



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ – ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Πρόγραμμα Μεταπτυχιακών Σπουδών
«Προηγμένα Συστήματα Πληροφορικής - Ανάπτυξη Λογισμικού και Τεχνητής Νοημοσύνης»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής:	Ανίχνευση ανωμαλιών με μη επιτηρούμενη μάθηση σε ήχους λειτουργίας μηχανημάτων Unsupervised Anomaly Detection in Mechanical Sounds
Όνοματεπώνυμο φοιτητή:	Χρήστος Βεζαλής
Πατρώνυμο:	Αχιλλέας
Αριθμός Μητρώου:	ΜΠΣΠ20006
Επιβλέπων:	Άγγελος Πικράκης, Επίκουρος Καθηγητής

Ημερομηνία Παράδοσης **Οκτώβριος 2022**

Τριμελής Εξεταστική Επιτροπή

Άγγελος Πικράκης
Επίκουρος Καθηγητής

Δημήτριος Αποστόλου
Καθηγητής

Θεμιστοκλής Παναγιωτόπουλος
Καθηγητής

Περίληψη

Τα μηχανήματα παραγωγής έχουν πολλές φορές εξαρτήματα όπως βαλβίδες, αντλίες, ανεμιστήρες και ράγες κύλισης. Όταν υπάρχει κάποια βλάβη στα εξαρτήματα, ο ήχος που παράγουν είναι διαφορετικός από τους ήχους της ομαλής λειτουργίας. Η παρούσα εργασία εξετάζει την ανίχνευση μη ομαλής λειτουργίας στους ήχους που παράγουν τα μηχανήματα με την χρήση μη επιτηρούμενων αλγορίθμων μηχανικής μάθησης.

Abstract

Production machines often have parts like valves, pumps, fans and sliders. When there is a malfunction the sound produced is different from the normal operation. Current thesis examines anomaly detection using sounds from production machines using unsupervised machine learning algorithms.

Περιεχόμενα

1	Εισαγωγή.....	7
1.1	Στόχοι της Εργασίας.....	7
1.2	Τεχνητή νοημοσύνη.....	7
1.3	Μηχανική μάθηση.....	8
1.3.1	Επιτηρούμενη και μη επιτηρούμενη μάθηση.....	9
1.3.2	Μέτρο απόδοσης.....	9
1.3.3	Υπερεκπαίδευση και Υποεκπαίδευση.....	9
1.4	Ελαχιστοποίηση με βάση την κλίση (gradient descent).....	10
1.5	Ανίχνευση ανωμαλιών.....	11
1.5.1	Έξοδος της ανίχνευσης ανωμαλιών.....	12
1.6	Ηχητικά σήματα.....	12
1.7	Εξαγωγή χαρακτηριστικών - Φασματογράφημα.....	13
1.7.1	Διακριτός μετασχηματισμός Fourier.....	13
1.7.2	Βραχέως χρόνου μετασχηματισμός Fourier (Short time Fourier transform STFT).....	14
1.7.3	mel φασματογράφημα.....	14
1.7.4	Συντελεστές MFCC.....	15
1.7.5	Τροχιές φασματικών κορυφών (Spectral peak trajectories).....	16
2	Αλγόριθμοι μηχανικής μάθησης.....	18
2.1	Βαθιά νευρωνικά δίκτυα πρόσθιας τροφοδότησης (deep feed forward networks).....	18
2.2	Οπισθοδιάδοση του σφάλματος (backpropagation).....	19
2.3	Συνελκτικά νευρωνικά δίκτυα.....	21
3	Αυτόματοι κωδικοποιητές.....	23
3.1.1	Μη επιτηρούμενη ανίχνευση ανώμαλων ήχων.....	24
3.1.2	Καμπύλη Receiver operating characteristic (ROC).....	25
3.1.3	Area under the curve (AUC).....	25
4	Εφαρμογές.....	27
4.1	Σύνολα δεδομένων.....	27
4.2	Πειράματα.....	29
4.3	Αποτελέσματα.....	33
5	Συμπεράσματα.....	37
5.1	Επεκτάσεις.....	37

Πίνακας εικόνων

Εικόνα 1: Ηχητικό σήμα μετατροπή σε mel φασματογράφημα.....	15
Εικόνα 2: Αυτόματος αποκωδικοποιητής.....	23
Εικόνα 3: Deep autoencoder.....	24
Εικόνα 4: Καμπύλη ROC (εικόνα από wikipedia).....	26
Εικόνα 5: Ομαλό και μη ομαλό φασματογράφημα από ανεμιστήρα.....	28
Εικόνα 6: Ομαλό και μη ομαλό φασματογράφημα από αντλία.....	28
Εικόνα 7: Ομαλό και μη ομαλό φασματογράφημα από ράγες κύλισης.....	29
Εικόνα 8: Ομαλό και μη ομαλό φασματογράφημα από βαλβίδες.....	29
Εικόνα 9: Μοντέλο - βαθύς αυτόματος κωδικοποιητής.....	30
Εικόνα 10: Μοντέλο - Συνελικτικός αυτόματος κωδικοποιητής.....	31
Εικόνα 11: Ιστόγραμμα με τα σκορ ανακατασκευής για το μηχάνημα 0 από τις ράγες κύλισης...35	
Εικόνα 12: Confusion matrix για μικρό κατώφλι στο σκορ ανακατασκευής.....	36
Εικόνα 13: Confusion matrix για μεγάλο κατώφλι στο σκορ ανακατασκευής.....	36

Πίνακες

Πίνακας 1: Αποτελέσματα από την εκτέλεση με τον αυτόματο κωδικοποιητή με U σχήμα.....	33
Πίνακας 2: Αποτελέσματα από τον συνελικτικό αυτόματο κωδικοποιητή.....	33
Πίνακας 3: Αποτελέσματα από τη χρήση των mfcc συντελεστών με τον U αυτόματο κωδικοποιητή.....	33
Πίνακας 4: Αποτελέσματα από τη χρήση των mfcc συντελεστών μαζί με το 4p-MSD με τον U αυτόματο κωδικοποιητή.....	34
Πίνακας 5: Συγκριτικά αποτελέσματα με το baseline σύστημα.....	35

1 Εισαγωγή

1.1 Στόχοι της Εργασίας

Τα μηχανήματα παραγωγής έχουν πολλές φορές εξαρτήματα όπως βαλβίδες, αντλίες, ανεμιστήρες και ράγες κύλισης που παράγουν ήχους κατά την λειτουργία τους. Όταν υπάρχει κάποια βλάβη στα εξαρτήματα, ο ήχος που παράγουν είναι διαφορετικός από τους ήχους της ομαλής τους λειτουργίας. Για τη διάγνωση των βλαβών απαιτείται προσωπικό που να παρακολουθεί τα παραγωγικά μηχανήματα κατά τη διάρκεια της λειτουργίας τους.

Η παρούσα εργασία εξετάζει την παρακολούθηση των ήχων που παράγουν τα εξαρτήματα των μηχανημάτων από συστήματα τεχνητής νοημοσύνης με τη χρήση αλγόριθμων μηχανικής μάθησης για την αυτόματη ανίχνευση των βλαβών και την έγκαιρη ειδοποίηση του προσωπικού. Οι ήχοι των εξαρτημάτων τους έχουν καταγραφεί από μικρόφωνα. Λόγο του ότι οι ήχοι των μηχανημάτων την ώρα που έχουν βλάβη είναι διαφορετικοί κάθε φορά, ανάλογα και με το είδος της βλάβης, αλλά και επειδή οι βλάβες και δύσκολο να ηχογραφηθούν και να συλλέγουν γιατί είναι σπάνιες, οι αλγόριθμοι θα εκπαιδευτούν μόνο με ήχους ομαλής λειτουργίας με τη χρήση μη επιτηρούμενων αλγορίθμων μηχανικής μάθησης (unsupervised machine learning algorithms).

1.2 Τεχνητή νοημοσύνη

Πράκτορας (agent) ορίζεται οτιδήποτε μπορεί να αντιληφθεί το περιβάλλον που το περιβάλλει με τη χρήση αισθητήρων και μπορεί να επηρεάσει αυτό το περιβάλλον μέσω ενεργειών. Ένας ανθρώπινος πράκτορας έχει μάτια, αυτιά και άλλα όργανα για αισθητήρες και χέρια, πόδια, φωνητικές χορδές κ.α. για την πραγματοποίηση ενεργειών. Ένας πράκτορας λογισμικού δέχεται σαν ερεθίσματα αρχεία, δικτυακά πακέτα και ανθρώπινη αλληλεπίδραση (πληκτρολόγιο, ποντίκι, οθόνη αφής, ήχο κ.α.) και ενεργεί στο περιβάλλον γράφοντας σε αρχεία, στέλνοντας δικτυακά πακέτα, δείχνοντας πληροφορίες σε μια οθόνη ή παράγοντας ήχους.

Ο ρόλος της Τεχνητής Νοημοσύνης (T.N.) είναι η σχεδίαση ενός προγράμματος για τον πράκτορα (agent program) που υλοποιεί την συνάρτηση του πράκτορα (agent function) - την αντιστοίχιση δηλαδή των ερεθισμάτων (percepts) σε πράξεις (actions). Υποθέτουμε ότι το πρόγραμμα θα τρέξει σε κάποιου είδους υπολογιστική συσκευή (computing device) με φυσικούς αισθητήρες (physical sensors) αλλά και ενεργοποιητές (actuators) - που ονομάζεται αρχιτεκτονική του πράκτορα [1].

Στην παρούσα εργασία θα ασχοληθούμε με την σχεδίαση ενός προγράμματος, που έχει σαν στόχο την αντιστοίχιση αρχείων ήχου μηχανημάτων παραγωγής, που δέχεται ο πράκτορας από

μικρόφωνα που παρακολουθούν τα μηχανήματα σαν ερεθίσματα από το περιβάλλον, σε ειδοποιήσεις σε οθόνες, μηνύματα (emails, sms κ.α.) ή ήχους στην περίπτωση που ανιχνευθεί ανωμαλία-βλάβη στην λειτουργία. Οι ειδοποιήσεις αυτές είναι και οι ενέργειες που πραγματοποιεί ο πράκτορας στο περιβάλλον. Πιο συγκεκριμένα το πρόγραμμα θα δέχεται σαν είσοδο ένα ηχητικό αρχείο και θα επιστρέφει σαν έξοδο αν το μηχανήμα που παράγει τον ήχο είναι σε ομαλή ή όχι λειτουργία.

1.3 Μηχανική μάθηση

Ο πράκτορας μαθαίνει (learn), αν βελτιώνει την απόδοσή του αφότου πραγματοποιεί παρατηρήσεις για τον κόσμο που τον περιβάλλει. Όταν ο πράκτορας είναι ηλεκτρονικός υπολογιστής η διαδικασία αυτή ονομάζεται Μηχανική μάθηση (machine learning): Ο υπολογιστής παρατηρεί δεδομένα, φτιάχνει ένα μοντέλο βασισμένο στα δεδομένα και χρησιμοποιεί το μοντέλο αυτό σαν υπόθεση για την κατάσταση που βρίσκεται ο κόσμος που τον περιβάλλει και ως λογισμικό που λύνει προβλήματα [1]

Παράδειγμα ορίζεται μια συλλογή από χαρακτηριστικά (features) που έχουν ποσοτικοποιηθεί από κάποιο αντικείμενο ή γεγονός που θέλουμε το σύστημα μηχανικής μάθησης να επεξεργαστεί. Συνήθως χρησιμοποιούμε ένα διάνυσμα (vector) $x \in \mathbb{R}^n$ για να αναπαραστήσουμε ένα παράδειγμα όπου κάθε στοιχείο x_i του διανύσματος είναι άλλο χαρακτηριστικό [2].

Ορίζοντας το σύνολο των παραδειγμάτων ή γεγονότων ως τις εμπειρίες E (experiences) από το περιβάλλον, την εργασία T (task) που έχει να εκτελέσει ο αλγόριθμος και το μέτρο απόδοσης P (performance) του αλγόριθμου στην εργασία T , μπορούμε να πούμε ότι ο αλγόριθμος μαθαίνει όταν η επίδοση του στην T όπως μετριέται από το P μεγαλώνει όταν ο αλγόριθμος επεξεργάζεται παραδείγματα από το E . [3].

Το πιο χαρακτηριστικό παράδειγμα εργασίας είναι η ταξινόμηση (classification). Σε αυτήν την εργασία ο αλγόριθμος μηχανικής μάθησης ρωτάτε για το σε ποια κατηγορία ανήκει το κάθε παράδειγμα. Η λύση αυτού του προβλήματος είναι μια συνάρτηση που αντιστοιχεί το διάνυσμα χαρακτηριστικών του κάθε παραδείγματος σε μια τιμή που αντιπροσωπεύει την κατηγορία. Άλλο παράδειγμα έργου είναι η παλινδρόμηση (regression) όπου στόχος είναι η πρόβλεψη μιας τιμής μετά την επεξεργασία των παραδειγμάτων.

Η εργασία μηχανικής μάθησης που θα ασχοληθούμε είναι η ανίχνευση ανωμαλιών (anomaly detection). Σε αυτή την εργασία ο αλγόριθμος μετά την επεξεργασία ενός αριθμού γεγονότων-παραδειγμάτων θα δώσει σε κάποια από αυτά σήμανση ως μη αναμενόμενα ή ανώμαλα και στα υπόλοιπα ως ομαλά. Πιο συγκεκριμένα σε κάθε αλληλουχία ήχων που θα δώσουμε σαν είσοδο

στον αλγόριθμο με την μορφή διανύσματος θέλουμε να έχουμε στην έξοδο μια ετικέτα για το αν είναι ήχος από ομαλή ή προβληματική λειτουργία του μηχανήματος που τον παράγει.

1.3.1 Επιτηρούμενη και μη επιτηρούμενη μάθηση

Ανάλογα με τον τύπο των παραδειγμάτων-γεγονότων η μάθηση χωρίζεται σε 2 βασικές κατηγορίες. Στην μη επιτηρούμενη μάθηση (unsupervised learning) το σύνολο δεδομένων - παραδειγμάτων συνήθως περιέχει πολλά χαρακτηριστικά και στόχος είναι η εκμάθηση χρήσιμων ιδιοτήτων από τη δομή του συνόλου δεδομένων ή/και την κατανομή τους. Στην επιτηρούμενη μάθηση (supervised learning) τα παραδείγματα συνδέονται και με μια ετικέτα ή τιμή και στόχος είναι η πρόγνωση της ετικέτας σε δεδομένα που δεν έχουν ή η πρόγνωση μιας τιμής [2].

1.3.2 Μέτρο απόδοσης

Για να αξιολογήσουμε τις δυνατότητες ενός αλγορίθμου Μηχανικής μάθησης σχεδιάζουμε ένα μέτρο (metric) που είναι ικανό να ποσοτικοποιήσει την επίδοση του. Για εργασίες π.χ. όπως η ταξινόμηση συχνά υπολογίζουμε την ακρίβεια (accuracy) δηλαδή το ποσοστό των παραδειγμάτων όπου το μοντέλο υπολογίζει την σωστή ετικέτα. Αντίστοιχο μέτρο είναι το ποσοστό λάθους (error rate), το οποίο για την ταξινόμηση είναι το ποσοστό των παραδειγμάτων με λάθος έξοδο από τον αλγόριθμο. Οι δύο αυτοί τρόποι ονομάζονται και 0-1 λάθος (0-1 loss) ορίζοντας ως 0 το μέτρο όταν έχουμε σωστή ετικέτα και 1 όταν δεν έχουμε.

Σε άλλες εργασίες όπως η παλινδρόμηση δεν έχει νόημα το 0-1 λάθος και εκεί πρέπει να υπολογίσουμε μια συνεχή τιμή που να ποσοτικοποιεί κατάλληλα την απόσταση που έχει η έξοδος από την αληθινή τιμή (ground truth) που έχει η ετικέτα του παραδείγματος.

Επειδή ενδιαφερόμαστε για το πόσο καλά τα πάει ο αλγόριθμος σε δεδομένα που δεν τα έχει ξαναδεί υπολογίζουμε την απόδοσή του σε ένα σύνολο δεδομένων διαφορετικό από το σύνολο δεδομένων της εκπαίδευσης που το ονομάζουμε σύνολο δοκιμής (test set) [2].

1.3.3 Υπερεκπαίδευση και Υποεκπαίδευση

Σημαντικό πρόβλημα στην Μηχανική μάθηση είναι να μπορεί ο αλγόριθμος να έχει καλή απόδοση σε παραδείγματα που δεν έχει ξαναδεί, όχι μόνο σε αυτά που έχει εκπαιδευτεί. Η ικανότητα του να έχει καλή απόδοση σε άγνωστα δεδομένα ονομάζεται γενίκευση (Generalization) [2].

Όταν εκπαιδεύουμε ένα μοντέλο μηχανικής μάθησης έχουμε πρόσβαση στο σύνολο εκπαίδευσης και σε αυτό το σύνολο υπολογίζουμε ένα μετρικό λάθους και προσπαθούμε να μειώσουμε αυτό το μετρικό. Ως εδώ η διαδικασία είναι σαν μια βελτιστοποίηση. Αυτό όμως που

διαχωρίζει την μηχανική μάθηση από τη βελτιστοποίηση είναι ότι θέλουμε ο αλγόριθμος ταυτόχρονα να έχει όσο μικρότερο γίνεται και το αντίστοιχο μετρικό λάθους στο σύνολο δοκιμής, δηλαδή σε δεδομένα που δεν τα έχει ξαναδεί ο αλγόριθμος το οποίο ονομάζεται και λάθος γενίκευσης (generalization error).

Όταν το λάθος στο σύνολο εκπαίδευσης είναι πολύ μεγάλο και όχι αποδεκτό λέμε ότι το μοντέλο έχει υποεκπαιδευτεί. Όταν το μοντέλο τα πάει καλά στο σύνολο εκπαίδευσης αλλά υπάρχει μεγάλη διαφορά με το λάθος στο σύνολο δοκιμής λέμε ότι το μοντέλο έχει υπερεκπαιδευτεί. Αλλάζοντας την χωρητικότητα (capacity) - αριθμός των παραμέτρων που μοντέλου - μπορούμε να επηρεάσουμε τον βαθμό της υποεκπαίδευσης ή της υπερεκπαίδευσης. Μοντέλα με μικρή χωρητικότητα δυσκολεύονται να προσαρμοστούν (fit) στα δεδομένα και μοντέλα με πολύ μεγάλη χωρητικότητα υπερεκπαιδεύονται απομνημονεύοντας τα δεδομένα από το σύνολο εκπαίδευσης με αποτέλεσμα να έχουν πολύ κακή απόδοση στο σύνολο δοκιμής και μεγάλο λάθος γενίκευσης.

Το μοντέλο της μηχανικής μάθησης έχει και παραμέτρους που δεν γίνεται να μάθει ο αλγόριθμος. Αυτές ονομάζονται υπερπαραμέτροι (hyperparameters). Για τη βελτιστοποίηση αυτών των παραμέτρων που μπορεί να είναι π.χ. η χωρητικότητα του μοντέλου δεν μπορεί να χρησιμοποιηθεί το σύνολο δεδομένων για την εκπαίδευση γιατί θα επιλεγεί π.χ. για την περίπτωση της χωρητικότητας η μεγαλύτερη δυνατή και θα έχουμε ως αποτέλεσμα πάλι υπερεκπαίδευση του μοντέλου και μεγάλο σφάλμα γενίκευσης. Για τον λόγο αυτό ένα τμήμα από το σύνολο δεδομένων που ονομάζεται σύνολο επικύρωσης (validation set) χρησιμοποιείται για τη βελτιστοποίηση μόνο των υπερπαραμέτρων.

1.4 Ελαχιστοποίηση με βάση την κλίση (gradient descent)

Η βελτιστοποίηση είναι η διαδικασία της ελαχιστοποίησης ή της μεγιστοποίησης κάποιας συνάρτησης $f(x)$ αλλάζοντας το x . Συνήθως αντιμετωπίζουμε τα περισσότερα προβλήματα σαν προβλήματα ελαχιστοποίησης. Η μεγιστοποίηση μπορεί να γίνει με την ελαχιστοποίηση της $-f(x)$. Η συνάρτηση που θέλουμε να ελαχιστοποιήσουμε ή να μεγιστοποιήσουμε ονομάζεται αντικειμενική συνάρτηση (objective function). Όταν ελαχιστοποιούμε μπορούμε να την πούμε και συνάρτηση λάθους (error/loss function) ή συνάρτηση κόστους (cost function).

Αν υποθέσουμε ότι έχουμε μια συνάρτηση $y = f(x)$, όπου το x και y είναι πραγματικοί αριθμοί, η παράγωγος της συνάρτησης ορίζεται σαν $f'(x)$ ή $\frac{dx}{dy}$. Η παράγωγος $f'(x)$ δίνει την κλίση της $f(x)$ στο σημείο x . Ορίζει δηλαδή το πόσο πρέπει να αλλάξουμε την είσοδο ώστε να έχουμε την επιθυμητή τιμή στην έξοδο:

$$f(x + \epsilon) \approx f(x) + \epsilon f'(x) \quad (1.1)$$

Η παράγωγος άρα μπορεί να χρησιμοποιηθεί ώστε αλλάζοντας το x να έχουμε μικρή βελτίωση στο y . Άρα μπορούμε να μειώσουμε την $f(x)$ κάνοντας μικρά βήματα στην αντίθετη τιμή της παραγώγου. Η μέθοδος αυτή ονομάζεται βελτιστοποίηση με βάση την κλίση (gradient descent) [4].

Για συναρτήσεις με πολλαπλές εισόδους χρησιμοποιούμε τις μερικές παραγώγους (partial derivatives). Η μερική παράγωγος $\frac{\partial}{\partial x_i} f(x)$ μετράει πως αλλάζει η συνάρτηση f όταν αλλάζει μόνο η παράμετρος x_i στο σημείο x . Η κλίση ή διανυσματική παράγωγος (gradient) γενικεύει την έννοια της παραγώγου στην περίπτωση που η παράγωγος είναι με βάση ένα διάνυσμα. Η κλίση της συνάρτησης f είναι το διάνυσμα που περιέχει όλες τις μερικές παραγώγους και συμβολίζεται $\nabla_x f(x)$. Το κάθε i της κλίσης είναι η μερική παράγωγος της f ως προς το x_i .

Για την περίπτωση των πολλαπλών εισόδων ένα νέο σημείο της συνάρτησης είναι το:

$$x' = x - \epsilon \nabla_x f(x) \quad (1.2)$$

όπου ϵ είναι το μήκος του βήματος για την ανανέωση που ονομάζεται και ρυθμός μάθησης (learning rate) [2].

1.5 Ανίχνευση ανωμαλιών

Η ανίχνευση ανωμαλιών (anomaly detection) αναφέρεται στο πρόβλημα της εύρεσης μοτίβων (patterns) στα δεδομένα που δεν συνάδουν με την αναμενόμενη συμπεριφορά. Οι ανωμαλίες είναι μοτίβα που δεν συμμορφώνονται με μια καλά καθορισμένη έννοια για το τι είναι ομαλά μοτίβα στα δεδομένα. Μια πρώτη αντιμετώπιση του προβλήματος είναι το να ορίσουμε μια περιοχή των δεδομένων ως ομαλή συμπεριφορά και όλα τα υπόλοιπα σημεία που δεν ανήκουν σε αυτή την περιοχή ορίζονται σαν ανώμαλα [5]. Η αντιμετώπιση αυτή όμως έχει προβλήματα όπως:

- Το να οριστεί μια περιοχή των δεδομένων ως ομαλή είναι πολύ δύσκολο και η διαχωριστική γραμμή πολλές φορές δεν μπορεί να είναι πάντα ακριβής.
- Σε διαφορετικά πεδία (domains) είναι πολλές φορές διαφορετικό το τι μπορούμε να ορίσουμε ως ανωμαλία. Έτσι μια τεχνική που έχει αναπτυχθεί σε ένα πεδίο δεν μπορεί να εφαρμοστεί σε ένα άλλο διαφορετικό.
- Η διαθεσιμότητα δεδομένων με ετικέτες για το τι είναι ομαλό και τι όχι για τα σύνολα εκπαίδευσης και δοκιμής που θα χρησιμοποιηθούν για την κατασκευή μοντέλων ανίχνευσης ανωμαλιών είναι συχνά σημαντικό πρόβλημα.

- Συχνά τα δεδομένα περιέχουν θόρυβο που μπορεί να μοιάζει με μοτίβα ανωμαλιών στα δεδομένα τα οποία είναι δύσκολο να ανιχνευτούν και να αφαιρεθούν.

Τα ανώμαλα δεδομένα μπορούν να χωριστούν σε 3 βασικές κατηγορίες [5]:

1. Σημειακές ανωμαλίες (point anomalies). Αν ένα σημείο στα δεδομένα (data point) μπορεί να οριστεί σαν ανώμαλο σε σχέση με τα υπόλοιπα δεδομένα τότε μπορούμε να πούμε ότι έχουμε μια σημειακή ανωμαλία.
2. Υπό συνθήκη ανωμαλίες (contextual anomalies). Αν ένα σημείο στα δεδομένα είναι ανώμαλο όταν βρίσκεται σε ένα συγκεκριμένο πλαίσιο (context) αλλά όχι έξω από αυτό. Σε αυτή την περίπτωση ένα σημείο των δεδομένων περιγράφεται με 2 κατηγορίες χαρακτηριστικών:
 1. Χαρακτηριστικά πλαισίου (contextual attributes). Τα χαρακτηριστικά αυτά περιγράφουν το πλαίσιο μέσα στο οποίο θα εξετάσουμε το δεδομένο.
 2. Χαρακτηριστικά συμπεριφοράς (behavioral attributes). Περιγράφουν χαρακτηριστικά που δεν αφορούν το πλαίσιο που εξετάζουμε αλλά τα ίδια τα δεδομένα και είναι οι τιμές τους αυτές που καθορίζουν αν θα είναι ένα δεδομένο ομαλό ή όχι.
3. Συλλογικές ανωμαλίες (Collective Anomalies). Αν ένα σύνολο από δεδομένα είναι ανώμαλο σε σχέση με όλα τα υπόλοιπα δεδομένα. Υπάρχουν μόνο σε σύνολο δεδομένων που τα δεδομένα είναι συσχετισμένα.

1.5.1 Έξοδος της ανίχνευσης ανωμαλιών

Ένα σημαντικό θέμα της ανίχνευσης ανωμαλιών είναι ο τρόπος που τους εξάγει το σύστημα. Αυτό μπορεί να γίνει με δύο τρόπους [5]:

1. Σκορ. Με τη χρήση τεχνικών που δίνουν ένα σκορ ανωμαλίας σε κάθε σημείο των δεδομένων ανάλογο με τον βαθμό που πιστεύει το μοντέλο ότι πρόκειται για ανωμαλία.
2. Ετικέτες. Με αυτή την τεχνική το μοντέλο δίνει μια ετικέτα στο κάθε σημείο (ανώμαλο ή ομαλό).

Με την τεχνική του σκορ μπορούμε να θέσουμε ένα κατώφλι πάνω από το οποίο όλα τα σημεία είναι ανώμαλα.

1.6 Ηχητικά σήματα

Η ταλάντωση-δόνηση των εξαρτημάτων στα μηχανημάτων δημιουργεί στα μόρια του αέρα μια διαταραχή-πίεση. Η διαταραχή αυτή μεταφέρεται στον αέρα με σφαιρικά κύματα. Τα σφαιρικά

αυτά κύματα με την σειρά τους όταν φτάνουν στο μέσο της ηχογράφησης δημιουργούν μια μεταβαλλόμενη πίεση, σε κάποια π.χ. ευαίσθητη μεμβράνη ενός μικροφώνου, την ένταση της οποίας καταγράφει σε τακτά χρονικά διαστήματα η συσκευή και μας την παρέχει στην έξοδο της. Η ακολουθία αυτών των τιμών μας δίνουν το ψηφιακό ηχητικό σήμα.

Ένα ψηφιακό ηχητικό σήμα $x(n), n = 0, \dots, N - 1$ μπορεί να αναπαρασταθεί ως ένα διάνυσμα πραγματικών αριθμών N δειγμάτων. Τα σήματα αυτά είναι δείγματα του ήχου ανά τακτά πολύ μικρά χρονικά διαστήματα. Η διαδικασία αυτή ονομάζεται δειγματοληψία. Η διαφορά μεταξύ 2 συνεχόμενων χρονικών στιγμών ονομάζεται περίοδος της δειγματοληψίας T_s (sampling period). Το αντίστροφο της περιόδου δειγματοληψίας είναι η συχνότητα δειγματοληψίας $F_s = \frac{1}{T_s} Hz$. Η συχνότητα μετριέται σε Hz και δείχνει τον αριθμό των δειγμάτων που έχουν ληφθεί σε ένα δευτερόλεπτο. Για να είναι επιτυχής μια δειγματοληψία σε ένα συνεχές σήμα η δειγματοληψία πρέπει να έχει συχνότητα τουλάχιστον διπλή από την μέγιστη συχνότητα του σήματος. Αυτό το κάτω όριο στην συχνότητα δειγματοληψίας λέγεται ρυθμός Nyquist και εξασφαλίζει ότι δεν θα έχουμε αλλοίωση στο ηχητικό σήμα [6].

Η ανάλυση του ήχου πολλές φορές δεν γίνεται σε όλο το σήμα αλλά διαμερίζουμε το σήμα σε σύντομου χρόνου τμήματα που μπορεί να είναι και επικαλυπτόμενα (overlapping short-term windows - frames). Η διαδικασία αυτή ονομάζεται ανάλυση ήχου βραχέως χρόνου (short term audio processing).

1.7 Εξαγωγή χαρακτηριστικών - Φασματογράφημα

Την εξαγωγή χαρακτηριστικών μπορούμε να τη δούμε και από την οπτική γωνία της μείωσης του ρυθμού παραγωγής των δεδομένων (data rate) ώστε οι αλγόριθμοι να βασιστούν σε ένα αριθμό χαρακτηριστικών που είναι τα πιο αντιπροσωπευτικά για το ηχητικό σήμα αλλά και να μειώνουν τον όγκο των δεδομένων.

Χρησιμοποιώντας την τεχνική της ανάλυσης βραχέως χρόνου χωρίζουμε το σήμα σε επικαλυπτόμενα τμήματα που από το καθένα θα φτιάξουμε ένα διάνυσμα χαρακτηριστικών.

1.7.1 Διακριτός μετασχηματισμός Fourier

Ο διακριτός μετασχηματισμός Fourier (Discrete Fourier Transform-DFT) χρησιμοποιείται για την απεικόνιση του σήματος από το πεδίο του χρόνου στο πεδίο της συχνότητας (frequency domain) ή φασματική (spectral) απεικόνιση του σήματος. Έχοντας ένα σήμα διακριτού χρόνου $x(n), n = 0, \dots, N - 1, N$ δειγμάτων ο DFT ορίζεται ως:

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp\left(-j \frac{2\pi}{N} kn\right), k = 0, \dots, N-1 \quad (1.3)$$

όπου $j = \sqrt{-1}$. Ο μετασχηματισμός αυτός είναι μια ακολουθία από N όρους, τα $X(k)$ s, όπου γενικά είναι μιγαδικοί αριθμοί. Οι όροι αυτοί είναι η αναπαράσταση του σήματος στο πεδίο των συχνοτήτων.

Ο αντίστροφος διακριτός μετασχηματισμός Fourier έχει σαν είσοδο την ακολουθία των όρων από τον DFT και επιστρέφει το αρχικό σήμα:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) \exp\left(j \frac{2\pi}{N} kn\right), n = 0, \dots, N-1 \quad (1.4)$$

Αν η παράσταση (1.4) γραφεί ως:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) \gamma_k(n), n = 0, \dots, N-1 \quad (1.5)$$

όπου $\gamma_k(n) = \exp\left(j \frac{2\pi}{N} kn\right)$, $n = 0, \dots, N-1$, τότε μπορούμε να δούμε ότι το αρχικό σήμα, $x(n)$, μπορεί να γραφεί ως ένας σταθμισμένος μέσος όρος από βασικά σήματα όπου το κάθε βασικό σήμα, $\gamma_k(n)$ είναι ένας μιγαδικός εκθετικός και το βάρος του είναι ίσο με τον k th DFT όρο [6]. Το μέγεθος του k th DFT όρου, $|X(k)|$ μπορεί να είναι και ένα μέτρο της έντασης με την οποία συμμετέχει η αντίστοιχη συχνότητα στο αρχικό σήμα $x(n)$. Το τελευταίο μας οδηγεί στην φασματική αναπαράσταση του σήματος.

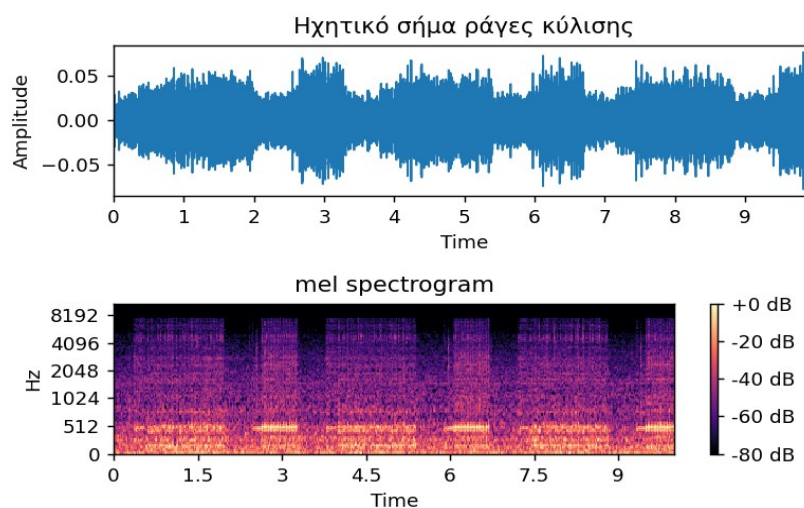
1.7.2 Βραχέως χρόνου μετασχηματισμός Fourier (Short time Fourier transform STFT)

Ο στόχος του βραχέως χρόνου μετασχηματισμού Fourier είναι να σπάσει το σήμα σε τμήματα που μπορεί να είναι και επικαλυπτόμενα με τη χρήση ενός κυλιόμενου παραθύρου και τον υπολογισμό του DFT σε κάθε τμήμα. Μεγαλύτερα παράθυρα οδηγούν σε πιο λεπτομερή απεικόνιση των συχνοτήτων αλλά χειρότερη απεικόνιση στο πεδίο του χρόνου και αντίστροφα. Στις εφαρμογές ανάλυσης ηχητικών σημάτων το παράθυρο έχει συνήθως μήκος από 10 μέχρι και 50ms.

1.7.3 μετ φασματογράφημα

Αν τοποθετήσουμε τους συντελεστές από το κάθε τμήμα σε διαφορετικές στήλες σε ένα πίνακα ο STFT μπορεί να αναπαρασταθεί από ένα πίνακα όρων όπου ο δείκτης της κάθε στήλης

αντιπροσωπεύει χρόνο και το δείκτης της κάθε γραμμής αντιπροσωπεύει την συχνότητα του κάθε DFT συντελεστή. Αν για τον κάθε όρο υπολογιστεί το μέγεθος (μέτρο του μιγαδικού) ο πίνακας μπορεί να χρησιμοποιηθεί και σαν μια εικόνα. Η εικόνα αυτή ονομάζεται φασματογράφημα του σήματος και αντιπροσωπεύει την εξέλιξη του σήματος στα πεδία του χρόνου και των συχνοτήτων. Για την κατασκευή του χρησιμοποιούμε το μέγεθος των STFT όρων σε γραμμική ή και λογαριθμική κλίμακα (dB) [6]. Πριν από τον κάθε υπολογισμό DFT στο κάθε παράθυρο εφαρμόζεται μια συνάρτηση που ασθενεί το σήμα στις άκρες του παραθύρου και το ενισχύει στο κέντρο του παραθύρου όπως είναι η συνάρτηση Hamming.



Εικόνα 1: Ηχητικό σήμα μετατροπή σε mel φασματογράφημα

Η mel (από το melody) κλίμακα εφαρμόζεται στους STFT συντελεστές με στόχο τον διαχωρισμό των τόνων (pitch) του ηχητικού σήματος με τρόπο που ο κάθε τόνος να έχει την ίδια απόσταση από τον άλλον όπως τους αντιλαμβάνεται το ανθρώπινο αυτί. Υπάρχουν πολλοί τρόποι για να γίνει αυτό. Ένας από αυτούς τους τρόπους είναι η επόμενη εξίσωση:

$$m = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (1.6)$$

1.7.4 Συντελεστές MFCC

Οι Mel-Frequency Cepstrum Coefficients (MFCC) είναι μια φασματική αναπαράσταση όπου οι συχνότητες έχουν κατανομή σύμφωνα με την mel κλίμακα. Ο υπολογισμός τους γίνεται με τον εξής τρόπο:

1. Υπολογίζονται οι συντελεστές με τον διακριτό μετασχηματισμό Fourier (DFT).
2. Οι συντελεστές δίνονται σαν είσοδος σε μια συνάρτηση που τους μετατρέπει στην mel κλίμακα χρησιμοποιώντας L φίλτρα που συνήθως έχουν τριγωνική συχνοτική απόκριση.
3. Αν $\tilde{O}_k, k = 1, \dots, L$ είναι η ένταση από την έξοδο που k th φίλτρου τότε οι συντελεστές MFCC δίνονται από την εξίσωση:

$$c_m = \sum_{k=1}^L \log \tilde{O}_k \cos \left[m \left(k - \frac{1}{2} \right) \frac{\pi}{L} \right], m = 1, \dots, L \quad (1.7)$$

Άρα οι συντελεστές MFCC είναι ο διακριτός συνημιτονοειδής μετασχηματισμός των συντελεστών στην mel κλίμακα του λογαριθμικού φάσματος της έντασης (log-power spectrum) [6].

1.7.5 Τροχιές φασματικών κορυφών (Spectral peak trajectories)

Οι μηχανικές ταλαντώσεις από την κίνηση και την τριβή των διαφόρων εξαρτημάτων και τμημάτων του μηχανήματος πολλές φορές προκαλούν αντηχήσεις. Όταν οι συχνότητες των αντηχήσεων αυτών είναι κοντά σε κάποια βασική ή αρμονική συχνότητα ο συντονισμός αυτός προκαλεί αύξηση της έντασης της βασικής ή αρμονικής συχνότητας. Αν παρακολουθήσουμε τις φασματικές αυτές κορυφές στη διάρκεια του χρόνου έχουμε τις τροχιές των φασματικών κορυφών (Spectral peak trajectories - SPT) [7]. Οι τροχιές αυτές πολλές φορές περιέχουν μοναδικά μοτίβα και χαρακτηριστικά από τον ήχο που παράγει κατά την λειτουργία του το μηχάνημα. Απόκλιση από αυτά τα μοτίβα του ηχητικού σήματος μπορεί να αναδεικνύει μη ομαλή λειτουργία.

Έστω $X_l[k]; k = 0, \dots, N_f - 1$ οι συντελεστές DFT για το τμήμα του ήχου l αν χωρίσουμε το ηχητικό σήμα σε L τμήματα.

Από αυτούς βρίσκουμε τις θέσεις που οι συχνότητες έχουν τις τοπικά μέγιστες τιμές-κορυφές και φτιάχνουμε το:

$$H_l = \{k : (|X_l[k-1]| < |X_l[k]|) \wedge (|X_l[k]| > |X_l[k+1]|)\} \quad (1.8)$$

όπου $0 \leq k < (N_f - 1)$ και $|X_l[k]|$ είναι το μέγεθος του DFT συντελεστή $X_l[k]$.

Από αυτές τις κορυφές χρησιμοποιούμε μόνο τις p μεγαλύτερες τις οποίες και ταξινομούμε από την μεγαλύτερη προς την μικρότερη. Αν δεν υπάρχουν p κορυφές συμπληρώνουμε στην αρχή του διανύσματος όσες φορές χρειάζεται την μεγαλύτερη από αυτές.

$$\tilde{H}_l = \{k_0, k_1, \dots, k_p\} \quad (1.9)$$

Τα στοιχεία του \tilde{H}_l τα ταξινομούμε ώστε να φτιάξουμε το fH_l που περιέχει τις θέσεις σε αύξουσα σειρά.

$$fH_l[0] \geq fH_l[1] \geq \dots \geq fH_l[p-1] \quad (1.10)$$

Τα διανύσματα fH_l , ($l = 0, \dots, L-1$) τα χρησιμοποιούμε για να φτιάξουμε τον πίνακα PLM για το ηχητικό σήμα. Η l th στήλη του πίνακα ορίζεται ως

$$PLM_l = fH_l^T \quad (1.11)$$

Με τον ίδιο τρόπο φτιάχνουμε και τον πίνακα PAM με το πλάτος (amplitude) της συχνότητας στην κάθε κορυφή

$$PAM[r, l] = X_l[h] \quad (1.12)$$

όπου $h = PLM[r, l]$, $r = 0, \dots, (p-1)$ και $l = 0, \dots, (L-1)$.

Η κάθε γραμμή του PLM ορίζεται ως η ακολουθία των θέσεων των φασματικών κορυφών (Location Sequence of Peak Traces - LSPT). Όμοια η κάθε γραμμή του PAM ορίζεται σαν η ακολουθία του πλάτους των φασματικών κορυφών (Amplitude Sequence of Peak Traces - ASPT).

Αν πάρουμε τον μέσο όρο και την τυπική απόκλιση από κάθε ακολουθία των θέσεων των φασματικών κορυφών, για κάθε δηλαδή γραμμή του PLM και του PAM πίνακα φτιάχνουμε δύο διανύσματα χαρακτηριστικών για όλο ηχητικό σήμα.

$$MSD - LSPT = [\mu_0^{PLM}, \dots, \mu_{p-1}^{PLM}, \sigma_0^{PLM}, \dots, \sigma_{p-1}^{PLM}] \quad (1.13)$$

Όμοια και με τον ASPT:

$$MSD - ASPT = [\mu_0^{PAM}, \dots, \mu_{p-1}^{PAM}, \sigma_0^{PAM}, \dots, \sigma_{p-1}^{PAM}] \quad (1.14)$$

Η ένωση των MSD-LSPT και MSD-ASPT είναι ένα διάνυσμα με 4-p διαστάσεις (4p-MSD) και είναι το διάνυσμα που χρησιμοποιήσαμε για την εκπαίδευση των αυτόματων κωδικοποιητών.

2 Αλγόριθμοι μηχανικής μάθησης

2.1 Βαθιά νευρωνικά δίκτυα πρόσθιας τροφοδότησης (deep feed forward networks)

Τα νευρωνικά δίκτυα πρόσθιας τροφοδότησης (feed forward neural networks) ή πολλών επιπέδων perceptron (multilayer perceptron MLP) ορίζουν μια αντιστοίχιση από μια είσοδο x σε μια έξοδο y , την προσέγγιση δηλαδή μιας συνάρτησης f^* . Στην περίπτωση π.χ ενός ταξινομητή η έξοδος μπορεί να είναι μια κατηγορία ενώ στην περίπτωση της παλινδρόμησης (regression) μια τιμή. Κατά την εκμάθηση ο στόχος είναι να μάθουν τους παραμέτρους θ που έχουν το καλύτερο αποτέλεσμα $y = f(x; \theta)$. Ονομάζονται δίκτυα γιατί τυπικά ορίζονται σαν την σύνθεση συναρτήσεων. Π.χ. μπορεί να έχουμε 3 συναρτήσεις $f^{(1)}, f^{(2)}, f^{(3)}$ που συνδέονται αλυσιδωτά $f(x) = f^{(3)}(f^{(2)}(f^{(1)}(x)))$. Το $f^{(1)}$ ονομάζεται πρώτο επίπεδο του δικτύου, το $f^{(2)}$ δεύτερο επίπεδο κτλ. Ο συνολικός αριθμός των επιπέδων είναι το βάθος του δικτύου (depth). Το πρώτο επίπεδο είναι το επίπεδο εισόδου ενώ το τελευταίο επίπεδο ονομάζεται και επίπεδο εξόδου (output layer). Επειδή τα παραδείγματα για την εκπαίδευση δίνουν πληροφορίες μόνο για τα επίπεδα εισόδου και εξόδου τα υπόλοιπα επίπεδα ονομάζονται κρυφά επίπεδα. Τα επίπεδα έχουν πολλές μονάδες (units), οι οποίες εκφράζουν μια συνάρτηση από ένα διάνυσμα σε μια τιμή και ενεργούν παράλληλα [2].

Η εκπαίδευση των MLP's περιλαμβάνει την επιλογή ενός αλγορίθμου βελτιστοποίησης (optimizer), της συνάρτησης σφάλματος αλλά και την επιλογή των μονάδων του επιπέδου εξόδου. Η εκμάθηση των βαθιών δικτύων χρειάζεται επίσης τον υπολογισμό των κλίσεων (gradients) πολύπλοκων συναρτήσεων.

Η κάθε μονάδα (unit) στο κάθε επίπεδο που ονομάζεται και perceptron αντιπροσωπεύει έναν γραμμικό μετασχηματισμό των εισόδων. Για το πρώτο επίπεδο ο μετασχηματισμός είναι της μορφής:

$$a_j = \sum_{i=1}^D w_{ji}^{(1)} x_i + w_{j0}^{(1)} \quad (2.1)$$

Όπου x_1, \dots, x_D είναι το διάνυσμα χαρακτηριστικών του παραδείγματος. Οι παράμετροι $w_{ji}^{(1)}$ ονομάζονται βάρη (weights) και οι παράμετροι $w_{j0}^{(1)}$ πολώσεις (biases). Οι ποσότητες a_j ονομάζονται ενεργοποιήσεις (activations). Κάθε μια από αυτές μετασχηματίζεται με μια παραγωγίσιμη μη γραμμική συνάρτηση ενεργοποίησης $h(\cdot)$:

$$z_j = h(a_j) \quad (2.2)$$

Οι ποσότητες z_j ονομάζονται και κρυμμένες μονάδες (hidden units). Το ίδιο συμβαίνει και στα επόμενα επίπεδα όπου αντί για το διάνυσμα των εισόδων έχουμε το διάνυσμα με τις ενεργοποιήσεις από το προηγούμενο επίπεδο. Τέλος στο επίπεδο εξόδου οι συναρτήσεις ενεργοποίησης καθορίζονται από τον τύπο του προβλήματος. Για προβλήματα παλινδρόμησης έχουμε την ταυτοτική (identity) συνάρτηση έτσι ώστε η έξοδος $y_k = a_k$. Για προβλήματα πολλαπλής δυαδικής ταξινόμησης σε κάθε έξοδο έχουμε την σιγμοειδή συνάρτηση:

$$\sigma(a) = \frac{1}{1 + \exp(-a)} \quad (2.3)$$

Για προβλήματα ταξινόμησης με πολλές κλάσεις έχουμε για συνάρτηση ενεργοποίησης την softmax συνάρτηση.

Η μη γραμμική συνάρτηση ενεργοποίησης που είναι προεπιλογή για τις μονάδες των κρυμμένων επιπέδων στα βαθιά δίκτυα είναι η Rectified linear unit (ReLU) ή παραλλαγές της όπως η leaky ReLU [2].

$$h(x) = \max(0, x) \quad (2.4)$$

Αν συνδυάσουμε όλα τα προηγούμενα για όλο το δίκτυο με την σιγμοειδή συνάρτηση ενεργοποίησης έχουμε [8]:

$$y_k(x, w) = \sigma \left(\sum_{j=1}^M w_{kj}^{(2)} h \left(\sum_{i=1}^D w_{ji}^{(1)} x_i + w_{j0}^{(1)} \right) + w_{k0}^{(2)} \right) \quad (2.5)$$

Άρα συνολικά ένα δίκτυο πρόσθιας τροφοδότησης που ονομάζεται πολλές φορές και μοντέλο νευρωνικού δικτύου είναι απλά μιλώντας μια μη γραμμική συνάρτηση με είσοδο ένα σύνολο από μεταβλητές εισόδου $\{x_i\}$ και έξοδο ένα σύνολο μεταβλητών $\{y_k\}$ που ρυθμίζονται από ένα σύνολο παραμέτρων $\{w\}$.

2.2 Οπισθοδιάδοση του σφάλματος (backpropagation)

Για την εκπαίδευση του δικτύου με τη χρήση του αλγορίθμου gradient descent χρειάζεται ο υπολογισμός της κλίσης της συνάρτησης σφάλματος με βάση τις παραμέτρους του δικτύου. Για αυτόν τον υπολογισμό γίνεται χρήση του αλγορίθμου οπισθοδιάδοσης του σφάλματος (backpropagation) [9].

Πρώτα υπολογίζουμε τις εξόδους όλων των μονάδων σε όλα τα επίπεδα με τη χρήση της πρόσθιας τροφοδότησης και το διάνυσμα εισόδου.

$$a_j = \sum_i w_{ji} z_i \quad (2.6)$$

όπου a_j είναι η έξοδος της μονάδας, w_{ji} είναι το βάρος της σύνδεσης και z_i η ενεργοποίηση της μονάδας της σύνδεσης ή η είσοδος.

$$z_j = h(a_j) \quad (2.7)$$

όπου $h(\cdot)$ η συνάρτηση ενεργοποίησης της μονάδας.

Για το επίπεδο εξόδου υπολογίζουμε την κλίση δ_k .

$$\delta_k = y_k - t_k \quad (2.8)$$

όπου y_k είναι η έξοδος και t_k η ετικέτα για το παράδειγμα.

Αμέσως μετά μπορούμε να υπολογίσουμε τα δ για το προηγούμενο επίπεδο:

$$\delta_j = h'(a_j) \sum_k w_{kj} \delta_k \quad (2.9)$$

Η κλίση για το error ενός παραδείγματος εξαρτάται από το βάρος w_{ji} μόνο μέσω του αθροίσματος της εισόδου a_j (2.6) της μονάδας j . Αν εφαρμόσουμε τον κανόνα της αλυσίδας έχουμε:

$$\frac{\partial E_n}{\partial w_{ji}} = \frac{\partial E_n}{\partial a_j} \frac{\partial a_j}{\partial w_{ji}} \quad (2.10)$$

Αν ορίσουμε:

$$\delta_j \equiv \frac{\partial E_n}{\partial a_j} \quad (2.11)$$

Επίσης από την (2.6) έχουμε:

$$\frac{\partial a_j}{\partial w_{ji}} = z_i \quad (2.12)$$

Και από τον (2.10) μπορούμε να υπολογίσουμε την κλίση της συνάρτησης λάθους E_n με βάση το βάρος w_{ji} .

$$\frac{\partial E_n}{\partial w_{ji}} = \delta_j z_i \quad (2.13)$$

Για τη χρήση με μια συστάδα n παραδειγμάτων (batch) η κλίση για το συνολικό σφάλμα E μπορεί να υπολογιστεί επαναληπτικά υπολογίζοντας την κλίση για το κάθε παράδειγμα και αθροίζοντας στο τέλος [8]:

$$\frac{\partial E}{\partial w_{ji}} = \sum_n \frac{\partial E_n}{\partial w_{ji}} \quad (2.14)$$

Η παραπάνω διαδικασία γενικεύεται και σε δίκτυα με περισσότερα επίπεδα υπολογίζοντας επαναληπτικά τις κλίσεις της συνάρτησης λάθους ως προς τα βάρη και τις πολώσεις για κάθε μονάδα κάθε επιπέδου.

2.3 Συνελικτικά νευρωνικά δίκτυα

Τα συνελικτικά νευρωνικά δίκτυα είναι εξειδικευμένα νευρωνικά δίκτυα για την επεξεργασία δεδομένων που έχουν τοπολογία πλέγματος. Το όνομά τους προδίδει ότι χρησιμοποιούν μια μαθηματική διαδικασία που ονομάζεται συνέλιξη. Τα συνελικτικά νευρωνικά δίκτυα χρησιμοποιούν στην θέση του πολλαπλασιασμού πινάκων την συνέλιξη. Για διακριτές τιμές του t η συνέλιξη για 2 συναρτήσεις x και w ορίζεται ως [2]:

$$s(t) = (x * w)(t) = \sum_{a=-\text{inf}}^{\text{inf}} x(a)w(t-a) \quad (2.15)$$

Η συνάρτηση x ονομάζεται είσοδος και είναι πολυδιάστατος πίνακας και η συνάρτηση w είναι πολυδιάστατος πίνακας με παραμέτρους που ρυθμίζονται από τον αλγόριθμο και ονομάζεται πυρήνας (kernel). Οι πολυδιάστατοι αυτοί πίνακες ονομάζονται και tensors.

Για συνέλιξη σε 2 διαστάσεις έχουμε:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i-m, j-n) \quad (2.16)$$

Η συνέλιξη έχει 3 σημαντικές ιδιότητες για ένα σύστημα μηχανικής μάθησης: αραιές αλληλεπιδράσεις (sparse interactions) ανάμεσα στους νευρώνες, κοινή χρήση των παραμέτρων (parameter sharing) και ισοδυναμία στις αναπαραστάσεις (equivariant representations).

Τα κλασσικά νευρωνικά δίκτυα χρησιμοποιούν τον πολλαπλασιασμό πίνακα με έναν πίνακα παραμέτρων που η κάθε παράμετρος περιγράφει την αλληλεπίδραση ανάμεσα σε κάθε μονάδα εισόδου με κάθε μονάδα εξόδου. Αυτό σημαίνει ότι η κάθε μονάδα εισόδου αλληλεπιδρά με την κάθε μονάδα εξόδου. Στα συνελικτικά νευρωνικά δίκτυα οι αλληλεπιδράσεις είναι αραιές (sparse) και αυτό επιτυγχάνεται με τον πυρήνα (kernel) να είναι μικρότερης διάστασης από την είσοδο.

Οι παράμετροι του πυρήνα (kernel) στα συνελκτικά δίκτυα χρησιμοποιούνται σε κάθε είσοδο (parameter sharing). Αυτό σημαίνει ότι αντί το δίκτυο να μάθει άλλο σετ παραμέτρων για κάθε περιοχή θα μάθει μόνο ένα σετ. Αυτό μειώνει τον χώρο αποθήκευσης των παραμέτρων και άρα κάνει τα συνελκτικά δίκτυα πιο αποδοτικά. Η ιδιότητα του διαμοιρασμού των παραμέτρων κάνει τα συνελκτικά επίπεδα να έχουν και μια ισοδυναμία στις μετατροπές (equivariance to translation). Αν αλλάξουμε την είσοδο η έξοδος θα αλλάξει με αντίστοιχο-ισοδύναμο τρόπο. Αν π.χ. σε μια εικόνα μετακινήσουμε ένα αντικείμενο οι αναπαραστάσεις που έχουν οι πίνακες χαρακτηριστικών στην έξοδο θα έχουν αντίστοιχες μετακινήσεις. Δεν συμβαίνει όμως το ίδιο σε όλες τις αλλαγές όπως π.χ. αν αλλάξουμε την κλίμακα ή περιστρέψουμε ένα αντικείμενο δεν θα έχουμε αντίστοιχες αλλαγές στην έξοδο.

Μπορούμε να αποφύγουμε κάποιες θέσεις του πυρήνα με σκοπό να μειώσουμε το κόστος υπολογισμού αλλά με την λιγότερο λεπτομερή απεικόνιση των χαρακτηριστικών. Η συνάρτηση αν δειγματίσουμε μόνο κάθε s pixels σε κάθε κατεύθυνση θα γίνει [2]:

$$Z_{i,j,k} = c(K, V, s)_{i,j,k} = \sum_{l,m,n} [V_{l,(j-1)xs+m,(k-1)xs+m} K_{i,l,m,n}] \quad (2.17)$$

όπου s είναι το βήμα (stride) της συνέλιξης. Μπορούμε να ορίσουμε και διαφορετικό βήμα σε κάθε κατεύθυνση.

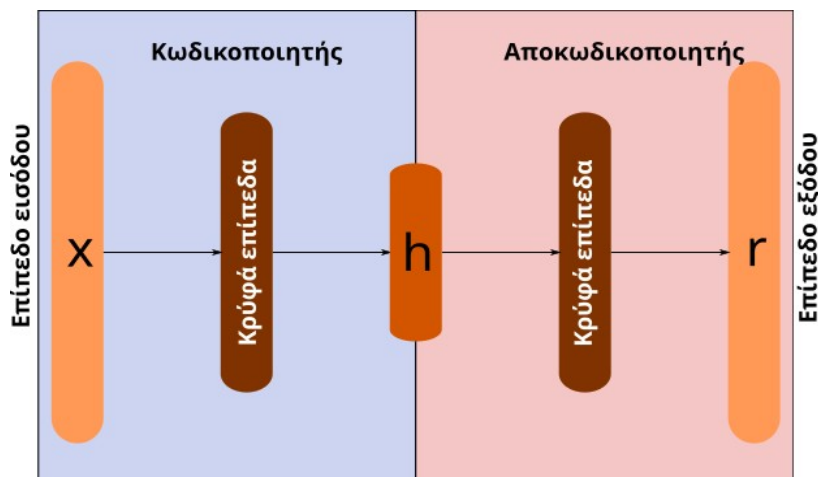
Επειδή σε κάθε εφαρμογή της συνέλιξης της εισόδου με τον πυρήνα η έξοδος είναι ένα pixel μικρότερη πολλές φορές γεμίζουμε (padding) τις έξω πλευρές της εισόδου με μηδενικά ώστε το αποτέλεσμα να μην έχει ένα pixel λιγότερο. Η διαδικασία αυτή λέγεται όμοια συνέλιξη (same convolution) ενώ αν δεν κάνουμε καθόλου γέμισμα ονομάζεται έγκυρη συνέλιξη (valid convolution).

Συνήθως τα συνελκτικά επίπεδα ακολουθούνται από ένα επίπεδο που συνοψίζει (pooling) τις γειτονικές μονάδες με τις συναρτήσεις μέγιστου (max pooling) ή μέσου όρου (average pooling). Στην περίπτωση όμως των αυτόματων κωδικοποιητών δεν βοηθούν στην εκπαίδευση και συχνά δίνουν χειρότερα αποτελέσματα. Για τον λόγο αυτόν δεν χρησιμοποιήσαμε pooling επίπεδα στα πειράματα με τους συνελκτικούς αυτόματους κωδικοποιητές.

Ο πολλαπλασιασμός με τον αντίστροφο (transpose) του πίνακα που ορίζεται από την συνέλιξη είναι η αντίστροφη διαδικασία του συνελκτικού επιπέδου (transpose convolution) και τη χρησιμοποιούμε στους συνελκτικούς αυτόματους κωδικοποιητές για την ανακατασκευή των ορατών μονάδων από τις κρυφές μονάδες στο τμήμα του αποκωδικοποιητή [2].

3 Αυτόματοι κωδικοποιητές

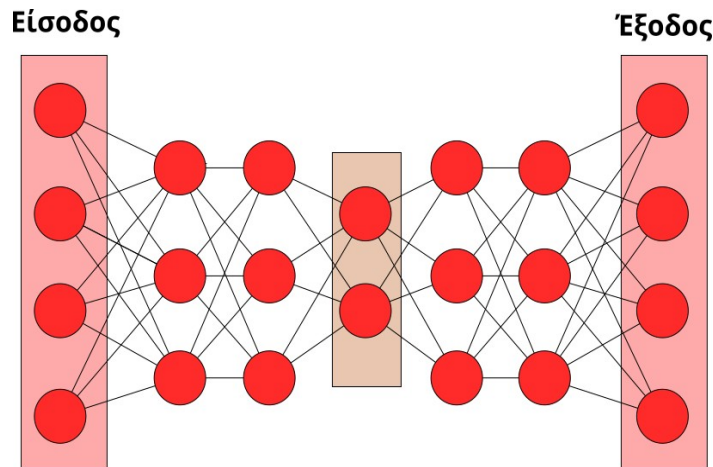
Ένας αυτόματος κωδικοποιητής (autoencoder) είναι ένα νευρωνικό δίκτυο που εκπαιδεύεται να αντιγράψει την είσοδο του στην έξοδο. Εσωτερικά έχει ένα κρυφό επίπεδο που αναπαριστά την είσοδο με έναν κώδικα. Το δίκτυο αποτελείται από δύο τμήματα. Το πρώτο είναι μια συνάρτηση κωδικοποίησης (encoder) $h = f(x)$ και το δεύτερο είναι μια συνάρτηση αποκωδικοποίησης (decoder) που ανακατασκευάζει την είσοδο $r = g(h)$ [2]. Συνήθως είναι περιορισμένοι στο να ανακατασκευάζουν κατά προσέγγιση την έξοδο και να δίνουν σημασία στα χαρακτηριστικά των εισόδων που είναι τα πιο σημαντικά. Η έξοδος του κωδικοποιητή έχει διάσταση μικρότερη από τη διάσταση της εισόδου.



Εικόνα 2: Αυτόματος αποκωδικοποιητής

Το κάθε τμήμα είναι ένα νευρωνικό δίκτυο ή MLP και η συνάρτηση σφάλματος είναι το μέσο τετραγωνικό σφάλμα (mean squared error) ανάμεσα στην είσοδο και την έξοδο που ονομάζεται και σφάλμα ανακατασκευής [10].

Τα επίπεδα του κωδικοποιητή αλλά και του αντίστοιχου αποκωδικοποιητή μπορεί να είναι περισσότερα από ένα. Σε αυτή την περίπτωση έχουμε βαθιούς αυτόματους κωδικοποιητές (deep autoencoders).



Εικόνα 3: Deep autoencoder

3.1.1 Μη επιτηρούμενη ανίχνευση ανώμαλων ήχων

Για να ανιχνεύσουμε ανώμαλους ήχους θα υπολογίσουμε ένα σκορ με τη χρήση βαθιών αυτόματων κωδικοποιητών για το κάθε αρχείο ήχου που όσο πιο μεγάλο θα είναι το μέγεθος του σκορ τόσο πιο πιθανό είναι το μηχάνημα που τον παράγει να έχει προβλήματα στην λειτουργία του.

Έστω ένα αρχείο ήχου $x \in \mathbb{R}^l$ ως διάνυσμα x μήκους l . Το πρόβλημα ορίζεται ως την αναγνώριση αν το μηχάνημα που παράγει τον x έχει ομαλή ή όχι λειτουργία αναλύοντας το x . Για να πετύχουμε τον στόχο θα υπολογίσουμε ένα σκορ με τη χρήση μιας συνάρτησης A με παραμέτρους θ ώστε αν το σκορ είναι μεγαλύτερο από ένα κατώφλι ϕ να θεωρείται ανώμαλο [11].

$$A_{\theta}(x) > \phi \quad (3.1)$$

Από την 3.1 εξάγεται το συμπέρασμα ότι η A πρέπει να σχεδιαστεί ώστε τα ανώμαλα ηχητικά κομμάτια να έχουν μεγάλο σκορ και τα ομαλά μικρό.

Ένας βαθύς αυτόματος κωδικοποιητής θα εκπαιδευτεί μόνο με ομαλούς ήχους ώστε να ελαχιστοποιεί το σφάλμα ανακατασκευής. Όταν μετά την εκπαίδευση θα δοθούν στον αυτόματο κωδικοποιητή ανώμαλοι ήχοι το σφάλμα ανακατασκευής αναμένουμε να είναι μεγάλο γιατί δεν θα έχει εκπαιδευτεί με αυτούς.

Πρώτα υπολογίζουμε το mel-φασματογράφημα για την είσοδο $X \in \mathbb{R}^{F \times T}$ όπου το F είναι ο αριθμός των mel φίλτρων και T τα χρονικά τμήματα. Αμέσως μετά το φασματογράφημα στον

χρόνο t ενώνεται με πριν και μετά P τμήματα και φτιάχνουμε το $\phi = (X_{t-P}, \dots, X_{t+p})$. Το σκορ τότε υπολογίζεται ως [11]:

$$A_{\theta}(x) = \frac{1}{T} \sum_{t=1}^T \|\phi_t - D_{\theta_D}(E_{\theta_E}(\phi_t))\|_2^2 \quad (3.2)$$

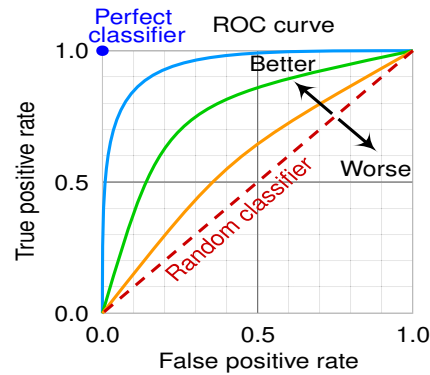
όπου $\|\cdot\|_2$ είναι η l_2 νόρμα (Ευκλείδεια απόσταση) και τα E και D είναι ο αποκωδικοποίησης και ο κωδικοποιητής του βαθιού αυτόματου κωδικοποιητή (deep autoencoder) όπου οι παράμετροί τους είναι θ_E και θ_D αντίστοιχα. Άρα συνολικά οι παράμετροι είναι $\theta = \{\theta_E, \theta_D\}$. Στα κρυφά επίπεδα οι συναρτήσεις ενεργοποίησης είναι η συνάρτηση ReLU (rectified linear unit) και για τη βελτιστοποίηση της συνάρτησης σφάλματος έγινε χρήση του αλγορίθμου Adam που είναι βασισμένος στον Gradient descent αλγόριθμο [12].

3.1.2 Καμπύλη Receiver operating characteristic (ROC)

Αληθώς θετική (true positive - TP) ονομάζεται μια θετική έξοδος αν είναι ίδια με την πραγματική τιμή (ground truth) που συνοδεύει το παράδειγμα σε ένα πρόβλημα ταξινόμησης. Ψευδώς θετική (false positive - FP) ονομάζεται η θετική έξοδος όταν είναι διαφορετική από την πραγματική τιμή. Ο ρυθμός των αληθώς θετικών (true positive rate - TPR) είναι ο αριθμός των αληθώς θετικών προς όλα τα θετικά παραδείγματα (positive - P) $TPR = \frac{TP}{P}$. Αντίστοιχα ο ρυθμός των ψευδώς θετικών είναι ο αριθμός των ψευδώς θετικών (false positive - FP) προς το σύνολο των αρνητικών εξόδων (negative - N) εξόδων $FPR = \frac{FP}{N}$. Η καμπύλη που σχηματίζεται από τον ρυθμό των αληθώς θετικών προς τον ρυθμό των ψευδώς θετικών (false positive rate) καθώς μετακινούμε το κατώφλι πάνω από το οποίο ορίζουμε κάτι θετικό (ομαλό παράδειγμα για την περίπτωση τις ανίχνευσης ανωμαλιών) ή αρνητικό, ονομάζεται receiver operating characteristic curve (ROC curve) [13]. Η ROC καμπύλη αποτελεί ένα δείκτη για το πόσο καλός είναι ένας δυαδικός ταξινομητής σε σχέση με έναν τυχαίο ταξινομητή ή άλλον ταξινομητή.

3.1.3 Area under the curve (AUC)

Όσο μεγαλύτερη είναι η περιοχή κάτω από την καμπύλη (area under the curve AUC) τόσο καλύτερος είναι ο ταξινομητής.



Εικόνα 4: Καμπύλη ROC (εικόνα από wikipedia)

Πιο συγκεκριμένα για την περίπτωση της ανίχνευσης ανωμαλιών έχουμε:

$$AUC = \frac{1}{N_- N_+} \sum_{i=1}^{N_-} \sum_{j=1}^{N_+} H(A_\theta(x_j^+) - A_\theta(x_i^-)) \quad (3.3)$$

όπου $H(a)$ είναι η συνάρτηση κατωφλίου που επιστρέφει 1 όταν $a > 0$ και μηδέν αλλιώς, τα $\{x_i^-\}_{i=1}^N$ και $\{x_j^+\}_{j=1}^{N_+}$ είναι τα ομαλά και ανώμαλα παραδείγματα αντίστοιχα και τα N_- και N_+ είναι το πλήθος των ομαλών και ανώμαλων παραδειγμάτων. Τα παραδείγματα είναι ταξινομημένα με βάση το σκορ από την έξοδο του αυτόματου κωδικοποιητή. Επειδή ένα σύστημα ανίχνευσης ανωμαλιών δεν θα πρέπει να έχει μεγάλο αριθμό λάθος ειδοποιήσεων γιατί δεν θα θεωρείται αξιόπιστο θα χρησιμοποιήσουμε και ένα δεύτερο μετρικό για την αξιολόγηση που θα χρησιμοποιεί ένα κομμάτι από τα θετικά παραδείγματα ώστε να υπάρχει καλύτερος ρυθμός αληθώς θετικών (TPR) κάτω από λιγότερα ψευδώς θετικά (FPR) αποτελέσματα [11]:

$$pAUC = \frac{1}{\lfloor pN_- \rfloor N_+} \sum_{i=1}^{\lfloor pN_- \rfloor} \sum_{j=1}^{N_+} H(A_\theta(x_j^+) - A_\theta(x_i^-)) \quad (3.4)$$

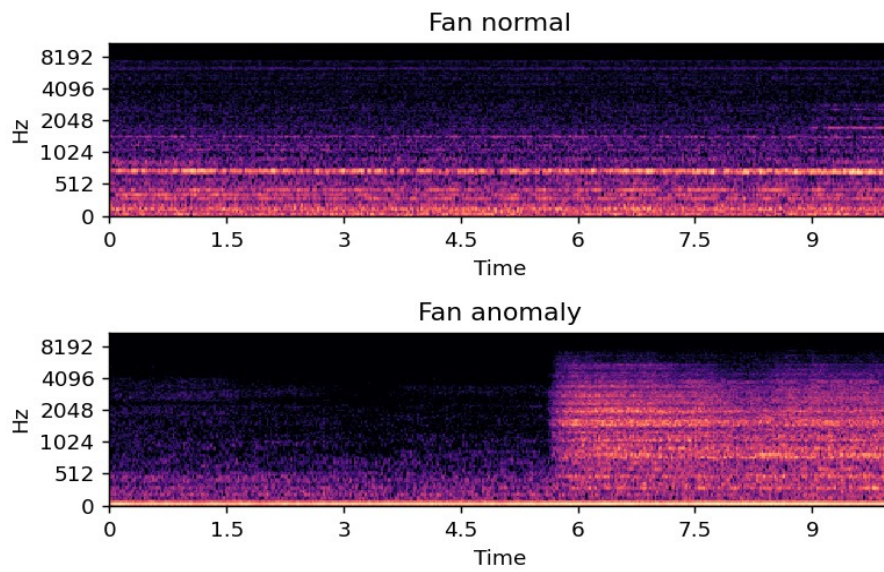
όπου το p είναι αριθμός που δηλώνει ποσοστό π.χ. 0.1. και το $\lfloor \cdot \rfloor$ είναι ο κοντινότερος μικρότερος ακέραιος.

4 Εφαρμογές

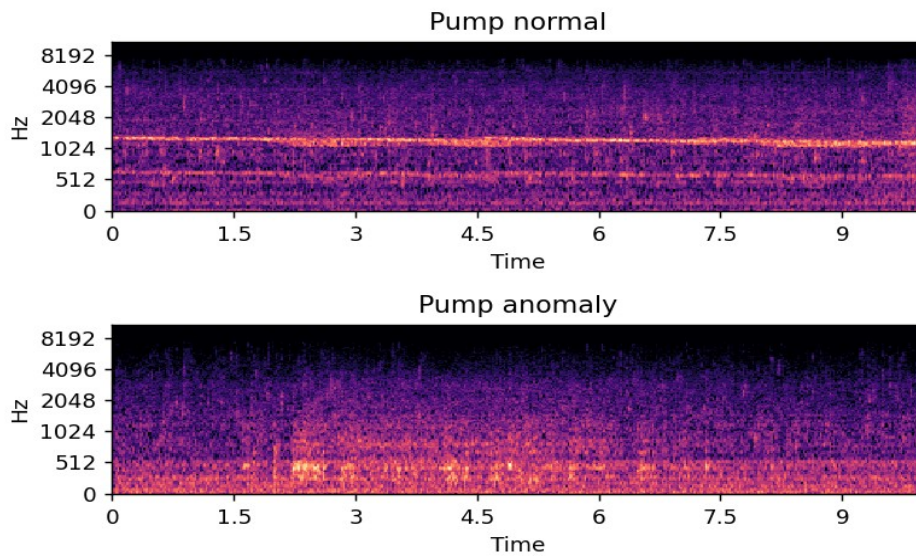
4.1 Σύνολα δεδομένων

Το σύνολο δεδομένων που χρησιμοποιήσαμε για τα πειράματα είναι το MIMII (malfunctioning industrial machine investigation and insertion) [14]. Σε αυτό υπάρχουν ομαλοί και ανώμαλοι ήχοι από 4 κατηγορίες μηχανημάτων: βαλβίδες, αντλίες, ανεμιστήρες αλλά και ράγες κύλισης. Για την κάθε κατηγορία μηχανήματος το σύνολο δεδομένων έχει ήχους από 4 διαφορετικούς τύπους μηχανημάτων. Οι ανώμαλοι ήχοι περιλαμβάνουν προβλήματα στην λειτουργία των εξαρτημάτων όπως διαρροές, ανισορροπία στην περιστροφή και προβλήματα στις ράγες. Η διάρκεια του κάθε ήχου είναι 10 δευτερόλεπτα. Το σύνολο δεδομένων περιέχει 26 χιλιάδες ήχους από την ομαλή λειτουργία και των τεσσάρων κατηγοριών για την εκπαίδευση των μοντέλων και ήχους από ομαλή αλλά και ανώμαλη λειτουργία για την αξιολόγηση του μοντέλου που κατασκευάσαμε. Οι ήχοι από την μη ομαλή λειτουργία είναι 6 χιλιάδες και είναι λιγότεροι σε αριθμό γιατί το σύνολο δεδομένων έχει σχεδιαστεί για την εκπαίδευση μη επιτηρούμενων αλγορίθμων. Η δειγματοληψία είναι στα 16kHz και περιέχουν και θορύβους που έχουν ηχογραφηθεί από πραγματικά εργοστάσια και ενωθεί με τους ήχους των μηχανημάτων.

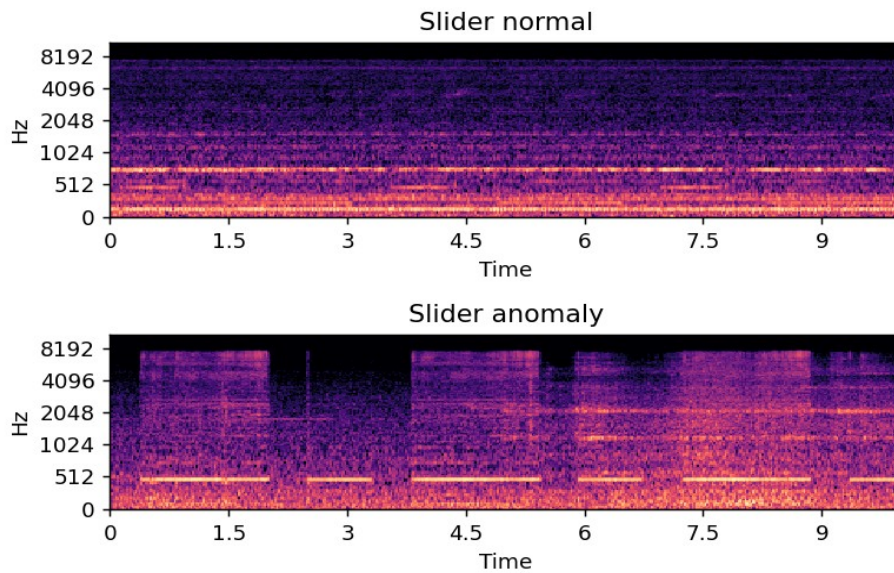
Οι βαλβίδες είναι ηλεκτροβάνες που ανοιγοκλείνουν συνεχόμενα, οι αντλίες είναι νερού και ρουφάνε και αμέσως μετά αδειάζουν το νερό σε μια πισίνα, οι ανεμιστήρες είναι επαγγελματικοί και χρησιμοποιούνται για τη δημιουργία συνεχής ροής αέρα σε εργοστάσια και τέλος οι ράγες κύλισης είναι γραμμικές-ευθύγραμμες. Οι προβληματικές λειτουργίες που έχουν ηχογραφηθεί είναι πάνω από 2 διαφορετικοί τύποι διαρροών στις βαλβίδες, διαρροές και βουλώματα στις αντλίες, μη ισορροπημένη λειτουργία, αλλαγή στην τάση του ρεύματος και βουλώματα στους ανεμιστήρες, και τέλος, βλάβες στις ράγες, λύσιμο του ιμάντα και έλλειψη γράσου στις ράγες κύλισης.



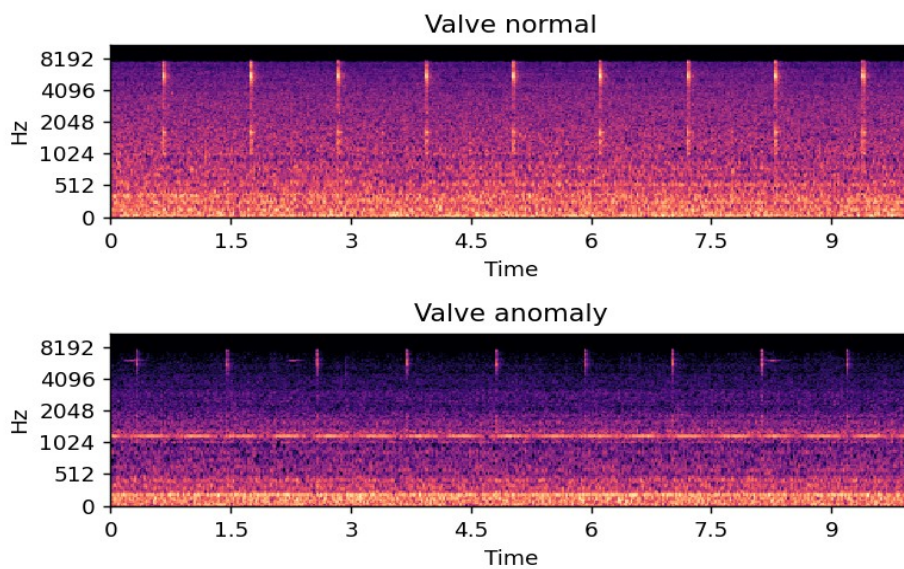
Εικόνα 5: Ομαλό και μη ομαλό φασματογράφημα από ανεμιστήρα



Εικόνα 6: Ομαλό και μη ομαλό φασματογράφημα από αντλία



Εικόνα 7: Ομαλό και μη ομαλό φασματογράφημα από ράγες κύλισης

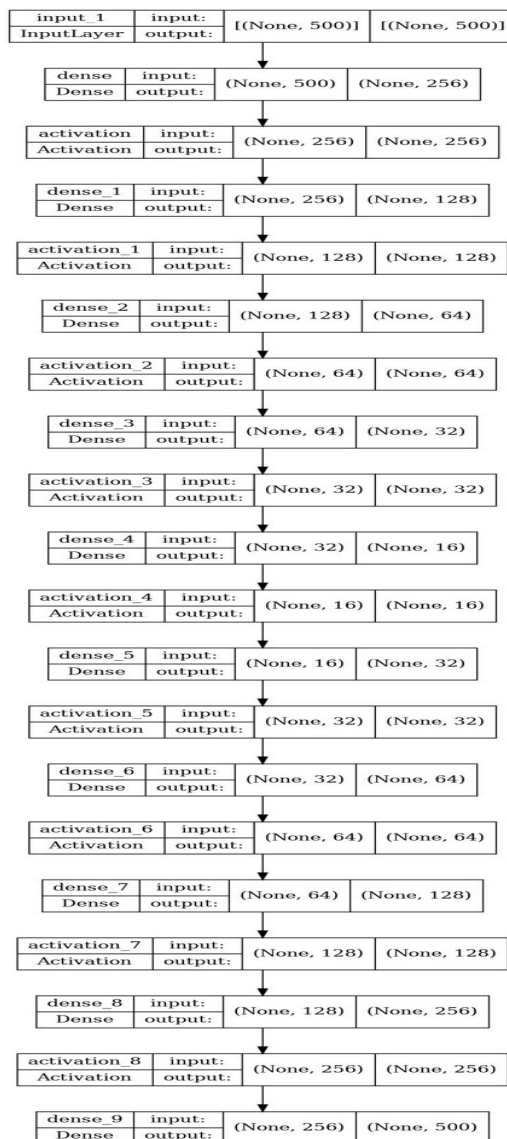


Εικόνα 8: Ομαλό και μη ομαλό φασματογράφημα από βαλβίδες

4.2 Πειράματα

Τα mel φασματογραφήματα έγιναν με 1024 συντελεστές fft, 512 μήκος βήματος και 128 mels. Ενώσαμε 5 τμήματα για να κατασκευάσουμε τα χαρακτηριστικά για τους απλούς αυτόματους κωδικοποιητές.

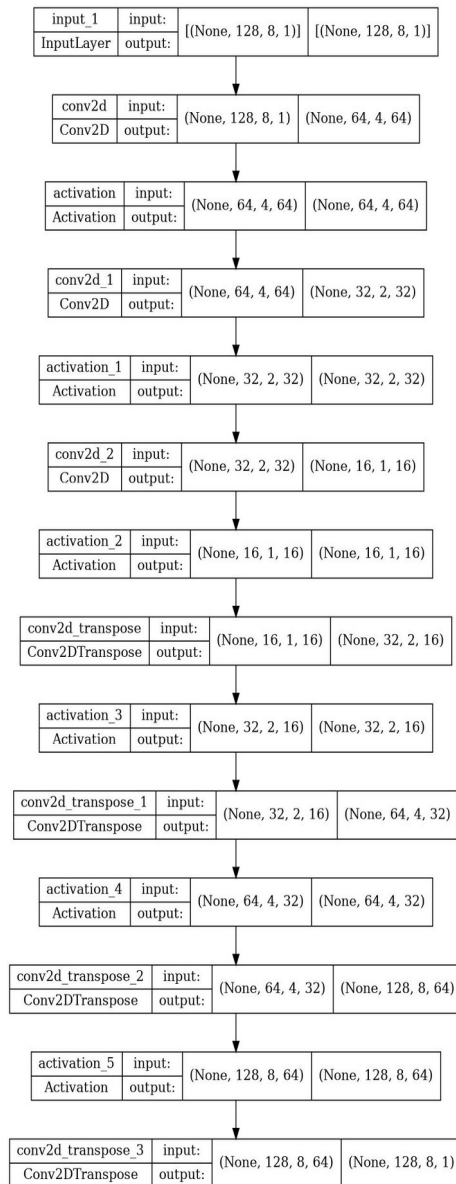
Το μοντέλο του αυτόματου κωδικοποιητή του χρησιμοποιήσαμε στο πρώτο πείραμα μας βασίστηκε στο βασικό (baseline) σύστημα που συνοδεύει το MIMII σύνολο δεδομένων. Η εκδοχή του χρησιμοποιήσαμε έχει βελτιωμένα αποτελέσματα στα AUC/pAUC μετρικά κατά 1-5% από το baseline μοντέλο.



Εικόνα 9: Μοντέλο - βαθύς αυτόματος κωδικοποιητής

Αποτελείται από βαθύ αυτόματο κωδικοποιητή (deep autoencoder) με πλήρως συνδεδεμένα επίπεδα που έχει ένα σχήμα που μοιάζει στο αγγλικό γράμμα U. Ξεκινάει με ένα κρυφό επίπεδο που έχει 256 μονάδες που ακολουθούνται από επίπεδα με 128-64-32-16 και 12 μονάδες για τον κώδικα στην άκρη του κωδικοποιητή (bottleneck). Αντίστοιχα ακολουθεί ένας αποκωδικοποιητής με τα εξής επίπεδα: 16-32-64-128-256. Η συνάρτηση για τον υπολογισμό του σφάλματος ανακατασκευής που είναι και το σκορ για την ανίχνευση ή όχι ανωμαλίας στον ήχο είναι το μέσο τετραγωνικό σφάλμα. Ο αλγόριθμος για την βελτιστοποίηση είναι ο Adam και για την αξιολόγηση του ταξινομητή χρησιμοποιήθηκε η επιφάνεια κάτω από την καμπύλη ROC (ROC-AUC).

Στο δεύτερο πείραμα ο συνελκτικός αυτόματος κωδικοποιητής έχει 3 συνελκτικά επίπεδα με φίλτρα (64, 32, 32) και kernel (5x5) με βήμα (stride) 2. Το φασματογράφημα είχε τις ίδιες ρυθμίσεις με τους απλούς αυτόματους κωδικοποιητές αλλά ενώσαμε 8 τμήματα και τη δημιουργία των χαρακτηριστικών.



Εικόνα 10: Μοντέλο - Συνελικτικός αυτόματος κωδικοποιητής

Το τρίτο πείραμα έγινε με τους mfcc συντελεστές ως χαρακτηριστικά και με νευρωνικό δίκτυο με ρυθμίσεις ίδιες με το πρώτο πείραμα. Στους ανεμιστήρες και στις αντλίες χρησιμοποιήσαμε 80 συντελεστές mfcc ενώ για τις ράγες κύλισης και τις βαλβίδες 40 συντελεστές mfcc.

Το τέταρτο πείραμα έγινε με τους mfcc συντελεστές μαζί με το 4p-MSD ως χαρακτηριστικά και με νευρωνικό δίκτυο και ρυθμίσεις ίδιες με το πρώτο πείραμα. Για το τους ανεμιστήρες και τις

αντλίες χρησιμοποιήσαμε 100 mfcσ συντελεστές ενώ για τις ράγες κύλισης και τις βαλβίδες 80 συντελεστές και ο αριθμός των φασματικών κορυφών ήταν 10.

4.3 Αποτελέσματα

Από την εκτέλεση των μοντέλων είχαμε τα εξής αποτελέσματα:

Πίνακας 1: Αποτελέσματα από την εκτέλεση με τον αυτόματο κωδικοποιητή με U σχήμα

ID	Ανεμιστήρας		Αντλία		Ράγες κύλισης		Βαλβίδες	
	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC
0	53.95	49.47	67.51	56.22	95.35	77.03	71.92	53.08
2	76.44	58.15	60.66	56.69	79.56	59.71	67.7	52.03
4	60.43	51.95	95.86	82.17	92.44	66.88	76.03	52
6	87.11	63.68	73.85	57.8	69.42	49.32	60.92	48.86

Έγιναν 4 ανεξάρτητες εκτελέσεις. Το αποτέλεσμα είναι ο μέσος όρος.

Πίνακας 2: Αποτελέσματα από τον συνελκτικό αυτόματο κωδικοποιητή

ID	Ανεμιστήρας		Αντλία		Ράγες κύλισης		Βαλβίδες	
	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC
0	52.70	48.70	55.26	51.05	94.80	76.39	70.24	54.36
2	75.42	53.11	69.78	55.26	80.85	58.55	87.58	60.77
4	54.06	53.22	98.78	93.82	96.80	83.26	77.03	51.79
6	87.50	62.31	72.66	52.00	82.38	53.44	58.97	48.52

Έγιναν 4 ανεξάρτητες εκτελέσεις. Το αποτέλεσμα είναι ο μέσος όρος.

Πίνακας 3: Αποτελέσματα από τη χρήση των mfcc συντελεστών με τον U αυτόματο κωδικοποιητή

ID	Ανεμιστήρας		Αντλία		Ράγες κύλισης		Βαλβίδες	
	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC
0	54.07	49.43	69.78	51.59	94.06	72.28	73.44	51.66
2	78.78	55.95	66.77	59.62	77.36	58.30	81.30	53.71
4	61.91	52.43	92.83	76.54	95.67	77.31	83.47	51.42
6	82.58	60.86	79.79	59.84	77.74	49.69	63.74	49.42

Έγιναν 4 ανεξάρτητες εκτελέσεις. Το αποτέλεσμα είναι ο μέσος όρος.

Πίνακας 4: Αποτελέσματα από τη χρήση των mfcc συντελεστών μαζί με το 4p-MSD με τον U αυτόματο κωδικοποιητή

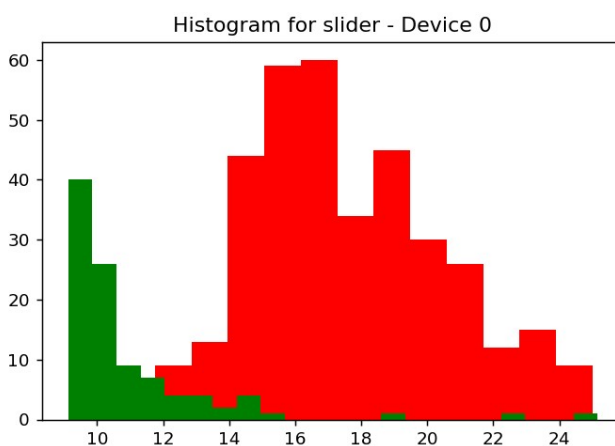
ID	Ανεμιστήρας		Αντλία		Ράγες κύλισης		Βαλβίδες	
	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC
0	69.81	59.35	77.42	61.81	98.00	91.12	68.91	52.10
2	72.90	51.56	67.79	58.61	73.53	50.58	68.41	50.77
4	59.74	50.71	79.91	62.77	92.01	69.43	70.19	49.84
6	85.99	63.68	73.89	60.87	54.27	50.33	52.57	49.08

Έγιναν 4 ανεξάρτητες εκτελέσεις. Το αποτέλεσμα είναι ο μέσος όρος.

Πίνακας 5: Συγκριτικά αποτελέσματα με το baseline σύστημα

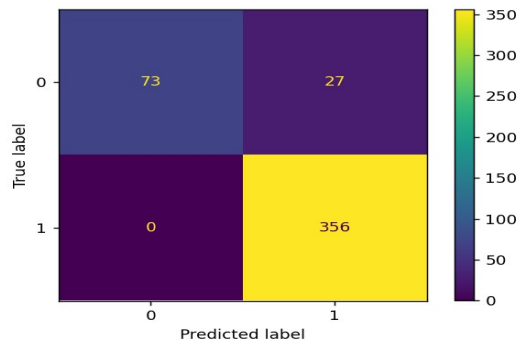
Τύπος μηχανής	Μέσος όρος AUC					Μέσος όρος pAUC				
	Baseline	U	Conv	mfcc	mfcc 4pMSD	Baseline	U	Conv	mfcc	mfcc 4pMSD
Ανεμιστήρας	65.83	69.48	67.42	69.33	72.11	52.45	55.81	54.33	54.67	56.57
Αντλία	72.89	74.47	72.66	77.3	74.75	59.99	63.22	63.03	61.9	61.01
Ράγες κύλισης	84.76	84.19	88.7	86.21	79.45	66.53	63.23	67.91	64.4	65.36
Βαλβίδες	66.28	69.14	73.45	75.49	65.36	50.98	51.49	53.86	51.55	50.45
Μέσος όρος	72.44	74.32	75.55	77.08	72.91	57.48	58.43	59.78	58.13	58.34

Για την περίπτωση του μηχανήματος με κωδικό 0 από την κατηγορία ράγες κύλισης έχουμε πολύ καλά αποτελέσματα. Από το ιστόγραμμα των σκορ βλέπουμε ότι ο αυτόματος κωδικοποιητής έχει καταφέρει να ξεχωρίσει τα ομαλά από τα μη ομαλά με εμφανή καλό τρόπο.



Εικόνα 11: Ιστόγραμμα με τα σκορ ανακατασκευής για το μηχανήμα 0 από τις ράγες κύλισης

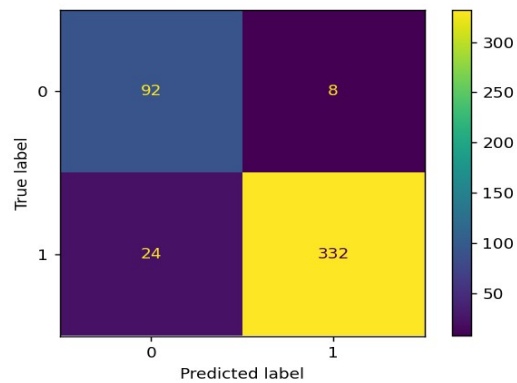
Επιλέγοντας σαν κατώφλι στο σκορ ανακατασκευής το 11 τα αποτελέσματα που έχει στην ανίχνευση ο αυτόματος κωδικοποιητής είναι:



Εικόνα 12: Confusion matrix για μικρό κατώφλι στο σκορ ανακατασκευής

Σε αυτό μπορούμε να δούμε ότι όλα τα 356 μη ομαλά έχουν ανιχνευθεί (true positive) ενώ από τους 100 ομαλούς ήχους έχουμε να ανιχνεύονται ψευδώς 27 ως μη ομαλοί (false positive).

Επειδή στα συστήματα ανίχνευσης ανωμαλιών δεν θέλουμε να έχουμε πολλά ψευδώς θετικά μπορούμε να ορίσουμε πιο μεγάλο κατώφλι με στόχο την μείωση τους. Έτσι αν θέσουμε ως κατώφλι το 14 στο σκορ ανακατασκευής έχουμε τα εξής αποτελέσματα:



Εικόνα 13: Confusion matrix για μεγάλο κατώφλι στο σκορ ανακατασκευής

Σε αυτή την περίπτωση μπορούμε να δούμε ότι το μοντέλο ανιχνεύει 332 από τα 356 σωστά (true positives) και 24 μη ομαλά του ξεφεύγουν κατά την ανίχνευση (false negative). Έχει μόνο 8 όμως από τα 100 ομαλά που τα έχει ανιχνεύσει ως μη ομαλά (false positive) πράγμα που είναι σημαντικό σε πολλές εφαρμογές.

5 Συμπεράσματα

Για την περίπτωση των μηχανημάτων που δεν έχουν σταθερά μοτίβα στα ομαλά ηχητικά σήματα τα αποτελέσματα δεν είναι πολύ καλά και είναι δύσκολη η ανίχνευση ανωμαλιών στην λειτουργία. Αντίθετα στην περίπτωση των μηχανημάτων όπου υπάρχουν σταθερά (stationary) μοτίβα στα ηχητικά σήματα η ανίχνευση ανωμαλιών με χρήση του αυτόματου κωδικοποιητή είναι αρκετά αποτελεσματική. Με τη χρήση αυτόματου κωδικοποιητή με μεγαλύτερο μέγεθος στις παραμέτρους ήταν εφικτή η βελτίωση του baseline συστήματος. Το ίδιο ήταν εφικτό με τη χρήση συνελκτικού αυτόματου κωδικοποιητή αλλά και τη χρήση των mfcc συντελεστών αλλά και σε συνδυασμό με το 4p-MSD.

5.1 Επεκτάσεις

Σε κάποιους τύπους μηχανημάτων η AUC μειώνεται γρήγορα όταν τα επίπεδα του θορύβου αυξάνουν. Αυτό δείχνει ότι η αντιμετώπιση του θορύβου πριν από την ανίχνευση των ανωμαλιών πιθανά να βελτιώνει τα αποτελέσματα. Άλλοι τύποι αυτόματων κωδικοποιητών όπως είναι ο Variational Autoencoder που μαθαίνει την κατανομή των παραδειγμάτων στον κωδικοποιητή και δειγματίζει από την κατανομή αυτή στον αποκωδικοποιητή, είναι μια επέκταση που θα μπορούσε ίσως να βελτιώσει τα αποτελέσματα.

Τέλος η χρήση προ-εκπαιδευμένων νευρωνικών δικτύων σε μεγάλα σύνολα δεδομένων όπως είναι το imagenet και εξαγωγή χαρακτηριστικών (transfer learning) από τα φασματογραφήματα και χρήση αυτών των χαρακτηριστικών με convolutional autoencoders είναι μια άλλη κατεύθυνση που πιθανά θα βελτιώνει τα αποτελέσματα.

Βιβλιογραφία, παραπομπές

- 1: Peter Norvig, Stuart J. Russell, Artificial Intelligence: A Modern Approach, 2020
- 2: Goodfellow et al, Deep learning, 2016
- 3: Michell Tom, Machine Learning, 1997
- 4: , Methode generale pour la resolution des systemes d'equations, 1847
- 5: Chandola et al., Anomaly Detection: A Survey, 2009
- 6: Theodoros Gianakopoulos, Aggelos Pikrakis, Introduction to Audio Analysis, 2014
- 7: Mrinmoy Bhattacharjee, Mahadeva Prasanna, Prithwjit Guha , Speech/Music Classification Using Features From Spectral Peaks, 2020
- 8: Christopher M. Bishop, Pattern Recognition and Machine Learning, 2006
- 9: David Rumelhart, Geoffrey Hinton and Ronald Williams, Learning representations by back-propagating errors, 1986
- 10: Sridhar Alla, Suman Kalyan Adari, Beginning Anomaly Detection Using Python-Based Deep Learning, 2019
- 11: Koizumi et al, Unsupervised anomalous sound detection for machine condition monitoring, 2020
- 12: Diederik Kingma and Jimmy Ba, Adam: A method for stochastic optimization, 2015
- 13: Charles Metz, Basic Principles of ROC Analysis, 1978
- 14: Harsh Purohit, Ryo Tanabe, Kenji Ichige, Takashi Endo, Yuki Nikaido, Kaori Suefusa, and Yohei Kawaguchi, MIMII Dataset: Sound dataset for malfunctioning industrial machine investigation and inspection, 2019