ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

«Μελέτη και ανάπτυξη συστήματος αναγνώρισης ηχητικών συμβάντων, χρησιμοποιώντας τεχνικές μηχανικής μάθησης»

ΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΚΩΝΣΤΑΝΤΙΝΟΥ ΑΘΑΝΑΣΙΟΣ

ΠΑΤΡΑ, 2023

UNIVERSITY OF PELOPONNESE

SCHOOL OF ENGINEERING

Department of Electrical and Computer Engineering

«Study and development of an audio event recognition system, using machine learning techniques»

GRADUATE THESIS

KONSTANTINOU ATHANASIOS

PATRA, 2023

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

Αθανάσιος Κούτρας, Αναπληρωτής καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Πανεπιστήμιο Πελοποννήσου, ***Επιβλέπων***

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ

Αυτή η διατριβή διερευνά τη σφαίρα της ταξινόμησης ήχου εξετάζοντας σχετικές εργασίες και διεξάγοντας βασικά πειράματα χρησιμοποιώντας αρχιτεκτονικές Συνελικτικού Νευρωνικού Δικτύου (CNN) στο σύνολο δεδομένων Urban Sound 8k. Ξεκινάμε παρέχοντας μια επισκόπηση διαφόρων μεθόδων και συνόλων δεδομένων που χρησιμοποιήθηκαν σε προηγούμενες μελέτες στον τομέα της ταξινόμησης ήχου. Μετά από αυτήν την ανάλυση, στρέφουμε την εστίασή μας στον πρακτικό πειραματισμό, εφαρμόζοντας απλά μοντέλα CNN για να διερευνήσουμε την ικανότητά τους στην ταξινόμηση των αστικών ήχων. Τα μοντέλα μας αξιολογούνται χρησιμοποιώντας μετρικες όπως το F1-score και η ακρίβεια, παρέχοντας μια ποσοτική αξιολόγηση της απόδοσής τους. Τα αποτελέσματα, αν και είναι προκαταρκτικά, δείχνουν ενθαρρυντικά σημάδια, υποδεικνύοντας ότι ακόμη και απλά μοντέλα CNN μπορούν να είναι κάπως αποτελεσματικά σε εργασίες ταξινόμησης ήχου.

ABSTRACT

This thesis explores the realm of sound classification by examining related works and conducting basic experiments using Convolutional Neural Network (CNN) architectures on the UrbanSound8k dataset. We begin by providing an overview of various methods and datasets utilized in prior studies within the sound classification domain. Following this analysis, we shift our focus towards practical experimentation, implementing straightforward CNN models to explore their capability in classifying urban sounds. Our models are evaluated using metrics such as the F1 Score and accuracy, providing a quantitative assessment of their performance. The results, while preliminary, show encouraging signs, indicating that even simple CNN models can be somewhat effective in sound classification tasks.

1. Εισαγωγή

Η ανίχνευση περιβαλλοντικών ήχων αποσκοπεί στο να κατανοήσει και να κατηγοριοποιήσει την ακατέργαστη ηχητική πληροφορία που προκύπτει σε πραγματικά περιβάλλοντα. Αυτός ο σκοπός έχει ως αποτέλεσμα την ανάπτυξη αυτοματοποιημένων συστημάτων που μπορούν να παράγουν σημασιολογική ανάλυση του περιεχομένου των ηχητικών σημάτων περιβαλλοντικών ήχων. Αυτό επιτυγχάνεται μέσω της χρήσης αλγορίθμων μηχανικής μάθησης. Η ανίχνευση αυτής της ηχητικής πληροφορίας μπορεί να περιλαμβάνει ένα ευρύ φάσμα απρόβλεπτων και ανασταλτικών συμβάντων και θορύβων που υπάρχουν στο περιβάλλον. Αυτά τα ηχητικά σήματα μπορεί να παράγονται από διάφορες πηγές, με κινούμενες ή στατικές ιδιότητες, που ενδεχομένως να επικαλύπτονται από άλλα σήματα ή θορύβους, ενισχύοντας έτσι τον ποικιλόμορφο χαρακτήρα τους. Για παράδειγμα, σε αστικά περιβάλλοντα, τα ηχητικά σήματα μπορεί να προέρχονται από διάφορες πηγές, που ενδέχεται να είναι κινούμενες ή σταθερές, ενισχυμένες ή μη από άλλες πηγές, ενώ ο παρασκηνιακός θόρυβος μπορεί να μεταβάλλει τον ηχητικό χαρακτήρα. Συνεπώς, ο τομέας αυτός συναντά πραγματικούς φυσικούς περιορισμούς, αντανακλώντας τη φύση της γενικής αναγνώρισης προτύπων. Παράλληλα, υπάρχει και η δυσκολία να καλύψουμε το πλήρες εύρος δυνατών συμβάντων που μπορεί να συμβούν. Σε αντίθεση, με άλλους κλάδους, η κατηγοριοποίηση περιβαλλοντικών κατηγοριών προκαθορίζεται σε έναν περιορισμένο κατάλογο, προσφέροντας ένα προσεγγιστικό και ιεραρχικό πλαίσιο για τη σύλληψη των βασικών ηχητικών στοιχείων που απαρτίζουν μια περιβαλλοντική ηχητική συνθήκη. Ο κλάδος της αναγνώρισης περιβαλλοντικών ήχων αναπτύχθηκε ως ένας νέος τομέας σε σχέση με τους παρεμφερείς κλάδους της αναγνώρισης σημάτων ομιλίας και της ανάκτησης μουσικής πληροφορίας. Σύμφωνα με Cowling M. κ.ά. 2003 [1] , ο αρχικός στόχος ήταν η επέκταση αυτών των παραδοσιακών πεδίων, με σκοπό τη δημιουργία νέων πεδίων εφαρμογών, χρησιμοποιώντας τις υπάρχουσες τεχνικές ανάπτυξης και υλοποίησης συστημάτων ταξινόμησης. Στα πρώτα στάδια, ο κλάδος επικεντρώθηκε στην ταξινόμηση περιορισμένου αριθμού περιβαλλοντικών κατηγοριών, με έμφαση στην αναγνώριση μεμονωμένων ηχητικών συμβάντων, όπως η αναγνώριση επιμέρους θορύβων. Στη συνέχεια, η έρευνα εξελίχθηκε παρόμοια με τους προηγούμενους κλάδους, επεκτείνοντας την ανάλυση σε πιο σύνθετα περιβαλλοντικά σενάρια.

Ένα ακόμη σημαντικό βήμα ήταν η αντιμετώπιση της ταξινόμησης επικαλυπτόμενων ηχητικών συμβάντων, όπου πολλά ηχητικά στοιχεία μπορεί να εκδηλώνονται ταυτόχρονα, όπως συμβαίνει σε περιβαλλοντικές συνθήκες πολλών φωνών. Αυτό είναι ανάλογο με τον διαχωρισμό της ομιλίας ή τη

μεταγραφή πολυφωνικής μουσικής. Συνεπώς, ο κλάδος ακολούθησε μια εξελικτική πορεία παρόμοια με προηγούμενους κλάδους, επεκτείνοντας ωστόσο την αναγνώριση πέρα από μονοδιάστατες κατηγορίες και αντιμετωπίζοντας την πολυπλοκότητα της ανίχνευσης επικαλυπτόμενων ήχων. Έπειτα, από εκτεταμένες μελέτες και πειράματα, τα οποία σαφώς κέντρισαν το ενδιαφέρον της επιστημονικής κοινότητας, καθώς επίσης και λόγω της προοπτικής της σύνθεσης ποικίλων σετ δεδομένων για την κάλυψη των αναγκών του κλάδου, έχουν αναπτυχθεί πρωτοποριακές μέθοδοι βασισμένες σε διάφορες αρχιτεκτονικές προσεγγίσεις. Αυτές οι νέες μέθοδοι αποκλίνουν από τις προηγούμενες προσεγγίσεις, στοχεύοντας στην επίτευξη υψηλής απόδοσης, ειδικά για τα συστήματα ταξινόμησης περιβαλλοντικών ήχων. Αυτές οι προηγμένες προσεγγίσεις εκμεταλλεύονται την εμπειρία αλλά και την γνώση που αποκτήθηκε από την έρευνα και την πρακτική εφαρμογή. Επιπλέον, η εισαγωγή διαφορετικών αρχιτεκτονικών προσεγγίσεων επιτρέπει την ανάπτυξη πρωτοποριακών τεχνικών που προσαρμόζονται στις ιδιαιτερότητες των περιβαλλοντικών ηχητικών δεδομένων. Οι νέες αυτές προσεγγίσεις αποτελούν ένα βήμα προς τη βελτιστοποίηση της απόδοσης των συστημάτων ταξινόμησης περιβαλλοντικών ήχων, και συμβάλλουν στη συνεχή εξέλιξη του τομέα. Ο τομέας του κλάδου σήμερα παρέχει αξιόλογες εφαρμογές σε πολλούς τομείς, κερδίζοντας χρηστική αξία και ενσωματώνοντας την ηχητική πληροφορία σε ποικίλες εφαρμογές.

Φυσικά, θα πρέπει να τονίσουμε ότι ο ήχος, λόγω της χαμηλότερης διάστασής του σε σύγκριση με την εικόνα, μπορεί να αποτελέσει σημαντική συμπληρωματική πληροφορία σε μεθόδους που ασχολούνται με την ανάκτηση οπτικής πληροφορίας. Αυτό είναι ιδιαίτερα σημαντικό στον τομέα της ανάπτυξης συστημάτων αναγνώρισης περιεχομένου, καθώς η συμπερίληψη του ήχου μπορεί να οδηγήσει σε πιο οικονομικές αλλά ταυτόχρονα αποδοτικές μεθόδους υλοποίησης, μειώνοντας τις υπολογιστικές δαπάνες και ενισχύοντας την ακρίβεια των αποτελεσμάτων (S Sigtia κ.ά. 2016) [4] . Σχετικά με τη διάκριση των προβλημάτων ταξινόμησης περιβαλλοντικών ηχητικών σημάτων, προσδιορίζονται δύο βασικές κατηγορίες προβλημάτων, ανάλογα με τον τύπο της πληροφορίας που αποσκοπούμε να ανακτήσουμε. Στην πρώτη κατηγορία, επιδιώκεται η αναγνώριση του γενικού περιβάλλοντος στο σύνολό του και ταυτόχρονα τη διάκριση ανάμεσα σε διάφορα αυτόνομα σενάρια που παρουσιάζουν διαφορετικά πρότυπα. Με άλλα λόγια, ο στόχος είναι να αναγνωρίσει ηχητικά αρχεία και να τα ταξινομήσει σε προκαθορισμένες κατηγορίες περιβαλλοντικών. Από την άλλη, η δεύτερη κατηγορία προβλημάτων επικεντρώνεται στην ανάλυση του ηχητικού περιβάλλοντος με σκοπό την αναγνώριση διακριτών συμβάντων που λαμβάνουν χώρα σε ένα αυτόνομο περιβάλλον. Σε αυτήν την περίπτωση, το βασικό ενδιαφέρον είναι η ταξινόμηση διαφορετικών ηχητικών σημάτων περιβάλλοντος σε

προκαθορισμένες κλάσεις. Αυτές οι δύο κατηγορίες προσεγγίσεων αποσκοπούν στη διάκριση διαφορετικών σεναρίων μέσω διαφορετικών τεχνικών, οι οποίες έχουν εκτενώς μελετηθεί μέχρι σήμερα. Θα πρέπει να εξεταστεί η δυνατότητα συνδυασμού αυτών των δύο προσεγγίσεων, με κύριο σκοπό την επίτευξη μιας πιο ολοκληρωμένης κατανόησης του ηχητικού περιβάλλοντος.
Στα πλαίσιο της παρούσας εργασίας, θα εξεταστούν προβλήματα ανάκτησης μονοφωνικών θορύβων που εντάσσονται στην δεύτερη κατηγορία που προαναφέρθηκε, δηλαδή της αναγνώρισης διακριτών ηχητικών γεγονότων.

1.1 Related work

Εισαγωγή στην Κατηγοριοποίηση Ήχου σε Αστικά και Περιβαλλοντικά Πλαίσια

Τα τελευταία χρόνια παρατηρείται αυξανόμενο ενδιαφέρον για την ταξινόμηση των ήχων στο πλαίσιο των αστικών και περιβαλλοντικών συνθηκών, λόγω των δυνητικών εφαρμογών της σε έξυπνες πόλεις, συστήματα παρακολούθησης και άλλες έξυπνες λύσεις. Ο Salamon κ.ά. (2014) [2] συνέβαλαν σε αυτόν τον τομέα με την εισαγωγή του Urban Sound dataset [26], αντιμετωπίζοντας την έλλειψη δεδομένων για την έρευνα των αστικών ήχων. Επίσης, ο Abdoli κ.ά. (2019) [3] παρουσίασαν ένα 1D Convolutional Neural Network (CNN) για την ταξινόμηση περιβαλλοντικών ήχων, ενώ οι Sigtia κ.ά. (2016) [4] διερεύνησαν τον υπολογιστικό κόστος των αλγορίθμων αυτόματης αναγνώρισης περιβαλλοντικών ήχων. Πιο πρόσφατα, ο Mu κ.ά. (2021) [5] πρότειναν ένα CNN με εστίαση στα χρονικά-συχνοτικά πεδία για την ταξινόμηση περιβαλλοντικών ήχων, και ο Mushtaq κ.ά. (2021) [6] διερεύνησαν τη χρήση εικόνων Mel spectrogram και CNNs, επιτυγχάνοντας εντυπωσιακή ακρίβεια σε datasets όπως τα ESC-10 και Us8k. Ο Nasiri (2021) [7] προήγαγε επιπλέον αλγορίθμους βαθιάς μάθησης για την αναγνώριση ήχου σε υπολογιστές.

Επεκτείνοντας την ταξινόμηση των ήχων στον αστικό χώρο, ο Özseven (2023) [8] διαπίστωσε την αποτελεσματικότητα της συνδυασμένης χρήσης των Mel-spectrograms και ακουστικών χαρακτηριστικών με μοντέλα βαθιάς μάθησης. Ο Li κ.ά. (2019) [9] χρησιμοποίησαν διάφορα μοντέλα μηχανικής μάθησης, συμπεριλαμβανομένων νευρωνικών δικτύων και CNNs, για την ταξινόμηση ήχων στο dataset Urban Sound 8k. Ο Rahman και η Jothi (2021) [10] εξέτασαν επίσης το UrbanSound8k dataset, χρησιμοποιώντας βαθιά CNNs και τεχνικές data augmentation. Επιπλέον, οι Nanni κ.ά. (2021)

[11] διενήργησαν μελέτη για τα μοντέλα συνόλου CNN για την ταξινόμηση περιβαλλοντικών ήχων, επιτυγχάνοντας εντυπωσιακά αποτελέσματα σε πολλά datasets.

Οι Mesaros κ.ά. (2016) [13] παρουσίασαν και συζήτησαν μετρικές για την αξιολόγηση συστημάτων εντοπισμού πολυφωνικών συμβάντων ήχου, προσαρμοσμένες από τον τομέα της αναγνώρισης ομιλίας και της διακριτοποίησης ομιλητή. Το έργο τους υπογράμμισε την ανάγκη για την αντιμετώπιση των επικαλυπτόμενων συμβάντων, παρέχοντας εισαγωγή στις μετρικές βασισμένες σε τμήματα και συμβάντα, και στην κλάση μέσου όρου χρησιμοποιώντας εφαρμογές στον πραγματικό κόσμο.
Ο Adapa (2019) [14] παρουσίασε ένα πλαίσιο για την ταξινόμηση περιβαλλοντικών ήχων, κυρίως κατάλληλο για σενάρια με περιορισμένα ετικετοφορημένα δεδομένα. Χρησιμοποιώντας προ-εκπαιδευμένα μοντέλα ταξινόμησης εικόνας και διακυμάνσεις δεδομένων, το μοντέλο του, βασισμένο σε ένα τροποποιημένο CNN MobileNetV2, επέτυχε κορυφαία αποτελέσματα στην ταξινόμηση ήχων Urban Sound Tagging του DCASE2019.
Οι Fonseca κ.ά. (2019) [15] συμμετείχαν στο Task 2 του DCASE2019 [16] , το οποίο αφορά την ετικέτωση ήχου με θορυβώδεις ετικέτες και ελάχιστη επίβλεψη, εστιάζοντας στην αξιολόγηση συστημάτων για πολλαπλή ετικέτωση ήχου χρησιμοποιώντας ένα μεγάλο σύνολο δεδομένων με θορυβώδεις ετικέτες. Το προτεινόμενο σύνολο δεδομένων δημιουργεί ένα πρόβλημα ακουστικής αναντιστοιχίας μεταξύ των συνόλων εκπαίδευσης και δοκιμών, προσφέροντας παράλληλα το σύνολο δεδομένων FSDKaggle2019 και ένα βασικό σύστημα βασισμένο σε συνελικτικό νευρωνικό δίκτυο ως ελεύθερα διαθέσιμους πόρους. Επίσης, οι Ebbers και Häb-Umbach (2019) [17] παρουσίασαν ένα σύστημα για το Task 2 του DCASE 2019, χρησιμοποιώντας συνελικτικά και αναδρομικά νευρωνικά δίκτυα για την επεξεργασία ακολουθιών μεταβλητού μήκους και χρησιμοποιώντας log-mel-energies για την ταξινόμηση ήχου. Το μοντέλο τους, εκπαιδευμένο με περιορισμένα ετικετοφορημένα δεδομένα και δεδομένα από τον Ιστό με θόρυβο ετικετών, έθεσε νέα υψηλά πρότυπα απόδοσης στο σύνολο δεδομένων.
Οι Adavanne κ.ά. (2018) [18] προτείνουν τη χρήση χωρικών και αρμονικών χαρακτηριστικών σε συνδυασμό με τα δίκτυα νευρώνων LSTM (Long Short Term Memory) για την αυτόματη ανίχνευση συμβάντων ήχου (SED). Αντλώντας έμπνευση από τον τρόπο με τον οποίο οι ανθρώπινοι ακροατές αναγνωρίζουν επιτυχώς το μείγμα επικαλυπτόμενων ηχητικών συμβάντων χρησιμοποιώντας ενδείξεις τόνου και εκμεταλλευόμενοι το στερεοφωνικό ήχο, οι συγγραφείς προτείνουν την επέκταση των συστημάτων SED ώστε να χρησιμοποιούν πολυκαναλικό ήχο. Το προτεινόμενο σύστημα SED

συγκρίνεται με διάφορες μεθόδους στο σύνολο βάσης δεδομένων TUT Sound Events Detection 2016, δείχνοντας ότι η χρήση χωρικών και αρμονικών χαρακτηριστικών βελτιώνει την απόδοση του SED.

Αυτές οι εργασίες αντιπροσωπεύουν σημαντικές συνεισφορές στον τομέα της ανάλυσης ήχου και αποτελούν σημεία αναφοράς για την πρόοδο και την ανάπτυξη των σχετικών εφαρμογών.

1.2 Σκοπός

Ο σκοπός αυτής της διπλωματικής εργασίας είναι η ανάλυση και η εφαρμογή τεχνικών μηχανικής μάθησης σε ηχητικά δεδομένα. Κατά τη διάρκεια της έρευνας, θα εξετάσουμε τον τρόπο με τον οποίο μπορούμε να συλλέξουμε, να προετοιμάσουμε και να αναλύσουμε αυτά τα ηχητικά δεδομένα. Στόχος είναι να κατανοήσουμε τα διάφορα χαρακτηριστικά των ηχητικών δεδομένων και τις προκλήσεις που προκύπτουν κατά την επεξεργασία τους.
Κατά τη διαδικασία αυτή, θα χρησιμοποιήσουμε αλγόριθμους μηχανικής μάθησης και θα εκπαιδεύσουμε μοντέλα που να είναι ικανά να αναγνωρίζουν και να αναλύουν τα ηχητικά δεδομένα. Στη συνέχεια, θα αξιολογήσουμε την απόδοση των μοντέλων μας χρησιμοποιώντας μετρικές όπως η accuracy και F1-score.
Τέλος, θα προσφέρουμε συζήτηση και συμπεράσματα για τα ευρήματα της έρευνας και τη σημασία της για τον τομέα της ανάλυσης ηχητικών δεδομένων.

1.3 Ήχος

Ο ήχος είναι ένα φαινόμενο που προκαλείται από τις δονήσεις των μορίων σε ένα μέσο, όπως ο αέρας, το νερό ή ακόμη και στερεά αντικείμενα. Αυτές οι δονήσεις δημιουργούν περιοδικές περιπτύξεις και αραιώσεις του μέσου, προκαλώντας μια διαδοχή συμπιεστικών και αραιωτικών περιοχών που διαδίδονται μέσω αυτού. Στο πλαίσιο της ανάλυσης του ήχου, υπάρχουν πολλές πτυχές που μπορούν να εξεταστούν, και αυτές είναι σημαντικές για τον τομέα της ανάλυσης ηχητικών δεδομένων. Ο ήχος μπορεί να χαρακτηριστεί από διάφορες ιδιότητες, όπως η συχνότητα, αυτή η παράμετρος σχετίζεται με

το πόσες δονήσεις λαμβάνει χώρα κάθε δευτερόλεπτο, το πλάτος, που αφορά την απόσταση μεταξύ της μέγιστης και ελάχιστης τιμής μιας κυματομορφής και σχετίζεται με την ένταση του ήχου, την φάση, όπου αφορά την αναλογία μεταξύ της θέσης του ακροδέκτη μιας κυματομορφής και της αρχικής τιμής του αλλά και την χρονική διάρκεια που χρειάζεται μια κυματομορφή για να ολοκληρώσει έναν κύκλο.

Ο ήχος διαδίδεται μέσω διαφόρων μέσων, όπως τον αέρα, τα υγρά και στερεά. Η κατανόηση του ήχου και η ανάλυσή του με χρήση μηχανικής μάθησης είναι ένας σημαντικός τομέας έρευνας, όπου οι προκλήσεις και οι εφαρμογές είναι ευρείες. Ο τρόπος με τον οποίο ο ήχος διαδίδεται μπορεί να επηρεαστεί από παράγοντες όπως η θερμοκρασία, η πυκνότητα του μέσου και η πίεση. Εκτός από την φυσική του πτυχή, ο ήχος έχει μεγάλη σημασία στην καθημερινή μας ζωή. Χρησιμοποιείται στη μουσική, στον κινηματογράφο, στην τηλεόραση, στις τηλεπικοινωνίες και σε πολλούς άλλους τομείς. Επίσης, η μελέτη του ήχου και της αντίληψής μας γι' αυτόν συνδέεται με την ακουστική, την επεξεργασία ήχου και την ανάπτυξη νέων τεχνολογιών όπως οι ηχητικοί αλγόριθμοι και οι ακουστικές συσκευές .

Εκτός από τα βασικά χαρακτηριστικά του ήχου, όπως η συχνότητα και το πλάτος, υπάρχουν και άλλες σημαντικές έννοιες που σχετίζονται με τον ήχο, όπως η αντήχηση, η ηχητική ταχύτητα, η ηχητική ένταση, το φάσμα του ήχου και η ηχογράφηση. Η ανάλυση αυτών των πτυχών του ήχου είναι αναγκαία για πολλές εφαρμογές, συμπεριλαμβανομένων των συστημάτων αναγνώρισης και επεξεργασίας ηχητικών δεδομένων. Η αντήχηση (Reverberation) είναι η παρατεταμένη αντανάκλαση του ήχου από επιφάνειες σε ένα χώρο. Η αντήχηση παίζει σημαντικό ρόλο στη δημιουργία της ατμόσφαιρας και της ποιότητας του ήχου σε διάφορα περιβάλλοντα. Η ηχητική ταχύτητα αφορά τον ήχο που διαδίδεται με μια συγκεκριμένη ταχύτητα, η οποία εξαρτάται από το μέσο μέσα από το οποίο διαδίδεται. Όσον αφορά την ηχητική ένταση (Sound Intensity), είναι η ποσότητα της ενέργειας του ήχου που διαδίδεται ανά μονάδα επιφάνειας κάθετης στην κατεύθυνση της διάδοσης. Το φάσμα του ήχου (Sound Spectrum), περιλαμβάνει την ανάλυση της συνιστώσας συχνοτήτων ενός ήχου. Οι διάφορες συχνότητες συνιστούν το φάσμα του ήχου και επηρεάζουν την αντίληψη του ανθρώπινου αυτιού για τον ήχο. Από την άλλη, η ηχογράφηση (Sound Recording) αφορά την αποτύπωση των ηχητικών κυμάτων σε μέσα, όπως τον ήχο μέσω της χρήσης μικροφώνων και ηχητικών εγγραφικών συσκευών.

Η κατανόηση του ήχου έχει εφαρμογές σε πολλούς τομείς, όπως:

- **Μουσική:** Η αντίληψη των ηχητικών κυμάτων και των μουσικών τόνων είναι θεμελιώδης για τη δημιουργία και την απόλαυση της μουσικής .
- **Τηλεπικοινωνίες:** Η μετάδοση και η λήψη ήχου σε τηλεφωνικές συνομιλίες και ακουστικά μηνύματα είναι θεμελιώδης για τις επικοινωνίες.
- **Κινηματογράφος και Τηλεόραση:** Ο ήχος συνοδεύει τις εικόνες και δημιουργεί ατμόσφαιρα, δραματικότητα και αληθοφάνεια σε ταινίες και τηλεοπτικές εκπομπές.
- **Ιατρική:** Η ακουστική χρησιμοποιείται για τη διάγνωση και τη θεραπεία των ακουστικών παθήσεων.

Συνολικά, ο ήχος αποτελεί μια σημαντική πτυχή της καθημερινής μας ζωής και έχει ευρείες εφαρμογές σε πολλούς τομείς, ενώ η μελέτη του συνεχίζεται για τη βελτίωση της τεχνολογίας και της κατανόησης του φυσικού κόσμου.

1.4 Ηχητικά συμβάντα

Τα ηχητικά συμβάντα αντιπροσωπεύουν κάθε είδους ακουστικό γεγονός που συμβαίνει στο περιβάλλον μας. Αυτά μπορεί να περιλαμβάνουν φωνητικές εντολές, μουσική, περιβαλλοντικούς ήχους, θόρυβο, και πολλά άλλα. Η κατανόηση και αναγνώριση αυτών των ηχητικών συμβάντων έχει εφαρμογές σε πολλούς τομείς, όπως η αυτόματη ανίχνευση επικίνδυνων ήχων, η αναγνώριση φωνητικών εντολών, η παρακολούθηση περιβαλλοντικών συνθηκών και η βελτίωση της αλληλεπίδρασης ανθρώπου-μηχανής. Παραδείγματα ηχητικών συμβάντων μπορεί να είναι τα φωνητικά συμβάντα, όπως οι φωνητικές εντολές, οι διάλογοι, οι συνομιλίες, οι ομιλίες, κ.ά., η μουσική, δηλαδή μουσικά κομμάτια, μελωδίες, ρυθμοί και μουσικά όργανα. Επίσης, περιλαμβάνουν περιβαλλοντικούς ήχους, όπως η βροχή, ανέμους, πουλιά, θάλασσα, ηχητικα εφέ, όπως για παράδειγμα ήχους που παράγονται από ειδικά εφέ σε ταινίες, παιχνίδια και άλλα μέσα και φυσικά συμπεριλαμβάνεται και ο θόρυβος, δηλαδή ο ανεπιθύμητος ήχος που μπορεί να προκαλείται από παρεμβολές ή διαταραχές.

Τα ηχητικά συμβάντα έχουν ευρεία εφαρμογή σε πολλούς τομείς της καθημερινότητάς μας. Για παράδειγμα η επικοινωνία ανθρώπου-μηχανή. Η επικοινωνία ανθρώπου – μηχανής, αφορά την

αναγνώριση φωνητικών εντολών και την αλληλεπίδραση με τις μηχανές μέσω της ανθρώπινης φωνής. Φυσικά, τα ηχητικά συμβάντα εφαρμόζονται και στον τομέα της υγείας και ιατρικής. Η ανίχνευση ηχητικών συμβάντων στον ιατρικό τομέα μπορεί να βοηθήσει σε μεγάλο βαθμό τόσο στη διάγνωση όσο και στην παρακολούθηση των ασθενειών. Επίσης, μια ακόμη σημαντική εφαρμογή τους, αφορά την ευρύτερη ασφάλεια, όπως την αναγνώριση επικίνδυνων ήχων, πυροβολισμών, κραυγές για βοήθεια κ.ά .Σαφώς, ο τομέας της ψυχαγωγίας είναι ένας τομέας όπου εφαρμόζονται πολλά ηχητικά συμβάντα, με στόχο την δημιουργία συναρπαστικών ήχων σε βιντεοπαιχνίδια, ταινίες, κ.ά. Μια ακόμη εφαρμογή των ηχητικών συμβάντων αφορά τον τομέα της περιβαλλοντικής παρακολούθησης, δηλαδή την ανίχνευση ήχων από φυσικά φαινόμενα όπως σεισμοί, κύματα, κ.ά.

1.5 Συστήματα αναγνώρισης ηχητικών συμβάντων

Τα συστήματα αναγνώρισης ηχητικών συμβάντων αποτελούν ένα είδος τεχνολογίας που επιτρέπει στις μηχανές και τους υπολογιστές να αναγνωρίζουν, ταξινομούν και ερμηνεύουν ηχητικά γεγονότα από το περιβάλλον τους. Η βασική ιδέα είναι τα υπολογιστικά συστήματα να εξοπλιστούν με την ικανότητα να αναλύουν ηχητικά σήματα και να τα μετατρέπουν σε κατηγορίες ή πληροφορίες που έχουμε καθορίσει.Η ανάπτυξη των τεχνικών αναγνώρισης ηχητικών συμβάντων έχει γίνει δυναμική χάρη στην πρόοδο στον τομέα της τεχνητής νοημοσύνης και της μηχανικής μάθησης. Αρχικά, τα συστήματα βασίζονταν σε χειροκίνητους κανόνες και προγραμματισμένες λογικές διαδικασίες. Ωστόσο, η πολυπλοκότητα των ηχητικών δεδομένων καθιστούσε δύσκολη την ανάπτυξη ευέλικτων και ακριβών συστημάτων.

Με την εισαγωγή των αλγορίθμων μηχανικής μάθησης, όπως τα νευρωνικά δίκτυα, τα συστήματα αναγνώρισης ηχητικών συμβάντων έχουν βελτιωθεί σημαντικά. Τα βαθιά νευρωνικά δίκτυα, ειδικά τα αναδραστικά νευρωνικά δίκτυα (RNNs) και τα συνελικτικά νευρωνικά δίκτυα (CNNs), έχουν τη δυνατότητα να εξάγουν συνθέτες χαρακτηριστικά από τα ηχητικά δεδομένα και να μάθουν τα μοτίβα που αναγνωρίζουν διάφορα ηχητικά συμβάντα.Η ανάπτυξη ενός αποτελεσματικού συστήματος αναγνώρισης ηχητικών συμβάντων χρησιμοποιώντας τεχνικές μηχανικής μάθησης απαιτεί προσεκτική σχεδίαση, επιλογή κατάλληλων μεθόδων και εκπαίδευση με υψηλής ποιότητας δεδομένα. Η συνεχής

εξέλιξη των τεχνικών μηχανικής μάθησης ανοίγει νέες προοπτικές για τη βελτίωση της ακρίβειας και της απόδοσης των συστημάτων αναγνώρισης ηχητικών συμβάντων σε ποικίλα περιβάλλοντα.

Για την αναγνώριση των ηχητικών συμβάντων, χρησιμοποιούνται διάφορες τεχνολογίες, όπως:

- **Σήματα και Συστήματα:** Ανάλυση σήματος και συστημάτων για την κατανόηση των χαρακτηριστικών των ηχητικών συμβάντων
- **Τεχνητή Νοημοσύνη:** Χρήση αλγορίθμων μηχανικής μάθησης, βαθιών νευρωνικών δικτύων και άλλων μεθόδων για την αναγνώριση και ταξινόμηση ηχητικών προτύπων

Τα συστήματα αναγνώρισης ηχητικών συμβάντων περνούν συνήθως από τρια βασικά στάδια. Το πρώτο στάδιο αφορά την **π**ροεπεξεργασία σήματος.Σε αυτό το στάδιο, τα ηχητικά σήματα υποβάλλονται σε προεπεξεργασία για τη μείωση του θορύβου, την εξαγωγή χρήσιμων χαρακτηριστικών και την προσαρμογή της ανάλυσης τους. Έπειτα, ακολουθεί το δεύτερο στάδιο, το οποίο περιλαμβάνει την **ε**ξαγωγή χαρακτηριστικών**.** Σε αυτό το στάδιο, τα χρήσιμα χαρακτηριστικά εξάγονται από τα προ επεξεργασμένα σήματα. Αυτά τα χαρακτηριστικά μπορεί να είναι συχνότητες, χρονικές διακυμάνσεις κ.ά.Το τρίτο βασικό στάδιο αφορά την κατηγοριοποίηση. Σε αυτό το στάδιο, τα εξαγόμενα χαρακτηριστικά χρησιμοποιούνται από μοντέλα μηχανικής μάθησης (όπως νευρωνικά δίκτυα) για την αναγνώριση και ταξινόμηση των ηχητικών συμβάντων σε προκαθορισμένες κατηγορίες.

Πρέπει να τονίσουμε ότι παρά την πρόοδο της τεχνολογίας, υπάρχουν ακόμη αρκετές προκλήσεις που αφορούν την αναγνώριση ηχητικών συμβάντων.Οι προκλήσεις αυτές, περιλαμβάνουν την πολυπλοκότητα των ηχητικών περιβαλλόντων, τις παρεμβολές και τον θόρυβο, καθώς και την ανάγκη για πιο μεγάλα αλλά και ποιοτικά σύνολα δεδομένων για την εκπαίδευση ακριβών μοντέλων.

2. Μηχανική μάθηση

Η μηχανική μάθηση αποτελεί τομέα της τεχνητής νοημοσύνης και επικεντρώνεται στην ανάπτυξη αλγορίθμων και μοντέλων που επιτρέπουν στις μηχανές να μαθαίνουν από δεδομένα και να βελτιώνουν την απόδοσή τους με την εμπειρία. Ο βασικός στόχος της μηχανικής μάθησης είναι να δημιουργήσει μοντέλα που μπορούν να γενικεύουν από δεδομένα, ώστε να μπορούν να προβλέπουν

και να λαμβάνουν αποφάσεις για νέα και απρόβλεπτα δεδομένα. Αυτό επιτυγχάνεται με τη χρήση διαφόρων αλγορίθμων και τεχνικών που επιτρέπουν στα μοντέλα να εντοπίζουν μοτίβα, χαρακτηριστικά και συσχετίσεις στα δεδομένα εκπαίδευσης.

Η μηχανική μάθηση χωρίζεται σε τρεις βασικές κατηγορίες:

1. **Επιβλεπόμενη Μάθηση (Supervised Learning):** Σε αυτό το είδος, το μοντέλο εκπαιδεύεται με ζεύγη εισόδου-εξόδου, όπου για κάθε είσοδο υπάρχει μια αντίστοιχη επιθυμητή έξοδο. Ο στόχος είναι να μάθει να προβλέπει την έξοδο για νέες εισόδους. Οι αλγόριθμοι που ανήκουν σε αυτήν την κατηγορία είναι ικανοί να εφαρμόσουν τις γνώσεις που έχουν αποκτήσει από το παρελθόν σε νέα σύνολα δεδομένων. Ξεκινώντας με την ανάλυση ενός γνωστού σετ δεδομένων, αυτοί οι αλγόριθμοι δημιουργούν μια συνάρτηση με σκοπό την πρόβλεψη των αντίστοιχων εξόδων. Επιπλέον, οι αλγόριθμοι έχουν τη δυνατότητα να συγκρίνουν τα αποτελέσματά τους με τα σωστά αποτελέσματα και να ανιχνεύουν τυχόν σφάλματα, με σκοπό να προχωρήσουν σε τροποποιήσεις στο μοντέλο τους, ώστε να βελτιώσουν την ακρίβεια των προβλέψεών τους.
2. **Μη επιβλεπόμενη Μάθηση (Unsupervised Learning):** Το μοντέλο εκπαιδεύεται χωρίς τη χρήση ετικετών εξόδου. Ο στόχος είναι να ανακαλύψει μοτίβα ή δομές στα δεδομένα, χωρίς να γνωρίζει εκ των προτέρων τις επιθυμητές εξόδους. Η μη επιβλεπόμενη μηχανική μάθηση χρησιμοποιείται σε συστήματα που μπορούν να εξάγουν μια συνάρτηση, η οποία αναδεικνύει μια κρυφή δομή, χρησιμοποιώντας δεδομένα που δεν έχουν ετικέτες. Αυτά τα συστήματα δεν είναι σε θέση να προβλέψουν ακριβώς την επιθυμητή έξοδο, αλλά μέσα από την εξερεύνηση των δεδομένων, μπορούν να βγάλουν συμπεράσματα σχετικά με την ολική δομή των δεδομένων.
3. **Ενισχυτική Μάθηση (Reinforcement Learning):** Σε αυτήν την κατηγορία, το μοντέλο μαθαίνει να προσαρμόζεται σε ένα περιβάλλον αντλώντας εμπειρία μέσω δοκιμών και σφαλμάτων. Η ενισχυτική μηχανική μάθηση αποτελεί μια προσέγγιση που αλληλεπιδρά με το περιβάλλον, αναλαμβάνοντας μη προκαθορισμένες ενέργειες, τις οποίες ακολουθεί αντιδρώντας σε απροσδιόριστες καταστάσεις. Μέσω αυτής της διαδικασίας, η μέθοδος καταφέρνει να αυτοματοποιήσει την αναζήτηση των βέλτιστων στρατηγικών, με στόχο τη μεγιστοποίηση της απόδοσης του αλγορίθμου. Αυτό επιτυγχάνεται μέσω της συνεχούς αλληλεπίδρασης με το περιβάλλον και της χρήσης ανατροφοδότησης για την κατεύθυνση της μάθησης.

2.1 Τεχνικές μηχανικής μάθησης

Οι τεχνικές μηχανικής μάθησης έχουν αποδειχθεί ως πολύ αποτελεσματικές για την αναγνώριση ηχητικών συμβάντων. Αυτές περιλαμβάνουν:

1. **Κατηγοριοποίηση (Classification):** Η ταξινόμηση ή κατηγοριοποίηση των ηχητικών συμβάντων σε προκαθορισμένες κατηγορίες, όπως "φωνητικές εντολές," "μουσική," "θόρυβος," γίνεται με την εκπαίδευση ενός μοντέλου που μπορεί να αναγνωρίσει τα χαρακτηριστικά που διακρίνουν κάθε κατηγορία
2. **Ανίχνευση Ανωμαλιών (Anomaly Detection):** Αντί να επικεντρωθεί στην κανονική αναγνώριση κατηγοριών, αυτή η προσέγγιση επικεντρώνεται στον εντοπισμό μη συνηθισμένων ήχων που μπορεί να υποδεικνύουν ανωμαλίες σε ένα σύστημα.
3. **Νευρωνικά Δίκτυα (Neural Networks):** Οι βαθιές νευρωνικές δομές όπως τα αναδραστικά νευρωνικά δίκτυα (RNNs) και τα συνελικτικά νευρωνικά δίκτυα (CNNs) έχουν χρησιμοποιηθεί ευρέως για την αναγνώριση ηχητικών συμβάντων. Τα RNNs είναι κατάλληλα για χρονική σειρά ήχων, ενώ τα CNNs μπορούν να εξάγουν χαρακτηριστικά από χρονικά και συχνοτικά περιβάλλοντα.
4. **Συνδυασμός Πληροφοριών (Fusion of Information):** Ορισμένες φορές, η χρήση πολλαπλών αισθητήρων μπορεί να βοηθήσει στην αύξηση της ακρίβειας της αναγνώρισης. Συνδυάζοντας τις πληροφορίες από διάφορες πηγές μπορεί να δοθεί πληρέστερη εικόνα του ακουστικού περιβάλλοντος.

Φυσικά, η ανάπτυξη ενός αξιόπιστου συστήματος αναγνώρισης ηχητικών συμβάντων απαιτεί την αντιμετώπιση πολλών προκλήσεων. Ορισμένες από αυτές περιλαμβάνουν:

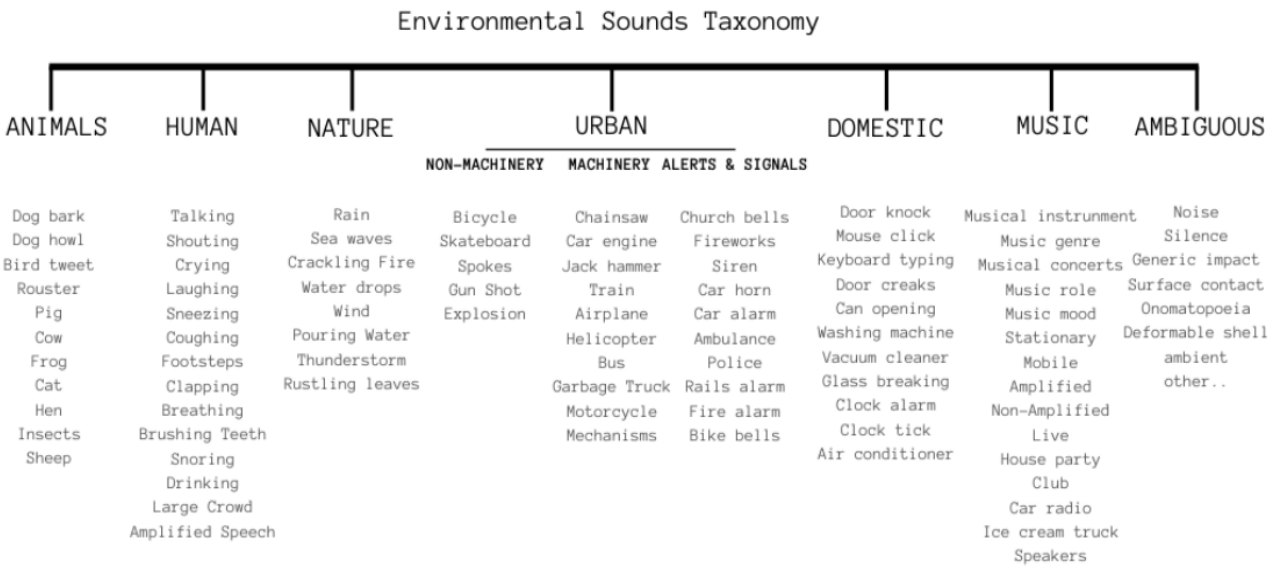
1. **Δεδομένα Εκπαίδευσης:** Η συλλογή αρκετών και ποιοτικών δεδομένων είναι απαραίτητη για την εκπαίδευση του μοντέλου. Η δημιουργία και επεξεργασία αυτών των δεδομένων μπορεί να είναι χρονοβόρα.
2. **Επιλογή Χαρακτηριστικών:** Η επιλογή των κατάλληλων χαρακτηριστικών των ηχητικών συμβάντων είναι κρίσιμη. Αυτά τα χαρακτηριστικά θα πρέπει να αντιπροσωπεύουν τα διακριτικά χαρακτηριστικά των κατηγοριών.

3. **Διαχείριση Θορύβου:** Ο θόρυβος και οι παρεμβολές μπορούν να επηρεάσουν την ακρίβεια του συστήματος. Οι μέθοδοι για τη μείωση του θορύβου και τη βελτίωση της ποιότητας των σημάτων είναι σημαντικές.
4. **Προσαρμογή σε Νέα Σενάρια:** Το σύστημα θα πρέπει να είναι ευέλικτο ώστε να μπορεί να αναγνωρίζει ήχους σε νέα περιβάλλοντα ή σενάρια.

2.2 Σύνολα Δεδομένων

Τα δεδομένα αποτελούν κρίσιμο και αναπόσπαστο κομμάτι σε κάθε ανάπτυξη συστήματος ταξινόμησης. Τα δεδομένα παρέχονται από τους παρόχους σε μια οργανωμένη μορφή, συνθέτοντας ένα σύνολο δεδομένων. Στο πλαίσιο της επιβλεπόμενης μάθησης που χρησιμοποιείται σε αυτήν την εργασία, τα δεδομένα περιλαμβάνουν ηχητικά αρχεία που έχουν προηγουμένως ταξινομηθεί σε διάφορες κλάσεις, με κάθε αρχείο να συνοδεύεται από μια ετικέτα που υποδεικνύει την αντίστοιχη κλάση οπως βλέπουμε παρακάτω στο σχήμα (1).

Για τις ανάγκες της υλοποίησης, το αρχικό σύνολο δεδομένων χωρίζεται σε φακέλους, οι οποίοι είναι οργανωμένοι ανάλογα με τα στάδια της εκπαίδευσης και της επικύρωσης. Η διανομή των αρχείων σε αυτούς τους φακέλους είναι καίριας σημασίας, για να διατηρηθεί μια ομοιόμορφη κατανομή των κλάσεων σε κάθε φάκελο. Αυτός ο τρόπος διανομής προαναγγέλλεται εκ των προτέρων, λαμβάνοντας υπόψη το μέγεθος του συνόλου δεδομένων. Αυτό το βήμα συμβάλλει σημαντικά στην αντικειμενική σύγκριση της μέσης απόδοσης των φακέλων από τα αναπτυσσόμενα συστήματα, καθώς και στην αντιμετώπιση του προβλήματος υπερ εκπαίδευσης για υποσύνολα κλάσεων που παρατηρείται όταν τα δεδομένα έχουν ανομοιόμορφη κατανομή. Σύμφωνα με (JP Bello κ.ά. 2005) [24] σε πολλές περιπτώσεις, τα μεταδεδομένα που παρέχονται περιέχουν σημαντικές πληροφορίες σχετικά με τη χρήση του συγκεκριμένου σετ δεδομένων.



Σχημα 1: ΓΕΝΙΚΗ ΤΑΞΙΝΟΜΙΑ ΤΩΝ ΠΕΡΙΒΑΛΛΟΝΤΙΚΩΝ ΚΛΑΣΕΩΝ

Σήμερα, υπάρχουν διάφορα σετ δεδομένων που μπορούν να χρησιμοποιηθούν για το πρόβλημα της ανάκτησης περιβαλλοντικών ήχων. Αυτά τα σετ δεδομένων διαφέρουν σε πολλούς παράγοντες, όπως το μέγεθος, το περιεχόμενο που καλύπτεται από τις κλάσεις, η ιεραρχία των κλάσεων, η δομή των φακέλων, η διαθεσιμότητα μεταδεδομένων και τα τεχνικά χαρακτηριστικά των ηχητικών αρχείων.

Ένα από τα πιο γνωστά σετ δεδομένων είναι το "Urban Sound 8k". Αυτό το σετ δεδομένων έχει χρησιμοποιηθεί στην παρούσα εργασία και περιγράφονται παρακάτω.

Το σετ δεδομένων UrbanSound [26] αντιπροσωπεύει μια πρώιμη έκδοση του UrbanSound8K και περιλαμβάνει 1302 αρχεία που έχουν ταξινομηθεί σε 10 διαφορετικές κλάσεις, αντιπροσωπεύοντας περιβαλλοντικούς ήχους αστικού περιεχομένου. Αυτά τα δεδομένα προέρχονται από το σύστημα ταξινομίας urban sound taxonomy. Όσον αφορά τα τεχνικά χαρακτηριστικά των αρχείων που περιέχει, αυτά μπορεί να είναι είτε μονοφωνικά είτε στερεοφωνικά, με ποικίλους ρυθμούς δειγματοληψίας, βαθμούς κβάντισης και χρονικές διάρκειες. Αυτά τα χαρακτηριστικά αντλήθηκαν από τον ιστότοπο Freesound [25] . Παρ' όλα αυτά, τα αρχεία διαθέτουν αξιόπιστες ετικέτες αναφοράς για τις κλάσεις.

Επιπλέον, τα μεταδεδομένα περιέχουν ετικέτες που δείχνουν τα σημεία έναρξης και λήξης των αρχείων πηγής από τα οποία αντλήθηκαν τα ηχητικά αρχεία. Η βασική διαφορά σε σύγκριση με το

UrbanSound8K είναι ότι τα ηχητικά αρχεία του UrbanSound μπορεί να περιλαμβάνουν πολυφωνικά συμβάντα, παρά το γεγονός ότι διαθέτουν μοναδικές ετικέτες για τις κλάσεις.

Αντίθετα, το σετ δεδομένων Urban-SED [27] αποτελείται από συνθετικά δεδομένα που προέρχονται από τα αρχικά δεδομένα του UrbanSound8K όπως περιγράφεται στο έργο (Abdoli, κ.ά. 2019) [3] . Αυτό το σετ επικεντρώνεται στην ενίσχυση των δεδομένων μέσω τεχνικών σύνθεσης, με τις οποίες δημιουργούνται νέα αρχεία που περιέχουν επικαλυπτόμενα συμβάντα. Κάθε αρχείο μπορεί να αντιστοιχεί σε μια ή περισσότερες ετικέτες και να ανήκει είτε σε γενικές είτε σε ειδικές κλάσεις σε μια ιεραρχική δομή. Ο σκοπός αυτών των τεχνικών είναι να αντικατοπτρίζουν με μεγαλύτερη ακρίβεια την πραγματική ποικιλία των ηχητικών συμβάντων σε πραγματικά περιβάλλοντα.

Στο Urban-SED, τα αρχεία δεδομένων περιλαμβάνουν επιπλέον παρασκηνιακό θόρυβο, ο οποίος παρουσιάζει μικρό εύρος σηματοθορυβικού λόγου. Επιπλέον, τα αρχικά δεδομένα μετασχηματίζονται χρονικά και σε ύψος τόνου για να προσομοιώσουν διαφορετικές εκδοχές των ηχητικών συμβάντων σε διάφορα περιβάλλοντα. Αυτές οι τεχνικές επίδρασης στα δεδομένα αυξάνουν την ικανότητα γενίκευσης των αλγορίθμων ταξινόμησης, καθιστώντας τα πιο αποδοτικά στην αντιμετώπιση άγνωστων δεδομένων. Συνολικά, το σετ δεδομένων Urban-SED αποτελείται από 10,000 ηχητικά αρχεία περιβαλλοντικών σκηνών (soundscapes) διάρκειας 30 ωρών, περιλαμβάνοντας περίπου 50,000 ηχητικά συμβάντα.

Κάθε αρχείο παρέχει ετικέτες που περιλαμβάνουν τα χρονικά στιγμιότυπα έναρξης και λήξης του συμβάντος μέσα στην προκαθορισμένη διάρκεια των 10 δευτερολέπτων για κάθε αρχείο. Αυτό επιτρέπει την ακριβή τοποθέτηση των ηχητικών συμβάντων εντός της χρονικής διάρκειας των αρχείων. Σημειώνεται επίσης ότι, όπως και στο UrbanSound8k, ο ρυθμός δειγματοληψίας, ο βαθμός κβάντισης και ο παράγοντας στερεοφωνίας για τα αρχεία παρουσιάζουν μεγάλη ποικιλομορφία, αντικατοπτρίζοντας τη διαφοροποίηση στις τεχνικές παραμέτρους των ηχογραφήσεων.

Πολλά σετ δεδομένων προσφέρουν αξιόπιστες ετικέτες για τα αρχεία τους. Ωστόσο, η παροχή αυτών των ετικετών δεν είναι καθολική και μπορεί να αντιμετωπίζεται με δυσκολία, ειδικά όταν απαιτείται χειροκίνητη ταξινόμηση από πολλαπλούς ακροατές. Ο παράγοντας αυτός μπορεί να οδηγήσει σε σφάλματα ετικετοποίησης, λόγω της δυσκολίας στην αναγνώριση ή ταξινόμηση ορισμένων ήχων.

Για να αντιμετωπιστεί αυτό το θέμα, υπάρχουν προσεγγίσεις όπως η εφαρμογή φίλτρων για την απομάκρυνση δειγμάτων θορύβου από τα δεδομένα. Επίσης, μπορεί να γίνει προσπάθεια διόρθωσης σφαλμάτων ετικετοποίησης με τη χρήση τεχνικών, όπως η εφαρμογή μοντέλων για τη βελτίωση της αξιοπιστίας των ετικετών. Παρόμοια με τον κλάδο του "big data", όπου η έννοια του θορύβου

δεδομένων αναφέρεται σε ανεπιθύμητες και παραμορφωμένες πληροφορίες, ο περιορισμός αυτός, στην παροχή αξιόπιστων ετικετών μπορεί να αντιμετωπιστεί με παρόμοιες πρακτικές.
Τέλος, προσπάθειες για την ενίσχυση της αξιοπιστίας των ετικετών μπορούν να βελτιώσουν τις επιδόσεις των μοντέλων ταξινόμησης, καθώς η ποιότητα των ετικετών επηρεάζει απευθείας την απόδοσή τους.

3. Περιγραφή δεδομένων - Πειράματα

Οι στόχοι της παρούσας μελέτης, σχετίζονται κυρίως στην ανάπτυξη τεχνικών για την ταξινόμηση συστημάτων στο πλαίσιο της μηχανικής μάθησης, ειδικότερα όσον αφορά την αναγνώριση περιβαλλοντικών ήχων.
Το σετ δεδομένων που χρησιμοποιήθηκε στην εργασία είναι το UrbanSound8k .Όπως βλέπουμε απο το σχήμα (2) το UrbanSound8k , περιλαμβάνει 8732 μονοφωνικά ηχητικά αρχεία θορύβων.Τα αρχεία αυτά, ταξινομούνται σε 10 κλάσεις περιβαλλοντικών ήχων αστικού περιεχομένου. Το σύνολο δεδομένων παρέχεται σε 10 προκαθορισμένους φακέλους, με μη ομοιόμορφη κατανομή κλάσεων και πρόκειται για αρχεία είτε μονοφωνικά είτε στερεοφωνικά, με διάφορα επίπεδα δειγματοληψίας (από 8 kHz έως 96 kHz) και βαθμού κβάντισης (από 8-bit έως 32-bit) κωδικοποίησης. Επίσης, παρέχονται αξιόπιστες ετικέτες αναφοράς για τις κλάσεις.
Το σύνολο δεδομένων παρέχει αξιόπιστες ετικέτες για τα αρχεία τους. Το UrbanSound8k περιλαμβάνει μονοφωνικά ηχητικά αρχεία με ποικιλία σε χαρακτηριστικά όπως δειγματοληψία και κβάντιση. Η χρήση αυτού του σετ δεδομένων, επιτρέπει την αξιολόγηση αλγορίθμων ταξινόμησης ή μοντέλα μηχανικής μάθησης, για την αναγνώριση και ταξινόμηση ηχητικών σημάτων στις αντίστοιχες κλάσεις.

Τεχνικά Χαρακτηριστικά του σετ δεδομένων	
Χαρακτηριστικά	**UrbanSound8k**
Πλήθος αρχείων	**8732**
Πλήθος κλάσεων	**10**
Δομή κλάσεων	**UrbanSound taxonomy**
Μέση διάρκεια αρχείων	**Mean=3.6secs(std=1 sec)**
Εύρος διάρκειας αρχείων	**0.05-4.11 seconds**
Φάκελοι	**10(806-990/fold)**
Στερεοφωνία	**Mono/Stereo**
Δειγματοληψία	**8 KHz-96 KHz**
Βαθμός κβάντισης	**8-24-bit,FLOAT,ADPCM**

Σχημα 2.Τεχνικά Χαρακτηριστικά του σετ δεδομένων

4. Μεθοδολογία Πειράματος

Σε αυτήν την εργασία, εξετάζουμε τη μεθοδολογία για την αναγνώριση ήχων, ένα σημαντικό πρόβλημα στον χώρο της επεξεργασίας ηχητικών δεδομένων. Ο σκοπός είναι να αναπτυχθεί ένα νευρωνικό δίκτυο που μπορεί να αναγνωρίζει αυτόματα διάφορους τύπους ήχων, όπως τον ήχο ενός αυτοκινήτου, μουσική, σφυροκόπησης, και πολλούς άλλους.

Η πρώτη φάση του πειράματος είναι η συλλογή και προεπεξεργασία των ηχητικών δεδομένων. Αυτά τα δεδομένα αντλούνται από το UrbanSound8K dataset, ένα σύνολο ήχων που καλύπτει διάφορες πραγματικές καταστάσεις. Είναι σημαντικό να τα προεπεξεργαστούμε με προσοχή, ώστε να εξάγουμε χρήσιμα χαρακτηριστικά που θα χρησιμοποιήσουμε στο μοντέλο μας.

Ένα σημαντικό βήμα στην προεπεξεργασία είναι η κωδικοποίηση των κατηγοριών των ήχων. Αυτό μας επιτρέπει να μετατρέψουμε τις ετικέτες των ήχων σε μια μορφή που μπορεί να επεξεργαστεί από το νευρωνικό δίκτυο.

Το κύριο εργαλείο που χρησιμοποιείται σε αυτήν την εργασία είναι το νευρωνικό δίκτυο. Αυτό το δίκτυο έχει σχεδιαστεί με προσοχή και αποτελείται από διάφορα επίπεδα, συμπεριλαμβανομένων συνελίξεων και επιπέδων MaxPooling για την εξαγωγή χαρακτηριστικών, ενώ χρησιμοποιεί και επίπεδα Dropout για την αποφυγή υπερ εκπαίδευσης. Το τελευταίο επίπεδο είναι πλήρως συνδεδεμένο και χρησιμοποιείται για την τελική ταξινόμηση των ήχων σε διάφορες κατηγορίες.

Μετά την εκπαίδευση του μοντέλου σε ένα σύνολο δεδομένων, ακολουθεί η αξιολόγησή του σε ένα ανεξάρτητο σύνολο δεδομένων. Αυτός είναι ο τρόπος με τον οποίο μετράμε την απόδοση του μοντέλου μας.

Τέλος, γίνεται μια εκτενής ανάλυση των αποτελεσμάτων, περιλαμβανομένης της απόδοσης του μοντέλου σε κάθε κατηγορία ήχου, με τη χρήση πινάκων σύγχυσης και αναφορών ταξινόμησης. Αυτό μας επιτρέπει να κατανοήσουμε ποιες κατηγορίες ήχων αναγνωρίζονται καλύτερα από το μοντέλο και πού ενδέχεται να υπάρχουν προβλήματα.

4.1 Βιβλιοθήκες

Κάποιες από τις κύριες βιβλιοθήκες που χρησιμοποιήθηκαν είναι η librosa για την ανάλυση ηχητικών δεδομένων, το numpy για την επεξεργασία δεδομένων, το tensorflow για την κατασκευή και την εκπαίδευση του μοντέλου, καθώς και άλλες βιβλιοθήκες για την ανάλυση και την οπτικοποίηση αποτελεσμάτων.

4.2 Προεπεξεργασία Δεδομένων (Data Preprocessing)

Στο ακόλουθο στάδιο, τα ηχητικά αρχεία υποβάλλονται σε προεπεξεργασία προκειμένου να είναι κατάλληλα για την εκπαίδευση του μοντέλου. Αυτό περιλαμβάνει τη χρήση των χαρακτηριστικών

MFCC (Mel-Frequency Cepstral Coefficients) για την αναπαράσταση των ηχητικών σημάτων. Η επεξεργασία γίνεται με τη χρήση της βιβλιοθήκης librosa. Τα MFCC παρέχουν σημαντικές πληροφορίες για το περιεχόμενο των ηχητικών αρχείων.

Στα πειράματά μας, χρησιμοποιούμε όπως είπαμε το σύνολο δεδομένων Urban Sound 8K, το οποίο αποτελείται από 8732 αποσπάσματα ήχου κατηγοριοποιημένα σε 10 μοναδικές κατηγορίες αστικών ήχων. Κάθε αρχείο ήχου στο σύνολο δεδομένων είναι ετικετοποιημένο με την αντίστοιχη κλάση του, όπως 'dog bark', 'children playing', 'car horn' κ.λπ. Το πρώτο βήμα στα πειράματα μας περιελάμβανε τον μετασχηματισμό των αρχικών αρχείων ήχου σε μια πιο διαχειρίσιμη και χρήσιμη μορφή για ταξινόμηση. Το κάνουμε αυτό εξάγοντας τα MFCCs από τα ηχητικά σήματα και αλλάζοντας το μέγεθος των MFCCs σε ένα ομοιόμορφο σχήμα [40, 173]. Επίσης κωδικοποιούμε τις ετικέτες των κλάσεων σε μια κατηγορική μορφή κατάλληλη για να χρησιμοποιηθεί από το μοντέλο ταξινόμησης μας.

Τέλος,, τα δεδομένα διαχωρίζονται σε σύνολα εκπαίδευσης και ελέγχου. Αυτό γίνεται για την αξιολόγηση της απόδοσης του μοντέλου σε ανεξάρτητα δεδομένα. Ο διαχωρισμός αυτός βασίζεται στην αρχή της αξιοπιστίας του μοντέλου, διότι ένα μοντέλο πρέπει να δοκιμαστεί σε δεδομένα που δεν έχει ξαναδεί κατά τη διάρκεια της εκπαίδευσής του.

4.3 Αρχιτεκτονική Μοντέλου (Model Building)

Το μοντέλο ταξινόμησης κατασκευάστηκε χρησιμοποιώντας ένα Συνελικτικό Νευρωνικό Δίκτυο (CNN), λαμβάνοντας υπόψη τις χωρικές συσχετίσεις στον τομέα συχνότητας των προ επεξεργασμένων ηχητικών δεδομένων. Η αρχιτεκτονική του μοντέλου είναι η εξής:

- Τρία συνελικτικά επίπεδα (Conv2D) με συναρτήσεις ενεργοποίησης rectified linear unit (ReLU), ακολουθούμενα από επίπεδα μέγιστης υποδοχής (MaxPooling2D) για τη μείωση των χωρικών διαστάσεων των εισερχόμενων δεδομένων.
- Τα επίπεδα Dropout χρησιμοποιήθηκαν μετά τα δεύτερο και τρίτο συνελευτικά επίπεδα για τη μείωση του overfitting, ορίζοντας τυχαία ένα κλάσμα των εισερχόμενων μονάδων σε 0 κατά τη διάρκεια κάθε ενημέρωσης κατά την εκπαίδευση.

- Το επίπεδο GlobalAveragePooling2D χρησιμοποιήθηκε για τη μείωση των χωρικών διαστάσεων πριν από τα πλήρως συνδεδεμένα (Dense) επίπεδα.
- Το τελικό επίπεδο Dense χρησιμοποιεί μια συνάρτηση ενεργοποίησης softmax για την εξαγωγή της κατανομής πιθανότητας πάνω στις 10 κλάσεις

4.4 Υπερ-παράμετροι για την Εκπαίδευση του Μοντέλου (Model Training)

Το μοντέλο μας χρησιμοποιεί τον βελτιστοποιητή Adam, με ρυθμό μάθησης 0.001, και categorical crossentropy ως συνάρτηση απώλειας (loss function), η οποία είναι κατάλληλη για προβλήματα ταξινόμησης πολλαπλών κλάσεων. Για τη βελτίωση της διαδικασίας εκπαίδευσης, υλοποιήθηκε ένας μηχανισμός πρόωρης διακοπής (early stopping) για την παρακολούθηση του loss function και διακοπή της εκπαίδευσης μόλις σταματήσει να μειώνεται, προλαμβάνοντας τυχόν πιθανό overfitting. Το μοντέλο εκπαιδεύτηκε για έως και 15 εποχές, με batch size = 8.

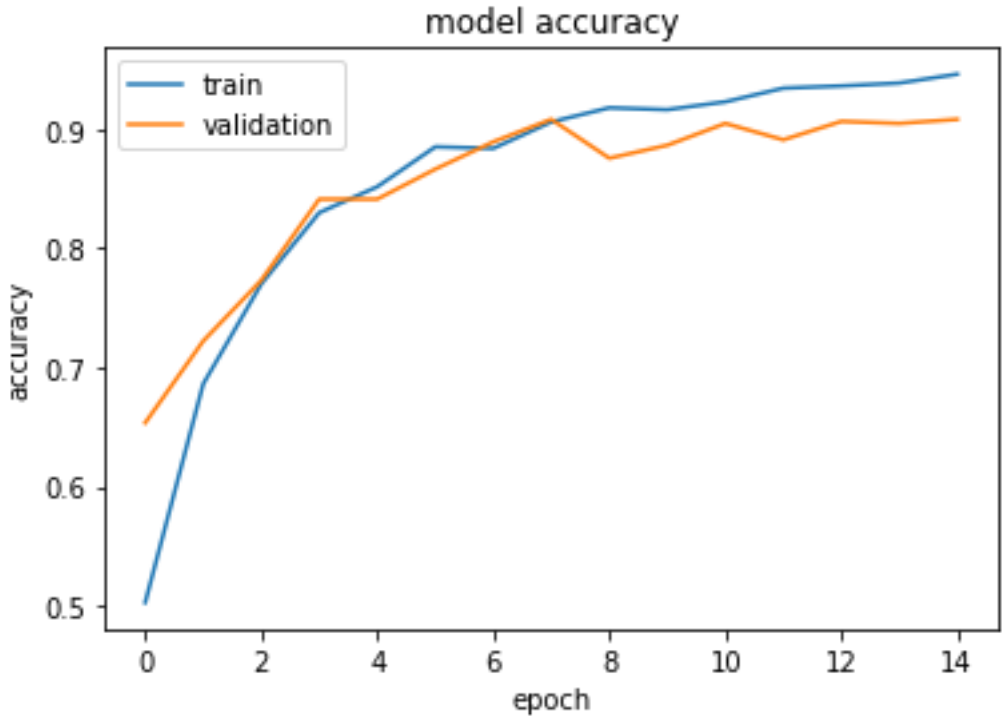
5. Αποτελέσματα

Τα αποτελέσματα που αντλούμε αφορούν τη διαδικασία εκπαίδευσης του μοντέλου μας. Οι εκπαιδευτικές εποχές είναι σημαντικές στην εκπαίδευση ενός νευρωνικού δικτύου, και αυτά τα αποτελέσματα παρουσιάζουν την εξέλιξη της ακρίβειας (accuracy) και της συνάρτησης απώλειας (loss) κατά τη διάρκεια των 15 εκπαιδευτικών εποχών.

5.1 Accuracy Evaluation - Διαγράμματα

Το accuracy είναι μια σημαντική μετρική που μετρά το ποσοστό των σωστών προβλέψεων του μοντέλου στα δεδομένα ελέγχου. Βλέπουμε πώς η ακρίβεια του μοντέλου υπολογίζεται και παρουσιάζεται για το σύνολο εκπαίδευσης και το σύνολο ελέγχου κατά τη διάρκεια των εποχών.
Η γραφική αναπαράσταση του accuracy μεταξύ των εποχών γίνεται μέσω γραφημάτων που δείχνουν την εξέλιξη του κατά την εκπαίδευση για το training και το validation set. Βλέπουμε τις αντίστοιχες καμπύλες στο παρακάτω σχήμα (3). Οι γραφικές αυτές αναπαραστάσεις είναι χρήσιμες για τον

εντοπισμό τυχόν προβλημάτων υπερ εκπαίδευσης ή υποεκπαίδευσης καθώς και για την επιλογή του βέλτιστου αριθμού εποχών για την εκπαίδευση.



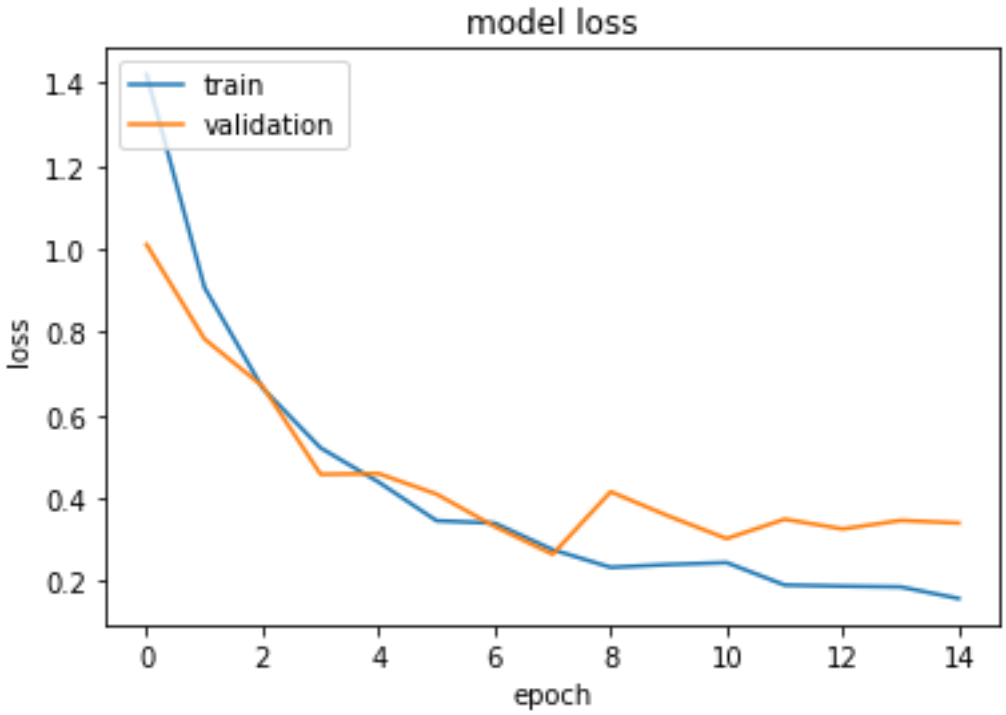
Σχημα 3. καμπύλες για το accuracy στο train και validation set

5.2 Loss Evaluation - Διαγράμματα

Το Loss είναι μια μετρική που μετρά το σφάλμα των προβλέψεων του μοντέλου σε σχέση με τις πραγματικές τιμές. Στον κώδικα που παρέχουμε, υπολογίζεται και αναπαρίσταται το Loss για το σύνολο εκπαίδευσης και ελέγχου κατά τη διάρκεια των εποχών.

Η γραφική αναπαράσταση όπως βλέπουμε στο σχήμα (4) του Loss επίσης είναι χρήσιμη για τον έλεγχο της σύγκλισης του μοντέλου κατά τη διάρκεια της εκπαίδευσης. Στόχος είναι η ελαχιστοποίηση του, καθώς αυτό υποδηλώνει καλύτερη απόδοση.

Συνολικά, η αξιολόγηση της ακρίβειας και της απώλειας αποτελεί κρίσιμη διαδικασία για την κατανόηση και τη βελτίωση της απόδοσης του μοντέλου, καθώς και για την παρακολούθηση της σύγκλισης της εκπαίδευσης.



Σχημα 4. καμπύλες για το loss στο train και validation set

Γενικά, το μοντέλο μας φαίνεται να εκπαιδεύεται καλά και να βελτιώνει την απόδοσή του καθώς περνάει ο χρόνος εκπαίδευσης. Το οτι αυξανεται το accuracy αλλά και μειώνεται παράλληλα το loss είναι θετικά σημάδια. Το μοντέλο μοιάζει να έχει καλή απόδοση στο σύνολο ελέγχου (validation set).

5.3 Αξιολόγηση Μοντέλου (Model Evaluation)

Τα αποτελέσματα που λάβαμε,παρουσιάζονται στο σχημα (5) και είναι από την αξιολόγηση του μοντέλου μας χρησιμοποιώντας τις μετρικές: precision, recall, και F1-σκορ. Αυτές οι μετρικές χρησιμοποιούνται για να αξιολογήσουν την απόδοση του μοντέλου μας σε κάθε κατηγορία της κατηγοριοποίησης.

Ας εξηγήσουμε κάθε μετρική:

- Το precision (ακρίβεια) μετρά το ποσοστό των πραγματικά θετικών προβλέψεων σε σχέση με το συνολικό πλήθος των προβλέψεων που έγιναν για μια συγκεκριμένη κατηγορία. Έχετε υψηλή πρόσοψη όταν οι προβλέψεις μας είναι ακριβείς και δεν υπάρχουν πολλές ψευδείς θετικές προβλέψεις.

- Η recall μετρά το ποσοστό των πραγματικά θετικών προβλέψεων σε σχέση με το συνολικό πλήθος των πραγματικά θετικών δεδομένων στην κατηγορία. Έχουμε υψηλή ανάκληση όταν το μοντέλο μας εντοπίζει αποτελεσματικά όλα τα πραγματικά θετικά δεδομένα.

- Το F1-score είναι ένας συνδυασμός του precision και του recall και χρησιμοποιείται για να αξιολογήσει την ακρίβεια του μοντέλου μας. Υψηλό F1-score σημαίνει ότι το μοντέλο μας έχει και υψηλό precision και υψηλό recall.

classes	precision	recall	f1-score	support
0	0.92	0.91	0.91	03
1	0.96	0.91	0.93	86
2	0.87	0.91	0.89	183
3	0.94	0.92	0.93	201
4	0.91	0.87	0.89	206
5	0.85	0.98	0.91	193
6	0.97	0.94	0.96	72
7	0.95	0.88	0.91	208
8	0.83	0.99	0.91	165
9	0.95	0.83	0.89	230
accuracy			0.91	1747
macro avg	0.92	0.91	0.91	1747

Σχημα 5. Μετρικές precision recall και f1-score

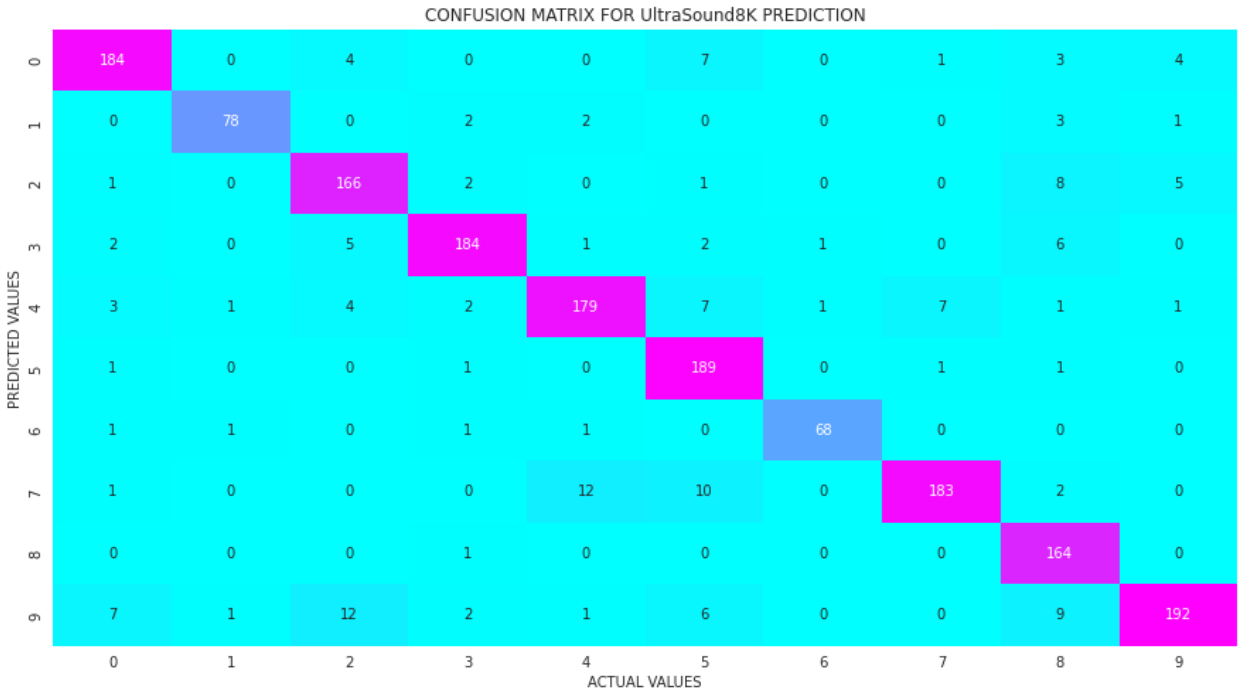
Στο σύνολο των αποτελεσμάτων που λάβαμε:

- Η ακρίβεια (accuracy) του μοντέλου μας είναι περίπου 91%, που είναι ένα πολύ καλό αποτέλεσμα. Αυτό σημαίνει ότι το μοντέλο μας έχει τη δυνατότητα να κατηγοριοποιεί σωστά περίπου το 91% των δεδομένων.

- Αν κοιτάξουμε τις μετρικές για κάθε κατηγορία, βλέπουμε ότι το μοντέλο μας έχει καλή απόδοση σε πολλές από αυτές τις κατηγορίες. Ορισμένες κατηγορίες, όπως η κατηγορία "6" (πιθανώς αντιστοιχεί σε μια συγκεκριμένη ήχου) έχουν πολύ υψηλή ακρίβεια και recall, πράγμα που σημαίνει ότι το μοντέλο μας καταφέρνει πολύ καλά στην κατηγοριοποίηση αυτών των δεδομένων.

- Είναι σημαντικό να επισημάνουμε ότι το F1-score είναι μια ισορροπημένη μετρική που λαμβάνει υπόψη και το precision και το recall. Σε γενικές γραμμές, το μοντέλο μας φαίνεται να καταφέρνει καλά στην κατηγοριοποίηση των δεδομένων.

Συνοπτικά, τα αποτελέσματα δείχνουν ότι το μοντέλο μας έχει καλή απόδοση στην κατηγοριοποίηση των ήχων, με υψηλή ακρίβεια και ανάκληση σε πολλές κατηγορίες.



σχημα 7. πίνακας σύγχυσης (Confusion Matrix)

Ο πίνακας σύγχυσης (Confusion Matrix) που παρουσιάζεται στο σχήμα (7) παρέχει μια οπτική απεικόνιση της απόδοσης του μοντέλου σε όλες τις κατηγορίες ήχου. Κάθε στοιχείο του πίνακα αντιπροσωπεύει τον αριθμό των προβλέψεων που το μοντέλο έκανε από μία πραγματική κλάση (rows) προς μια προβλεπόμενη κλάση (columns). Τα διαγώνια στοιχεία δείχνουν τον αριθμό των σωστά ταξινομημένων περιπτώσεων για κάθε κατηγορία, ενώ τα εκτός διαγωνίου στοιχεία αποτελούν τις περιπτώσεις όπου το μοντέλο έκανε λάθος πρόβλεψη.

6. Αποτελέσματα και Συμπεράσματα

Οι παρακάτω ερμηνείες και συμπεράσματα αντανακλούν την αξιολόγηση των αποτελεσμάτων που προέκυψαν από την εκπαίδευση του μοντέλου. Τα αποτελέσματα καταδεικνύουν ότι το μοντέλο παρουσιάζει εξαιρετική απόδοση στην αυτόματη κατηγοριοποίηση ηχητικών σημάτων. Με μια ακρίβεια που φτάνει περίπου το 91%, το μοντέλο είναι σε θέση να προβλέπει σωστά τις κατηγορίες για το 91% των δεδομένων ελέγχου. Είναι σημαντικό να σημειώσουμε ότι αυτό το επίπεδο ακρίβειας είναι εντυπωσιακό, ειδικά όταν πρόκειται για προβλέψεις σε πολύπλοκα ηχητικά σήματα.

Επιπλέον, η ανάλυση των μετρικών precision, recall και F1-σκορ δείχνει ότι το μοντέλο είναι ικανό να εντοπίζει αποτελεσματικά τις πραγματικές θετικές προβλέψεις και να παράγει ακριβείς προβλέψεις. Το F1-σκορ, που λαμβάνει υπόψη την πρόσοψη και την ανάκληση, επίσης δείχνει μια ισορροπημένη απόδοση του μοντέλου. Είναι αξιοσημείωτο ότι το μοντέλο παρουσιάζει καλή απόδοση σε πολλές κατηγορίες, όπως αποδεικνύεται από τα υψηλά ποσοστά ακρίβειας και ανάκλησης σε αυτές τις κατηγορίες. Αυτό υποδηλώνει ότι το μοντέλο είναι ευέλικτο και ικανό να αντιμετωπίζει διάφορα είδη ηχητικών δεδομένων.

Οι μετρικές που καταγράφηκαν κατά τη διάρκεια της εκπαίδευσης αναδεικνύουν την σημαντική βελτίωση της απόδοσης του μοντέλου μας. Συγκεκριμένα, παρατηρήσαμε την αύξηση της ακρίβειας τόσο στο σύνολο εκπαίδευσης όσο και στο σύνολο επικύρωσης κατά τη διάρκεια των 15 εκπαιδευτικών εποχών. Η μείωση της συνάρτησης κόστους επίσης υποδεικνύει ότι το μοντέλο βελτιώνει την απόδοσή του με την πάροδο του χρόνου.

Σημαντικό είναι το γεγονός ότι το μοντέλο μας κατά την τελευταία εκπαιδευτική εποχή επέτυχε ακρίβεια εκπαίδευσης 94.63% και ακρίβεια επικύρωσης 90.84%. Αυτό υποδεικνύει ότι το μοντέλο έχει μάθει να αναγνωρίζει αποτελεσματικά τα χαρακτηριστικά των ηχητικών σημάτων και διαθέτει καλή ικανότητα γενίκευσης σε νέα δεδομένα.

Βάσει των αποτελεσμάτων που παρουσιάστηκαν, μπορούμε να συναγάγουμε τα εξής συμπεράσματα:

- Η εκπαίδευση ενός προηγμένου μοντέλου μηχανικής μάθησης για την κατηγοριοποίηση ηχητικών σημάτων είναι εφικτή και επιφέρει εξαιρετικά αποτελέσματα.
- Οι εκπαιδευτικές εποχές αποδείχθηκαν κρίσιμες για την ανάπτυξη του μοντέλου, με την ακρίβεια και την απώλεια να βελτιώνονται σταδιακά κατά τη διάρκεια της εκπαίδευσης.
- Το μοντέλο μας επιδεικνύει υψηλή ικανότητα γενίκευσης, καθώς επιτυγχάνει αξιόλογη ακρίβεια σε νέα, μη εκπαιδευτικά δεδομένα.
- Οι παραπάνω παρατηρήσεις επιβεβαιώνουν τη σημασία της μηχανικής μάθησης στην ανάλυση και κατηγοριοποίηση ηχητικών σημάτων και υποδεικνύουν προοπτικές για μελλοντικές εφαρμογές και επεκτάσεις στον ευρύτερο τομέα της έρευνας.

Συνολικά, η παρούσα διπλωματική εργασία παρουσιάζει ένα εντυπωσιακό μοντέλο για την αυτόματη κατηγοριοποίηση ηχητικών σημάτων. Τα αποτελέσματα δείχνουν ότι το μοντέλο είναι αποτελεσματικό και μπορεί να χρησιμοποιηθεί με επιτυχία σε διάφορες εφαρμογές, όπως η αναγνώριση ήχων. Η επιτυχία αυτή αποδεικνύει τη σημασία της μηχανικής μάθησης στον τομέα της επεξεργασίας ηχητικών δεδομένων και ανοίγει τον δρόμο για περαιτέρω έρευνα και ανάπτυξη σε αυτόν τον τομέα.

7. ΠΗΓΕΣ ΚΑΙ ΒΙΒΛΙΟΓΡΑΦΙΑ

1. M Cowling, R Sitte, (2003), « Comparison of techniques for environmental sound », Pattern Recognition Letters, (vol:24, no:15, pp: 2895-2907)

2. Salamon, J., Jacoby, C., & Bello, J. P. (2014). "A dataset and taxonomy for urban sound research." In Proceedings of the 22nd ACM international conference on Multimedia. ACM.

3. Abdoli, S., Cardinal, P., & Koerich, A. L. (2019). "End-to-end environmental sound classification using a 1D convolutional neural network." ScienceDirect.

4. Sigtia, S., Stark, A. M., Krstulović, S., & Plumbley, M. D. (2016). "Automatic environmental sound recognition: Performance versus computational cost." IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24(11), 2096-2107.

5. Mu, W., Yin, B., Huang, X., Xu, J., & Du, Z. (2021). "Environmental sound classification using temporal-frequency attention-based convolutional neural network." Nature.

6. Mushtaq, Z., Su, S. F., & Tran, Q. V. (2021). "Spectral images based environmental sound classification using CNN with meaningful data augmentation." ScienceDirect.

7. Nasiri, A. (2021). "Deep Learning Based Sound Event Detection and Classification." University of South Carolina.

8. Özseven, T. (2023). "Investigation of the effectiveness of time-frequency domain images and acoustic features in urban sound classification." ScienceDirect.

9. Li, J., Wang, Y., Zhu, H., & Zhang, Y. (2019). "What’s That Sounds? Machine Learning for Urban Sound Classification." NoiseLab, UC San Diego.

10. Rahman, A. A., & Jothi, J. A. A. (2021). "Classification of UrbanSound8k: A Study Using Convolutional Neural Network and Multiple Data Augmentation Techniques." Springer.

11. Nanni, L., Maguolo, G., Brahnam, S., & Paci, M. (2021). "An Ensemble of Convolutional Neural Networks for Audio Classification." MDPI.

12. O'Shea, K., & Nash, R. (2015). "An introduction to convolutional neural networks." CoRR.

13. Mesaros, A., Heittola, T., & Virtanen, T. (2016). "Metrics for polyphonic sound event detection." Applied Sciences, 6(6), 162.

14. Adapa, S. (2019). "Urban sound tagging using convolutional neural networks." CoRR.

15. 15.Fonseca, E., Plakal, M., Font, F., Ellis, D. P. W., & Serra, X. (2019). "Audio tagging with noisy labels and minimal supervision." CoRR.

16. DCASE 2019 TASK 2

17. Ebbers, J., & Häb-Umbach, R. (2019). "Convolutional recurrent neural network and data augmentation for audio tagging with noisy labels and minimal supervision." In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019). New York University.

18. Adavanne, S., Politis, A., Nikunen, J., & Virtanen, T. (2018). "Sound event localization and detection of overlapping sources using convolutional recurrent neural networks." IEEE Journal of Selected Topics in Signal Processing, 13(1), 34–48.

19. 19 Adavanne, S., Parascandolo, G., Pertilä, P., Heittola, T., & Virtanen, T. (2017). "Sound event detection in multichannel audio using spatial and harmonic features." CoRR.

20. Cho, J., Yun, S., Park, H., Eum, J., & Hwang, K. (2019). "Acoustic scene classification based on a large-margin factorized CNN." CoRR.

21. Choi, K., Fazekas, G., Cho, K., & Sandler, M. B. (2019). "A comparison on audio signal preprocessing methods for deep neural networks on music tagging."

22. Cao, Y., Kong, Q., Iqbal, T., An, F., Wang, W., & Plumbley, M. (2019). "Polyphonic sound event detection and localization using a two-stage strategy." In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019). New York University.

23. Chu, S., Narayanan, S., & Kuo, C.-C. J. (2009). "Environmental sound recognition with time-frequency audio features." Transactions on Audio, Speech and Language Processing, 17(6), 1142–1158

24. JP Bello, L Daudet, S Abdallah, C. Duxbury, M. Davies, M.B. Sandler, (2005), «A tutorial on onset detection in music signals», IEEE Transactions on Speech and Audio Processing

25. Freesound website - Σύνδεσμος.

26. UrbanSound dataset - Σύνδεσμος.

27. Urban-SED - Σύνδεσμος.