία για την μουσική, την αρμονία και τον ρυθμό. Τα συμβολικά δεδομένα είναι ουσιαστικά σαν να έχουμε ψηφιακή παρτιτούρα.

Διαβάσαμε επίσης για τα διάφορα είδη μουσικής, και πως αυτά δεν είναι ευδιάκριτα. Αυτό καθιστά το πρόβλημα της ταξινόμησης δύσκολο.

Από τη μελέτη παρεμφερών εργασιών και ερευνών προέκυψε ότι για την ανάπτυξη ενός τέτοιου συστήματος υπάρχουν διαφορετικές προσεγγίσεις. Δεν υπάρχει μέχρι στιγμής κάποιο σύστημα που να μπορεί να κάνει με μεγάλη επίδοση αναγνώριση είδους σε πολλαπλές κατηγορίες. Υπάρχουν πολλές τεχνικές προσέγγισης αυτού του προβλήματος. Η εξαγωγή χαρακτηριστικών φαίνεται να είναι κυριάρχουσα τακτική ως πρώτο βήμα για την ταξινόμηση. Τέτοια χαρακτηριστικά μπορεί να είναι ο ρυθμός, το τονικό ύψος των νοτών, κ.α. Το κύριο πλεονέκτημα των συνελικτικών νευρωνικών δικτύων CNN είναι ότι ανιχνεύει αυτόματα τα σημαντικά χαρακτηριστικά χωρίς καμία ανθρώπινη επίβλεψη. Με γνώμονα τα παραπάνω κάναμε

πολλές δοκιμές πάνω σε συνελικτικά νευρωνικά δίκτυα.

Αφού καταλείξαμε να χρησιμοποιήσουμε συνελικτικά νευρωνικά δίκτυα, έπρεπε να βρούμε ένα κατάλληλο σετ δεδομένων για να χρησιμοποιήσουμε, και να ελέγξουμε εάν η επιλογή μας ήταν σωστή. Χρησιμοποιήσαμε διάφορα dataset που ήταν σε μορφή pianoroll που προέρχεται από την MIDI κωδικοποίηση, η οποία είναι η πιο συνήθης στην αναγνώρισης είδους μουσικής με ανάλυση κομματιών από συμβολικά δεδομένα.

Έπρεπε να κάνουμε επεξεργασία αυτών των δεδομένων για να τα φέρουμε σε κατάλληλη μορφή για να μπορούμε να τα χρησιμοποιήσουμε στο νευρωνικό μας δίκτυο.

Ύστερα σχεδιάσαμε διάφορες αρχιτεκτονικές συνελικτικού νευρωνικού δικτύου και τρέξαμε πολλά πειράματα μέχρι να καταλείξουμε στις καλύτερες. Θεωρούμε πως υπάρχουν ακόμα καλύτερες αρχιτεκτονικές που δεν έχουν βρεθεί ή που δεν ελέγχθηκαν στην παρούσα διπλωματική. Αφού επιλέξαμε τις αρχιτεκτονικές κάναμε και πειράματα συνδυάζοντας τα νευρωνικά για να κατασκευάσουμε ένα ακόμη καλύτερο νευρωνικό δίκτυο.

7.2 Σύγκριση αποτελεσμάτων με παρεμφερείς εργασίες

Σε σύγκριση με τα αποτελέσματα των Ferraro και Lemstroem (2019) [15], τα οποία θεωρούνται τα καλύτερα κατά τη διάρκεια συγγραφής αυτής της διπλωματικής, πετύχαμε αποτελέσματα με μικρότερο Accuracy και F1-score. Παρόλα αυτά πετύχαμε αρκετά κοντινά ποσοστά στις μετρικές αυτές.

Ακόμα, η προσέγγισή μας ήταν η αυτόματη αναγνώριση από το νευρωνικό δίκτυο των χαρακτηριστικών γνωρίσματων των δεδομένων μας. Έτσι, είχαμε μία τελείως διαφορετική προσέγγιση στην αναζήτηση λύσης του προβλήματος αναγνώρισης είδους. Αυτό θα μπορούσε να σημαίνει πως εάν το νευρωνικό μας δίκτυο χρησιμοποιούνταν επικουρικά στο μοντέλο των Ferraro και Lemstroem να πετύχαινε καλύτερη επίδοση από τα δικά τους μοντέλα.

Στον πίνακα 7.1 φαίνεται η σύγκριση των αποτελεσμάτων μας, με των καλύτερων των ενΦερραρο και Lemstroem. Βέβαια, εμείς χρησιμοποιήσαμε το lpd-cleansed των MASD και TopMAGD, δηλαδή κάποια κομμάτια των αρχικών MASD και TopMAGD έχουν αφαιρεθεί σύμφωνα με τους κανόνες που αναφέραμε στο 5.1

Πίνακας 7.1: Σύγκριση αποτελεσμάτων με Ferraro, Lemstroem

	Top-MAGD		MASD	
	f1-measure	Accuracy	f1-measure	Accuracy
Ferraro, Lemstroem	0.662	0.620	0.455	0.342
Παρούσα διπλωματική	0.207	0.595	0.182	0.246

7.3 Μελλοντικές Επεκτάσεις και Προτάσεις

Το σύστημα ταξινόμησης του είδους που αναπτύχθηκε εδώ θα μπορούσε εύκολα να προσαρμοστεί σε εργασίες όπως ταυτοποίηση συνθέτη, αναγνώριση ερμηνευτή, ταξινόμηση συ-

ναισθήματος ή ταξινόμηση με βάση τη χρονική περίοδο, αλλάζοντας απλά την διαδικασία ταξινόμησης και τα δεδομένα εκπαίδευσης. Μελλοντικά πειράματα με αυτήν την επέκταση του πεδίου θα μπορούσαν να διερευνήσουν τους πολλαπλούς τρόπους με τους οποίους θα μπορούσε να χρησιμοποιηθεί αυτό το σύστημα.

Το σύστημα θα μπορούσε επίσης να επεκταθεί ώστε να περιλαμβάνει δυνατότητες που εξάγονται από δεδομένα ήχου χαμηλού επιπέδου απευθείας, τα οποία δεν είναι απαραίτητα άμεσα μεταφράσιμα σε συμβολικούς όρους, αλλά παρόλα αυτά βοηθούν κάποιον να διακρίνει μεταξύ των ειδών μουσικής. Τα ηχητικά δεδομένα περιέχουν τις πληροφορίες που χρησιμοποιούν οι περισσότεροι άνθρωποι για να κάνουν ταξινομήσεις βάσει περιεχομένου, οπότε θα ήταν επωφελές να κάνουν χρήση αυτών των ενδείξεων χαμηλού επιπέδου, καθώς και της υψηλότερης στάθμης μουσικής αντίληψης που διαθέτουν οι εκπαιδευμένοι μουσικοί.

Η χρήση τόσο συμβολικών όσο και ηχητικών δεδομένων θα καθιστούσε δυνατή την αξιοποίηση της έρευνας αντιστοίχισης παρτιτούρας-ηχου (score alignment research), προκειμένου να ταιριάξει τις παρτιτούρες με τις ηχογραφήσεις και να αφαιρέσει θορυβώδη μεταγραφή και σφάλματα απόδοσης. Αυτό θα έδινε επίσης ένα μέτρο του ποσοστού απόκλισης μιας εκτέλεσης ενός κομματιού σε σχέση με αυτή την πληροφορία από τα συμβολικά δεδομένα , η οποία θα μπορούσε να είναι ένα χρήσιμο χαρακτηριστικό από μόνο του. Ίσως το ιδανικό θα ήταν η εκμετάλλευση των μορφών μουσικής, όπως αυτή που προτείνεται στο πρότυπο MPEG-21, τα οποία μπορούν να συσκευάσουν ηχητικά δεδομένα μαζί με συμβολικά δεδομένα που περιγράφουν τις νότες και τις παραμέτρους παραγωγής της μουσικής. Αυτό θα καθιστούσε ένα εξαιρετικά ευρύ φάσμα δυνατοτήτων διαθέσιμο σε ένα πακέτο.

Μια άλλη αντιμετώπιση του προβλήματος αυτού θα μπορούσε να είναι η εκπαίδευση ενός μοντέλου μη-επιβλεπόμενης μάθησης. Σε αυτή την περίπτωση τα δεδομένα δεν είναι απαραίτητο να συνοδεύονται από ετικέτες (labels) και μπορούν να ανακτηθούν σε μεγάλους όγκους.

Ακόμα θα μπορούσε να γίνει συνδυασμός ηχητικών και συμβολικών δεδομένων, ακόμη και στίχους τραγουδιών για την αναγνώριση είδους. Αυτό θα μπορούσε να παράξει καλύτερη επίδοση στην ταξινόμηση.

Τέλος θα μπορούσαμε να εξάγουμε κάποια χαρακτηριστικά των κομματιών, αρμονίας, μελωδίας, ρυθμού και χρησιμοποιώντας αυτά να κάνουμε ταξινόμηση, όπως έχουν γίνει σε άλλες έρευνες, όπως στους Ferraro, Lemstroem (2019) [15]. Επειδή στα δικά μας πειράματα τα χαρακτηριστικά τα ανιχνεύε αυτόματα το νευρωνικό δίκτυο, θα ήταν ενδιαφέρον ο συνδυασμός αυτών των δύο τεχνικών.

Έτσι, καταλείγουμε πως είχε χρησιμότητα αυτή η μελέτη για μελοντικές χρήσεις.

Βιβλιογραφία

[1] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[2] Igor Aleksander. *An introduction to neural computing*. London: Chapman and Hall, 1990. ISBN: 0412377802.

[3] Geoffrey E. Hinton Alex Krizhevsky Ilya Sutskever. «ImageNet classification with deep convolutional neural networks». In: *Communication of the ACM* 60 (2012), pp. 84–90. DOI: 10.1145/3065386.

[4] Ethem Alpaydin. *Introduction to machine learning*. Cambridge, Mass: MIT Press, 2010. ISBN: 9780262012430.

[5] Roberto Basili, Alfredo Serafini, and Armando Stellato. «Classification of musical genre: a machine learning approach.» In: Jan. 2004.

[6] Roberto Basili, Alfredo Serafini, and Armando Stellato. «EXTRACTING MUSIC FEATURES WITH MIDXLOG». In: (Jan. 2005).

[7] Thierry Bertin-Mahieux et al. «The Million Song Dataset». In: *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)*. 2011.

[8] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag New York Inc., Aug. 17, 2006. 738 pp. ISBN: 0387310738. URL: https://www.ebook.de/de/product/5324937/christopher_m_bishop_pattern_recognition_and_machine_learning.html.

[9] Wei Chai and Barry Vercoe. «Folk Music Classification Using Hidden Markov Models». In: (Nov. 2020).

[10] Ronald J. Williams David E. Rumelhart Geoffrey E. Hinton. «Learning representations by back-propagating errors». In: *Nature* 323 (1986), pp. 533–536. ISSN: 1476-4687. DOI: 10.1038/323533a0. URL: https://doi.org/10.1038/323533a0.

[11] T.N. Wiesel D.H. Hubel. «Receptive fields and functional architecture of monkey striate cortex». In: *J Physiol* 195 (1968), pp. 215–243. DOI: 10.1113/jphysiol.1968.sp008455.

[12] Hao-Wen Dong and Yi-Hsuan Yang. «Convolutional Generative Adversarial Networks with Binary Neurons for Polyphonic Music Generation». In: (Apr. 25, 2018). arXiv: 1804.09399 [cs.LG].

[13] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. Wiley John + Sons, Oct. 1, 2000. ISBN: 0471056693. URL: https://www.ebook.de/de/product/3244086/richard_o_duda_peter_e_hart_david_g_stork_pattern_classification.html.

[14] Laurene Fausett. *Fundamentals of neural networks : architectures, algorithms, and applications*. Englewood Cliffs, NJ Delhi Dorling Kindersley: Prentice-Hall, 1994. ISBN: 0133341860.

[15] Andres Ferraro and Kjell Lemström. «On large-scale genre classification in symbolically encoded music by automatic identification of repeating patterns». In: *Proceedings of the 5th International Conference on Digital Libraries for Musicology (DLfM '18). ACM, New York, NY, USA, 34-37. 2018* (Oct. 21, 2019). DOI: 10.1145/3273024.3273035. arXiv: 1910.09242 [cs.IR].

[16] Jonathan Fieldsend and Richard Everson. «Visualisation of multi-class ROC surfaces». In: *Proceedings of the ICML 2005 Workshop on ROC Analysis in Machine Learning* (2005), pp. 49–56. URL: http://dmip.webs.upv.es/ROCML2005/papers/fieldsend2CRC.pdf.

[17] Dr.Kunihiko Fukushima. «Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position». In: *Biological Cybermetrics* 36 (1980), pp. 193–202. DOI: doi:10.1007/BF00344251.

[18] A. Gabura. «Undergraduate research in computer sciences: Computer analysis of musical style». In: (Jan. 1965), pp. 303–314. DOI: 10.1145/800197.806054.

[19] Ian Goodfellow et al. *Deep Learning*. MIT Press Ltd, Nov. 18, 2016. 800 pp. ISBN: 0262035618. URL: https://www.ebook.de/de/product/26337726/ian_goodfellow_joshua_bengio_aaron_courville_francis_bach_deep_learning.html.

[20] Wen-Yi Hsiao Hao-Wen Dong and Yi-Hsuan Yang. «Pypianoroll: Open Source Python Package for Handling Multitrack Pianorolls». In: *in Late-Breaking Demos of the 19th International Society for Music Information Retrieval Conference (ISMIR)* (2018).

[21] Simon Haykin. *Neural Networks and Learning Machines*. 3rd edition. Translator: Gkagkatsiou, Eleni. Papasotiriou, 2010. ISBN: 9789607182647. URL: http://dai.fmph.uniba.sk/courses/NN/haykin.neural-networks.3ed.2009.pdf.

[22] Hal Daume III. *A Course in Machine Learning*. ciml.info, Jan. 2017. URL: http://ciml.info/.

[23] Jose Iñesta. «Musical Style Identification Using Self-Organising Maps». In: (May 2003).

[24] Thomas Kautz, Bjoern M. Eskofier, and Cristian F. Pasluosta. «Generic performance measure for multiclass-classifiers». In: *Pattern Recognition* 68 (2017), pp. 111–125. DOI: 10.1016/j.patcog.2017.03.008.

[25] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2014. arXiv: 1412.6980 [cs.LG].

[26] Raymond L. Knapp and Julian Medforth Budden. «Ludwig van Beethoven». In: *Encyclopædia Britannica* (2020). URL: https://www.britannica.com/biography/Ludwig-van-Beethoven.

[27] Qiuqiang Kong, Keunwoo Choi, and Yuxuan Wang. «Large-Scale MIDI-based Composer Classification». In: (Oct. 28, 2020). arXiv: arXiv:2010.14805v1 [cs.SD].

[28] Frank Krüger. «Activity, Context, and Plan Recognition with Computational Causal Behaviour Models». PhD thesis. Dec. 2016.

[29] Olivier Lartillot et al. «Automatic Modeling of Musical Style». In: (Sept. 2001).

[30] P. D. León. «MIREX 2005 : Symbolic Genre classification with an ensemble of parametric and lazy classifiers». In: 2005.

[31] Agnes Lydia and Sagayaraj Francis. «Adagrad - An Optimizer for Stochastic Gradient Descent». In: (May 2019).

[32] Spyros Makridakis. «The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms». In: *Futures* 90 (2017), pp. 46–60. DOI: 10.1016/j.futures.2017.03.006.

[33] C. McKay. «Automatic genre classification of MIDI recordings». MA thesis. McGill University, Canada., 2004.

[34] C McKay. «Automatic music classification with jMIR». PhD thesis. McGill University, Canada., 2010.

[35] C. McKay and Ichiro Fujinaga. «THE BODHIDHARMA SYSTEM AND THE RESULTS OF THE MIREX 2005 SYMBOLIC GENRE CLASSIFICATION CONTEST». In: 2005.

[36] Tom Mitchell. *Machine Learning*. New York: McGraw-Hill, 1997. ISBN: 0070428077.

[37] Hélio de Oliveira and Raimundo Oliveira. «Understanding MIDI: A Painless Tutorial on Midi Format». In: (May 2017).

[38] Colin Raffel. «Learning-Based Methods for Comparing Sequences, with Applications to Audio-to-MIDI Alignment and Matching». PhD thesis. 2016.

[39] Colin Raffel and Daniel P. W. Ellis. «Intuitive analysis, creation and manipulation of MIDI data with pretty midi». In: *ISMIR Late Breaking Demo Paper*. 2014. URL: http://www.terasoft.com.tw/conf/ismir2014/LBD/LBD29.pdf.

[40] Stuart Russell. *Artificial intelligence : a modern approach*. 2nd edition. Translators: Alvas T., Kartsaklis D., Skoularikis F. Klidarithmos, 2005. ISBN: 9788120323827.

[41] Albrecht Schmidt. «A Modular Neural Network Architecture with Additional Generalization Abilities for High Dimensional Input Vectors». In: (Nov. 2000).

[42] Hendrik Schreiber. «Improving Genre Annotations for the Million Song Dataset». In: *In Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR),Málaga, Spain*. Oct. 2015, pages 241–247.

[43] Man-Kwan Shan and Fang-Fei Kuo. «Music Style Mining and Classification by Melody». In: *IEICE Transactions on Information and Systems* E86-D (Mar. 2003).

[44] Edward Smith. *Categories and concepts*. Cambridge, Mass: Harvard University Press, 1981. ISBN: 9780674102750.

[45] Nitish Srivastava et al. «Dropout: A Simple Way to Prevent Neural Networks from Overfitting». In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

[46] Bob Sturm. «A Survey of Evaluation in Music Genre Recognition». In: Jan. 2012. DOI: 10.1007/978-3-319-12093-5_2.

[47] Gerald Tesauro. «TD-Gammon, a Self-Teaching Backgammon Program, Achieves Master-Level Play». In: *Neural Computation* 6.2 (1994), pp. 215–219. DOI: 10.1162/neco.1994.6.2.215.

[48] Bassiliades N. Kokkoras F. Sakellariou I. Vlahavas I. Kefalas P. *Artificial Intelligence*. University of Macedonia, 2011. ISBN: 978-960-8396-64-7. URL: https://www.researchgate.net/publication/27378851_Technete_Noemosyne.

[49] Ludwig Wittgenstein. *Philosophical investigations*. Oxford: Basil Blackwell, 1968. ISBN: 0631119000.

[50] Matthew D. Zeiler. *ADADELTA: An Adaptive Learning Rate Method*. 2012. arXiv: 1212.5701 [cs.LG].

[51] Xiaojin Zhu. *Introduction to semi-supervised learning*. San Rafael, Calif: Morgan & Claypool Publishers, 2009. ISBN: 9781598295481.

7.3.0.0.0.1