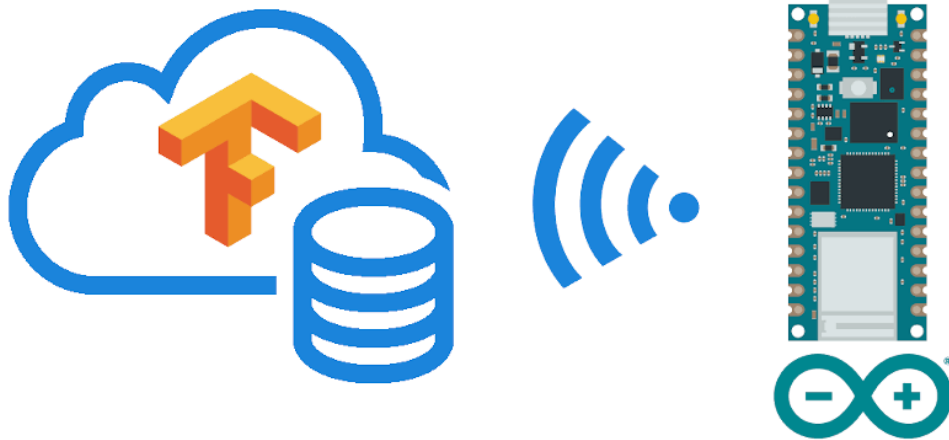


Πανεπιστήμιο Πελοποννήσου
Σχολή Μηχανικών
Τμήμα Ηλεκτρολόγων Μηχανικών και
Μηχανικών Υπολογιστών

Πτυχιακή Εργασία

Σύστημα αναγνώρισης λέξης-κλειδιού με
χρήση τεχνικών βαθιάς μάθησης σε
ενσωματωμένη συσκευή χαμηλής ισχύος



Όνοματεπώνυμο: Βασίλειος Θεοδώροπουλος

ΑΜ: 2444

Επιβλέπων Καθηγητής

Παναγιώτης Ζέρβας

Φεβρουάριος 2022

Ευχαριστίες

Με την ολοκλήρωση της πτυχιακής μου εργασίας, θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες σε όλους όσους συνέβαλλαν στην εκπόνησή της. Ευχαριστώ θερμά τον επιβλέπων καθηγητή μου, Ζέρβα Παναγιώτη, για την εμπιστοσύνη που μου έδειξε εξ' αρχής, αναθέτοντάς μου το συγκεκριμένο θέμα, την επιστημονική του καθοδήγηση, τις υποδείξεις του, την επιμονή του, τη συμπαράστασή του, τη συνεχή του υποστήριξη και το αμείωτο ενδιαφέρον που έδειξε από την αρχή μέχρι το τέλος. Τέλος, θα ήθελα εκφράσω την ευγνωμοσύνη μου στην οικογένειά μου για όλη τη στήριξη, τη συμπαράσταση και την κατανόησή τους, καθ' όλη τη διάρκεια των σπουδών μου.

Περίληψη

Η παρούσα πτυχιακή εργασία ασχολείται με την ανάπτυξη ενσωματωμένης εφαρμογής η οποία θα εκτελείτε σε αισθητήρα βασισμένο στη πλατφόρμα Arduino Nano 33 BLE Sense. Η εφαρμογή αυτή θα επιτρέπει την αυτόματη αναγνώριση φωνητικών εντολών περιορισμένου λεξιλογίου (voice command recognition). Για την εκπαίδευση του μοντέλου φωνητικής αναγνώρισης θα χρησιμοποιηθεί η βιβλιοθήκη ανάπτυξης νευρωνικών δικτύων βαθιάς μάθησης με δυνατότητα εκτέλεσης σε ενσωματωμένες συσκευές, TensorFlow Lite.

Η υλοποίηση που θα προκύψει θα μπορεί να χρησιμοποιηθεί σε σενάρια καταγραφής ηχητικών δεδομένων σε δίκτυο αισθητήρων όπου η διαδικασία της αναγνώρισης θα γίνεται πάνω στον αισθητήρα (network edge sensing). Αναφορικά με τη συνδεσιμότητα του αισθητήρα, αυτή θα πραγματοποιείται μέσω bluetooth σε ενδιάμεσο ΗΥ ο οποίος στη συνέχεια θα στέλνει τα δεδομένα σε εξυπηρετητή νέφους

Περιεχόμενα

Περιεχόμενα	6
1 Εισαγωγή.....	9
1.1 Αυτόματη αναγνώριση ομιλίας.....	10
1.1.1 Ενσωματωμένες συσκευές και διαδίκτυο των πραγμάτων	12
2 Μηχανική μάθηση και αναγνώριση προτύπων	13
2.1 Περιγραφή συστήματος αναγνώρισης προτύπων.....	17
2.2 Αλγόριθμοι στην αναγνώριση προτύπων	19
2.3 Εφαρμογές αναγνώρισης προτύπων.....	21
3 Νευρωνικά δίκτυα βαθιάς μάθησης	22
3.1 Μηχανική Μάθηση σε μικροσκοπικές κλίμακες	24
3.1.1 ΜΜσΜΚ: Εφαρμογές & περιπτώσεις χρήσης.....	26
4 Υλοποίηση συστήματος αναγνώρισης λέξεων κλειδιών ...	27
4.1 Πλατφόρμα υλοποίησης	27
4.1.1 Αυτόματη Μάθηση (AutoML).....	35
4.2 Υλικό υλοποίησης (hardware) Arduino Nano 33 BLE	37
4.3 Βάση ηχητικών δεδομένων	38
5 Πειράματα.....	39
5.1 Μέτρα αξιολόγησης	41
5.2 Αποτελέσματα	42
6 Επίλογος.....	45
Βιβλιογραφία	47

Κατάλογος εικόνων

Εικόνα 1: Διαδίκτυο των πραγμάτων (Internet of things).....	10
Εικόνα 2: Διάγραμμα ροής αλγορίθμου αναγνώρισης.....	11
Εικόνα 3: Εκμάθηση με επίβλεψη.....	15
Εικόνα 4: Μάθηση χωρίς επίβλεψη.....	16
Εικόνα 5: Ημι-εποπτευόμενη μάθηση.....	17
Εικόνα 6: (α) Σε μη-πραγματικό χρόνο φ(β) σε πραγματικό χρόνο φάση	17
Εικόνα 7: Διάγραμμα ροής αναγνώρισης προτύπων.....	18
Εικόνα 8: Δεδομένα δοκιμής και δεδομένα εκπαίδευσης	19
Εικόνα 9: Τεχνητός νευρώνας (perceptron)	22
Εικόνα 10: Νευρωνικό δίκτυο	23
Εικόνα 11: Βαθύ Νευρωνικό Δίκτυο	24
Εικόνα 12: Τρόποι συλλογής δεδομένων στην πλατφόρμα Edge Impulse.....	28
Εικόνα 13: (α) Επιτυχής σύνδεση , (β) Συλλογή δεδομένων	29
Εικόνα 14: Συλλογή δεδομένων ήχου	29
Εικόνα 15: Παράδειγμα φασματογραφήματος (Spectrogram).....	31
Εικόνα 16: Παράδειγμα απεικόνισης συντελεστών MFCC.....	32
Εικόνα 17: Παράδειγμα απεικόνισης συντελεστών MFE	33
Εικόνα 18: Ρυθμίσεις μοντέλου εκπαίδευσης	34
Εικόνα 19: Αυτοματοποιημένη εκμάθηση	35
Εικόνα 20: Παράδειγμα απεικόνισης του εργαλείου AutoML EON Tuner ...	36
Εικόνα 21: Στιγμιότυπο ενσ. συσκευής Arduino Nano 33 BLE Sense.....	37
Εικόνα 22: Αριθμός ομιλητών.....	38
Εικόνα 23: Αριθμός ηχογραφήσεων ανά κατηγορία.....	39
Εικόνα 24: Αρχιτεκτονική νευρωνικού δικτύου.....	40
Εικόνα 25: Αριθμός δειγμάτων στα σύνολα εκπαίδευσης και αξιολόγησης ...	40
Εικόνα 26: Παράδειγμα ενός πίνακα σύγκρισης για δυαδική κατηγοριοποίηση	42
Εικόνα 27: Ακρίβεια κατηγοριοποίησης ως προς τη στάθμη θορύβου.....	45

Κατάλογος Πινάκων

Πίνακας 1: Αποτελέσματα κατηγοριοποίησης δεδομένων χωρίς θόρυβο (καθαρή ομιλία).....	43
Πίνακας 2: Αποτελέσματα κατηγοριοποίησης δεδομένων με θόρυβο SNR 15	44
Πίνακας 3: Αποτελέσματα κατηγοριοποίησης δεδομένων με θόρυβο SNR 10	44
Πίνακας 4: Αποτελέσματα κατηγοριοποίησης δεδομένων με θόρυβο SNR 5	44
Πίνακας 5: Αποτελέσματα κατηγοριοποίησης δεδομένων με θόρυβο SNR 0	45

1 Εισαγωγή

Η παρούσα πτυχιακή εργασία καταπιάνεται με το πρόβλημα της αυτόματης αναγνώρισης λέξεων κλειδιών από σήμα ομιλίας με χρήση ενσωματωμένων συσκευών. Το πρόβλημα αυτό ανήκει στον ευρύτερο χώρο της τεχνολογίας ομιλίας και πολλές φορές συναντάται ως αυτόματη αναγνώριση απομονωμένων λέξεων (isolated word recognition) (Bilinger, 1999) ή αυτόματη αναγνώριση λέξεων κλειδιών (keyword recognition from speech) (Furui, 1986).

Στο πλαίσιο της παρούσας εργασίας, αρχικά αναπτύχθηκε ένα σώμα ομιλίας με τις λέξεις κλειδιά που θέλουμε να αναγνωρίσουμε και στη συνέχεια βασιζόμενοι στη πλατφόρμα ανάπτυξης έξυπνων εφαρμογών σε ενσωματωμένες συσκευές, Impulse Edge (ZDNet, 2022), αναπτύχθηκαν μοντέλα βαθιάς μάθησης τα οποία θα μπορούν να εκτελεστούν αποτελεσματικά στην συσκευή Arduino Nano 33 BLE. Δεδομένου ότι είναι δύσκολο κανείς να βρει πόρους ομιλίας για την Ελληνική γλώσσα προχωρήσαμε στην δημιουργία μιας δικιάς μας με ηχογραφήσεις που έγιναν από διάφορους ομιλητές μέσα από την χρήση των έξυπνων κινητών τους. Για την εκπαίδευση των μοντέλων χρησιμοποιήθηκε η βιβλιοθήκη Tensorflow lite (David et al., 2021), και μελετήθηκε η απόδοση ως είσοδος του αλγορίθμου αναγνώρισης χαρακτηριστικών όπως τα MFCC (Mansour, et al. 2015), MFE (Impulse Edge Docs, 2021) και το φασματογράφημα (Newman, 1966).

Η παρούσα εργασία έχει οργανωθεί ως εξής, στο κεφάλαιο 1 περιγράφεται το θέμα με το οποίο διαπραγματεύεται και δίνεται μια εισαγωγή στο πρόβλημα της αυτόματης αναγνώρισης ομιλίας και τη χρήση ενσωματωμένων συσκευών για την υλοποίηση τέτοιων τεχνολογιών. Το κεφάλαιο 2 γίνεται μια αναφορά σχετικά με το χώρο της μηχανικής μάθησης, της αναγνώρισης προτύπων και παρατίθεται μια εισαγωγή στην αυτόματη μηχανική μάθηση (autoML). Με τις λεπτομέρειες υλοποίησης, ανάπτυξης βάσης δεδομένων ομιλίας, εξαγωγή χαρακτηριστικών και προγραμματιστικών εργαλείων που χρησιμοποιήθηκαν καταπιάνεται το κεφάλαιο 3. Τέλος στο 4 μελετώνται τα πειράματα και τα αποτελέσματα που λάβαμε από το σύστημα μας, καθώς και δίνονται οι λεπτομέρειες λειτουργίας και απόδοσης του συστήματος μας σε Arduino Nano 33 BLE. Το κεφάλαιο κλείνει με τα συμπεράσματα και τις παρατηρήσεις μας καθώς και με προτάσεις για μελλοντικές βελτιώσεις και επεκτάσεις του συστήματος μας.

1.1 Αυτόματη αναγνώριση ομιλίας

Τα συστήματα αυτόματης αναγνώρισης ομιλίας στηρίζονται στην υλοποίηση αλγορίθμων με σκοπό την αντιστοίχιση του ακουστικού σήματος σε μια σειρά λέξεων. Η έρευνα για την ανάπτυξη τέτοιων συστημάτων απαιτεί τη συνεισφορά διαφόρων επιστημών η γλωσσολογία, η επιστήμη των υπολογιστών και η ακουστική. Τέτοιου είδους συστήματα μας επιτρέπουν να αλληλοεπιδρούμε με τις μηχανές απλά με τη χρήση της φωνής μας. Έτσι μπορούμε για παράδειγμα, να κρατήσουμε σημειώσεις, υπενθυμίσεις, να κάνουμε κλήσεις και άλλα χωρίς τη χρήση πολύπλοκων διεπαφών μεταξύ ανθρώπων και μηχανών.



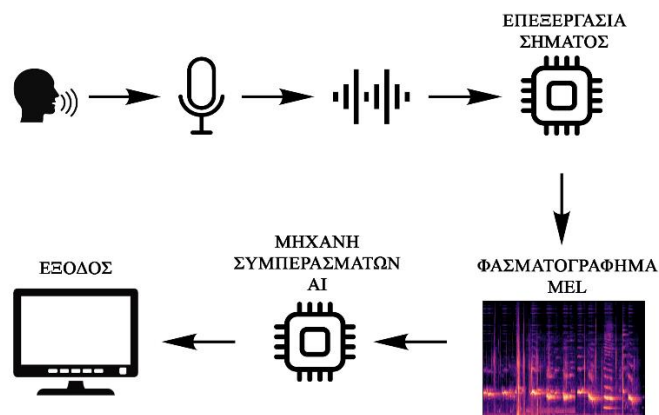
Εικόνα 1: Διαδίκτυο των πραγμάτων (Internet of things)

Στις μέρες τέτοιες τεχνολογίες έχουν βρει εμπορικές εφαρμογές σε προϊόντα όπως οι έξυπνοι βοηθοί (Li, Yuan, 2018) (π.χ. Amazon Alexa, Apple Siri, Google Assistant και Microsoft Cortana). Επίσης υπάρχουν και πολλές σχετικές εφαρμογές ανοιχτού κώδικα που επιτρέπουν την ανάπτυξη τέτοιων υπηρεσιών με χρήση αναγνώρισης και σύνθεσης ομιλίας.

Η τεχνολογία έχει γίνει αναπόσπαστο κομμάτι της ζωής μας. Νέες εξελίξεις συνεχίζουν να σημειώνονται στον τομέα της πληροφορικής και της τεχνολογίας. Οι επιστήμονες σε όλο τον κόσμο καταβάλλουν προσπάθειες για να μιμηθούν την ανθρώπινη νοημοσύνη στις μηχανές. Η αναγνώριση λέξεων και ομιλίας είναι μια τέτοια μέθοδος που επιτρέπει στις μηχανές να επιδεικνύουν ανθρώπινη νοημοσύνη. Τέτοιες τεχνολογίες βρίσκουν πρόσφορο έδαφος στην βιομηχανία αλλά και στη καθημερινή μας ζωή και αποδεικνύονται πολύ σημαντικές σε άτομα με ειδικές ικανότητες όπως, προβλήματα όρασης μιας και το πληκτρολόγιο αποτελεί εμπόδιο στην πρόσβαση στον υπολογιστή και την πληροφορία.

Σε ένα σύστημα αναγνώρισης μεμονωμένο λέξεων, ως είσοδος λαμβάνονται λέξεις ή μια λίστα λέξεων που χωρίζονται με από παύση. Για κάθε λέξη γίνεται ξεχωριστή επεξεργασία και η αναγνώριση της δεν επηρεάζει λέξεις που προηγούνται ή έπονται από αυτή. Αντίθετα σε ένα σύστημα αναγνώρισης συνεχούς ομιλίας ως είσοδος λαμβάνεται μια συνεχής ροή λέξεων η αναγνώριση λέξεων μπορεί να απαιτεί την κατανόηση των συμφραζομένων παρόμοιων λέξεων που ακούγονται.

Τα συστατικά από τα οποία αποτελείται ένα σύστημα αναγνώρισης μεμονωμένων λέξεων φαίνεται στην εικόνα 2. Αρχικά ως είσοδος λαμβάνεται το σήμα της λέξης προς αναγνώρισης. Αυτό μπορεί να προέρχεται είτε από ένα προηχογραφημένο αρχείο ήχου είτε είναι η έξοδος ενός μικροφώνου αφού προηγουμένως έχει μεσολαβήσει ένα στάδιο αυτόματης εύρεσης ορίων λέξεων. Σε επόμενο βήμα το σήμα ομιλίας υφίσταται επεξεργασία για την εξαγωγή χαρακτηριστικών παραμέτρων τα οποία θα αποτελέσουν την είσοδο του αλγορίθμου αναγνώρισης προτύπων τόσο για τη φάση εκπαίδευσης όσο και για την σε πραγματικές συνθήκες χρήση του. Ως αλγόριθμος αναγνώρισης έχουν χρησιμοποιηθεί με επιτυχία κρυμμένα μοντέλα Μαρκόβ (Rabiner, Juang, 1986), μηχανές διανυσμάτων στήριξης (Cristianini, Shawe-Taylor, 2000), νευρωνικά δίκτυα (McCulloch, Pitts, 1943) κ.α. Κατά την διαδικασία εκπαίδευσης ο αλγόριθμος αναγνώρισης «μαθαίνει» να συσχετίζει το σήμα εισόδου με τη σχετική κατηγορία (ποια είναι η λέξη) και αυτό γίνεται μέσω της χρήσης των δεδομένων εκπαίδευσης.



Εικόνα 2: Διάγραμμα ροής αλγορίθμου αναγνώρισης

1.1.1 Ενσωματωμένες συσκευές και διαδίκτυο των πραγμάτων

Ένα ενσωματωμένο σύστημα είναι ένα υπολογιστικό σύστημα χαμηλής ισχύος, ένας συνδυασμός επεξεργαστή, μνήμης και περιφερικών συσκευών εισόδου και εξόδου όπου έχει μια αποκλειστική λειτουργία. Τα ενσωματωμένα συστήματα είναι στην καρδιά πολλών και διαφορετικών προϊόντων από μηχανές απλής χρήσης όπως ένα πλυντήριο ρούχων έως εφαρμογές μηχανικής μάθησης και τεχνητής νοημοσύνης. Το 2009 υπολογίστηκε πως το συντριπτικό ποσό 98% των μικροεπεξεργαστών που κατασκευάζεται χρησιμοποιείται σε ενσωματωμένα συστήματα.

Από τα εργοστασιακά ρομπότ σε γραμμές παραγωγής μέχρι την απλή χρήση που κάνουμε τακτικά στα ΑΤΜ, τα ενσωματωμένα συστήματα μπόκαν κι άλλαξαν δραστικά τη ζωή του ανθρώπου.

Το Internet είναι ένα άλλο κομμάτι όπου άλλαξε μια για πάντα τον τρόπο ζωής μας. Το διαδίκτυο δεν επηρέασε μόνο την επικοινωνία, άλλαξε σημαντικά την εκπαίδευση, την οικονομία, τον τρόπο δουλείας κι πολλά ακόμα. Δημιούργησε πολλά προβλήματα όπως την ασφάλεια και απόρρητο, αλλά έφερε στο φως ανάγκες κι έννοιες όπως το διαδίκτυο των πραγμάτων ή αλλιώς Internet of things (IoT).

Το διαδίκτυο των πραγμάτων είναι φυσικές συσκευές ή ομάδες τέτοιων συσκευών οι οποίες είναι ενσωματωμένες με αισθητήρες, επεξεργαστική ισχύ, λογισμικό κι άλλες λειτουργίες που τους επιτρέπει να συνδέονται με άλλες συσκευές κι να ανταλλάσσουν δεδομένα μέσω του διαδικτύου ή άλλων δικτύων επικοινωνίας.

Μπορούμε να πούμε πως στην έκφραση ‘ διαδίκτυο των πραγμάτων’ η λέξη πραγμάτων αναφέρετε σε ένα ενσωματωμένο σύστημα, έτσι κατά κάποιον τρόπο τα ενσωματωμένα συστήματα αποτελούν ένα υποσύνολο του διαδικτύου των πραγμάτων. Όλες οι συσκευές στο διαδίκτυο των πραγμάτων έχουν ενσωματωμένα συστήματα αλλά δεν είναι όλα τα ενσωματωμένα συστήματα στο διαδίκτυο των πραγμάτων. Μια απλή τηλεόραση ένα ενσωματωμένο σύστημα αλλά μια έξυπνη τηλεόραση είναι μια συσκευή στο διαδίκτυο των πραγμάτων.

2 Μηχανική μάθηση και αναγνώριση προτύπων

Ένα επιστημονικό πεδίο μπορεί να οριστεί καλύτερα με βάση το κεντρικό ερώτημα το οποίο μελετά. Το πεδίο της μηχανικής μάθησης επιδιώκει να απαντήσει στο ερώτημα: *«Πως μπορούμε να δημιουργήσουμε υπολογιστικά συστήματα τα οποία βελτιώνονται αυτόματα με την εμπειρία, και ποιοι είναι οι θεμελιώδεις νόμοι που διέπουν αυτή την διαδικασία μάθησης;»* Το ερώτημα αυτό καλύπτει ένα ευρύ φάσμα μαθησιακών διαδικασιών, όπως το πως να σχεδιαστεί ένα ρομπότ το οποίο να μαθαίνει να κινείται από την δικιά του εμπειρία, πως να εξορύσσετε πληροφορία από παλαιότερα ιστορικά δεδομένα ασθενών με σκοπό την εξακρίβωση του ποια θεραπεία θα άρμοζε καλύτερα σε ένα μελλοντικό ασθενή, και πως να αναπτύσσονται μηχανές αναζήτησης οι οποίες να προσαρμόζονται κατάλληλα στις ανάγκες των χρηστών. Στο βιβλίο του (Mitchel, 2017) ο Mitchel αναφορικά με το τι σημαίνει μηχανική μάθηση είχε δώσει τον εξής ορισμό, *«μια μηχανή μαθαίνει σε σχέση με μια επιτελούμενη διεργασία Δ , μέτρο απόδοσης A , και εμπειρία E , αν το σύστημα αξιόπιστα βελτιώνει την απόδοση του A για την διεργασία Δ , δεδομένης της εμπειρίας E . Ανάλογα με το πως ορίζουμε τις Δ , A και E , η διαδικασία εκπαίδευσης μπορεί να ονόματα όπως εξόρυξη δεδομένων, αναγνώριση προτύπων, ανανέωση πληροφορίας βάσεων δεδομένων, αυτόνομη εύρεση, κλπ».*

Η μηχανική μάθηση είναι μια παραφυάδα του τομέα της επιστήμης των υπολογιστών και της στατιστικής. Όσον αφορά τη μηχανική μάθηση, η επιστήμη των υπολογιστών καλείται να απαντήσει στο ερώτημα, *“Πως μπορούμε να υλοποιήσουμε μηχανές που να επιλύουν προβλήματα και ποια από τα προβλήματα αυτά μπορούν να θεωρηθούν εγγενώς επιλύσιμα και όχι»* από την μεριά της στατιστικής γίνεται προσπάθεια να δωθεί απάντηση στο *«Τι μπορεί να εξαχθεί από τα δεδομένα καθώς και από ένα σύνολο υποθέσεων και με ποια αξιοπιστία»*. Με βάση τη γνώση λοιπόν που προκύπτει από τους συγκεκριμένους τομείς η μηχανική μάθηση επικεντρώνεται στο πως οι υπολογιστές μπορούν να “προγραμματιστούν” μόνοι τους (με βάση την εμπειρία και κάποια αρχική κατάσταση/δομή), με άλλα λόγια να μπορούν να εξάγουν κανόνες μέσα από την εμπειρία (δεδομένα). Πέρα λοιπόν από τις απαντήσεις/συμπεράσματα που μπορεί να μας δώσει η επιστήμη της στατιστικής μέσα από την ανάλυση δεδομένων, η μηχανική μάθηση μελετάει και δίνει απαντήσεις σε πρόσθετα ερωτήματα σχετικά με το τι υπολογιστικές αρχιτεκτονικές μπορούν

να χρησιμοποιηθούν για πιο αποτελεσματική συλλογή, αποθήκευση και αρχειοθέτηση, ανάκτηση και συνένωση αυτής της πληροφορίας, πως διαφορετικές υλοποιήσεις εκπαίδευσης μπορούν να αλληλοεπιδρούν και να εκτελούνται σε ένα ευρύτερο σύστημα, καθώς και ερωτήματα υπολογιστικής επιλυσιμότητας.

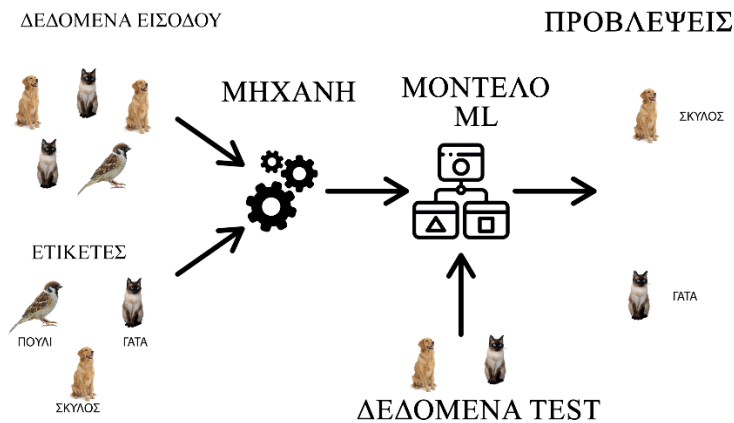
Μία από τις πιο κοινές εφαρμογές της μηχανικής μάθησης είναι η αναγνώριση προτύπων. Οι υπολογιστές που χρησιμοποιούν καλά εκπαιδευμένους αλγόριθμους αναγνωρίζουν το περιεχόμενο εικόνων (π.χ. αντικείμενα, ανθρώπους, ζώα), ανωμαλίες σε ακτινογραφίες για την αναγνώριση ύπαρξης κακοηθιών, αναγνώριση ηχητικού περιεχομένου (π.χ. ακουστικά γεγονότα, ομιλία), υπολογιστική όραση κ.α. Η αναγνώριση προτύπων είναι η διαδικασία αναγνώρισης κανονικότητας στα δεδομένα από μια μηχανή που χρησιμοποιεί αλγόριθμους μηχανικής μάθησης. Στην καρδιά της διαδικασίας βρίσκεται η ταξινόμηση των γεγονότων με βάση στατιστικές πληροφορίες, ιστορικά δεδομένα ή τη μνήμη του μηχανήματος. Ένα μοτίβο είναι μια κανονικότητα στον κόσμο ή σε αφηρημένες έννοιες. Αν μιλάμε για βιβλία ή ταινίες, μια περιγραφή ενός είδους θα ήταν ένα μοτίβο. Αυτός ο τύπος αναγνώρισης μπορεί να γίνει σε διάφορους τύπους εισόδου, όπως η βιομετρική αναγνώριση, τα χρώματα, η αναγνώριση εικόνας και η αναγνώριση προσώπου. Η αναγνώριση προτύπων έχει τα ακόλουθα χαρακτηριστικά:

- Έχει μεγάλη ακρίβεια στην αναγνώριση μοτίβων
- Μπορεί να αναγνωρίσει άγνωστα αντικείμενα
- Μπορεί να ανακτήσει μοτίβα σε περιπτώσεις που λείπουν δεδομένα
- Ένα σύστημα αναγνώρισης προτύπων μπορεί να ανακαλύψει μοτίβα που είναι εν μέρει κρυμμένα.

Προκειμένου το μηχανήμα να αναζητήσει μοτίβα σε δεδομένα, θα πρέπει να υποβληθεί σε προεπεξεργασία και να μετατραπεί σε μια μορφή που να μπορεί να κατανοήσει ο υπολογιστής. Στη συνέχεια, χρησιμοποιούνται αλγόριθμοι ταξινόμησης, παλινδρόμησης ή ομαδοποίησης ανάλογα με τις διαθέσιμες πληροφορίες σχετικά με το πρόβλημα για να λάβει πολύτιμα αποτελέσματα. Παρακάτω θα δώσουμε τον ορισμό για κάθε μια από αυτές τις κατηγορίες μηχανικής μάθησης.

Εποπτευόμενη μάθηση (Supervised learning): όταν ένας αλγόριθμος μαθαίνει από τα παραδείγματα δεδομένων και τις σχετικές απαντήσεις στόχου που μπορεί να

αποτελούνται από αριθμητικές τιμές ή κειμενικές ετικέτες, όπως κλάσεις ή ετικέτες, προκειμένου να προβλεφθεί αργότερα η σωστή απάντηση όταν υποβάλλεται με νέα παραδείγματα. Αυτή η προσέγγιση είναι πράγματι παρόμοια με την ανθρώπινη μάθηση υπό την επίβλεψη ενός δασκάλου. Ο δάσκαλος παρέχει καλά παραδείγματα για να απομνημονεύσει ο μαθητής και ο οποίος αντλεί γενικούς κανόνες από αυτά τα συγκεκριμένα παραδείγματα. Παράδειγμα εποπτευόμενης μάθησης είναι η ταξινόμηση. Με τον όρο αυτό αναφερόμαστε στη διαδικασία πρόβλεψης της κατηγορίας των σημείων των δεδομένων των οποίων μας δοθήκαν. Οι κλάσεις πολλές φορές ονομάζονται ετικέτες ή κατηγορίες. Όταν οι εισόδοι χωρίζονται σε δύο ή περισσότερες τάξεις και ο αλγόριθμος μηχανικής μάθησης θα πρέπει να παράγει ένα μοντέλο που εκχωρεί άρατες εισόδους σε μία ή περισσότερες κλάσεις (ταξινόμηση πολλαπλών ετικετών). Αυτό αντιμετωπίζεται συνήθως με εποπτευόμενο τρόπο.

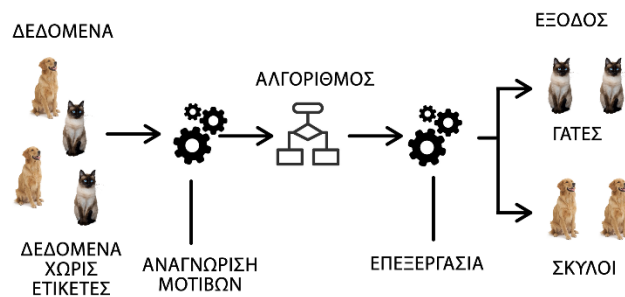


Εικόνα 3: Εκμάθηση με επίβλεψη

Μάθηση χωρίς επίβλεψη (Unsupervised learning): όταν ένας αλγόριθμος μαθαίνει από απλά παραδείγματα χωρίς καμία σχετική απάντηση, αφήνει στον αλγόριθμο να καθορίσει τα μοτίβα δεδομένων από μόνος του. Αυτός ο τύπος αλγορίθμου τείνει να μετασχηματίζει τα δεδομένα σε κάτι άλλο, όπως νέα χαρακτηριστικά που μπορεί να αντιπροσωπεύουν μια κλάση ή μια νέα σειρά μη συσχετισμένων τιμών. Είναι αρκετά χρήσιμα για την παροχή πληροφοριών στους ανθρώπους για την έννοια των δεδομένων και νέες χρήσιμες εισροές σε εποπτευόμενους αλγόριθμους μηχανικής μάθησης. Ως είδος μάθησης, μοιάζει με τις μεθόδους που χρησιμοποιούν οι

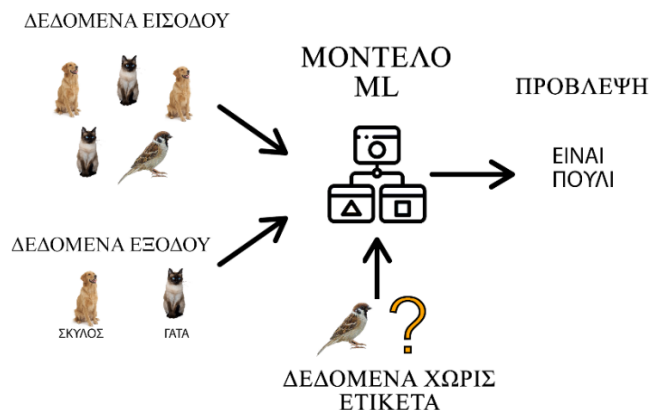
άνθρωποι για να καταλάβουν ότι ορισμένα αντικείμενα ή γεγονότα ανήκουν στην ίδια τάξη, όπως παρατηρώντας τον βαθμό ομοιότητας μεταξύ αντικειμένων. Ορισμένα συστήματα συστάσεων που βρίσκονται στον ιστό με τη μορφή αυτοματισμού μάρκετινγκ βασίζονται σε αυτόν τον τύπο μάθησης.

Ένας βασικός τύπος μη εποπτευόμενης μάθησης είναι το πρόβλημα της ομαδοποίησης. Κατά την ομαδοποίηση, ένα σύνολο εισόδων πρόκειται να διαιρεθεί σε ομάδες. Σε αντίθεση με την ταξινόμηση, οι ομάδες δεν είναι γνωστές εκ των προτέρων, καθιστώντας αυτό συνήθως ένα έργο χωρίς επίβλεψη. Η μηχανική μάθηση εμφανίζεται όταν τα προβλήματα δεν μπορούν να επιλυθούν με τυπικές προσεγγίσεις. (Mueller & Massaron, 2022)



Εικόνα 4: Μάθηση χωρίς επίβλεψη

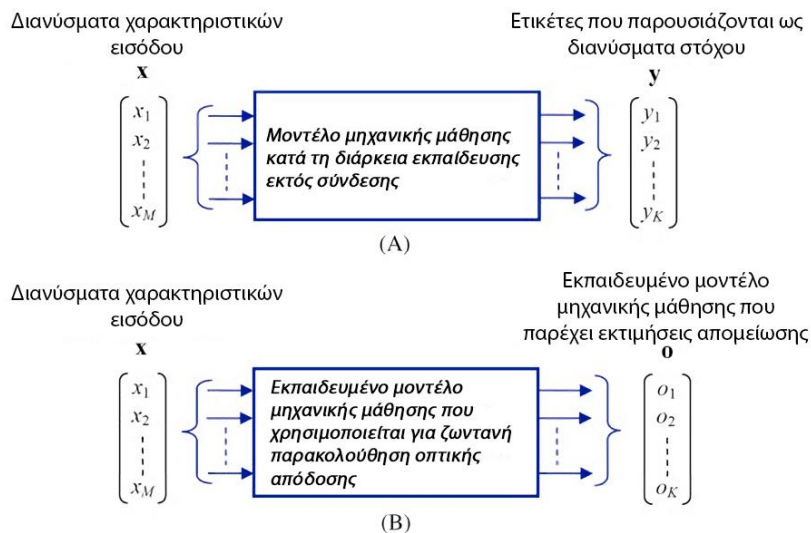
Ημι-εποπτευόμενη μάθηση (Semi-supervised learning): όταν ένα σήμα εκπαίδευσης είναι ημιτελές, σετ εκπαίδευσης με μερικούς ή και με πολλούς αντίστοιχα στόχους εξόδων να λείπουν. Υπάρχει μια ειδική περίπτωση αυτής της αρχής γνωστή ως Transduction (μεταγωγή) ή Transductive Inference (μεταγωγικό συμπέρασμα) όπου ολόκληρο το σύνολο προβλημάτων εμφανίζεται κατά την εκμάθηση, εκτός από το πρόβλημα με την έλλειψη των στόχων.



Εικόνα 5: Ημι-εποπτευόμενη μάθηση

2.1 Περιγραφή συστήματος αναγνώρισης προτύπων

Όπως είδαμε και στο κεφάλαιο 1, ένα σύστημα αναγνώρισης προτύπων ουσιαστικά αποτελείται από δύο φάσεις, (α) την σε μη-πραγματικό χρόνο (offline) και, (β) την σε πραγματικό χρόνο φάση, όπως φαίνεται στην **Error! Reference source not found.**



Εικόνα 6: (α) Σε μη-πραγματικό χρόνο φάση (offline) (β) σε πραγματικό χρόνο φάση (online)

Κατά την εκτέλεση της σε μη-πραγματικό χρόνο φάσης ουσιαστικά λαμβάνει χώρα η εκπαίδευση του αλγορίθμου μηχανικής μάθησης που έχει επιλεγεί. Συγκεκριμένα από τα δεδομένα εκπαίδευσης τα οποία έχουν μετασχηματιστεί σε διανύσματα και

έχουν αντιστοιχηθεί σε μια επιθυμητή έξοδο γίνεται η εκπαίδευση, το τέλος της οποίας ορίζεται από κριτήρια τα οποία ελέγχονται κατά την διεξαγωγή του κάθε βήματος. Για την εξακρίβωση του βαθμού εκπαίδευσης που έχει επιτευχθεί ο αλγόριθμος ελέγχεται με ένα σύνολο δεδομένων αξιολόγησης. Όταν η διαδικασία της εκπαίδευσης έχει ολοκληρωθεί επιτυχώς τότε λαμβάνει χώρα η σε πραγματικό χρόνο φάση λειτουργίας κατά την οποία το εκπαιδευμένο μοντέλο χρησιμοποιείται για την αναγνώριση άγνωστον δεδομένων εισόδου σε πραγματικό χρόνο.

Στην **Error! Reference source not found.**, παρουσιάζονται τα βήματα που λαμβάνουν χώρα κατά την ανάπτυξη ενός συστήματος αναγνώρισης προτύπων.

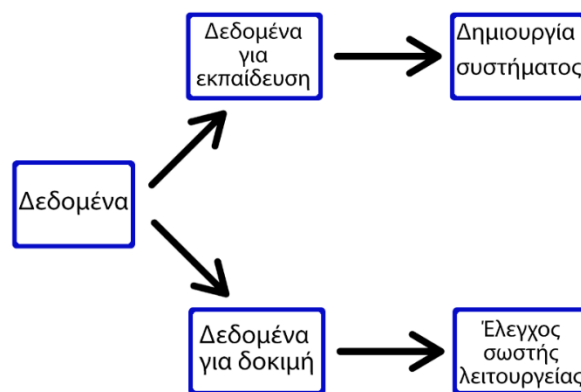
Αυτές οι φάσεις μπορούν να εξηγηθούν ως εξής:

1. Ανίχνευση: Σε αυτή τη φάση, το σύστημα αναγνώρισης προτύπων μετατρέπει τα δεδομένα εισόδου σε ανάλογα δεδομένα.
2. Τμηματοποίηση: Αυτή η φάση διασφαλίζει ότι τα αισθητά αντικείμενα απομονώνονται.
3. Εξαγωγή χαρακτηριστικών: Αυτή η φάση υπολογίζει τα χαρακτηριστικά ή τις ιδιότητες των αντικειμένων και τα στέλνει για περαιτέρω ταξινόμηση.
4. Ταξινόμηση: Σε αυτή τη φάση, τα αισθητά αντικείμενα κατηγοριοποιούνται ή τοποθετούνται σε ομάδες ή περιπτώσεις.
5. Μετα-επεξεργασία: Εδώ, λαμβάνονται περαιτέρω σκέψεις πριν ληφθεί μια απόφαση.



Εικόνα 7: Διάγραμμα ροής αναγνώρισης προτύπων

Η μάθηση επιτρέπει στο σύστημα αναγνώρισης προτύπων να εκπαιδευτεί και να γίνει προσαρμοστικό για να παρέχει πιο ακριβή αποτελέσματα. Για την επιτυχή υλοποίηση τέτοιων συστημάτων είναι απαραίτητη η ύπαρξη πολλών και ποιοτικών δεδομένων. Όπως αναφέρθηκε και προηγουμένως ένα τμήμα του συνόλου δεδομένων χρησιμοποιείται για την εκπαίδευση του συστήματος ενώ το υπόλοιπο χρησιμοποιείται για τη δοκιμή του. Η παρακάτω εικόνα δείχνει πώς χρησιμοποιούνται τα δεδομένα για εκπαίδευση και δοκιμές.



Εικόνα 8: Δεδομένα δοκιμής και δεδομένα εκπαίδευσης

Το σύνολο εκπαίδευσης περιέχει τα δεδομένα (π.χ. εικόνες, αρχεία ήχου, κείμενο κ.α.) που χρησιμοποιούνται για την κατασκευή (εκπαίδευση) του μοντέλου. Οι αλγόριθμοι εκπαίδευσης χρησιμοποιούνται για να αντιστοιχίσουν ένα διάνυσμα εισόδου (στιγμιότυπο) με μια αντίστοιχη απόφαση εξόδου. Τέλος το σύνολο αξιολόγησης χρησιμοποιείται για την επικύρωση της ακρίβειας του συστήματος. Αυτά τα δεδομένα αντιπροσωπεύουν περίπου το 20% του συνόλου των δεδομένων στο σύστημα αναγνώρισης προτύπων.

2.2 Αλγόριθμοι στην αναγνώριση προτύπων

Ακολουθούν μερικοί από τους αλγόριθμους που χρησιμοποιούνται στην αναγνώριση προτύπων.

Αλγόριθμοι βασισμένοι στη στατιστική: χρησιμοποιείται για την κατασκευή ενός στατιστικού μοντέλου. Αυτό είναι ένα μοντέλο του οποίου τα μοτίβα περιγράφονται χρησιμοποιώντας χαρακτηριστικά. Το μοντέλο μπορεί να προβλέψει την πιθανολογική φύση των προτύπων. Τα επιλεγμένα χαρακτηριστικά χρησιμοποιούνται για το σχηματισμό συμπλεγμάτων. Αναλύεται η κατανομή

πιθανοτήτων του προτύπου και το σύστημα προσαρμόζεται ανάλογα. Τα μοτίβα υποβάλλονται σε περαιτέρω επεξεργασία. Στη συνέχεια, στο μοντέλο εφαρμόζονται μοτίβα αξιολόγησης με σκοπό την εύρεση της ακρίβειας του συστήματος.

Δομικοί αλγόριθμοι: είναι αποτελεσματικοί όταν η διαδικασία αναγνώρισης προτύπων είναι πολύπλοκη. Είναι σημαντικά όταν χρησιμοποιούνται πολυδιάστατες οντότητες. Τα μοτίβα ταξινομούνται σε υποκατηγορίες, σχηματίζοντας έτσι μια ιεραρχική δομή. Το δομικό μοντέλο ορίζει τη σχέση μεταξύ των στοιχείων του συστήματος.

Αλγόριθμοι αντιστοίχισης προτύπων: χρησιμοποιούνται για τη δημιουργία ενός μοντέλου αντιστοίχισης προτύπων, το οποίο είναι ένα απλό μοντέλο αναγνώρισης προτύπων. Το μοντέλο χρησιμοποιεί δύο πρότυπα για να καθορίσει την ομοιότητα και το αντίστοιχο μοτίβο αποθηκεύεται με τη μορφή προτύπων. Το μειονέκτημα αυτού του μοντέλου είναι ότι δεν είναι αποτελεσματικό στην αναγνώριση μοτίβων σε παρουσία θορύβου.

Ασαφής αλγόριθμοι: βασίζονται σε ασαφή βάση εφαρμόζουν την έννοια της ασαφούς λογικής, η οποία χρησιμοποιεί τιμές αλήθειας μεταξύ 0 και 1. Σε ένα ασαφές μοντέλο, ορισμένοι κανόνες μπορούν να εφαρμοστούν για να αντιστοιχίσουν μια δεδομένη είσοδο με την αντίστοιχη έξοδο. Αυτό το μοντέλο παράγει καλά αποτελέσματα επειδή είναι κατάλληλο για αβέβαιους τομείς.

Υβριδικό αλγόριθμοι: χρησιμοποιούνται για την κατασκευή ενός υβριδικού μοντέλου, το οποίο χρησιμοποιεί πολλαπλούς ταξινομητές για την αναγνώριση προτύπων. Κάθε συγκεκριμένος ταξινομητής υποβάλλεται σε εκπαίδευση με βάση τους χώρους χαρακτηριστικών. Ένα σύνολο συνδυαστών και ταξινομητών χρησιμοποιείται για να εξαχθεί το συμπέρασμα. Μια συνάρτηση απόφασης χρησιμοποιείται για να αποφασίσει την ακρίβεια των ταξινομητών.

Αλγόριθμοι βασισμένοι σε νευρωνικά δίκτυα: σχηματίζουν ένα μοντέλο που αποτελείται από παράλληλες δομές (νευρώνες). Αυτό το μοντέλο είναι πιο ικανό από άλλα μοντέλα αναγνώρισης προτύπων λόγω των ανώτερων μαθησιακών του ικανοτήτων. Ένα καλό παράδειγμα νευρωνικού δικτύου που χρησιμοποιείται στην αναγνώριση προτύπων είναι το νευρωνικό δίκτυο Feed-Forward Backpropagation (FFBPNN).

2.3 Εφαρμογές αναγνώρισης προτύπων

Η αναγνώριση προτύπων μπορεί να εφαρμοστεί στους ακόλουθους τομείς:

Ανάλυση εικόνας: Η αναγνώριση προτύπων χρησιμοποιείται στην ανάλυση ψηφιακών εικόνων για την αυτόματη μελέτη εικόνων για τη συλλογή σημαντικών πληροφοριών από αυτές. Δίνει στα μηχανήματα την ευφυΐα αναγνώρισης που απαιτείται για την επεξεργασία εικόνας.

Σεισμική ανάλυση: Η σεισμική ανάλυση περιλαμβάνει τη μελέτη του τρόπου με τον οποίο φυσικά γεγονότα όπως οι σεισμοί επηρεάζουν τους βράχους, τα κτίρια και το έδαφος. Η αναγνώριση προτύπων χρησιμοποιείται για την ανακάλυψη και την ερμηνεία επαναλαμβανόμενων μοτίβων σε σεισμικά γεγονότα.

Υγειονομική περίθαλψη: Η αναγνώριση προτύπων χρησιμοποιείται στον τομέα της υγειονομικής περίθαλψης για τη βελτίωση των υπηρεσιών υγείας. Τα δεδομένα των ασθενών αποθηκεύονται και χρησιμοποιούνται από τους ιατρούς για περαιτέρω ανάλυση. Αυτή η τεχνική χρησιμοποιείται επίσης για την αναγνώριση όγκων και προβλημάτων στο ανθρώπινο σώμα.

Αναγνώριση δακτυλικών αποτυπωμάτων: Αυτή η διαδικασία χρησιμοποιείται για την αναγνώριση δακτυλικών αποτυπωμάτων σε υπολογιστές και έξυπνες συσκευές. Τα σύγχρονα έξυπνα τηλέφωνα διαθέτουν μια δυνατότητα αναγνώρισης δακτυλικών αποτυπωμάτων που μας επιτρέπει να αποκτήσουμε πρόσβαση στο τηλέφωνό μας αφού γίνει επαλήθευση του δακτυλικού μας αποτυπώματος.

Υπολογιστική όραση: Χρησιμοποιείται σε εφαρμογές υπολογιστών για την εξαγωγή χρήσιμων χαρακτηριστικών από δείγματα εικόνων. Έχει εφαρμοστεί στην όραση υπολογιστή για την εκτέλεση διαφόρων εργασιών όπως η αναγνώριση αντικειμένων.

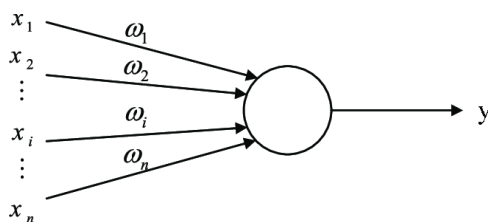
Τεχνολογία ομιλίας: Η αναγνώριση προτύπων στο τομέα της τεχνολογίας ομιλίας προσφέρει εργαλεία για την δημιουργία εύρωστων ταξινομητών για την αυτόματη αναγνώριση ομιλίας αλλά και σε άλλα θέματα της ψηφιακής επεξεργασίας σήματων ομιλίας (π.χ. αποθορυβοποίηση, αυτόματη κατάτμηση κ.α.)

3 Νευρωνικά δίκτυα βαθιάς μάθησης

Σε μια προσπάθεια να δημιουργηθούν συστήματα που μαθαίνουν παρόμοια με το πώς μαθαίνουν οι άνθρωποι, η υποκείμενη αρχιτεκτονική για τη βαθιά μάθηση εμπνεύστηκε από τη δομή του ανθρώπινου εγκεφάλου. Για αυτόν τον λόγο, αρκετές θεμελιώδεις ορολογίες στη βαθιά μάθηση μπορούν να αντιστοιχιστούν στη νευρολογία. Παρόμοια με το πώς οι νευρώνες σχηματίζουν τα θεμελιώδη δομικά στοιχεία του εγκεφάλου, η αρχιτεκτονική βαθιάς μάθησης περιέχει μια υπολογιστική μονάδα που επιτρέπει τη μοντελοποίηση μη γραμμικών συναρτήσεων που ονομάζεται perceptron [1].

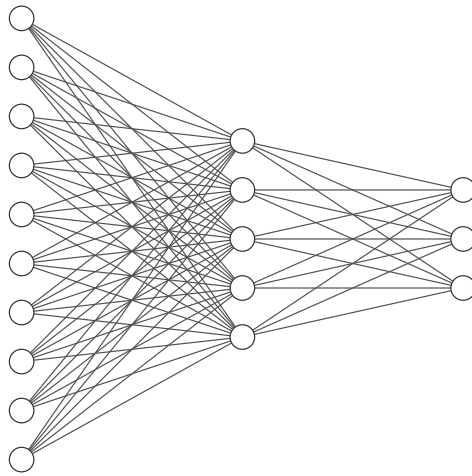
Η λειτουργία της βαθιάς μάθησης ουσιαστικά ξεκινά από τον αλγόριθμο perceptron. Παρόμοια με το πώς ένας «νευρώνας» του ανθρώπινου εγκεφάλου μεταδίδει ηλεκτρικούς παλμούς σε όλο το νευρικό μας σύστημα, ο αλγόριθμος perceptron λαμβάνει μια λίστα σημάτων εισόδου και τα μετατρέπει σε σήματα εξόδου.

Η βασική λειτουργία του perceptron είναι η κατανόηση της αναπαράστασης δεδομένων στοιβάζοντας μαζί πολλά επίπεδα, όπου κάθε επίπεδο είναι υπεύθυνο για την κατανόηση κάποιου μέρους της εισόδου. Ένα επίπεδο μπορεί να θεωρηθεί ως μια συλλογή υπολογιστικών μονάδων που μαθαίνουν να ανιχνεύουν μια επαναλαμβανόμενη αλληλουχία τιμών.



Εικόνα 9: Τεχνητός νευρώνας (perceptron)

Κάθε επίπεδο perceptron είναι υπεύθυνο για την ερμηνεία ενός συγκεκριμένου μοτίβου μέσα στα δεδομένα. Ένα δίκτυο από perceptron μιμείται τον τρόπο με τον οποίο οι νευρώνες στον εγκέφαλο σχηματίζουν ένα δίκτυο, επομένως η αρχιτεκτονική ονομάζεται νευρωνικά δίκτυα (ή τεχνητά νευρωνικά δίκτυα).



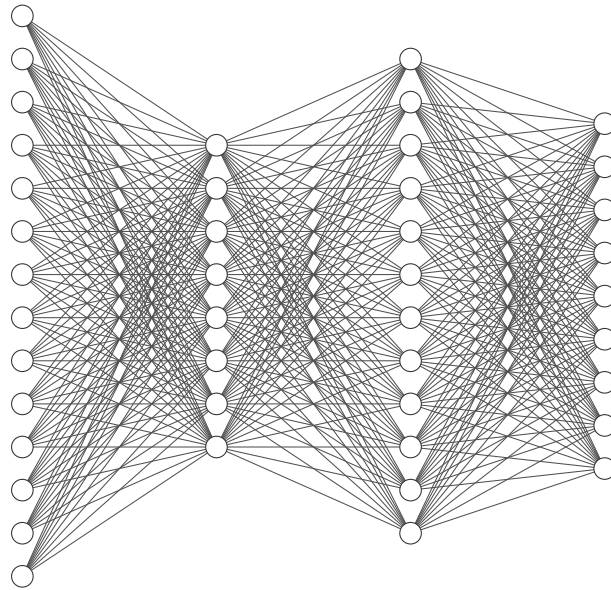
Εικόνα 10: Νευρωνικό δίκτυο

Μια πολυεπίπεδη αρχιτεκτονική perceptron περιγράφει ένα νευρωνικό δίκτυο (neural network). Στην πιο βασική του μορφή, ένα νευρωνικό δίκτυο περιέχει τρία επίπεδα: επίπεδο εισόδου, κρυφό επίπεδο και επίπεδο εξόδου. Όπως φαίνεται στο παρακάτω σχήμα, ένα δίκτυο με μόνο ένα κρυφό στρώμα ονομάζεται ρηχό νευρωνικό δίκτυο (shallow neural network).

Ένα βαθύ νευρωνικό δίκτυο είναι απλώς ένα ρηχό νευρωνικό δίκτυο με περισσότερα από ένα κρυφά στρώματα. Κάθε νευρώνας στο κρυφό στρώμα συνδέεται με πολλούς άλλους. Κάθε βέλος έχει μια ιδιότητα βάρους συνδεδεμένη σε αυτό, η οποία ελέγχει πόσο επηρεάζει η ενεργοποίηση αυτού του νευρώνα τους άλλους που συνδέονται με αυτό.

Η λέξη «βαθιά» στη βαθιά μάθηση αποδίδεται σε αυτά τα βαθιά κρυμμένα στρώματα και αντλεί την αποτελεσματικότητά της από αυτήν. Η επιλογή του αριθμού των κρυφών επιπέδων εξαρτάται από τη φύση του προβλήματος και το μέγεθος του συνόλου δεδομένων. Το παρακάτω σχήμα δείχνει ένα βαθύ νευρωνικό δίκτυο με δύο κρυφά στρώματα.

Αν και οι μέθοδοι βαθιάς μάθησης κέρδισαν τεράστια δημοτικότητα τα τελευταία 10 περίπου χρόνια, η ιδέα υπήρχε από τα μέσα της δεκαετίας του 1950 όταν ο Frank Rosenblatt ανακάλυψε το perceptron σε μια μηχανή IBM 704. Ήταν μια ηλεκτρονική συσκευή δύο επιπέδων που είχε την ικανότητα να ανιχνεύει σχήματα και να κάνει συλλογισμούς.



Εικόνα 11: Βαθύ Νευρωνικό Δίκτυο

Οι εξελίξεις σε αυτόν τον τομέα τα τελευταία χρόνια οφείλονται κυρίως στην αύξηση της υπολογιστικής ισχύος και των υψηλών επιδόσεων γραφικών μονάδων επεξεργασίας (GPU), σε συνδυασμό με τη μεγάλη αύξηση του πλούτου των δεδομένων που έχουν στη διάθεσή τους αυτά τα μοντέλα για μάθηση, καθώς και στο ενδιαφέρον και χρηματοδότηση από την κοινότητα για συνεχή έρευνα. Αν και η βαθιά μάθηση έχει απογειωθεί τα τελευταία χρόνια, έρχεται με το δικό της σύνολο προκλήσεων που οι ερευνητές εργάζονται για να επιλύσουν.

3.1 Μηχανική Μάθηση σε μικροσκοπικές κλίμακες

Η μηχανική μάθηση σε μικροσκοπικές κλίμακες (MMσMK, TinyML) είναι ένας τύπος μηχανικής μάθησης που συρρικνώνει τα δίκτυα βαθιάς εκμάθησης για να μπορέσουν να εκτελεστούν σε μικροσκοπικό υλικό (hardware). Ενσωματωμένες συσκευές εξαιρετικά χαμηλής ισχύος εισβάλλουν στον κόσμο μας και με νέα ενσωματωμένα πλαίσια μηχανικής μάθησης, θα επιτρέψουν περαιτέρω τον πολλαπλασιασμό συσκευών διαδικτύου των πραγμάτων (ΔτΠ) που λειτουργούν με τεχνητή νοημοσύνη. Όπως είδαμε η μηχανική μάθηση είναι μια έννοια που υπάρχει εδώ και λίγο καιρό, με πολλές χρήσιμες εφαρμογές για μεγάλα δεδομένα που χρειάζεται να κατανοήσουμε που όμως συνδέεται λιγότερο συχνά με το υλικό. Ωστόσο, οι εφαρμογές μηχανικής μάθησης σε μικρές/φορητές συσκευές δεν είναι

κάτι καινούργιο. Εδώ και μερικά χρόνια, τα περισσότερα από τα έξυπνα τηλέφωνα έχουν μια σειρά από έξυπνες λειτουργίες που έχουν υλοποιηθεί με αρχιτεκτονική νευρωνικού δικτύου. Παραδείγματα όπως, η αναγνώριση μουσικής, διάφορες προηγμένες λειτουργίες της ενσωματωμένης κάμερας (π.χ. η νυχτερινή όραση, αυτόματη επεξεργασία πορτρέτου) είναι μερικά μόνο παραδείγματα που βασίζονται στην ενσωματωμένη βαθιά εκμάθηση. Ωστόσο, υπάρχουν πολλές προκλήσεις για την ενσωματωμένη τεχνητή νοημοσύνη, μερικές από τις οποίες είναι περιορισμένη υπολογιστική ισχύς και ο χώρος αποθήκευσης. Σε αυτά τα προβλήματα έρχεται να δώσει λύση η ΜΜσΜΚ. (Bagnoli, 2021)

Η εξέλιξις στο χώρο της ΜΜσΜΚ ουσιαστικά οδηγούνται από την βιβλιοθήκη TensorFlow Lite και ενσωματωμένες συσκευές όπως αυτές που παράγονται από την εταιρία Arduino η οποία κατασκευάζει υλικό ανοιχτού κώδικα που επιτρέπει σε οποιονδήποτε να αγοράσει μια πλακέτα μικροελεγκτή και να δημιουργήσει τη δική του ψηφιακή συσκευή. Το TensorFlow Lite είναι βιβλιοθήκη μηχανικής εκμάθησης για ενσωματωμένες συσκευές που δημιουργήθηκε από την Google και έχει μια υποκατηγορία ειδικά σχεδιασμένη για μικροελεγκτές. Το 2019, εκτός από το TensorFlow Lite, άλλα πλαίσια άρχισαν να εστιάζουν στο να κάνουν τα μοντέλα βαθιάς εκμάθησης μικρότερα, γρηγορότερα και προσαρμοσμένα στο ενσωματωμένο υλικό, συμπεριλαμβανομένου του uTensor και του Arm's CMSIS-NN. Ταυτόχρονα, άρχισαν να εμφανίζονται πολυάριθμα σεμινάρια YouTube σχετικά με τον τρόπο χρήσης του ΜΜσΜΚ και παρόμοιων πλαισίων σε μικροελεγκτές που λειτουργούν με τεχνητή νοημοσύνη για την εκπαίδευση, την επικύρωση και στη συνέχεια την ανάπτυξη μικρών νευρωνικών δικτύων σε υλικό μέσω μηχανών συμπερασμάτων.

Τα οφέλη που φέρνει η χρήση της ΜΜσΜΚ σε ενσωματωμένες συσκευές είναι οι παρακάτω:

1. Εξοικονομεί ενέργεια: Οι μικροελεγκτές χρησιμοποιούν χαμηλή ισχύ και μπορούν να λειτουργήσουν με μπαταρίες για μεγάλο χρονικό διάστημα. Εξοικονομεί ενέργεια και το καθιστά οικονομικά αποδοτικό.
2. Χωρίς καθυστέρηση στη διαδρομή δεδομένων: Δεν απαιτείται να αποστέλλονται δεδομένα στον διακομιστή κάθε φορά, και τα δεδομένα που αποστέλλονται δεν είναι μεγάλου όγκου αφού η επεξεργασία γίνεται πάνω στον αισθητήρα (π.χ. δεν αποστέλλεται ο ήχος αλλά το γεγονός ως κείμενο).

3. Δεν απαιτείται Διαδίκτυο: Δεδομένου ότι τα δεδομένα δεν αποστέλλονται στον διακομιστή κάθε τόσο, η διαδικασία πραγματοποιείται σε λιγότερο εύρος ζώνης και μερικές φορές χωρίς χρήση διαδικτύου. Επομένως, δεν εξαρτάται από τη σύνδεση.
4. Απόρρητο δεδομένων: Εφόσον τα δεδομένα δεν αποστέλλονται σε εξωτερικούς χρήστες ή ιστότοπους, τα δεδομένα παραμένουν ασφαλή και το απόρρητο προστατεύεται.

3.1.1 ΜΜσΜΚ: Εφαρμογές & περιπτώσεις χρήσης

Η ΜΜσΜΚ έχει σημειώσει επιτυχία στην εκτέλεση σε συσκευές αιχμής και προσφέρει πολλές λύσεις. Μπορεί να απαντήσει σε ηχητικές εντολές για την εκτέλεση ενεργειών μέσω χημικών αλληλεπιδράσεων. Το Google Assistant και η Alexa είναι μερικά παραδείγματα του ΜΜσΜΚ. Οι συσκευές είναι πάντα ενεργοποιημένες και αναλύουν τη φωνή μας για να εντοπίσουν τη λέξη αφύπνισης. Μερικές άλλες εφαρμογές είναι:

Βιομηχανία: χρησιμοποιείται σε συσκευές χαμηλής κατανάλωσης, για να μπορεί να ανιχνεύει συνεχώς σφάλματα σε ένα μηχάνημα εκ των προτέρων. Συνεπάγεται συντήρηση με βάση τη προβλέψις. Ένα τέτοιο παράδειγμα είναι η εισαγωγή μιας συσκευής IoT από μια αυστραλιανή νεοσύστατη εταιρεία, την Ping Services, η οποία παρακολουθεί τις ανεμογεννήτριες προσαρτώντας τον εαυτό της στο εξωτερικό της τουρμπίνας. Ειδοποιεί τις αρχές εάν αντιληφθεί οποιοδήποτε πιθανό πρόβλημα ή δυσλειτουργία.

Γεωργία: Το εργαλείο του TensorFlow Lite επιτρέπει στους αγρότες να ανιχνεύουν ασθένειες σε ένα φυτό όταν το φωτογραφίζουν. Λειτουργεί σε οποιαδήποτε συσκευή και δεν χρειάζεται σύνδεση στο διαδίκτυο. Η διαδικασία επιτρέπει την προστασία των γεωργικών συμφερόντων και αποτελεί κρίσιμη απαίτηση για τους απομακρυσμένους αγρότες.

Νοσοκομεία: Ένα έργο που ονομάζεται Solar Scare Mosquito χρησιμοποιεί το ΜΜσΜΚ για να σταματήσει την εξάπλωση ασθενειών όπως ο δάγκειος πυρετός, η ελονοσία κ.λπ. Λειτουργεί με ηλιακή ενέργεια, ανιχνεύει τις συνθήκες αναπαραγωγής κουνουπιών και δίνει σήμα στο νερό για να αποτρέψει την αναπαραγωγή κουνουπιών. ("An Introduction to TinyML", 2020)

4 Υλοποίηση συστήματος αναγνώρισης λέξεων κλειδιών

4.1 Πλατφόρμα υλοποίησης

Η πλατφόρμα Edge Impulse μέσα από ένα φιλικό περιβάλλον επιτρέπει την επεξεργασία πολυμεσικών δεδομένων και την εκπαίδευση/αξιολόγηση αλγορίθμων μηχανικής μάθησης συμβατούς με ενσωματωμένες συσκευές. Δίνει την δυνατότητα στους προγραμματιστές να δημιουργήσουν την νέα γενιά έξυπνων συσκευών χρησιμοποιώντας μηχανική μάθηση. Με τη χρήση του Edge Impulse η συλλογή δεδομένων μέσα από πραγματικούς αισθητήρες, η επεξεργασία ζωντανών σημάτων από μη επεξεργασμένα δεδομένα από τα νευρωνικά δίκτυα καθώς η δοκιμή και η ανάπτυξη στην συσκευή της επιλογής μας γίνεται ακόμα πιο εύκολη.

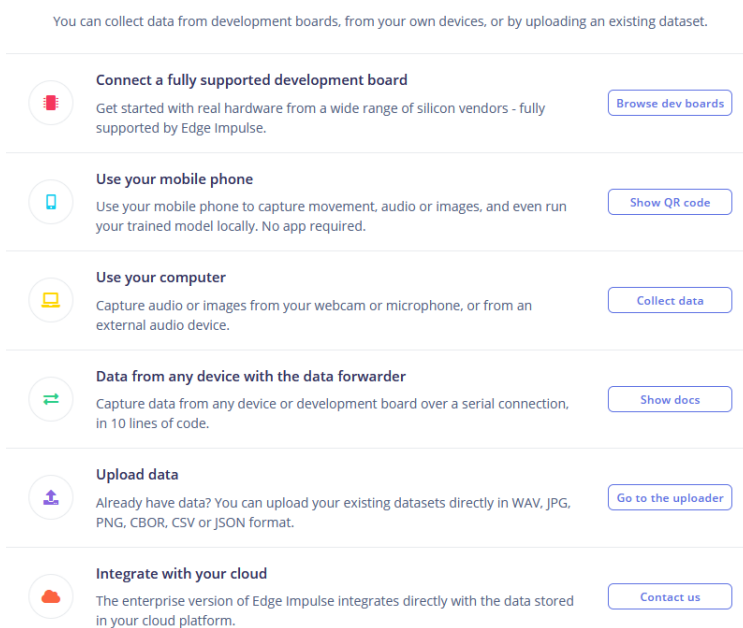
Για την συλλογή δεδομένων η πλατφόρμα του Edge Impulse μας δίνει τη δυνατότητα να συλλέξουμε δεδομένα με τη χρήση:

- Κάποιας ενσωματωμένης συσκευής όπως οι πλακέτες Arduino Nano 33 BLE Sense, Raspberry Pi4, Thunderboard Sense 2 κ.α.
- Μας επιτρέπει την συλλογή ακουστικών δεδομένων, εικονικών δεδομένων ακόμα και να εκτελέσουμε το εκπαιδευμένο μοντέλο μας τοπικά στο έξυπνο κινητό μας χωρίς τη χρήση κάποιας εφαρμογής.
- Με τη χρήση της WebCam του υπολογιστή μας μπορούμε να συλλέξουμε εικονικά και ηχητικά δεδομένα, ή με τη χρήση μικρόφωνου ή με τη χρήση κάποιας εξωτερικής συσκευής.
- Μπορούμε να χρησιμοποιήσουμε δεδομένα από οποιαδήποτε συσκευή με προώθηση δεδομένων. Μπορούμε να καταγράψουμε δεδομένα από οποιαδήποτε συσκευή ή πλακέτας ανάπτυξης μέσω σειριακής σύνδεσης μέσα απο λίγες γραμμές κωδικά.
- Ακόμα και στη περίπτωση που έχουμε συλλέξει ήδη τα δεδομένα μας, μπορούμε να τα ανεβάσουμε στη πλατφόρμα του Edge Impulse χωρίς να επηρεάζονται τα ήδη υπάρχων δεδομένα στη πλατφόρμα. Η πλατφόρμα υποστηρίζει δεδομένα τύπου WAV, JPG, PNG, CBOR, CSV και JSON.
- Τέλος το Edge Impulse στην εταιρική του έκδοση υποστηρίζει την ενσωμάτωση του, στο Cloud μας ώστε να δέχεται τα δεδομένα απευθείας από το Cloud μας.

4.1.2 Εκπαίδευση μοντέλων αναγνώρισης ηχητικών ροών Συλλογή ηχητικών δεδομένων

Στη συγκεκριμένη πτυχιακή εργασία η συλλογή των δεδομένων έγινε με τη χρήση έξυπνων κινητών, μια δυνατότητα η οποία υποστηρίζεται από το ρεπερτόριο λειτουργιών της πλατφόρμας Edge Impulse.

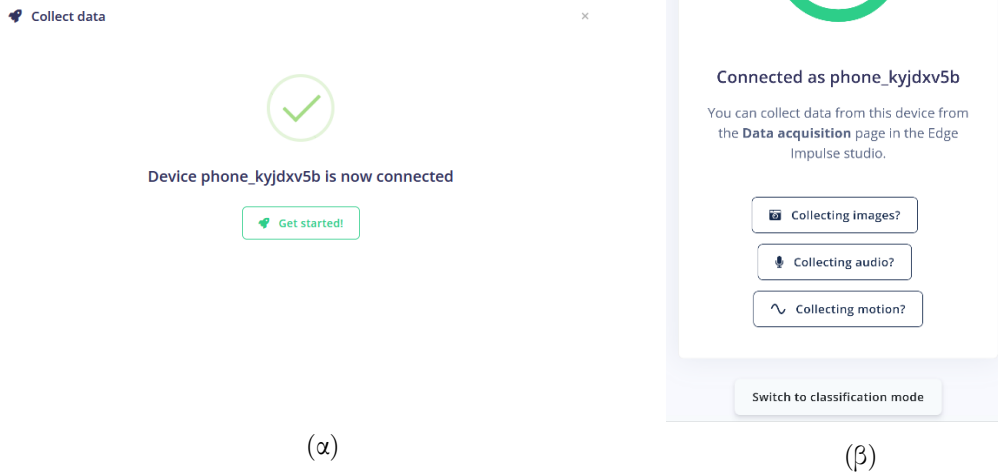
Για να μπορέσει κάποιος να συλλέξει και να ανεβάσει δεδομένα στην Database της εργασίας του στο Impulse Edge θα πρέπει να συνδέσει πρώτα το έξυπνο κινητό του με την πλατφόρμα. Για να γίνει επιτυχής η σύνδεση επιλέξαμε από την κυρία σελίδα την επιλογή 'Lets collect some data' και αμέσως μετά επιλέξαμε την δεύτερη επιλογή 'Use your mobile phone', ώστε να χρησιμοποιήσουμε το έξυπνο κινητό μας.



Εικόνα 12: Τρόποι συλλογής δεδομένων στην πλατφόρμα Edge Impulse

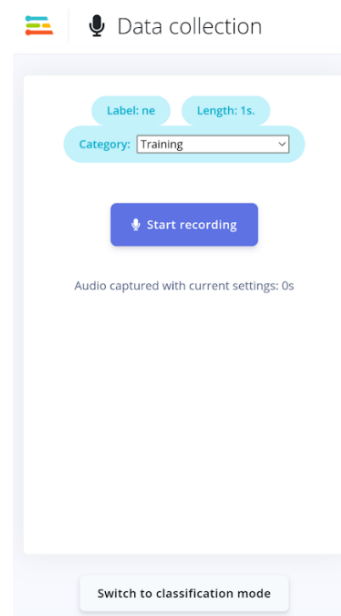
Αφού επιλέξουμε αυτή τη λειτουργία, το Edge Impulse εμφανίζει ένα QR Code που μας επιτρέπει την άμεση σύνδεση με το κινητό για να γίνουν οι ηχογραφήσεις. Μετά την σάρωση με του QR Code με το έξυπνο κινητό, το Edge Impulse επαληθεύει τη σωστή σύνδεση με τη πλατφόρμα με ένα μήνυμα πως το κινητό είναι πλέον συνδεδεμένο. Ταυτόχρονα εμφανίζεται και στο κινητό το μήνυμα πως το

κινητό αυτό είναι συνδεδεμένο με τη πλατφόρμα και π...
δεδομένα θα συλλέξουμε στη συγκεκριμένη συσκευή, κ...
κίνηση.



Εικόνα 13: (α) Επιτυχής σύνδεση , (β) Συλλογή δεδομένων

Αφου επιλεξουμε πως θελουμε να συλλεξουμε ηχους, η συσκευη μας ειναι ειναι και συνδεδειμεινη με τη πλατοφρμα του Edge Impulse. Αυτο οδηγει σε 3 επιλογες πριν ξεκινησει η ηχογραφηση, μας δινετε η επιλογη να βαλουμε ετικετα, να ορισουμε τη διαρκεια της ηχογραφησης και να ορισουμε σε ποια κατηγορια θα ενταχθουν τα αρχεια που θα ηχογραφησουμε. Ο ορισμος ετικετας ειναι πολυ σημαντικος καθως η βοηθεια του ειναι πολυτιμη στον διαχωρισμο των δεδομενων. Στην παρουμεναι εικαμε εξι ετικετες, μια για καθε λεξη, ναι, οχι, παμε, στοπ, ανοιξε, κλεισε. Οι λεξεις αυτες εχουν τις αντιστοιχες ετικετες ne, ohi, pame, stop,



Εικόνα 14: Συλλογή δεδομένων ήχου

ani3e, klise. Η διαρκεια καθε λεξης που ηχογραφηθηκε δεν ξεπερασε το ενα δευτερολεπτο. Υπαρχουν 3 επιλογες στις κατηγοριες, υπαρχει η επιλογη της

εκπαίδευσης όπου εάν επιλεγεί όλα τα αρχεία θα πάνε για την εκπαίδευση του μοντέλου. Η κατηγορία του τεστ όπου όλα τα δεδομένα θα πάνε για το τεστ του μοντέλου και τέλος η κατηγορία της αυτοματικής διαχωρισής. Σε αυτή τη περίπτωση τα δεδομένα θα διαχωριστούν σε δεδομένα εκπαίδευσης και δεδομένα για τεστ αυτομάτα από το Edge Impulse με την αναλογία 80/20.

4.1.3 Εξαγωγή παραμέτρων

Μετά την συλλογή δεδομένων το επόμενο βήμα είναι η εξαγωγή των δεδομένων αυτών. Όσον αφορά τη βαθμίδα εξαγωγής παραμέτρων χρησιμοποιήσαμε τις δυνατότητες που μας παρέχει η πλατφόρμα Impulse Edge. Συγκεκριμένα χρησιμοποιήσαμε τις παραμέτρους MFCC, MFE και φασματογράφημα.

Φασματογράφημα (Spectrogram): Το μπλοκ επεξεργασίας Spectrogram εξάγει χαρακτηριστικά χρόνου και συχνότητας από ένα σήμα. Οι επιδόσεις του είναι καλύτερες όταν γίνεται η χρήση του σε περιπτώσεις χωρίς αναγνώριση φωνής ή σε δεδομένα αισθητήρων με συνεχείς συχνότητες.

Οι παράμετροι ενός φασματογράμματος είναι τα εξής :

Μήκος Παραθύρου: Το μήκος κάθε καρέ σε δευτερόλεπτα

Ολισθήση παραθύρου: Το βήμα μεταξύ διαδοχικών καρέ σε δευτερόλεπτα

Ζώνες συχνότητων: Το μέγεθος FFT

Βαθμός θορύβου: Σήμα χαμηλότερο από αυτό το επίπεδο θα αποκοπεί.

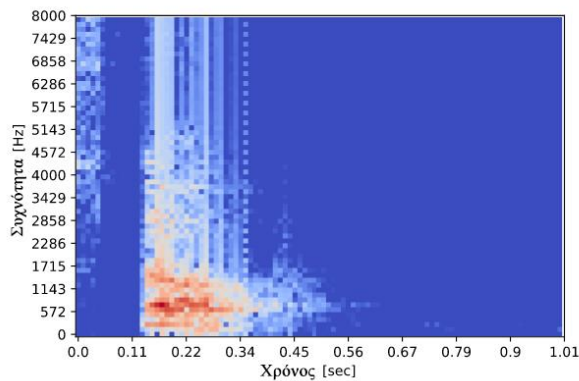
Το φασματογράφημα αρχικά χωρίζει το παράθυρο σε πολλαπλά επικαλυπτόμενα πλαίσια. Το μέγεθος όπως και αριθμός των πλαισίων μπορούν να ρυθμιστούν με τις παραμέτρους, μήκος πλαισίου και διασκελισμός πλαισίου.

Κάθε χρονικό πλαίσιο χωρίζεται στη συνέχεια σε πεδία συχνότητων χρησιμοποιώντας έναν μετασχηματισμό FFT και υπολογίζουμε το φάσμα ισχύος του. Ο αριθμός των πεδίων συχνότητας είναι ίσος με τη παράμετρο της ζώνης συχνότητας διαιρούμενα με το 2 συν 1. Συνιστάται να διατηρηθεί η τιμή ζωνών συχνότητας (είναι γνωστό και ως μέγεθος FFT) ως ισχύς για σκοπούς απόδοσης. Τέλος η τιμή κατωτάτου θορύβου εφαρμόζεται στο φάσμα ισχύος.

Τα χαρακτηριστικά που δημιουργούνται από το Spectrogram είναι ίσα με τον αριθμό των χρονικών πλαισίων που δημιουργούνται επί τον αριθμό των πλαισίων συχνότητας.

("Spectrogram", n.d.)

<https://docs.edgeimpulse.com/docs/spectrogram>



Εικόνα 15: Παράδειγμα φασματογραφήματος (Spectrogram)

MFE: Όπως και το Spectrogram το μπλοκ επεξεργασίας ήχου MFE εξάγει χαρακτηριστικά χρόνου και συχνότητας από ένα σήμα. Ωστόσο κάνει χρήση μιας μη γραμμικής κλίμακας στο τομέα της συχνότητας που ονομάζεται κλίμακα Mel. Έχει αρκετά καλή απόδοση σε δεδομένα ήχου κυρίως για χρήση μη αναγνώριση φωνής σε περιπτώσεις όπου οι ήχοι που πρέπει να ταξινομηθούν και μπορούν να αναγνωριστούν από το ανθρώπινο αυτί. Οι παράμετροι ενός φίλτρου Mel (Mel-filterbank) είναι τα εξής:

Μήκος παραθύρου: Το μήκος κάθε καρέ σε δευτερόλεπτα

Ολίσθηση παραθύρου: Το βήμα μεταξύ διαδοχικών καρέ σε δευτερόλεπτα

Αριθμός φίλτρου: Ο αριθμός τριγωνικών φίλτρων που εφαρμόζονται στο φασματόγραμμα.

Μήκος FFT: Το μέγεθος FFT

Χαμηλή συχνότητα: Χαμηλότερη άκρη ζώνης φίλτρων της κλίμακας Mel

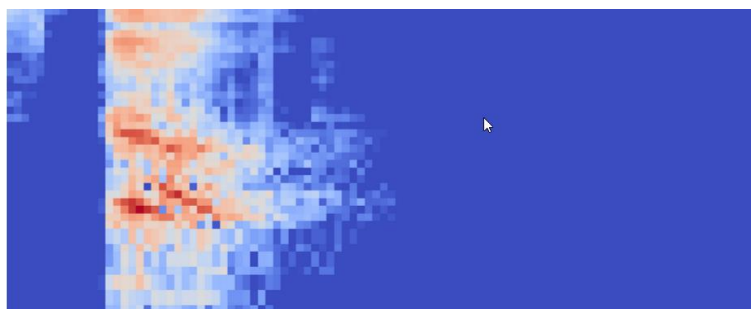
Υψηλή συχνότητα: Υψηλότερη άκρη ζώνης φίλτρων της κλίμακας Mel

Ομαλοποίηση

Βαθμός θορύβου: Το σήμα χαμηλότερο από αυτό το επίπεδο θα αποκοπεί
Οι εξαγωγές των χαρακτηριστικών είναι παρόμοιες με το φασματογράφημα οι παράμετροι, μήκος παραθύρου, ολίσθηση παραθύρου και μήκος FFT είναι ίδια αλλά προσθέτει δυο επιπλέον βήματα.

Μετά τον υπολογισμό του φασματογραφήματος γίνεται η εφαρμογή τριγωνικών φίλτρων της κλίμακας Mel για την εξαγωγή ζωνών συχνοτήτων. Έχει γίνει διαμόρφωση σχετικά με τις παραμέτρους του αριθμού φίλτρου, χαμηλή και υψηλή συχνότητα για την επιλογή της ζώνης συχνοτήτων και του αριθμού συχνοτήτων που θα εξαχθούν. Η κλίμακα Mel είναι μια αντιληπτική κλίμακα υψών που κρίνονται από τους ακροατές ως ίσες σε απόσταση μεταξύ τους. Η κεντρική ιδέα είναι να εξαχθούν περισσότερα χαρακτηριστικά στις χαμηλότερες συχνότητες και λιγότερα στις υψηλές συχνότητες επομένως αποδίδει καλά σε ήχους που μπορούν να διακριθούν από το ανθρώπινο αυτί.

Το τελευταίο βήμα είναι να εκτελεστεί μια τοπική μέση ομαλοποίηση του σήματος, εφαρμόζοντας την τιμή κατωτάτου θορύβου στο φάσμα ισχύος. ("Audio MFE", n.d.)



Εικόνα 16: Παράδειγμα απεικόνισης συντελεστών MFCC

MFCC: Το μπλοκ επεξεργασίας MFCC εξάγει συντελεστές από ένα ηχητικό σήμα. Όπως με το μπλοκ επεξεργασίας MFE κάνει χρήση της μη γραμμικής κλίμακας Mel. Είναι το μπλοκ αναφοράς για αναγνώριση ομιλίας και μπορεί να έχει εξίσου καλή επίδοση σε ορισμένες περιπτώσεις όπου η φωνή κατά την ομιλία δεν προέρχεται από άνθρωπο. Οι παράμετροι για τα φίλτρα MFCC (*Mel Frequency Cepstral Coefficients*) είναι οι εξής:

Αριθμός συντελεστών : Ο αριθμός συντελεστών cepstral που πρέπει να διατηρηθούν μετά την εφαρμογή διακριτού μετασχηματισμού συνημιτόνου.

Μήκος παραθύρου : Το μήκος κάθε καρέ σε δευτερόλεπτα

Ολίσθηση παραθύρου: Το βήμα μεταξύ διαδοχικών καρέ σε δευτερόλεπτα

Αριθμός φίλτρου: Ο αριθμός τριγωνικών φίλτρων που εφαρμόζονται στο φασματόγραμμα.

Μήκος FFT: Το μέγεθος FFT

Χαμηλή συχνότητα: Χαμηλότερη άκρη ζώνης φίλτρων της κλίμακας Mel

Υψηλή συχνότητα: Υψηλότερη άκρη ζώνης φίλτρων της κλίμακας Mel

Μέγεθος παραθύρου: Το μέγεθος του ολισθαίνοντάς παραθύρου για τον μέσο ορό του τοπικού αποτελέσματος του αντίστροφου μετασχηματισμού Fourier (cepstral). Το μέγεθος του παραθύρου πρέπει να είναι μονό.

Συντελεστής: Ο συντελεστής προ-έμφασης που θα εφαρμοστεί στο σήμα εισόδου (0 ισοδυναμεί με μηδενικό φιλτράρισμα)

Μετατόπιση: Η τιμή μετατόπισης προ-έμφασης για την αντιστροφή του σήματος εισόδου.

Η εξαγωγή των χαρακτηριστικών προσθέτει ένα επιπλέον βήμα στο μπλοκ MFE, με αποτέλεσμα τη συμπιεσμένη αναπαράσταση των filterbanks. Εφαρμόζεται ένας διακριτός μετασχηματισμός συνημιτόνου σε κάθε filterbank για την εξαγωγή συντελεστών cepstral. Συνήθως διατηρούνται 13 συντελεστές, ενώ οι υπόλοιποι απορρίπτονται καθώς αντιπροσωπεύουν γρήγορες αλλαγές που δεν είναι χρήσιμες για την αναγνώριση ομιλίας. ("Audio MFCC", n.d.)

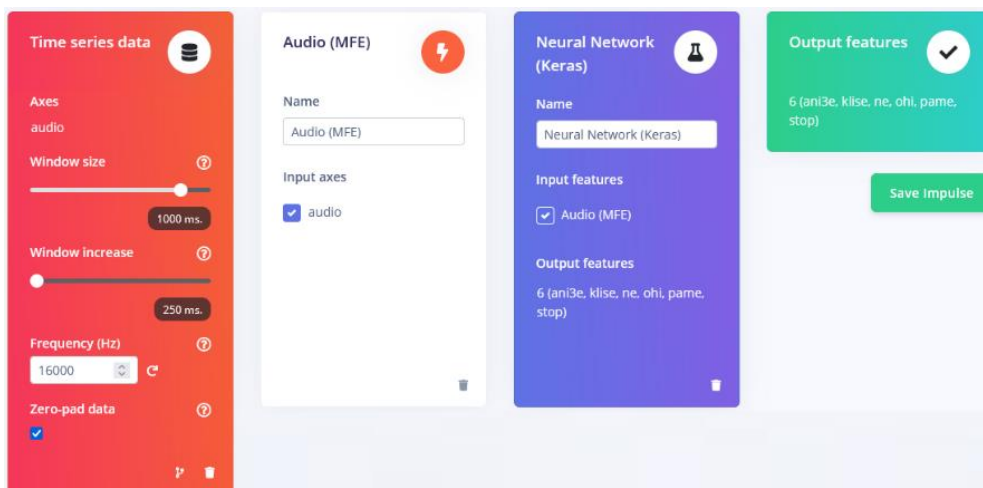


Εικόνα 17: Παράδειγμα απεικόνισης συντελεστών MFE

Το Edge Impulse μας επιτρέπει την εφαρμογή επαύξησης δεδομένων (data augmentation) μέσα από την εφαρμογή επεξεργασίας. Σε αυτό το κομμάτι γίνονται τις σχετικές επιλογές για τα δεδομένα μας καθώς και για τα μπλοκ επεξεργασίας καθώς και για τα μπλοκ μάθησης. Στα αρχεία ήχου που έχουμε επιλέγουμε το μέγεθος παραθύρου των ηχογραφήσεων μας δηλαδή αυτό θα είναι το μέγεθος δεδομένων που θα επεξεργαστεί ανά ταξινόμηση σε χιλιοστά του δευτερολέπτου. Η πλατφόρμα παρέχει λειτουργίες όπως το Windows Increase όπου εάν ένα δείγμα είναι μεγαλύτερο από το μέγεθος παραθύρου, τότε ένα ακόμα παράθυρο πάει πάνω

από το δείγμα αυτό. Τέλος για τα δεδομένα, μας δίνει να επιλέξουμε εάν θα εφαρμοστεί η τεχνολογία Zero-pad data. Η τεχνολογία αυτή εάν εντοπίσει ένα δείγμα από το σύνολο των δεδομένων μας να είναι μικρότερο από το μήκος παραθύρου που έχει οριστεί τότε μπαίνει σε εφαρμογή η τεχνική Zero padding αλλιώς το δείγμα θα απορριφθεί.

Αφού γίνουν οι σχετικές ρυθμίσεις για τα δεδομένα μας σειρά έχει η επιλογή των μπλοκ επεξεργασίας και η επιλογή μπλοκ εκμάθησης. Υπάρχουν πολλές επιλογές για επιλογή μπλοκ επεξεργασίας, για την αναγνώριση λέξεων κλειδιών την καλύτερη απόδοση την έχουν τα μπλοκ επεξεργασίας Spectrogram, MFE και MFCC. Οι επιλογές για τα μπλοκ εκμάθησης είναι αναμεσα σε ταξινόμηση (classification) χρησιμοποιώντας τη βιβλιοθήκη Keras όπου το μοντέλο θα μαθαίνει μοτίβα από τα δεδομένα και θα μπορέσει να τα εφαρμόσει σε νέα δεδομένα, μια αρκετά καλή επιλογή για κατηγοριοποίηση κινήσεων ή αναγνώριση ήχου. Η επίμονη διαθέσιμη επιλογή είναι η παλινδρόμηση (regression) κάνοντας χρήση της βιβλιοθήκης Keras όπου το μοντέλο μαθαίνει μοτίβα από τα δεδομένα και θα μπορέσει να τα εφαρμόσει σε νέα δεδομένα, η συγκριμένη επιλογή είναι ιδανική για πρόβλεψη αριθμητικών συνεχών τιμών. Η τελευταία και ιδανική επιλογή για τη παρουσία πτυχιακή εργασία, είναι η μεταφορά εκμάθησης (transfer learning) με χρήση της βιβλιοθήκης Keras. Με αυτό το μπλοκ εκμάθησης το μοντέλο θα συντονίσει με ακρίβεια ένα προ εκπαιδευμένο μοντέλο εντοπισμού λέξεων κλειδιών στα δεδομένα μας το οποίο έχει καλή απόδοση ακόμα και με μικρά σύνολα δεδομένων λέξεων κλειδιών.



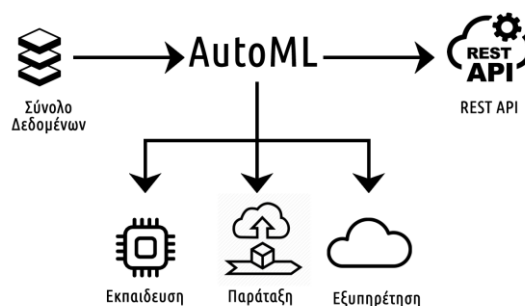
Εικόνα 18: Ρυθμίσεις μοντέλου εκπαίδευσης

1.1.1 Αυτόματη Μάθηση (AutoML)

Η αυτοματοποιημένη μάθηση (Automated Machine Learning – AutoML) είναι η διαδικασία αυτοματοποίησης των χρονοβόρων, επαναληπτικών εργασιών της ανάπτυξης ενός μοντέλου της μηχανικής μάθησης. Είναι πολύ σημαντικό καθώς επιτρέπει στους ερευνητές του πεδίου ανάλυσης δεδομένων, τους αναλυτές καθώς και στους προγραμματιστές να δημιουργήσουν μοντέλα μηχανικής μάθησης με μεγάλη κλίμακα, αποδοτικότητα, παραγωγικότητα και ταυτόχρονα διατηρείτε την ποιότητα του μοντέλου.

Η ανάπτυξη παραδοσιακών μοντέλων μηχανικής μάθησης απαιτεί πόρους, σημαντικό όγκο γνώσεων στον τομέα ώστε να μπορέσει να γίνει σωστή παραγωγή και σύγκριση δεκάδων μοντέλων. Με τη χρήση της αυτοματοποιημένης μηχανικής εκμάθησης, θα υπάρξει μεγάλη διαφορά στην ταχύτητα της διαδικασίας και κατά συνέπεια και στον χρόνο που απαιτείται για την απόκτηση μοντέλων μηχανικής μάθησης τα οποία θα είναι έτοιμα για παραγωγή με μεγάλη ευκολία και αποτελεσματικότητα. ("What is automated ML? AutoML - Azure Machine Learning", n.d.)

Η αυτοματοποιημένη μηχανική μάθηση είναι συνδεδεμένη με την παραγωγή λύσεων μηχανικής μάθησης για τους ερευνητές του πεδίου της ανάλυσης δεδομένων, χωρίς να γίνονται αμέτρητες έρευνες πάνω στη προετοιμασία δεδομένων για την επιλογή του μοντέλου, τις υπερ. παραμέτρους, καθώς και τις συμπυκνωμένες παραμέτρους του μοντέλου.

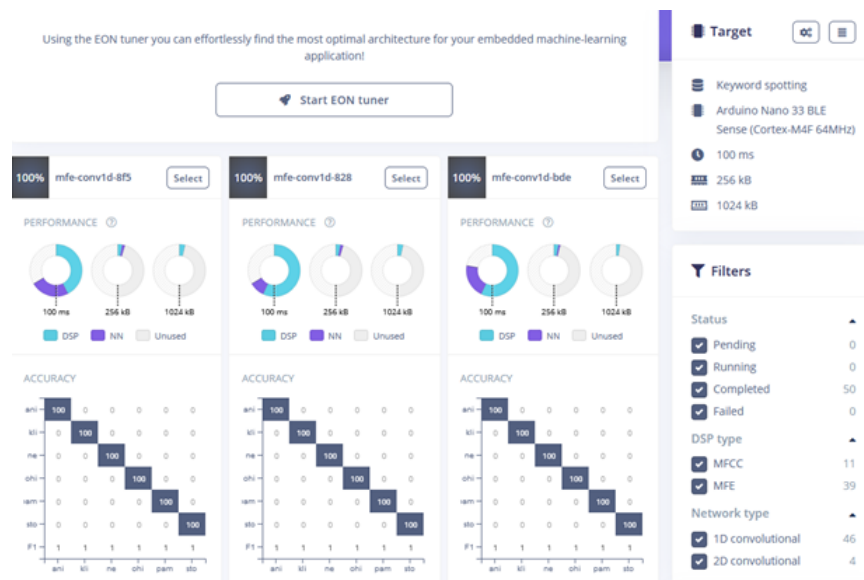


Εικόνα 19: Αυτοματοποιημένη εκμάθηση

Όσον αφορά τη προσέγγιση της αυτόματης μάθησης το Impulse Edge παρέχει σχετική λειτουργία μέσα από το εργαλείο του Eon Tuner. Όπως είδαμε η Edge Impulse δίνει τη δυνατότητα σε προγραμματιστές να δημιουργούν και να

αναπτύσσουν μοντέλα μηχανικής εκμάθησης που εκτελούνται σε ενσωματωμένες συσκευές. Η μνήμη σε τέτοιες συσκευές είναι αρκετά περιορισμένη για παράδειγμα μια τυπική μνήμη μπορεί να έχει λιγότερο από 128KB μνήμη RAM. Την λύση σε αυτό το πρόβλημα μας την έδωσε ο EON Tuner Compiler, ο οποίος μπορεί να κάνουμε χρήση νευρωνικών δικτύων με 25 έως 55% λιγότερη RAM καθώς και έως και 35% λιγότερη flash και διατηρεί την ίδια ακρίβεια σε σχέση με το TensorFlow Lite για μικροελεγκτές.

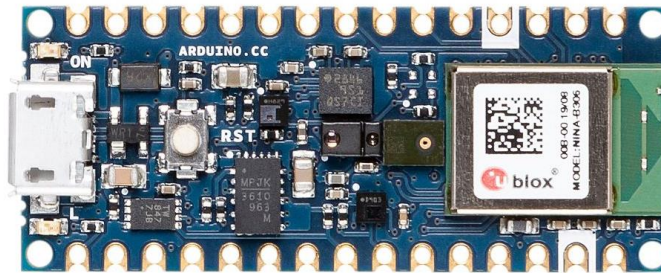
Αφού έχουμε συλλέξει τα δεδομένα μας και θέλουμε να ξεκινήσουμε τη διαδικασία εκπαίδευσης του μοντέλου πρέπει να επιλέξουμε στις ρυθμίσεις του EON Tuner πως θα χρησιμοποιήσουμε τα δεδομένα για παράδειγμα τα δεδομένα ήχου μπορεί να χρησιμοποιηθούν για την αναγνώριση ακουστικών γεγονότων, αναγνώριση λέξεων κλειδιών και αναγνώριση ήχων που δεν σταματάει. Μετά θα πρέπει να επιλέξουμε την συσκευή όπου θα χρησιμοποιηθεί το μοντέλο μας και τέλος τον χρόνο ανά συμπέρασμα. Η διαδικασία που ακολουθεί ο EON Tuner για την εκπαίδευση ενός μοντέλου είναι η εξής : Πρώτα αναλύει τα δεδομένα εισόδου, τα πιθανά μπλοκ επεξεργασίας σήματος και τις αρχιτεκτονικές νευρωνικών δικτύων. Αφού τελειώσει την ανάλυση μας δίνει μια επισκόπηση των πιθανών αρχιτεκτονικών μοντέλων που θα ταιριάζουν στις απαιτήσεις μνήμης της σκευής όπου επιλέξαμε.



Εικόνα 20: Παράδειγμα απεικόνισης αποτελεσμάτων του εργαλείου AutoML EON Tuner

4.2 Υλικό υλοποίησης (hardware) Arduino Nano 33 BLE

Η πλακέτα Arduino Nano 33 BLE Sense είναι σχεδιασμένη για να προσφέρει μια έξυπνη αλλά και αποδοτική λύση για όλους όσους επιθυμούν να έχουν συνδεσιμότητα Bluetooth με χαμηλή κατανάλωση ενέργειας στα project τους. Είναι βασισμένο στις προδιαγραφές επεξεργαστή NINA B306, ο οποίος φιλοξενεί έναν Nordic nRF52840 και περιέχει έναν μικροελεγκτή Cortex M4F στα 64 MHz. Έχει μνήμη 265 KB και 1 MB SRAM, υποστηρίζει Bluetooth 5 καθώς και Bluetooth Mesh που του επιτρέπει την δημιουργία δικτύων συσκευών μεγάλης κλίμακας και η τάση λειτουργίας του κυκλώματος είναι στα 3 V. Η διαφορά του Arduino Nano 33 BLE Sense με το Arduino Nano 33 BLE είναι ένα έξτρα σετ από αισθητήρες που είναι ενσωματωμένοι στην Arduino Nano 33 BLE Sense έκδοση.



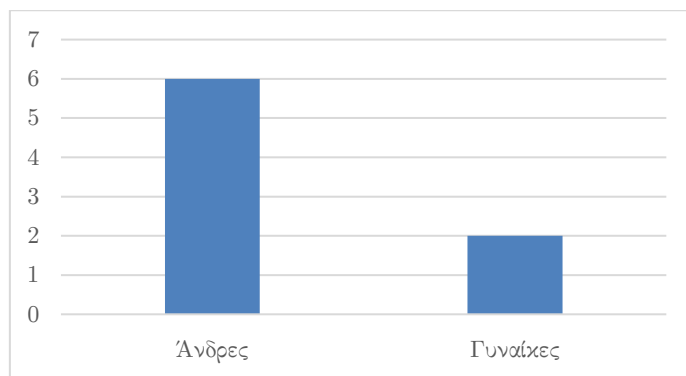
Εικόνα 21: Στιγμιότυπο ενσωματωμένης συσκευής Arduino Nano 33 BLE Sense
Ποιο συγκεκριμένα το η έκδοση Sense του συγκεκριμένου Arduino διαθέτει :

- IMU για ανίχνευση κίνησης: μονάδα αδρανειακής μέτρησης (IMU) LSM9DS1, διαθέτει τρισδιάστατο επιταχυνσιόμετρο, γυροσκόπιο και μαγνητόμετρο και επιτρέπει την ανίχνευση κίνησης, την ανίχνευση δόνησης και τον προσανατολισμό.
- Ακουστικός αισθητήρας (Το MP34DT05-A): ψηφιακό πυκνωτικό μικρόφωνο MEMS (Micro-Electro-Mechanical Systems) ικανό να ανιχνεύσει ακουστικά κύματα. Η κατασκευή του έχει προκύψει με τη χρήση εξειδικευμένης διαδικασία μικρο-κατεργασίας πυριτίου ειδικά για την παραγωγή αισθητήρων ήχου.
- Ανιχνευτής χειρονομίας και εγγύτητας: Το BLE Sense έχει ενσωματωμένο το τσιπ APDS-9960 που του επιτρέπει να μετρήσει τη ψηφιακή εγγύτητα και του περιβάλλοντα φωτισμοί, καθώς και την ανίχνευση RGB χρωμάτων και να ανιχνεύσει χειρονομίες.

- Βαρομετρικό αισθητήρα πίεσης: Ο αισθητήρας LPS22HB μπορεί να αντιληφθεί την βαρομετρική πίεση και επιτρέπει την έξοδο δεδομένων πίεσης 24 bit αναμεσα στα 0.26 και 1.26 bar. Τέτοιου είδους δεδομένα μπορούν να υποβληθούν σε επεξεργασία για τον υπολογισμό του ύψους πάνω από την επιφάνεια της θάλασσας της τρέχουσας τοποθεσίας.
- Αισθητήρας θερμοκρασίας και υγρασίας: Ο HTS221 είναι ένας αισθητήρας όπου μπορεί να μετρήσει τη σχετική υγρασία και θερμοκρασία. Έχει ακρίβεια θερμοκρασίας $\pm 0,5$ °C (μεταξύ 15 και 40 βαθμούς στη κλίμακα κελσίου) και ως εκ τούτου είναι απολυτά κατάλληλο για την μέτρηση θερμοκρασίας του περιβάλλοντα χώρου.

4.3 Βάση ηχητικών δεδομένων

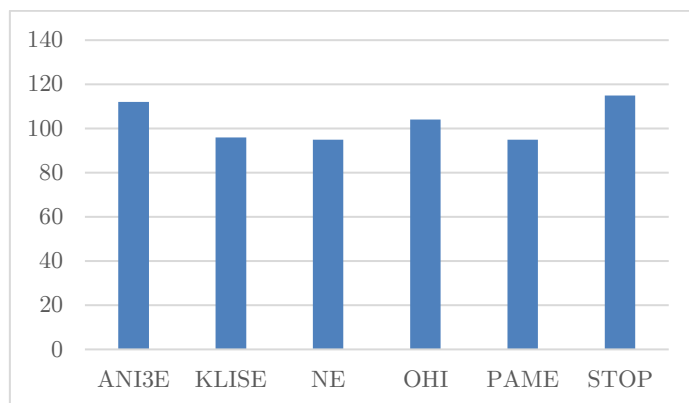
Όπως έγινε σαφές και από προηγούμενες παραγράφους της πτυχιακής για την εκπαίδευση εύρωστων μοντέλων αναγνώρισης προτύπων είναι σημαντική η ύπαρξη ποιοτικών δεδομένων. Η Ελληνική γλώσσα, ως γλώσσα που ομιλείτε από μικρό αριθμό ανθρώπων σε σχέση για παράδειγμα με την Αγγλική, δεν υπάρχουν πολλά δεδομένα διαθέσιμα για έρευνα. Για το λόγο αυτό προχωρήσαμε στην συλλογή και επεξεργασία δικών μας ηχογραφήσεων. Η συλλογή των ηχογραφήσεων έγινε με χρήση έξυπνων κινητών μέσα από την πλατφόρμα της Impulse Edge, Εικόνα 14. Οι ηχογραφήσεις που συλλέχθηκαν έγιναν από 4 άνδρες ομιλητές και 2 γυναίκες.



Εικόνα 22: Αριθμός ομιλητών

Οι κατηγορίες που επιλέχθηκαν ήταν 6 και αφορούσαν λέξεις οι οποίες προέτρεπαν στην εκτέλεση διαφόρων δραστηριοτήτων. Οι λέξεις αυτές ήταν ναι, όχι, πάμε, στοπ, άνοιξε, κλείσε. Όλες οι ηχογραφήσεις αποθηκεύτηκαν σε μορφή wav με συχνότητα δειγματοληψίας 16000Hz, 16bit και ένα κανάλι (mono). Στο παρακάτω

σχήματα απεικονίζεται η κατανομή των ηχογραφήσεων ανά κατηγορία και φύλλο ομιλητή.



Εικόνα 23: Αριθμός ηχογραφήσεων ανά κατηγορία

Όπως είδαμε και στο σχήμα παραπάνω ο συνολικός αριθμός των αρχικών ηχογραφήσεων ανέρχεται στις 617 ηχογραφήσεις από συνολικά 6 άνδρες και 2 γυναίκες. Για να γίνει χρήση μόνο των ποιοτικών ηχογραφήσεων ώστε να μην επηρεαστεί αρνητικά το σύστημα μας έγινε ξεκαθάρισμα των ηχογραφήσεων αυτών και κρατήσαμε μόνο τα αρχεία όπου η αντίστοιχη λέξη ακουγόταν σωστά και καθαρά. Σε αυτό το σημείο θα πρέπει να γίνει αναφορά πως στην αρχή έγινε η εκπαίδευση του μοντέλου και με αρχεία προερχόμενα από εργαλεία TTS (Text To Speech) καθώς και με όλες τις υπήρχε σύγχυση μεταξύ των ηχογραφήσεων, αυτό είχε σαν αποτέλεσμα να τη μη σωστή λειτουργία του μοντέλου μας διότι μπερδευε αρκετές λέξεις μεταξύ τους. Αυτό μας οδήγησε στην απόκλιση των αρχείων TTS από τη βάση δεδομένων μας και τον ξεκαθάρισμά των αρχικών ηχογραφήσεων.

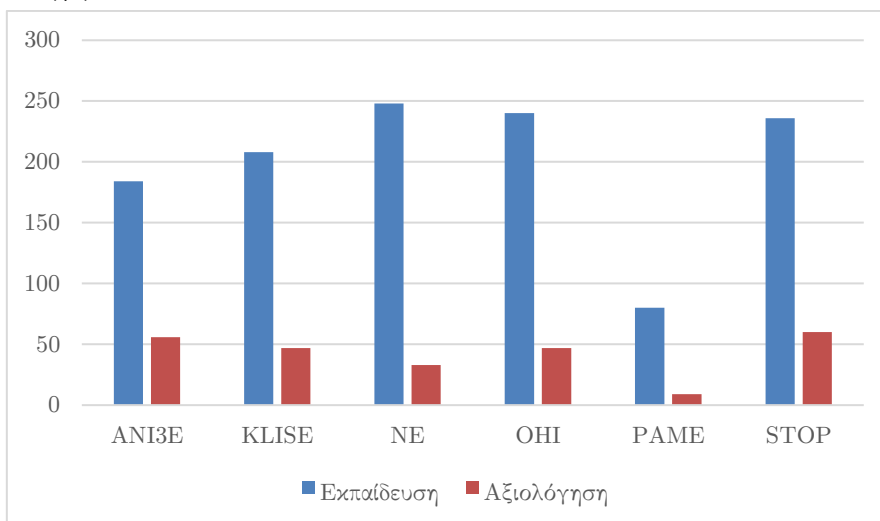
5 Πειράματα

Για την διεξαγωγή της πειραματικής διαδικασίας χρησιμοποιήθηκε η αρχιτεκτονική δικτύου (έπειτα από διεξαγωγή σειράς πειραμάτων) που απεικονίζεται στο παρακάτω σχήμα. Συγκεκριμένα κάθε διάνυσμα εισόδου αποτελείται από 800 χαρακτηριστικά και έπειτα από την εφαρμογή 2 κρυφών συνελικτικών επιπέδων, και ενός επιπέδου dropout για την αποφυγή του προβλήματος της υπερπροσαρμογής (overfit) καταλήγουμε στο επίπεδο εξόδου το οποίο απαρτίζεται από 6 νευρώνες, όσες και οι κατηγορίες προς αναγνώριση.



Εικόνα 24: Αρχιτεκτονική νευρωνικού δικτύου

Τα δεδομένα μας χωρίστηκαν σε δεδομένα εκπαίδευσης (train) και αξιολόγησης (test) όπως φαίνεται και στο παρακάτω σχήμα. Επίσης κατά την διαδικασία της εκπαίδευσης το 80% του συνόλου εκπαίδευσης χρησιμοποιήθηκε για την εκπαίδευση και το 20% χρησιμοποιήθηκε για αξιολόγηση (validation) της μάθησης σε κάθε επανάληψη.



Εικόνα 25: Αριθμός δειγμάτων στα σύνολα εκπαίδευσης και αξιολόγησης

Για τον σκοπό της αξιολόγησης της ευρωστίας έναντι του περιβαλλοντικού πρόσθετου θορύβου, στα αρχεία αξιολόγησης προστέθηκε θόρυβος ο οποίος περιείχε το ηχοτοπίο που συναντάει κανείς όταν βρίσκεται σε μια αίθουσα αναμονής π.χ. σε κάποιο σταθμό λεωφορίων δηλαδή διάφορες ομιλίες στο ηχητικό υπόβαθρο. Ο θόρυβος προστέθηκε στα δεδομένα δοκιμής με διαφορετικούς λόγους σήματος προς θόρυβο (SNR) συμπεριλαμβανομένων 0, 5, 10 και 15 dB, ενώ διατηρούνται

αυτοί των δεδομένων εκπαίδευσης. χωρίς θόρυβο. Η θορυβώδης ομιλία δημιουργήθηκε με τον ακόλουθο τρόπο: για κάθε αρχείο ομιλίας στο σώμα δοκιμής, ένα τμήμα θορύβου μήκους ίσο με το μήκος του αρχείου ομιλίας εξήχθη τυχαία, πολλαπλασιάστηκε με έναν παράγοντα κέρδους που εξαρτάται από το επιθυμητό SNR και προστέθηκε στο αρχείο ομιλίας.

5.1 Μέτρα αξιολόγησης

Καθώς ο σκοπός ενός ταξινομητή [5], είναι η όσο καλύτερη απόδοση γίνεται, εφαρμόζονται και οι αντίστοιχες μετρήσεις, όπου καθορίζουν την απόδοση αυτή, καθώς και τον υπολογισμό της πιθανότητας λάθους. Ο τρόπος με το οποίος γίνεται αυτό, είναι ότι η έξοδος του ταξινομητή είναι σε μορφή διακριτών αριθμών, με αποτέλεσμα να μπορεί να οριστεί ακριβώς αν μια πρόβλεψη έχει αποτύχει ή όχι, σε δυαδικό αριθμό. Κατά την δοκιμή ενός δείγματος, ο ταξινομητής δοκιμάζει μία-μία τις ετικέτες για να εντοπίζει αν το δείγμα ανήκει σε αυτές ή όχι. Εάν Εντοπίζει πως ανήκει, τότε το δείγμα ορίζεται σαν θετικό, ενώ, σε αντίθετη περίπτωση, το δείγμα ορίζεται σαν αρνητικό. Η πρόβλεψη αυτή, μπορεί να είναι είτε σωστή, άρα και χαρακτηρίζεται ως αληθής, είτε λανθασμένη, άρα χαρακτηρίζεται σαν ψευδής. Αυτό οδηγεί στους παρακάτω ορισμούς, οι οποίοι θα αναλυθούν στην συνέχεια

Ακρίβεια (Accuracy): Η απλούστερη και πιο διαισθητική εκτίμηση για την απόδοση του αλγορίθμου είναι η ακρίβεια (accuracy). Η τιμή της ακρίβειας προκύπτει από την διαίρεση του πλήθους των σωστών κατηγοριοποιήσεων με το πλήθος όλως των κατηγοριοποιήσεων. Στο παράδειγμα του παραπάνω πίνακα η τιμή αυτή ισούται με το άθροισμα των τιμές της κυρίας διαγώνιου δια το άθροισμα των τιμών σε όλα τα κελιά. Δηλαδή, $1030/1270 = 0.811$

Πίνακας σύγχυσης (Confusion matrix): Ένας πίνακας σύγχυσης αποτελεί μια οπτική αναπαράσταση της απόδοσης ενός αλγορίθμου κατηγοριοποίησης. Κάθε στήλη του πίνακα αντιπροσωπεύει τις ταξινομήσεις όπως προβλέφθηκαν από τον αλγόριθμο και κάθε σειρά την πραγματική τιμή των κατηγοριών. Αυτή η αναπαράσταση είναι ένας χρήσιμος τρόπος για την αξιολόγηση ενός μοντέλου ταξινόμησης. Ένα μοντέλο το οποίο έχει σωστή λειτουργία θα πρέπει να έχει έναν ισορροπημένο πίνακα σύγχυσης και θα περιγράφεται από καλές τιμές ακρίβειας (accuracy), ανάκλησης (recall), ακρίβειας (precision) και αρμονικό μέσο (F-

measure). Έστω ένας αλγόριθμος κατηγοριοποίησης ο οποίος ξεχωρίζει τις κλάσεις Positive (Pos), Negative (Neg) και Neutral (Neu) και έχει τον παρακάτω πίνακα σύγκυσης.

		Pos	Neg	Neu
Actual	Pos	15	10	100
	Neg	10	15	10
	Neu	10	100	1000

Εικόνα 26: Παράδειγμα ενός πίνακα σύγκυσης για δυαδική κατηγοριοποίηση

Η στήλη Pos λέει ότι ως Positive αναγνωρίστηκαν, 15 Positive, 10 Negative και 10 Neutral. Από την άλλη μεριά η γραμμή Pos δηλώνει ότι, επί του συνόλου 125 Positive, 15 αναγνωρίστηκαν ορθώς ως Positive, λανθασμένα 10 ως Negative και 100 ως Neutral. Στον πίνακα αυτό το σύνολο των δεδομένων κάθε κατηγορίας είναι ίσο με το άθροισμα των στοιχείων κάθε γραμμής (125 Positive, 35 Negative και 1110 Neutral).

Οι στήλες αναπαριστούν τις εκτιμήσεις που έκανε ο αλγόριθμος. Στην πρώτη στήλη, 15 περιπτώσεις κατηγοριοποιήθηκαν σωστά ως positive. 10 Negative κατηγορίες λανθασμένα κατηγοριοποιήθηκαν ως Positive και 10 Neutral περιπτώσεις λανθασμένα κατηγοριοποιήθηκαν ως Positive. Οι 20 λανθασμένα κατηγοριοποιημένες περιπτώσεις καλούνται false positives.

Διαβάζοντας από αριστερά προς τα δεξιά κατά μήκος της πρώτης γραμμής, επί του συνόλου των Positive (άθροισμα γραμμής), 15 αναγνωρίστηκαν σωστά ως positive, 10 αναγνωρίστηκαν ως negative και 100 ως Neutral. 110 Positive περιπτώσεις αναγνωρίστηκαν λάθος και καλούνται false negatives. Τέλος, οι τιμές στη κύρια διαγώνιο δείχνουν τις περιπτώσεις που ταξινομήθηκαν σωστά.

5.2 Αποτελέσματα

Στους παρακάτω πίνακες φαίνονται οι πίνακες σύγκυσης των πειραμάτων που διεξάγαμε με το μοντέλο σε διάφορες στάθμες θορύβου. Όσον αφορά τον ορισμό, του SNR ή η αναλογία σήματος προς θόρυβο είναι η αναλογία μεταξύ των επιθυμητών πληροφοριών ή της ισχύος ενός σήματος και του ανεπιθύμητου σήματος ή της ισχύος του θορύβου περιβάλλοντος. Επίσης, το SNR είναι μια παράμετρος μέτρησης που χρησιμοποιείται στους τομείς της επιστήμης και της μηχανικής που συγκρίνει το επίπεδο του επιθυμητού σήματος με το επίπεδο του θορύβου

περιβάλλοντος. Με άλλα λόγια, το SNR είναι ο λόγος της ισχύος του σήματος προς την ισχύ του θορύβου και η μονάδα έκφρασής του είναι συνήθως ντεσιμπέλ (dB). Επίσης, μια αναλογία μεγαλύτερη από 0 dB ή μεγαλύτερη από 1:1, σημαίνει περισσότερο σήμα από θόρυβο. Άρα όσο μικρότερο το SNR τόσο μεγαλύτερη η αναλογία θορύβου προς ωφέλιμο σήμα.

Το σύστημα μας μπορεί να χαρακτηριστεί ως εύρωστο μιας και φαίνεται να διατηρεί την απόδοση του άνω των 86% ακόμα και για SNR 0.

	ANI3E	KLISE	NE	OHI	PAME	STOP
ANI3E	100	0	0	0	0	0
KLISE	0	100	0	0	0	0
NE	0	0	100	0	0	0
OHI	0	0	0	100	0	0
PAME	11.1	0	0	0	88.9	0
STOP	0	0	0	3.3	0	96.7
F1 SCORE	0.99	1	1	0.98	0.94	0.98

Πίνακας 1: Αποτελέσματα κατηγοριοποίησης δεδομένων χωρίς θόρυβο (καθαρή ομιλία)

Όταν το σύστημα μας αξιολογήθηκε με καθαρό σήμα ομιλίας η συνολική ακρίβεια ήταν 98.8%. Το χαμηλότερο ποσοστό αναγνώρισης το είχαμε στη περίπτωση της λέξης κλειδί PAME για το οποίο είχε μια απόδοση της τάξης του 88.9% και όλες τις λάθος εκτιμήσεις τις έκανε μπερδεύοντας την με τη λέξη ANI3E κάτι το οποίο μπορεί να δικαιολογηθεί κατά ένα μέρος λόγω της ομοιότητας που έχουν οι λέξεις στη τελευταία τους συλλαβή.

	ANI3E	KLISE	NE	OHI	PAME	STOP
ANI3E	100	0	0	0	0	0
KLISE	0	100	0	0	0	0
NE	0	0	100	0	0	0
OHI	0	0	0	93.6	4.3	2.1
PAME	11.1	0	0	0	88.9	0
STOP	0	0	0	1.7	1.7	96.7
F1 SCORE	0.99	1	1	0.96	0.8	0.97

Πίνακας 2: Αποτελέσματα κατηγοριοποίησης δεδομένων με θόρυβο SNR 15

Όταν στο καθαρό σήμα προστέθηκε θόρυβος με τιμή SNR 15 το σύστημα μας συνέχισε να αποδίδει ικανοποιητικά. Μια ακόμα κατηγορία, συγκεκριμένα η OHI, κατηγοριοποιήθηκε λάθος ως PAME και STOP, οι οποίες είναι μάλιστα και οι κατηγορίες στις οποίες το σύστημα μας επίσης έχει κατηγοριοποιήσει λάθος.

	ANI3E	KLISE	NE	OHI	PAME	STOP
ANI3E	100	0	0	0	0	0
KLISE	0	100	0	0	0	0
NE	0	0	100	0	0	0
OHI	0	0	0	93.6	4.3	0
PAME	0	0	0	0	88.9	11.1
STOP	0	0	0	5	1.7	95
F1 SCORE	1	1	1	0.94	0.84	0.97

Πίνακας 3: Αποτελέσματα κατηγοριοποίησης δεδομένων με θόρυβο SNR 10

Σε επίπεδο θορύβου με SNR 10 το σύστημα μας αποδίδει όμοια με το SNR 10 με μια μικρή μείωση στη κατηγορία STOP.

Όταν προστέθηκε θόρυβος με SNR 5 όπως ήταν αναμενόμενο το σύστημα μας έκανε περισσότερες άστοχες εκτιμήσεις σε όλες τις κατηγορίες. Η κατηγορία με τη χαμηλότερη ακρίβεια ήταν η PAME.

	ANI3E	KLISE	NE	OHI	PAME	STOP
ANI3E	92.9	0	3.6	0	0	0
KLISE	0	97.9	2.1	0	0	0
NE	0	0	97	0	3	0
OHI	0	0	0	91.5	4.3	4.3
PAME	11.1	0	0	0	77.8	11.1
STOP	0	0	0	8.3	0	91.7
F1 SCORE	0.95	0.99	0.94	0.91	0.74	0.93

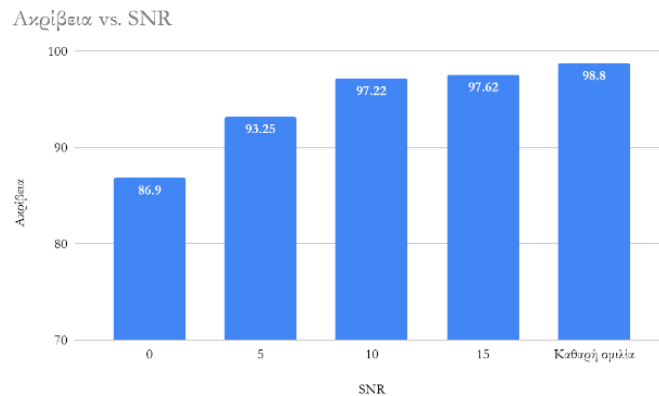
Πίνακας 4: Αποτελέσματα κατηγοριοποίησης δεδομένων με θόρυβο SNR 5

Στη περίπτωση που ο θόρυβος ήταν στάθμης με SNR 0 έχουμε σαφή μείωση της απόδοσης σε όλες τις κατηγορίες, παρόλα αυτά θα μπορούσαμε να πούμε ότι είναι ικανοποιητική η απόδοση δεδομένης του μικρού αριθμού δειγμάτων που είχαμε στα δεδομένα εκπαίδευσης.

	ANI3E	KLISE	NE	OHI	PAME	STOP
ANI3E	87.5	0	3.6	0	3.6	0
KLISE	0	93.6	6.4	0	0	0
NE	0	3	97	0	0	0
OHI	0	0	0	78.7	4.3	12.8
PAME	11.1	0	0	0	77.8	11.1
STOP	0	0	0	8.3	1.7	83.3
F1 SCORE	0.92	0.96	0.91	0.83	0.67	0.85

Πίνακας 5: Αποτελέσματα κατηγοριοποίησης δεδομένων με θόρυβο SNR 0

Στην εικόνα 24 παρατίθεται ένα διάγραμμα που απεικονίζει τη συνολική ακρίβεια αναγνώρισης για καθένα από τα πειράματα που πραγματοποιήθηκαν όπου φαίνεται ξεκάθαρα η καλή απόδοση του συστήματος μας.



Εικόνα 27: Συνολική ακρίβεια κατηγοριοποίησης ως προς τη στάθμη θορύβου

6 Επίλογος

Μετά από όλα τα πειράματα που πραγματοποιήθηκαν, γίνεται κατανοητό πως το αποτέλεσμα της ακρίβειας από όλους τους ταξινομητές, είναι αρκετά ικανοποιητικά. Όπως φαίνεται στα διαγράμματα σύγκρισης, αρκετές κατηγορίες μπλέκονται μεταξύ τους, πράγμα που οφείλεται στα όμοια μορφολογικά στοιχεία των ηχητικών κατηγοριών. Αυτό, έχει σε μεγάλο βαθμό να κάνει από την ποιότητα και ποσότητα των ηχογραφήσεων που διεξαγάγαμε.

Τα αποτελέσματά της εργασίας, αν και αρκετά ικανοποιητικά για την διαδικασία, θα μπορούσαν να ήταν ακόμα καλύτερα αν είχε δοθεί λίγο περισσότερο προσοχή στην οργάνωση του συνόλου δεδομένων κατά τα αρχικά στάδια, με πιο προσεκτικό σχεδιασμό των κατηγοριών, στις οποίες μετέπειτα θα γινόταν η ταξινόμηση.

Ενδεχομένως θα είχαμε ακόμα καλύτερα αποτελέσματα αν μελετούσαμε και άλλα ακουστικά χαρακτηριστικά πέραν αυτών που υποστηρίζονται από τη πλατφόρμα Impulse Edge όπως και αρχιτεκτονικές νευρωνικών δικτύων και υπερπαραμέτρων.

Βιβλιογραφία

1. M. Bilginer Gulmezoglu, V. Dzhafarov, M. Keskin and A. Barkana, "A novel approach to isolated word recognition," in *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 6, pp. 620-628, Nov. 1999, doi: 10.1109/89.799687.
2. S. Furui, "Speaker-independent isolated word recognition using dynamic features of speech spectrum," in *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 1, pp. 52-59, February 1986, doi: 10.1109/TASSP.1986.1164788.
3. ZDNet. 2022. *Edge Impulse wants to bring machine learning at the edge to everyone, announces \$34M Series B funding.* [online] Ανακτήθηκε από: <https://www.zdnet.com/article/edge-impulse-wants-to-bring-machine-learning-at-the-edge-to-everyone-announces-34m-series-b-funding/> [Ημερομηνία τελευταίας προσπέλασης: 29 Ιανουάριου 2022].
4. David, R., Duke, J., Jain, A., Reddi, V. J., Jeffries, N., Li, J., ... Warden, P. (2021). TensorFlow Lite Micro: Embedded Machine Learning on TinyML Systems. *arXiv [cs.LG]*. Opgehaal van <http://arxiv.org/abs/2010.08678>
5. H.Mansour, A., Zen Alabdeen Salh, G. and A. Mohammed, K., 2015. Voice Recognition using Dynamic Time Warping and Mel-Frequency Cepstral Coefficients Algorithms. *International Journal of Computer Applications*, 116(2), pp.34-41.
6. Newman, E., 1966. Speech Analysis, Synthesis and Perception. *Physics Bulletin*, 17(4), pp.127-127.
7. Li S., Yuan X. J., Overview of Intelligent Personal Assistants, HCI International 2018 – Posters' Extended Abstracts (pp.253-257)
8. Rabiner, L. and Juang, B., 1986. An introduction to hidden Markov models. *IEEE ASSP Magazine*, 3(1), pp.4-16.
9. Cristianini N., Shawe-Taylor J., 2000. *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press
10. McCulloch, W. S., Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133.
11. Mitchell, T., 2017. *Machine learning*. New York: McGraw Hill.
12. Mueller, J., & Massaron, L. (2022). dummies - Learning Made Easy. Ανακτήθηκε από: <https://www.dummies.com/article/technology/information-technology/ai/machine-learning/3-types-machine-learning-226663> [Ημερομηνία τελευταίας προσπέλασης: 26 Νοεμβρίου 2022].

13. Bagnoli, M. (2021). What is TinyML? Making Smart Devices Tinier than Ever. Ανακτήθηκε από: <https://www.pluginplaytechcenter.com/resources/tinyml-making-smart-devices-tinier-ever/> [Ημερομηνία τελευταίας προσπέλασης : 12 Νοεμβρίου 2021]
14. An Introduction to TinyML. (2020). Ανακτήθηκε από: <https://towardsdatascience.com/an-introduction-to-tinyml-4617f314aa79> [Ημερομηνία τελευταίας προσπέλασης: 4 Δεκέμβριου 2021]
15. Edge Impulse Docs. 2022. *Spectrogram*. [online] Ανακτήθηκε από: <https://docs.edgeimpulse.com/docs/spectrogram> [Ημερομηνία τελευταίας προσπέλασης: 13 Δεκέμβριου 2021].
16. Edge Impulse Docs. 2022. *Audio MFE* . [online] Ανακτήθηκε από: <https://docs.edgeimpulse.com/docs/audio-mfe> [Ημερομηνία τελευταίας προσπέλασης: 13 Δεκέμβριου 2021].
17. Edge Impulse Docs. 2022. *Audio MFCC*. [online] Ανακτήθηκε από: <https://docs.edgeimpulse.com/docs/audio-mfcc> [Ημερομηνία τελευταίας προσπέλασης: 13 Δεκέμβριου 2021].
18. He, X., Zhao, K., & Chu, X. (2021). AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems*, 212, 106622.