

Machine Learning for Anomaly Detection

Νομικός Γεώργιος

AEM: 9857

email: gnomikos@ece.auth.gr

Τσιμπλιαρίδης Νικόλαος

AEM: 9652

email: tenikola@ece.auth.gr

Σύνοψη — Η συγκεκριμένη εργασία πραγματοποιήθηκε στα πλαίσια του μαθήματος “Αξιοπιστία Συστημάτων” και είχε ως στόχο να παραθέσει παραδείγματα της χρήσης μοντέλων Machine Learning για την ανίχνευση ανωμαλιών. Πιο συγκεκριμένα τα παραδείγματα που μελετήθηκαν αφορούν δύο πολύ διαφορετικούς τομείς, με το ένα να αναλύει τον τρόπο λειτουργίας ενός μοντέλου Sparse Autoencoder για την ανίχνευση Σχιζοφρένειας, ενώ το δεύτερο μελετά την περίπτωση ανίχνευσης καταστροφικών συμβάντων σε επιταχυντές σωματιδίων.

I. Εισαγωγή

A. Τι είναι Anomaly Detection:

Με τον όρο “Anomaly detection” (“Ανίχνευση Ανωμαλιών”), αναφερόμαστε στο «πρόβλημα του εντοπισμού μοτίβων/προτύπων σε δεδομένα που δεν εμφανίζουν την αναμενόμενη συμπεριφορά. Για την επίλυση τέτοιου είδους προβλημάτων υπάρχει μία πληθώρα μεθόδων, οι οποίες έχουν αναπτυχθεί για ένα τεράστιο εύρος εφαρμογών που περιλαμβάνουν από την ανάλυση βιοϊατρικών και οικονομικών δεδομένων μέχρι την ασφάλεια δικτύων και την ανίχνευση σφαλμάτων σε αεροσκάφη

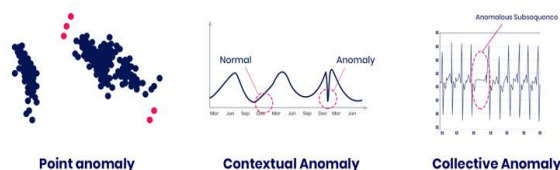
Είδη ανωμαλιών:

Ανωμαλία ορίζεται ως ένα δεδομένο ή ένα μοτίβο από δεδομένα των οποίων η συμπεριφορά είναι διαφορετική από την αναμενόμενη. Μπορούμε να την χωρίσουμε σε τρεις διαφορετικές κατηγορίες:

1. Σημειακές ανωμαλίες (Point anomalies): Θεωρείται η απλούστερη μορφή ανωμαλίας και αφορά τις περιπτώσεις που ένα μόνο δεδομένο ακολουθεί απρόσμενη συμπεριφορά.
2. Σχετικές ανωμαλίες (Contextual anomalies): Θεωρούνται οι ανωμαλίες, οι οποίες υπό διαφορετικές συνθήκες θα ήταν φυσιολογικές, αλλά δεν έχουν θέση στο σημείο που βρίσκονται. Για παράδειγμα μια τιμή

θερμοκρασίας 0 βαθμοί Κελσίου θα ήταν φυσιολογική τιμή για τον Χειμώνα, αλλά το καλοκαίρι θα θεωρούνταν contextual anomaly.

3. Συλλογικές ανωμαλίες (Collective anomalies): Αυτός ο τύπος ανωμαλίας αναφέρεται σε ένα σύνολο δεδομένων που σταδιακά εμφανίζουν διαφορετικό μοτίβο από τα κανονικά δεδομένα.
4. Άλλου είδους ανωμαλίες: οι οποίες έχουν να κάνουν με τις ιδιαιτερότητες που εμφανίζουν τα δεδομένα μίας συγκεκριμένης εφαρμογής.



Σχήμα 1. Είδη ανωμαλιών

Οι “παράδοσιακοί” τρόποι ανίχνευσης ανωμαλιών αποτελούνταν κυρίως από στατιστικές μεθόδους και μοντέλα. Τα τελευταία χρόνια όμως αρχίζουν να αναπτύσσονται όλο και περισσότερες μέθοδοι μέσω της χρήσης machine learning (μηχανικής μάθησης), οι οποίες χωρίζονται σε τρεις αρκετά γενικές κατηγορίες:

- Supervised Anomaly Detection: Σε αυτήν την κατηγορία υπάρχουν δυο ξεχωριστά training datasets για κανονικά και μη, τα οποία περιέχουν labeled δεδομένα. Σε αυτήν την κατηγορία, φτιάχνουμε ένα μοντέλο για πρόβλεψη στα κανονικά δεδομένα και ένα για πρόβλεψη στα δεδομένα που είναι classified ως ανώμαλα. Στη συνέχεια συγκρίνουμε τα δύο αυτά μοντέλα. Ωστόσο, έχουμε δύο κύρια προβλήματα: πρώτον, ο αριθμός των ανωμαλιών στο σετ εκπαίδευσης είναι πολύ χαμηλότερος σε σχέση με τις κανονικές. Δεύτερον, είναι δύσκολο να βρεθούν οι κατάλληλες ετικέτες, ιδιαίτερα για τις ανωμαλίες.
- Semi-supervised Anomaly Detection: Τα training datasets εδώ περιλαμβάνουν μόνο κλάσεις για τα κανονικά δεδομένα και οτιδήποτε δεν ανήκει σε κάποια από τις κλάσεις των κανονικών δεδομένων γίνεται classify σαν ανωμαλία. Επειδή δεν χρειάζονται κλάσεις για τις ανωμαλίες χρησιμοποιούνται πολύ πιο συχνά από τα supervised.
- Unsupervised Anomaly Detection: Δεν χρειάζονται labeled training datasets, αλλά για να χρησιμοποιηθεί αυτή η μέθοδος και να έχει

σωστά αποτελέσματα θέλουμε τα δεδομένα μας να περιέχουν πολύ περισσότερα φυσιολογικά δεδομένα απ' ό,τι ανωμαλίες. Πολλές semi-supervised μέθοδοι μπορούν να χρησιμοποιηθούν σαν unsupervised αν χρησιμοποιηθούν training datasets χωρίς labeled δεδομένα.

II. Medical Applications

To deep learning γίνεται ολοένα και πιο δημοφιλής για την ανίχνευση βιοϊατρικών ανωμαλιών για διάφορους λόγους:

- Μη γραμμικότητα: Τα μοντέλα βαθιάς μάθησης έχουν τη δυνατότητα να μοντελοποιούν μη γραμμικές σχέσεις στα δεδομένα, οι οποίες μπορούν να διαχωρίσουν καλύτερα κανονικά και ανώμαλα δείγματα και να μοντελοποιήσουν καλύτερα τις ασυνέπειες στα δεδομένα.
- Αυτόματη εκμάθηση χαρακτηριστικών: Τα μοντέλα βαθιάς μάθησης έχουν τη δυνατότητα να μαθαίνουν αυτόματα χαρακτηριστικά από τα δεδομένα, αποφεύγοντας την ανάγκη ρητού καθορισμού του τι συνιστά ανωμαλία.
- Χειρισμός μεγάλων δεδομένων: Με τη διαθεσιμότητα μεγάλων δεδομένων και αυξημένους υπολογιστικούς πόρους, τα μοντέλα βαθιάς μάθησης μπορούν να χειριστούν μεγάλους όγκους δεδομένων και να εκτελέσουν ιεραρχική εκμάθηση χαρακτηριστικών, η οποία μπορεί να βελτιώσει την απόδοση του μοντέλου.
- Χειρισμός πολύπλοκων δεδομένων: Τα ιατρικά δεδομένα έρχονται συχνά με τη μορφή εικόνων, χρονοσειρών και σημάτων, τα οποία μπορεί να είναι πολύπλοκα και δύσκολο να μοντελοποιηθούν. Τα μοντέλα βαθιάς μάθησης, με την ικανότητά τους να χειρίζονται πολύπλοκα δεδομένα και να εξάγουν σημαντικές συμπεράσματα.

Διάφορα παραδείγματα εφαρμογών του είναι η ανίχνευση ανωμαλιών σε εικόνες MRI, δεδομένα από ενδοσκόπηση, καρδιακούς ήχους αλλά και η πρόβλεψη επιληπτικών κρίσεων.

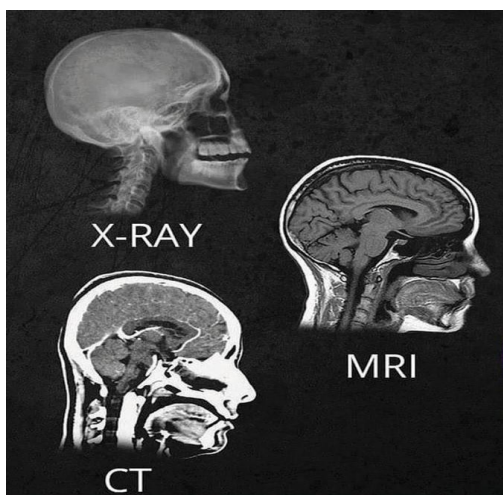
A. Είδη βιοϊατρικών δεδομένων

1) Βιοϊατρικές εικόνες:

- Ακτινογραφία ακτίνων X: Οι ακτίνες X έχουν μικρότερο μήκος κύματος από το ορατό φως και μπορούν να περάσουν από τους περισσότερους τύπους ιστών στο ανθρώπινο σώμα. Ωστόσο, το ασβέστιο που περιέχεται στα οστά είναι πιο πυκνό και διασκορπίζει τις ακτινογραφίες. Το φιλμ που βρίσκεται στην αντίθετη πλευρά της πηγής ακτίνων X είναι κατασκευασμένο έτσι ώστε οι περιοχές που εκτίθενται σε περισσότερο φως φαίνονται πιο σκούρες και αυτές που εκτίθενται σε λιγότερο φως πιο ανοιχτόχρωμες. Επομένως, καθώς περισσότερες ακτίνες X διεισδύουν σε ιστούς όπως οι πνεύμονες, αυτές οι περιοχές σκουραίνουν στο φιλμ και τα οστά εμφανίζονται ως φωτεινότερες περιοχές. Η απεικόνιση με ακτίνες X συνήθως χρησιμοποιείται για διάφορους διαγνωστικούς σκοπούς, συμπεριλαμβανομένης της ανίχνευσης καταγμάτων οστών, οδοντικών προβλημάτων, πνευμονίας και ορισμένων τύπων όγκων.
- Αξονική τομογραφία (CT): Στην αξονική τομογραφία, δημιουργούνται εικόνες διατομής του σώματος χρησιμοποιώντας μια στενή δέσμη ακτίνων X που εκπέμπονται ενώ ο ασθενής περιστρέφεται. Με αυτόν τον τρόπο αξιοποιούνται πολλές διαφορετικές εικόνες διατομής για να δημιουργήσουν μια τρισδιάστατη αναπαράσταση του ιστού, η οποία είναι πιο κατατοπιστική από μια συμβατική εικόνα ακτίνων X. Οι αξονικές τομογραφίες είναι ένα δημοφιλές διαγνωστικό εργαλείο κατά τον εντοπισμό ασθένειας ή τραυματισμού σε διάφορες περιοχές του σώματος. Οι εφαρμογές περιλαμβάνουν την ανίχνευση όγκων ή βλαβών στην κοιλιά και τον εντοπισμό τραυματισμών, όγκων και θρόμβων.
- Μαγνητική τομογραφία (MRI): Η μαγνητική τομογραφία χρησιμοποιεί ένα μαγνητικό πεδίο, αναγκάζοντας τα πρωτόνια στο σώμα να ευθυγραμμιστούν με αυτό. Συγκεκριμένα, τα πρωτόνια στο ανθρώπινο σώμα περιστρέφονται και δημιουργούν ένα μικρό μαγνητικό πεδίο. Όταν εισάγεται ένα εξωτερικό ισχυρό μαγνητικό πεδίο όπως από τη μηχανή MRI, τα πρωτόνια ευθυγραμμίζονται με αυτό το πεδίο. Στη συνέχεια εισάγεται ένας παλμός ραδιοσυχνότητας που διακόπτει την ευθυγράμμιση. Όταν ο παλμός ραδιοσυχνότητας απενεργοποιείται, τα

πρωτόνια προσπαθούν να ευθυγραμμιστούν εκ νέου με το μαγνητικό πεδίο. Η ενέργεια που απελευθερώνεται ποικίλλει για διαφορετικούς τύπους ιστών, επιτρέποντας στη μαγνητική τομογραφία να διαχωρίσει διαφορετικές περιοχές. Ως εκ τούτου, οι μαγνητικές τομογραφίες συνήθως χρησιμοποιούνται για την απεικόνιση περιοχών μη οστών ή μαλακών ιστών του ανθρώπινου σώματος. Συγκριτικές μελέτες έχουν δείξει ότι ο εγκέφαλος, ο νωτιαίος μυελός, τα νεύρα και οι μύες συλλαμβάνονται καλύτερα με μαγνητική τομογραφία παρά με αξονική τομογραφία. Ως εκ τούτου, η μαγνητική τομογραφία είναι η μέθοδος επιλογής για εργασίες όπως η ανίχνευση όγκων εγκεφάλου και ο εντοπισμός της βλάβης των ιστών.

- Εκτός από αυτές τις δημοφιλείς κατηγορίες αισθητήρων βιοϊατρικής απεικόνισης, υπάρχουν και άλλες κοινές πηγές δεδομένων όπως η Τομογραφία Εκπομπής Ποζιτρονίων (PET), ο Υπέρηχος και άλλα.



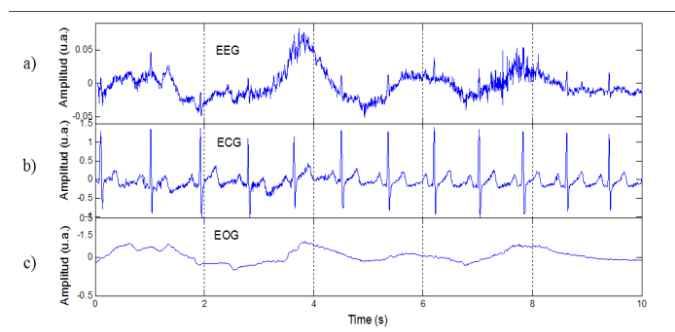
Σχήμα 2. Βιοϊατρικές εικόνες

2) Βιοϊατρικά σήματα:

- Ηλεκτροκαρδιογράφημα (ECG): Το ECG είναι ένα εργαλείο για την απεικόνιση του ηλεκτρικού ρεύματος που ρέει μέσω της καρδιάς και δημιουργεί τον καρδιακό παλμό ξεκινώντας από την κορυφή της καρδιάς και ταξιδεύοντας προς τα κάτω. Σε ηρεμία, τα καρδιακά κύτταρα είναι αρνητικά φορτισμένα σε σύγκριση με το εξωτερικό περιβάλλον και όταν εκπολώνονται φορτίζονται θετικά. Η διαφορά στην πόλωση αποτυπώνεται από το ECG. Υπάρχουν δύο τύποι πληροφοριών που

μπορούν να εξαχθούν με την ανάλυση του ECG. Πρώτον, μετρώντας τα χρονικά διαστήματα σε ένα ECG μπορεί κανείς να ελέγξει για ακανόνιστες ηλεκτρικές δραστηριότητες. Δεύτερον, η ισχύς της ηλεκτρικής δραστηριότητας παρέχει μια ένδειξη των περιοχών της καρδιάς που είναι υπερβολικά καταπονημένες.

- Ηλεκτροεγκεφαλογράφημα (EEG): Το EEG ανιχνεύει την ηλεκτρική δραστηριότητα στον εγκέφαλο, ο οποίος χρησιμοποιεί ηλεκτρικά ερεθίσματα μέσω νευρώνων. Για την αποτύπωση της ηλεκτρικής δραστηριότητας, τοποθετούνται μικροί μεταλλικοί δίσκοι (ηλεκτρόδια) στο τριχωτό της κεφαλής. Τα ηλεκτρικά σήματα που συλλαμβάνονται από αυτά τα ηλεκτρόδια ενισχύονται για την καλύτερη οπτικοποίηση της εγκεφαλικής δραστηριότητας. Συχνά χρησιμοποιούνται για τη μελέτη των προτύπων ύπνου, ψυχολογικών διαταραχών, εγκεφαλικών βλαβών από τραυματισμό στο κεφάλι και επιληψία.
- Μαγνητοεγκεφαλογραφία (MEG): Όπως περιγράφηκε παραπάνω, ένα EEG συλλαμβάνει τα ηλεκτρικά πεδία που δημιουργούνται από τα εξωκυτταρικά ρεύματα του ανθρώπινου εγκεφάλου, ενώ η MEG ανιχνεύει κυρίως τα μαγνητικά πεδία που προκαλούνται από αυτά τα εξωκυτταρικά ρεύματα.
- Ηλεκτρομυογραφία (EMG) όπου παρακολουθείται το ηλεκτρικό δυναμικό που παράγεται από μυϊκά κύτταρα για τη διάγνωση της υγείας των μυών και των κινητικών νευρώνων.



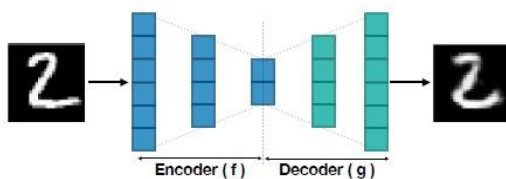
Σχήμα 3. Βιοϊατρικά σήματα

B. Μέθοδοι Machine Learning που χρησιμοποιούνται για Anomaly Detection

Auto-Encoders (AE):

Ένας ΑΕ είναι ένας τύπος νευρωνικού δικτύου που εκπαιδεύεται να ανακατασκευάζει τα δεδομένα εισόδου του. Αποτελείται συνήθως από τρία κύρια μέρη:

- Ο κωδικοποιητής: Αυτό είναι το τμήμα του δικτύου που αντιστοιχίζει τα δεδομένα εισόδου σε μια αναπαράσταση χαμηλότερης διάστασης, που ονομάζεται bottleneck ή λανθάνουσα αναπαράσταση. Ο κωδικοποιητής τυπικά αποτελείται από πολλά στρώματα νευρώνων που επεξεργάζονται τα δεδομένα εισόδου και μειώνουν τις διαστάσεις τους.
- Το σημείο συμφόρησης ή η λανθάνουσα αναπαράσταση: Αυτό είναι το μέρος του δικτύου όπου τα δεδομένα εισόδου συμπιέζονται σε μια αναπαράσταση χαμηλότερης διάστασης. Η αναπαράσταση του bottleneck καταγράφει τα πιο σημαντικά χαρακτηριστικά ή μοτίβα στα δεδομένα.
- Ο αποκωδικοποιητής: Αυτό είναι το τμήμα του δικτύου που αντιστοιχίζει την αναπαράσταση του bottleneck πίσω στον αρχικό χώρο εισόδου. Ο αποκωδικοποιητής τυπικά αποτελείται από πολλά στρώματα νευρώνων που «αποκωδικοποιούν» τη συμπιεσμένη αναπαράσταση και ανασυνθέτουν την αρχική είσοδο. Ο στόχος του αποκωδικοποιητή είναι να παράγει μια έξοδο που να είναι όσο το δυνατόν παρόμοια με την αρχική είσοδο.

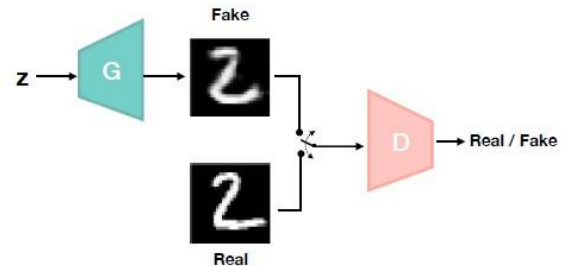


Σχήμα 4. Auto-Encoder

Κατά τη διάρκεια της εκπαίδευσης, ο ΑΕ παρουσιάζεται με ένα σύνολο δειγμάτων εισόδου και ο στόχος είναι να ρυθμιστούν τα βάρη του κωδικοποιητή και του αποκωδικοποιητή έτσι ώστε η έξοδος του αποκωδικοποιητή να είναι όσο το δυνατόν παρόμοια με την αρχική είσοδο. Η διαφορά μεταξύ της εισόδου και της εξόδου τυπικά μετριέται με μια συνάρτηση απώλειας όπως το μέσο τετράγωνο σφάλμα. Ελαχιστοποιώντας αυτή

την απώλεια, ο αυτόματος κωδικοποιητής μαθαίνει να αναδομεί τα δεδομένα εισόδου.

Generative Adversarial Networks (GANs):



Σχήμα 5. Generative Adversarial Networks

Τα Generative Adversarial Networks (GAN) είναι ένας τύπος μοντέλου βαθιάς εκμάθησης που χρησιμοποιείται για τη δημιουργία νέων δεδομένων που είναι παρόμοια με ένα σύνολο εισόδου. Τα GAN αποτελούνται από δύο κύρια μέρη: το generator network και το discrimination network

- ➔ Η αποστολή του generator είναι να δημιουργεί νέα δείγματα δεδομένων, παρόμοια με τα δεδομένα εισόδου. Αυτό το κάνει εισάγοντας με μη γραμμικό τρόπο μια είσοδο τυχαίου θορύβου στα δεδομένα.
- ➔ Η αποστολή του discrimination network είναι να προσδιορίσει εάν ένα δεδομένο δείγμα δεδομένων είναι πραγματικό (από το σύνολο δεδομένων εισόδου) ή ψεύτικο (που δημιουργείται από το generator).

Τα δύο δίκτυα εκπαιδεύονται μαζί σε μια αντίπαλη διαδικασία, όπου το ένα προσπαθεί να παράγει δείγματα που μπορούν να ξεγελάσουν το άλλο, ενώ το άλλο προσπαθεί να αναγνωρίσει σωστά τα ψεύτικα δείγματα. Καθώς η εκπαίδευση προχωρά, το generator γίνεται καλύτερο στην παραγωγή ρεαλιστικών δειγμάτων, ενώ η διάκριση γίνεται καλύτερη στην αναγνώριση ψεύτικων δειγμάτων. Τελικά, φτάνουμε σε σημείο το generator να παράγει δείγματα που δεν διακρίνονται από τα πραγματικά δεδομένα.

Recurrent Neural Networks (RNNs):

Το RNN είναι ένας τύπος νευρωνικού δικτύου που έχει σχεδιαστεί για την επεξεργασία “sequential

data”, όπως χρονοσειρές, κείμενο ή ομιλία. Τα RNN ονομάζονται "recurrent" επειδή χρησιμοποιούν συνδέσεις ανάδρασης, οι οποίες επιτρέπουν τη μετάδοση πληροφοριών από το ένα βήμα του δικτύου στο επόμενο.

Στα παραδοσιακά νευρωνικά δίκτυα τροφοδοσίας, η είσοδος περνά μέσα από το δίκτυο μόνο μία φορά και παράγει μια έξοδο, ενώ στα RNN, η είσοδος περνά μέσα από το δίκτυο πολλές φορές και η έξοδος εξαρτάται από την τρέχουσα είσοδο και τις προηγούμενες εξόδους. Αυτό επιτρέπει στα RNN να διατηρούν ένα είδος μνήμης των προηγούμενων εισόδων και να χρησιμοποιούν αυτή τη μνήμη για να ενημερώνουν την τρέχουσα έξοδο.

Συνελικτικά Νευρωνικά Δίκτυα (CNN):

Συνελικτικά Νευρωνικά Δίκτυα (CNN) είναι ένας τύπος νευρωνικών δικτύων βαθιάς μάθησης που είναι ιδιαίτερα κατάλληλος για εργασίες ανάλυσης εικόνας και βίντεο. Τα CNN αποτελούνται από πολλά επίπεδα, συμπεριλαμβανομένων των convolutional layers, των pooling layers και των fully connected layers.

- Τα convolutional layers είναι υπεύθυνα για την εκμάθηση τοπικών μοτίβων και χαρακτηριστικών στα δεδομένα εισόδου, όπως ακμές, σχήματα και υφές.
- Τα pooling layers είναι υπεύθυνα για τη μείωση των χωρικών διαστάσεων των δεδομένων
- Ενώ τα fully connected layers είναι υπεύθυνα για την τελική εργασία ταξινόμησης ή παλινδρόμησης.

Ένα από τα βασικά χαρακτηριστικά των CNN είναι η χρήση convolutional filters, τα οποία ολισθαίνουν πάνω από τα δεδομένα εισόδου και εξάγουν τοπικά χαρακτηριστικά.

Densely Connected Neural Network (DAN):

Densely Connected Neural Network (DAN) είναι ένας τύπος νευρωνικών δικτύων που χαρακτηρίζεται από πυκνή συνδεσιμότητα μεταξύ των layers. Με άλλα λόγια, κάθε νευρώνας σε ένα στρώμα συνδέεται με κάθε νευρώνα στο επόμενο στρώμα. Αυτό έρχεται σε αντίθεση με αρχιτεκτονικές όπως τα συνελικτικά νευρωνικά δίκτυα (CNN) όπου οι νευρώνες συνδέονται μόνο με ένα υποσύνολο νευρώνων στο επόμενο επίπεδο.

C. Ανίχνευση σχιζοφρένειας με χρήση Sparse Auto-encoders

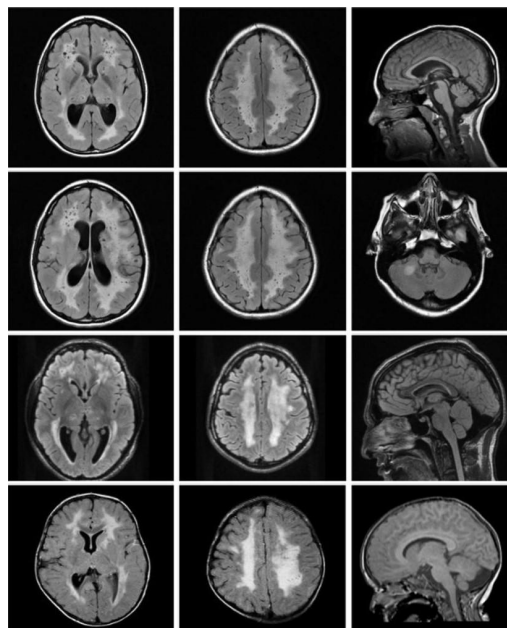
Λήψη δεδομένων

Η σχιζοφρένεια είναι μια σοβαρή ψυχική διαταραχή που επηρεάζει το πώς σκέφτεται, αισθάνεται και συμπεριφέρεται ένα άτομο. Χαρακτηρίζεται από μια σειρά συμπτωμάτων που μπορεί να περιλαμβάνουν παραισθήσεις, αυταπάτες, διαταραγμένη σκέψη και αλλαγές στη διάθεση και τη συμπεριφορά. Η ακριβής παθοφυσιολογία της σχιζοφρένειας δεν είναι πλήρως κατανοητή, αλλά πιστεύεται ότι περιλαμβάνει έναν συνδυασμό γενετικών, περιβαλλοντικών παραγόντων και χημείας του εγκεφάλου.

Στην συγκεκριμένη περίπτωση ελέγχεται η συσχέτιση της σχιζοφρένειας με την δομή του εγκεφάλου αφού διάφορες έρευνες έχουν δείξει ότι μπορεί να υπάρχουν δομικές και λειτουργικές αλλαγές στον εγκεφαλο ατόμων με σχιζοφρένεια. Αυτές οι αλλαγές περιλαμβάνουν μείωση του μεγέθους ορισμένων περιοχών του εγκεφάλου, όπως ο ιπόκαμπος και μείωση της δραστηριότητας ορισμένων νευρωνικών δικτύων.

Για την απεικόνιση του εγκεφάλου χρησιμοποιείται μαγνητική τομογραφία (MRI), η οποία είναι μια μη επεμβατική διαγνωστική τεχνική απεικόνισης που χρησιμοποιεί μαγνητικό πεδίο και ραδιοκύματα για την παραγωγή λεπτομερών εικόνων των εσωτερικών δομών του σώματος. Ο λόγος που χρησιμοποιείται MRI για την απεικόνιση του εγκεφάλου και του νωτιαίου μυελού είναι διότι μπορεί να παράγει λεπτομερείς εικόνες της λευκής και φαιάς ουσίας του εγκεφάλου, καθώς και του νωτιαίου μυελού και των νευρικών ριζών. Οι εικόνες που λαμβάνουμε συνεπώς από το MRI είναι αυτές που θα χρησιμοποιήσουμε σαν input στον auto-encoder για

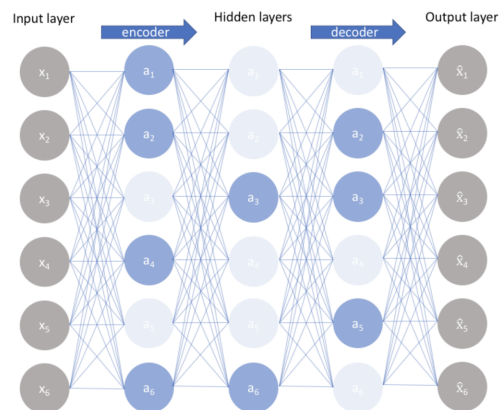
να τον εκπαιδεύσουμε.



Σχήμα 6. Εικόνες MRI

Unsupervised εκπαίδευση με Sparse Auto-Encoders (SAE)

Ο SAE είναι ένας τύπος AE που έχει εκπαιδευτεί για την ανακατασκευή των δεδομένων εισόδου. Η ειδοποιός διαφορά του με τον απλό AE είναι ότι δεν ενεργοποιούνται όλοι οι κόμβοι του νευρωνικού δικτύου των κρυφών layers σε κάθε επανάληψη. Ο περιορισμός αυτός σημαίνει ότι μόνο ένα υποσύνολο κρυφών κόμβων θα ενεργοποιηθεί για μια δεδομένη είσοδο. Αυτό επίσης συνεπάγεται ότι δεν είναι απαραίτητο οι κόμβοι στα hidden layers να είναι λιγότεροι από αυτούς στην είσοδο αρκεί κάθε φορά να επιλέγονται μόνο ορισμένοι από αυτούς. Οι εικόνες MRI συνεπώς που βάζουμε σαν είσοδο ανακατασκευάζονται με την χρήση του SAE ώστε να ελαχιστοποιείται το σφάλμα ανακατασκευής και οι πληροφορίες που παίρνουμε από το bottleneck μας βοηθάνε στο classification.

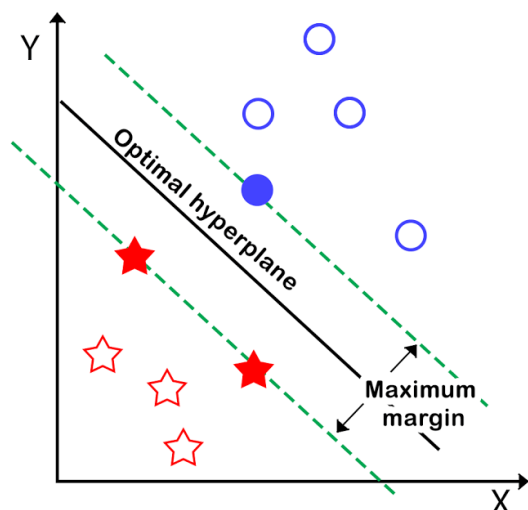


Σχήμα 7. Sparse Auto-Encoder

Classification with SVM

Ένα Support Vector Machine (SVM) είναι ένας τύπος εποπτευόμενου αλγόριθμου μάθησης που μπορεί να χρησιμοποιηθεί για εργασίες ταξινόμησης. Η κύρια ιδέα πίσω από το SVM είναι να βρεθεί το καλύτερο όριο (ή υπερεπίπεδο) που διαχωρίζει τις διαφορετικές κατηγορίες στα δεδομένα εκπαίδευσης. Αυτό το όριο επιλέγεται με τέτοιο τρόπο ώστε να μεγιστοποιείται το περιθώριο, το οποίο είναι η απόσταση μεταξύ του ορίου και των πλησιέστερων σημείων δεδομένων από κάθε κλάση. Αυτά τα πλησιέστερα σημεία δεδομένων ονομάζονται Support Vectors, εξ ου και η ονομασία Support Vector Machine.

Τα SVM μπορεί να είναι γραμμικά ή μη γραμμικά, ανάλογα με τη συνάρτηση πυρήνα που χρησιμοποιείται. Ένα γραμμικό SVM βρίσκει ένα γραμμικό όριο που χωρίζει τις κλάσεις, ενώ ένα μη γραμμικό SVM χρησιμοποιεί ένα “kernel trick” για να μετατρέψει τα δεδομένα σε έναν χώρο υψηλότερης διάστασης όπου ένα γραμμικό όριο μπορεί να διαχωρίσει τις κλάσεις.



Σχήμα 8. SVM Classifier

Στην περίπτωση των εικόνων MRI, αυτά τα χαρακτηριστικά μπορούν να προέρχονται από τις τιμές των pixels της εικόνας, όπως ο μέσος όρος, η τυπική απόκλιση ή τα ιστογράμματα των τιμών των pixels. Αυτά τα χαρακτηριστικά χρησιμοποιούνται στη συνέχεια για την εκπαίδευση ενός ταξινομητή SVM, ο οποίος στη συνέχεια να χρησιμοποιείται για την ταξινόμηση νέων εικόνων με βάση τα χαρακτηριστικά τους.

D. Interpretation of ML models

Η ερμηνεία της μηχανικής μάθησης είναι σημαντική στις ιατρικές εφαρμογές για διάφορους λόγους:

1. Ασφάλεια και ηθική: Οι ιατρικές αποφάσεις έχουν συχνά σοβαρές συνέπειες για την υγεία των ασθενών και είναι σημαντικό να διασφαλιστεί ότι τα μοντέλα μηχανικής μάθησης που χρησιμοποιούνται στην υγειονομική περίθαλψη κάνουν ακριβείς και ασφαλείς προβλέψεις.
2. Κατανόηση της διαδικασίας λήψης αποφάσεων του μοντέλου: Στις ιατρικές εφαρμογές, είναι συχνά σημαντικό να κατανοήσουμε το σκεπτικό πίσω από τις προβλέψεις ενός μοντέλου. Για παράδειγμα, σε μια διαγνωστική εργασία, η κατανόηση των χαρακτηριστικών των δεδομένων εισόδου που χρησιμοποιεί το μοντέλο για να κάνει την πρόβλεψή του μπορεί να βοηθήσει τους γιατρούς να κατανοήσουν την υποκείμενη παθολογία της νόσου.
3. Διαφάνεια και Επεξήγηση: Οι ιατροί, οι ασθενείς και οι ρυθμιστικοί φορείς μπορεί να απαιτούν σαφή κατανόηση του τρόπου με τον οποίο ένα μοντέλο κάνει τις προβλέψεις του, προκειμένου να οικοδομηθεί εμπιστοσύνη και αποδοχή της τεχνολογίας.
4. Βελτίωση της απόδοσης του μοντέλου: Η κατανόηση του τρόπου με τον οποίο ένα μοντέλο κάνει τις προβλέψεις του μπορεί να βοηθήσει στον εντοπισμό περιοχών όπου το μοντέλο έχει κακή απόδοση και στην πραγματοποίηση προσαρμογών για τη βελτίωση της απόδοσής του.

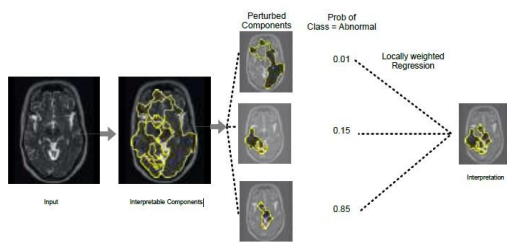
Interpretation methods

- Local Interpretable Model-agnostic Explanations(LIME):

Το Local Interpretable Model-agnostic Explanations (LIME) είναι μια τεχνική για την ερμηνεία των προβλέψεων σύνθετων μοντέλων μηχανικής μάθησης. Ονομάζεται "model-agnostic" γιατί μπορεί να εφαρμοστεί σε οποιοδήποτε τύπο μοντέλου, ανεξάρτητα από την αρχιτεκτονική ή την υλοποίησή του.

Το LIME λειτουργεί προσεγγίζοντας τη συμπεριφορά ενός σύνθετου μοντέλου γύρω από μια συγκεκριμένη πρόβλεψη, προσαρμόζοντας τοπικά ένα απλούστερο μοντέλο, όπως ένα γραμμικό μοντέλο ή ένα δέντρο αποφάσεων, στις προβλέψεις του μοντέλου σε ένα μικρό σύνολο σημείων δεδομένων εισόδου. Το απλούστερο μοντέλο, που ονομάζεται "τοπικό μοντέλο", εκπαιδεύεται στις εισόδους που είναι πιο παρόμοιες με τις εισόδους ενδιαφέροντος, όπως καθορίζεται από ένα μέτρο απόστασης.

Ο στόχος του LIME είναι να παρέχει ερμηνεύσιμες και κατανοητές εξηγήσεις των προβλέψεων του σύνθετου μοντέλου. Αυτό το επιτυγχάνει προσδιορίζοντας τα χαρακτηριστικά εισόδου που είναι πιο σημαντικά για την πρόβλεψη, όσον αφορά τα βάρη που τους αποδίδεται από το τοπικό μοντέλο. Αυτά τα χαρακτηριστικά μπορούν να γίνουν εύκολα κατανοητά από ειδικούς τομέα και μπορούν να χρησιμοποιηθούν για την κατανόηση της συμπεριφοράς του μοντέλου και τον εντοπισμό πιθανών σφαλμάτων ή προκαταλήψεων.



Σχήμα 9. LIME

- SHapley Additive exPlanations (SHAP):

Η SHapley Additive Explanations (SHAP) είναι μια μέθοδος για την ερμηνεία των προβλέψεων των μοντέλων μηχανικής μάθησης. Βασίζεται στην έννοια των τιμών Shapley από τη θεωρία παιγνίων, η οποία παρέχει έναν τρόπο δίκαιου διαμοιρασμού μιας τιμής σε άτομα με βάση τη συνεισφορά τους στην τιμή αυτή. Οι τιμές SHAP υπολογίζονται χρησιμοποιώντας έναν αλγόριθμο που εξετάζει όλους τους πιθανούς συνδυασμούς χαρακτηριστικών και τις αλληλεπιδράσεις τους. Ο αλγόριθμος λαμβάνει υπόψη τις εξαρτήσεις μεταξύ των χαρακτηριστικών, κάτι που είναι σημαντικό για μοντέλα όπως δέντρα αποφάσεων και τυχαία δάση, όπου η σημασία ενός χαρακτηριστικού μπορεί να εξαρτάται από τις τιμές άλλων χαρακτηριστικών. Οι τιμές SHAP μπορούν να παρέχουν πιο ακριβείς τιμές σημασίας χαρακτηριστικών επειδή λαμβάνουν υπόψη τις αλληλεπιδράσεις μεταξύ των χαρακτηριστικών.

III. Machine Learning ανίχνευση ανωμαλιών σε επιταχυντές σωματιδίων

A. Γενικά

Οι επιταχυντές σωματιδίων είναι πολύπλοκα συστήματα, αποτελούμενα από πολλαπλά υποσυστήματα, που πρέπει να συνεργαστούν μεταξύ τους, ώστε να παράξουν υψηλής ποιότητας ακτίνες. Οι ακτίνες αυτές χρησιμοποιούνται σε πειράματα φυσικής. Μια και μόνο δυσλειτουργία είναι αρκετή για να διαταράξει τη δέσμη ιόντων και να θέσει σε κίνδυνο το πείραμα, το οποίο συνήθως έχει ιδιαίτερα υψηλό κόστος.

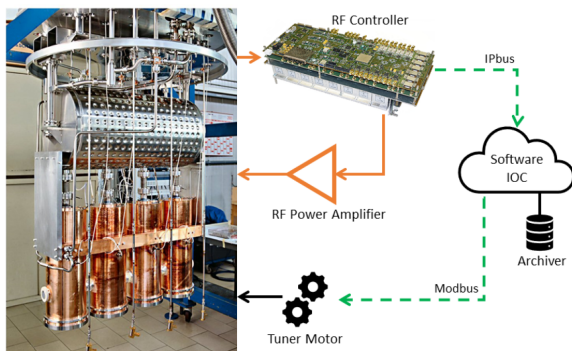
Δεδομένης της μεγάλης ποσότητας δεδομένων που παράγουν οι επιταχυντές και τα συστήματα ελέγχου τους, οι μέθοδοι που βασίζονται στην παρατήρηση των δεδομένων αυτών καθίστανται ιδιαίτερα βιώσιμοι ως προς την πρόληψη ενός συμβάντος που θα επηρεάσει την λειτουργία του επιταχυντή. Μάλιστα, η πρόοδος που έχει συντελεστεί στον τομέα του Machine Learning ευνοεί τις τεχνικές αυτές, οπότε, όλο και περισσότερο, βλέπουμε αυτές να εφαρμόζονται στα συστήματα ελέγχου των επιταχυντών σωματιδίων. Γενικά, η ανίχνευση ανωμαλιών σε επιταχυντές ανήκει σε μια ευρύτερη οικογένεια προβλημάτων διατήρησης της ορθής λειτουργίας ενός συστήματος, με κάποιες, βέβαια, ιδιαιτερότητες. Συγκεκριμένα, τα σφάλματα στην περίπτωση μας δεν προέρχονται από το ράγισμα ενός εξαρτήματος για παράδειγμα που πρέπει να φτιαχτεί, αλλά είναι απλά σφάλματα λειτουργίας που πρέπει να αποφευχθούν.

Ο επιταχυντής με τον οποίο θα ασχοληθούμε είναι ο “ALPI” (“Acceleratore Lineare per Ioni”, το οποίο στα αγγλικά μεταφράζεται ως “Linear Accelerator for Ions”) και βρίσκεται στο Legnaro National Laboratories στην Ιταλία. Ο εν λόγω επιταχυντής είναι βασισμένος σε υπεραγώγιμες κοιλότητες ραδιοσυχνότητας. Κάθε μία από αυτές τις κοιλότητες συνεισφέρει στην επιτάχυνση της δέσμης ιόντων, μεταφέροντας ενέργεια από την ισχύ των ραδιοσυχνοτήτων στα σωματίδια. Ειδικότερα, λόγω της γεωμετρίας των κοιλοτήτων, δημιουργείται ένα μαγνητικό πεδίο μιας συγκεκριμένης συχνότητας μέσα στις κοιλότητες, με αποτέλεσμα ένας αριθμός σωματιδίων να επιταχύνονται προς συγκεκριμένη κατεύθυνση, λόγω της δύναμης Lorentz. Ο επιταχυντής, προκειμένου να δώσει την απαιτούμενη ενέργεια στη δέσμη ιόντων, χρησιμοποιεί είκοσι κρυστάτες, καθένας εκ των οποίων αποτελείται από τέσσερις κοιλότητες.

Από την οπτική του ελέγχου του συστήματος αυτού, προκειμένου να είναι λειτουργική μία κοιλότητα θα πρέπει να βρίσκεται σε “locked state”, δηλαδή να έχει σταθερό πεδίο, συχνότητα και φάση. Για να το πετύχουμε αυτό, χρησιμοποιούνται δύο διαφορετικοί βρόχοι ελέγχου, όπως φαίνεται στο σχήμα 10. Ο πρώτος βρόχος υλοποιείται στον RF controller, δηλαδή ένα custom hardware με ένα FPGA και ένα analog front-end για την καταγραφή, διόρθωση και

παραγωγή του κατάλληλου σήματος. Ο βρόχος αυτός μπορεί να εφαρμόσει γρήγορες διορθώσεις σε πραγματικό χρόνο. Ο δεύτερος υλοποιείται με ένα software, το οποίο συλλέγει τις παραμέτρους λειτουργίας από τον RF controller και δύναται να ελέγξει τα μοτέρ του συστήματος.

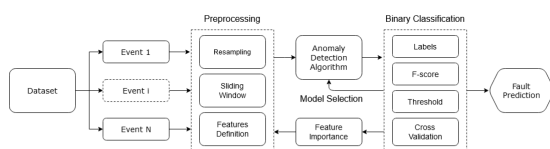
Στην περίπτωση που ένας από τους δύο βρόχους ελέγχου αποτύχει να διατηρήσει την ορθή λειτουργία της κοιλότητας, τότε έχουμε ένα “unlocked event”. Η λειτουργία του συστήματος τερματίζεται, οπότε αυτό έχει επίπτωση στην παραγωγή της ακτίνας ιόντων, η οποία, τελικά, δεν μπορεί να φτάσει στον στόχο της. Δεδομένου ότι υπάρχουν πολλαπλές κοιλότητες, η πιθανότητα σφάλματος είναι μη αμελητέα.



Σχήμα 10. Ένας κρυστάτης του ALPI, αποτελούμενος από τέσσερις κοιλότητες, και ένα σχήμα των βρόχων ελέγχου.

B. Εφαρμογή ανίχνευσης ανωμαλιών για την πρόβλεψη σφαλμάτων

Η προσέγγιση που προτείνεται για εφαρμογή στον συγκεκριμένο επιταχυντή ALPI συνοψίζεται στο σχήμα 11.



Σχήμα 11. Τα κύρια βήματα της δημιουργίας ενός μοντέλου ανίχνευσης ανωμαλιών για πρόβλεψη σφαλμάτων

Μπορούμε να χωρίσουμε την διαδικασία σε τρία στάδια: περιγραφή δεδομένων και προεπεξεργασία,

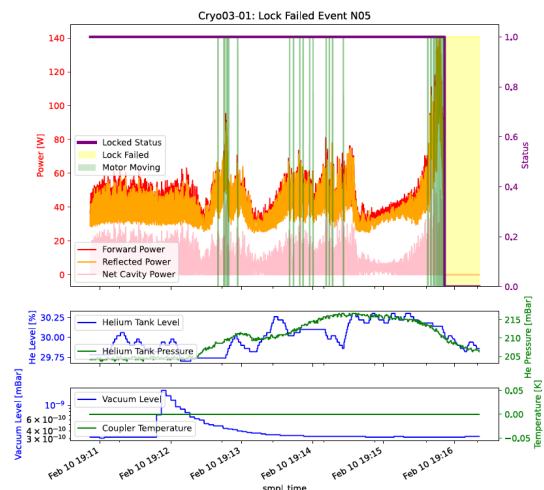
Feature Extraction και, τέλος, Unsupervised AD Algorithms, τα οποία παρουσιάζονται στη συνέχεια.

- Περιγραφή δεδομένων και προεπεξεργασία

Για κάθε κοιλότητα ενός κρυοστάτη ορίζονται πάνω από σαράντα Process μεταβλητές (Process Variables - PV). Κάθε μία από αυτές τις μεταβλητές, διαμορφώνει μία χρονοσειρά, αφού αποτελεί ένα σήμα από τα συστήματα ελέγχου της κοιλότητας, που παίρνει διάφορες τιμές στον χρόνο. Εφαρμόζεται μείωση της διάστασης του προβλήματος, οπότε, τελικά, αξιολογούνται έξι μεταβλητές. Αυτές είναι:

- Locked Status
- Forward, Reflected and Net RF power
- Tuning motor position, direction and moving flag
- Coupler temperature
- Cryostat liquid helium tank level and pressure
- Cryostat vacuum level

Η ύπαρξη της PV “locked status” μας επιτρέπει να χωρίσουμε τα δεδομένα σε πολλαπλά υποσύνολα, τα οποία τα ονομάζουμε events. Κάθε event αντιπροσωπεύει την κατάσταση στην οποία βρίσκεται μία κοιλότητα από τη στιγμή που έφτασε σε locked status (επιθυμητή λειτουργία), έως το επόμενο σφάλμα (unlocked event).



Σχήμα 12. Fault event και τάση των PVs κατά τα 5 λεπτά πριν συμβεί το σφάλμα.

Στο σχήμα 12 φαίνεται ένα παράδειγμα από την τάση των χρονοσειρών τα τελευταία πέντε λεπτά πριν συμβεί ένα σφάλμα. Όπως γίνεται εύκολα

αντιληπτό, κάποιες μεταβλητές, όπως οι “RF power” και “motor moving flag”, παρουσιάζουν υψηλή συσχέτιση με το σφάλμα που συνέβη, σε αντίθεση με τις PVs από τα cryogenic και vacuum υποσυστήματα, που παρουσιάζουν μικρή συσχέτιση.

- Feature Extraction

Αρχικά, όταν αναφερόμαστε σε feature extraction εννοούμε τη διαδικασία μετατροπής ακατέργαστων δεδομένων σε αριθμητικά χαρακτηριστικά που μπορούν να υποβληθούν σε επεξεργασία διατηρώντας παράλληλα τις πληροφορίες στο αρχικό σύνολο δεδομένων. Με τον τρόπο αυτόν, έχουμε καλύτερα αποτελέσματα συγκριτικά με την εφαρμογή machine learning απευθείας στα ακατέργαστα δεδομένα.

Στην περίπτωση μας, εφαρμόζεται ένα χρονικό παράθυρο στις χρονοσειρές, για το οποίο εξάγονται κάποια χαρακτηριστικά, ώστε να βγει ένα συμπέρασμα για τη τάση της χρονοσειράς. Για την ανάλυση και την απλούστευση των υπολογισμών, εφαρμόζεται υποδειγματοληψία στα δεδομένα των χρονοσειρών με τον σταθερό ρυθμό του ενός (1) Hz. Αυτή η τιμή επιλέγεται προκειμένου να μειώσει τον συνολικό αριθμό των δειγμάτων, ελατώνοντας τους χρόνους των υπολογισμών, ενώ, συνάμα, διατηρείται το μεγαλύτερο μέρος της πληροφορίας. Για κάθε sliding window που αναφέραμε, υπολογίζονται οι εξής τιμές:

- Μέση τιμή των RF forward, reflect, net power και του helium pressure
- Τυπική απόκλιση των ίδιων PVs
- Τιμή Delta, δηλαδή η διαφορά μεταξύ της τελευταίας και της πρώτης τιμής της χρονοσειράς για τις παραπάνω μεταβλητές, ώστε να βγει ένα συμπέρασμα για το trend της χρονοσειράς.
- Motor Moving time
- Motor Direction inversions

- Unsupervised AD Algorithms

Τα unsupervised anomaly detection μοντέλα που επιλέχθηκαν για την περίπτωση του ALPI, με βάση την αποδοτικότητά τους σε ανάλογα προβλήματα, ήταν τα εξής:

- k-Nearest Neighbors (KNN)
- Local Outlier Factor (LOF)

- Clustering-Based Local Outlier Factor (CBLOF)
- Feature Bagging with LOF base detector.
- Lightweight On-line Detector of Anomalies (LODA)
- Isolation Forest (iForest)

Από αυτά, παρουσιάζουμε σύντομα το Clustering-Based Local Outlier Factor (CBLOF).

Αρχικά, το CBLOF τρέχει έναν clustering αλγόριθμο, ώστε να ομαδοποιήσει τα δεδομένα σε ένα cluster και, έπειτα, τα clusters ταξινομούνται σε μικρά και μεγάλα. Συγκεκριμένα, για ένα σετ από clusters $C = \{C_1, C_2, \dots, C_k\}$, ταξινομημένα σε φθίνουσα σειρά μεγέθους, δηλαδή $|C_i| \geq |C_{i+1}|$, όπου $1 \leq i < k$. Επίσης, με δεδομένες αριθμητικές παραμέτρους $\alpha, \beta \in \mathbb{R}^+$, το όριο b μεταξύ μεγάλου και μικρού cluster ακολουθεί τον τύπο:

$$(|C_1| + |C_2| + \dots + |C_b|) \geq N \cdot \alpha, \quad \text{and} \quad \frac{|C_b|}{|C_{b+1}|} \geq \beta,$$

όπου N είναι ο συνολικός αριθμός των σημείων. Οπότε, τα μεγάλα clusters ορίζονται ως αυτά που έχουν δείκτη μικρότερο από b , ενώ τα υπόλοιπα είναι τα μικρά clusters.

Στη συνέχεια, ορίζουμε την μετρική CBLOF για ένα συγκεκριμένο σημείο x :

$$CBLOF(x) = \begin{cases} |C_i| \cdot \min(d(x, C_j)), & x \in C_i \\ |C_j| \cdot \min(d(x, C_j)), & x \in C_j \end{cases},$$

όπου d είναι η απόσταση μεταξύ cluster και σημείου, C_i είναι ένα μικρό cluster και C_j είναι ένα μεγάλο cluster. Η τιμή της μετρικής εξαρτάται και από το μέγεθος του cluster στο οποίο ανήκει το εκάστοτε σημείο x , αλλά και στην απόσταση του σημείου x από το κοντινότερο μεγάλο cluster (από το κέντρο του δικού του cluster στην περίπτωση που αυτό είναι μεγάλο). Αυτή η τιμή μπορεί να διακρίνει μεταξύ σημείων μέσα σε μεγάλα clusters και αυτών που ανήκουν σε μικρά απομακρυσμένα clusters, τα οποία πρόκειται να θεωρούνται ακραίες τιμές (outliers). Στη συγκεκριμένη υλοποίηση για τον ALPI, έχει αφαιρεθεί η στάθμιση που υπολογίζεται στον CBLOF αλγόριθμο ως προς το μέγεθος των clusters.

Στην περίπτωση μας, αξιοποιείται ένα διαθέσιμο σετ από labeled δεδομένα (το μοντέλο μας πλέον είναι semi-supervised) ώστε να γίνει η επιλογή του βέλτιστου threshold για το outlier score. Τονίζουμε ότι ένα σημείο από τα δεδομένα χαρακτηρίζεται

“ανώμαλο”, εφόσον το outlier score) του (μετρική CBLOF ξεπεράσει το threshold. Σαν βέλτιστο threshold, θεωρείται αυτό που μεγιστοποιεί τα ορθά ανιχνευμένα σφάλματα, ενώ, ταυτόχρονα, ελαχιστοποιεί και τα false positives. Η μετρική που συνυπολογίζει τα false positives και τα true positives ορίζεται ως το f-score. Τα labels στο dataset μας ορίζονται ως εξής: όλα τα φυσιολογικά δείγματα παίρνουν τιμή ετικέτας “-1”, ενώ όλα τα δείγματα που ανήκουν σε ένα χρονικό παράθυρο t_w δευτερολέπτων πριν από ένα unlocked event λαμβάνουν ως ετικέτα μια μη αρνητική τιμή, η οποία χρησιμοποιείται ως μοναδική ταυτοποίηση του εκάστοτε event, όπως φαίνεται παρακάτω.

$$Y_i = \begin{cases} -1 & t_i < t_{\text{event}} - t_w \\ \text{event.ID} & t_{\text{event}} - t_w \leq t_i \leq t_{\text{event}} \end{cases}$$

Οι τιμές των ετικετών για τα ανώμαλα δεδομένα ορίστηκαν κατά αυτόν τον τρόπο, προκειμένου να είναι δυνατή η ομαδοποίησή τους. Εμπειρικά, το t_w ορίστηκε στα οκτώ δευτερόλεπτα (8 s).

Η συνάρτηση που υπολογίζει true positives, false positives και true negatives ορίζεται ως:

$$Y_{\text{pred}} = \begin{cases} 1 & \text{score} \geq \text{thrs} \\ 0 & \text{otherwise} \end{cases}$$

```
all_ev= len(unique(Y[Y >= 0]))

TP = len(unique(Y[Y >=0][Y_pred]))
FP = Y[Y < 0][Y_pred]
FP = count_subsequences(FP)
FN = all_ev - TP
```

Με βάση τα παραπάνω, γίνεται υπολογισμός του f-score και ελέγχοντας μία πληθώρα τιμών για το threshold, επιλέγεται αυτή που μεγιστοποιεί το f-score. Ως εκ τούτου, αξιολογείται κατά πόσο ο αλγόριθμος είναι ικανός να πραγματοποιήσει ορθές προβλέψεις για ένα επικείμενο σφάλμα στον επιταχυντή.

BIBΛΙΟΓΡΑΦΙΑ

- [1] A. B. Nassif, M. A. Talib, Q. Nasir and F. M. Dakalbab, "Machine learning for anomaly detection: A systematic review", IEEE Access, vol. 9, pp. 78658-78700, 2021.
- [2] K. Choi, J. Yi, C. Park and S. Yoon, "Deep learning for anomaly detection in time-series data: Review analysis and guidelines", IEEE Access, vol. 9, pp. 120043-120065, 2021.
- [3] D. Marcato et al., "Machine Learning-based Anomaly Detection for Particle Accelerators," 2021 IEEE Conference on Control Technology and Applications (CCTA), 2021, pp. 240-246, doi: 10.1109/CCTA48906.2021.9658806.
- [4] Tharindu Fernando, Harshala Gammulle, Simon Denman, Sridha Sridharan, and Clinton Fookes. 2021. Deep Learning for Medical Anomaly Detection – A Survey. ACM Comput. Surv. 54, 7, Article 141 (September 2022)
- [5] L.-L. Zeng, H. Wang, P. Hu, B. Yang, W. Pu, H. Shen, X. Chen, Z. Liu, H. Yin, Q. Tan et al., “Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity mri,” EBioMedicine, vol. 30, pp. 74–85, 2018.