Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Συστηματων Μεταδοσης Πληροφοριας & Τεχνολογιας Υλικων

Σχεδίαση Αποδοτικών Μοντέλων Βαθιάς Μάθησης για Ανίχνευση Εισβολών σε Περιβάλλοντα IoT

Διπλωματικη Εργασια

του

ΝΙΚΗΤΑ ΤΣΙΝΝΑ

Επιβλέπων: Ιάκωβος Στ. Βενιέρης
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2024

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Συστημάτων Μετάδοσης Πληροφορίας & Τεχνολογίας Υλικών

Σχεδίαση Αποδοτικών Μοντέλων Βαθιάς Μάθησης για Ανίχνευση Εισβολών σε Περιβάλλοντα IoT

Διπλωματικη Εργασια

του

ΝΙΚΗΤΑ ΤΣΙΝΝΑ

Επιβλέπων: Ιάκωβος Στ. Βενιέρης
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 12η Ιουλίου 2024.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Ιάκωβος Στ. Βενιέρης	Δήμητρα-Θεοδώρα Κακλαμάνη	Εμμανουήλ Βαρβαρίγος
Καθηγητής Ε.Μ.Π.	Καθηγήτρια Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2024


Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Συστημάτων Μετάδοσης Πληροφορίας & Τεχνολογίας Υλικών

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Διπλωματικής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στη Διπλωματική μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων. Δηλώνω, συνεπώς, ότι αυτή η Διπλωματική Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

(Υπογραφή)

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

Περίληψη

Τα σύγχρονα δίκτυα επικοινωνιών παράγουν τεράστιο όγκο δεδομένων, απαιτώντας συνεχή παρακολούθηση για τη διατήρηση της απόδοσης και τη διαχείριση των τηλεπικοινωνιακών υποδομών. Οι τεχνικές Μηχανικής και Βαθιάς Μάθησης αποδεικνύονται ιδιαίτερα αποτελεσματικές στην ανίχνευση απειλών, παρά τις προκλήσεις που θέτει η έλλειψη επισημασμένων δεδομένων και τα ζητήματα απορρήτου. Η παρούσα εργασία επικεντρώνεται στη χρήση αυτών των τεχνικών για την ανίχνευση απειλών σε περιβάλλοντα IoT, με ιδιαίτερη έμφαση στη μείωση του μεγέθους των μοντέλων διατηρώντας ταυτόχρονα ικανοποιητική ακρίβεια. Συγκρίνονται τρεις δημοφιλείς αρχιτεκτονικές νευρωνικών δικτύων (Πολυεπίπεδα Perceptrons, Συνελικτικά Νευρωνικά Δίκτυα, Μετασχηματιστές) χρησιμοποιώντας δεδομένα από το σύνολο CICIoT2023, και περιλαμβάνονται βιβλιογραφική ανασκόπηση, αναλυτική διαδικασία προετοιμασίας δεδομένων και αξιολόγηση των εκπαιδευμένων μοντέλων. Η μελέτη αναδεικνύει τις προκλήσεις που προκύπτουν από την ανισορροπία των κλάσεων και την προστασία προσωπικών δεδομένων, με το εργαλείο NFStream να προσφέρει μια ενοποιημένη και αξιόπιστη προσέγγιση στην εξαγωγή χαρακτηριστικών. Από την αξιολόγηση των μοντέλων, τα μοντέλα μετασχηματιστών προσέφεραν την καλύτερη απόδοση, ενώ τα πολυεπίπεδα Perceptrons είχαν τη χαμηλότερη. Επιπλέον, η αξιολόγηση των μοντέλων δεν μπορεί να βασίζεται σε ένα μόνο κριτήριο, καθώς οι απαιτήσεις και οι στόχοι κάθε εφαρμογής διαφέρουν. Η εργασία προτείνει λύσεις για την εξισορρόπηση της απόδοσης και της αποδοτικότητας, παρουσιάζοντας συμπεράσματα και μελλοντικές κατευθύνσεις όπως η διερεύνηση δυαδικής και πολυταξικής ταξινόμησης, η μελέτη της καθυστέρησης ανίχνευσης σε ολοκληρωμένο σύστημα προσομοίωσης, η διερεύνηση διαφορετικών συνόλων δεδομένων ασφάλειας δικτύων, η χρήση τεχνικών υπερδειγματοληψίας για τη βελτίωση της ανισορροπίας των δεδομένων, και η επεξεργασία δεδομένων πακέτων σε επίπεδο bytes. Συνολικά, η εργασία συμβάλλει στην κατανόηση και βελτίωση των τεχνικών ανίχνευσης εισβολών σε περιβάλλοντα IoT, συμβάλλοντας στην ανάπτυξη πιο αποδοτικών και αποτελεσματικών συστημάτων ανίχνευσης.

Λέξεις Κλειδιά

Μηχανική Μάθηση, Βαθιά Μάθηση, Ανίχνευση Εισβολών, Συμπαγή Μοντέλα, Συσκευές Περιορισμένων Πόρων, Διαδίκτυο των Πραγμάτων, Ασφάλεια Δικτύων, Πολυστρωματικά Perceptrons, Συνελικτικά Νευρωνικά Δίκτυα, Μετασχηματιστές.

Abstract

Modern communication networks generate a massive volume of data, necessitating continuous monitoring to maintain performance and manage telecommunication infrastructures. Machine Learning (ML) and Deep Learning (DL) techniques have proven highly effective in threat detection, despite challenges posed by the lack of labeled data and privacy concerns. This study focuses on the application of these techniques for threat detection in Internet of Things (IoT) environments, with a particular emphasis on reducing model size while maintaining satisfactory accuracy. Three popular neural network architectures (Multilayer Perceptrons, Convolutional Neural Networks, Transformers) are compared using data from the CICIoT2023 dataset, including a literature review, a detailed data preparation process, and an evaluation of trained models. The study highlights challenges arising from class imbalance and data privacy, with the NFStream tool offering a unified and reliable approach to feature extraction. Model evaluations showed that transformer models provided the highest performance, while multilayer perceptrons had the lowest. Furthermore, model evaluation cannot rely on a single criterion, as application requirements and goals vary. The study proposes solutions for balancing performance and efficiency, presenting conclusions and future directions such as investigating binary and multiclass classification, studying detection delay in an integrated simulation system, exploring different network security datasets, using oversampling techniques to improve data imbalance, and processing packet data at the byte level. Overall, the study contributes to the understanding and improvement of intrusion detection techniques in IoT environments, aiding in the development of more efficient and effective detection systems.

Keywords

Machine Learning, Deep Learning, Intrusion Detection, Compact Models, Resource Constrained Devices, Internet of Things, Network Security, Multi-Layer Perceptrons (MLPs), Convolutional Neural Networks (CNNs), Transformer Encoders

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή κ. Ιάκωβο Στ. Βενιέρη για την ευκαιρία εκπόνησης της παρούσας διπλωματικής εργασίας. Το αντικείμενο της ασφάλειας δικτύων μέσω της ανάπτυξη μοντέλων βαθιάς μάθησης για την ανίχνευση εισβολών αποτέλεσε μια ενδιαφέρουσα πρόκληση.

Ιδιαίτερες ευχαριστίες θα ήθελα να απευθύνω στον υποψήφιο διδάκτορα της ΣΗΜΜΥ κ. Ιωάννη Πανόπουλο για την αμεσότητά του, την απεριόριστη στήριξη, το έντονο ενδιαφέρον και την πολύτιμη βοήθειά του. Οι διεξοδικές συζητήσεις που είχαμε στο εργαστήριο συνεχώς άνοιγαν νέα μέτωπα προς εξερεύνηση και καθιστούσαν την ερευνητική διαδικασία εξαιρετικά παραγωγική και ευχάριστη.

Ακόμα, δεν θα μπορούσα να παραλείψω τους συμφοιτητές και φίλους μου που ήταν κοντά μου όλα τα προηγούμενα έτη και μου προσέφεραν ο καθένας τους με τον δικό του τρόπο υποστήριξη αλλά κυρίως ευχάριστες αναμνήσεις που θα μείνουν ανεξίτηλες στο μυαλό μου. Αισθάνομαι ιδιαίτερα τυχερός που γνώρισα τέτοιους αξιόλογους και δυναμικούς ανθρώπους και ελπίζω να κρατήσουν οι φιλίες μας για πολλά χρόνια.

Τέλος, οι πιο θερμές ευχαριστίες απευθύνονται στους γονείς μου, Βλάση και Ρένα, καθώς και στον αδερφό μου, Κωνσταντίνο. Η ανιδιοτελής αγάπη τους είχε, και συνεχίζει να έχει, καθοριστικό ρόλο στη σταδιοδρομία μου. Χωρίς αυτούς αισθάνομαι πως δεν θα είχα καταφέρει όσα έχω μέχρι στιγμής και για αυτό θα τους είμαι πάντα ευγνώμων. Αυτή η διπλωματική εργασία αφιερώνεται σε αυτούς.

Αθήνα, Ιούλιος 2024

Νικήτας Τσίννας

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1

Εισαγωγή

Τα σημερινά δίκτυα επικοινωνιών παράγουν τεράστιο και ετερογενή όγκο από δεδομένα κίνησης (traffic data) λόγω της πληθώρας υπηρεσιών ή εφαρμογών και του μεγάλου αριθμού χρηστών που πρέπει να εξυπηρετήσουν. Λόγω της πολύπλοκης συμπεριφοράς τους, αυτά τα δεδομένα απαιτούν συνεχή παρακολούθηση για τη διατήρηση της επίδοσης, τη βελτιστοποίηση της κατανομής των δικτυακών πόρων, αλλά και τον έλεγχο και την αποτελεσματική διαχείριση των τηλεπικοινωνιακών υποδομών.

Τα τελευταία χρόνια, τεχνικές Μηχανικής Μάθησης (Machine Learning - ML) και Βαθιάς Μάθησης (Deep Learning - DL) έχουν αποδειχθεί εξαιρετικά αποτελεσματικές σε πολλούς τομείς χάρη στην ικανότητά τους να ανακαλύπτουν σύνθετες δομές σε μεγάλα σύνολα δεδομένων. Η όραση υπολογιστών, η αναγνώριση ομιλίας, η επεξεργασία φυσικής γλώσσας, η ιατρική και η φαρμακευτική αποτελούν λίγα παραδείγματα των τομέων που αυτές οι τεχνικές έχουν προσφέρει πρωτοφανείς βελτιώσεις. Επιπλέον, υπάρχει η κοινή πεποίθηση πως ο κλάδος της βαθιάς μάθησης θα γνωρίσει πολλές ακόμα επιτυχίες στο κοντινό μέλλον, καθώς χρειάζεται ελάχιστη χειροκίνητη μηχανική εργασία, οπότε μπορεί εύκολα να εκμεταλλευτεί τη διαθεσιμότητα μεγάλων όγκων δεδομένων. Νέοι αλγόριθμοι και αρχιτεκτονικές που αυτή την στιγμή αναπτύσσονται πρόκειται να επιταχύνουν αυτή την εξέλιξη [5].

Τα είδη προβλημάτων που μπορούν να αντιμετωπιστούν με μοντέλα μηχανικής και βαθιάς μάθησης, όπως η κατηγοριοποίηση, η πρόβλεψη και η λήψη αποφάσεων ταιριάζουν σε δικτυακά προβλήματα, όπως η πρόβλεψη της απόδοσης (π.χ. συμφόρησης), ο προγραμματισμός του δικτύου και κατηγοριοποίηση της κίνησης για λόγους ασφάλειας.

Ωστόσο, είναι σημαντικό να αναφερθεί πως σε σχέση με τους υπόλοιπους προαναφερθέντες τομείς, ο τομέας των δικτύων αντιμετωπίζει σημαντικές δυσκολίες στην ενσωμάτωση τεχνικών μηχανικής μάθησης. Βασικές δυσκολίες είναι η έλλειψη επισημασμένων δεδομένων, η αμφιβολία για την ποιότητα των δεδομένων καθώς και τα θέματα απορρήτου που σχετίζονται με τα δικτυακά δεδομένα [6].

Η διατήρηση της ασφάλειας και η προστασία των συνδεδεμένων συσκευών από επιθέσεις είναι απαραίτητες για τη διασφάλιση της ιδιωτικότητας των χρηστών, την εύρυθμη λειτουργία των συσκευών, τη διασφάλιση της ποιότητας υπηρεσίας (Quality of Service - QoS) και της ποιότητας εμπειρίας (Quality of Experience - QoE) των δικτύων. Προβλήματα ασφάλειας είναι συνήθως προβλήματα κατηγοριοποίησης της κίνησης, όπου κάθε εισερχόμενη ροή ταξινομείται ως καλοήθης (benign) ή κακόβουλη (malicious).

1.1 Κίνητρο

Η παρούσα εργασία εστιάζει στην εφαρμογή τεχνικών μηχανικής και βαθιάς Μάθησης για την ανίχνευση απειλών σε δίκτυα επικοινωνιών, και πιο συγκεκριμένα σε περιβάλλοντα Διαδικτύου των Πραγμάτων (Internet of Things - IoT). Έχει παρατηρηθεί, πως η πλειοψηφία των ερευνών παρουσιάζει επίτευξη υψηλών ποσοστών ακρίβειας (accuracy) στην κατηγοριοποίηση απειλών αλλά πολύ συχνά δεν ασχολείται με άλλες σημαντικές μετρικές επίδοσης και απόδοσης των μοντέλων. Για παράδειγμα, αν τα εκπαιδευμένα μοντέλα προορίζονται για συσκευές του άκρου (edge devices) (π.χ. έξυπνο κινητό, τηλεοράσεις, μικροελεγκτές με ικανότητα σύνδεσης στο διαδίκτυο, κάμερες κτλ) καθορίζεται επιτακτική η μελέτη των παρακάτω ζητημάτων:

- το Μέγεθος Μοντέλου: Η πολυπλοκότητα του μοντέλου επηρεάζει άμεσα την απαιτούμενη υπολογιστική ισχύ. Ως μετρικές για την εκτίμηση του μεγέθους και της πολυπλοκότητας του μοντέλου μπορεί να χρησιμοποιηθεί ο αριθμός των παραμέτρων (weights) και ο αριθμός των πράξεων κινητής υποδιαστολής ανά δευτερόλεπτο (Floating Point Operations Per Second - FLOPS). Αυτά βοηθούν στην αξιολόγηση της ισχύος και των πόρων που θα απαιτηθούν για την εκπαίδευση και την συμπερασματολογία (inference) του μοντέλου. Σε περιβάλλοντα με περιορισμένους πόρους, η υλοποίηση μεγάλων και περίπλοκων μοντέλων μπορεί να αποδειχθεί ασύμφορη. Επομένως, η επιλογή ενός κατάλληλου μεγέθους μοντέλου είναι κρίσιμη για την επίτευξη της ισορροπίας μεταξύ απόδοσης και αποδοτικότητας.

- την Καθυστέρηση (Latency): Η καθυστέρηση στην ανίχνευση απειλών μπορεί να έχει σημαντικές επιπτώσεις στην ασφάλεια του δικτύου. Είναι απαραίτητο η ανίχνευση να γίνεται έγκαιρα και με ελάχιστη καθυστέρηση.

- την Επιβάρυνση του Συστήματος (Workload): Η επιβάρυνση του συστήματος από τον τρόπο σχεδίασης μέχρι και την εκτέλεση του μοντέλου επηρεάζει την ομαλή λειτουργία της συσκευής. Η υπερβολική επιβάρυνση μπορεί να οδηγήσει σε υποβάθμιση της ποιότητας υπηρεσιών και εμπειρίας των χρηστών. Επιπλέον, η ευκολία εγκατάστασης (ease of deployment) αποτελεί σημαντικό παράγοντα, καθώς τα μοντέλα πρέπει να μπορούν να εγκατασταθούν και να λειτουργήσουν ομαλά σε διαφορετικά περιβάλλοντα και συσκευές. Η κατανάλωση μνήμης και ενέργειας είναι επίσης κρίσιμα ζητήματα, ειδικά για φορητές συσκευές. Τα αποδοτικά μοντέλα που απαιτούν λιγότερους πόρους μπορούν να λειτουργήσουν καλύτερα σε περιορισμένα περιβάλλοντα, ενώ ταυτόχρονα μειώνουν την κατανάλωση ενέργειας, παρατείνοντας τη διάρκεια ζωής της μπαταρίας των φορητών συσκευών.

Σε πολλές περιπτώσεις, έρευνες δεν δίνουν σημασία στα παραπάνω ζητήματα καθώς τα υπό διερεύνηση μοντέλα είτε αποσκοπούν απλώς στην επίτευξη της υψηλότερης δυνατής ακρίβειας, είτε αποτελούν μέρος μιας αρχιτεκτονικής συστήματος ανίχνευσης απειλών που βασίζεται σε κάποιον κεντρικό κόμβο ο οποίος υποθετικά έχει τους απαραίτητους πόρους για να αντέξει μεγάλα και πολύπλοκα μοντέλα. Ωστόσο, αυτή η προσέγγιση παρουσιάζει μερικά αξιόλογα μειονεκτήματα τα οποία αναλύονται εκτενώς στο Κεφάλαιο 3.

Υπάρχουν δύο βασικές μέθοδοι για την ελαχιστοποίηση του μεγέθους ενός μοντέλου βαθιού νευρωνικού δικτύου (Deep Neural Network - DNN).

1. Συμπίεση μοντέλων: Ο στόχος των τεχνικών συμπίεσης (compression techniques) είναι η μείωση του μεγέθους (με λιγότερες ή «μικρότερες» παραμέτρους) χωρίς να μειωθεί παράλληλα σημαντικά η ακρίβεια. Αυτές οι τεχνικές, όπως είναι το κλάδεμα παραμέτρων, η χβαντοποίηση, η απόσταξη γνώσης κ.ά, εφαρμόζονται σε ήδη γνωστές αρχιτεκτονικές με κάποιες από αυτές να μπορούν να εφαρμοστούν ακόμα σε εκπαιδευμένα μοντέλα.

2. Σχεδίαση συμπαγών μοντέλων: Μια διαφορετική προσέγγιση είναι ο σχεδιασμός νέων αρχιτεκτονικών νευρωνικών δικτύων με απώτερο στόχο τα μοντέλα που θα προκύψουν να είναι συμπαγή, δηλαδή μικρά και γρήγορα, και ταυτόχρονα να έχουν ικανοποιητική ακρίβεια.

Σε αυτή την εργασία χρησιμοποιούμε την δεύτερη μέθοδο, καθώς αναζητούμε συμπαγή μοντέλα που βασίζονται σε γνωστούς τύπους νευρωνικών δικτύων, δοκιμάζοντας πολλές διαφορετικές αρχιτεκτονικές. Με αυτή την λογική, η εργασία αποσκοπεί στο να διαφωτίσει το πεδίο, ώστε να μπορέσουμε να αξιολογήσουμε κατά πόσο είναι εφικτή μια λύση όπου οι δικτυακές συσκευές μπορούν αυτόνομα να εκτελούν μοντέλα βαθιάς μάθησης για την ανίχνευση απειλών χωρίς να θυσιάζεται η ακρίβεια ή η αποτελεσματικότητα.

1.2 Συνεισφορά

Όπως προαναφέρθηκε, η έρευνα εστιάζει στον συμβιβασμό μεταξύ ακρίβειας και μέγεθος μοντέλων βαθιάς μάθησης στην ανίχνευση εισβολών. Ειδικότερα, αφού έγινε η αντίστοιχη βιβλιογραφική έρευνα, επιλέχθηκε συγκεκριμένο σύνολο δεδομένων δικτυακής κίνησης για την ανάπτυξη και αξιολόγηση διαφορετικών αρχιτεκτονικών νευρωνικών δικτύων για το πρόβλημα της κατηγοριοποίησης.

Αναλυτικά, στην παρούσα εργασία:

- Πραγματοποιήθηκε μελέτη της βιβλιογραφίας σχετικά με την διεργασία ανίχνευσής εισβολών με χρήση τεχνικών μηχανικής και βαθιάς μάθησης. Δόθηκε έμφαση στην επιλογή συνόλου δεδομένων, τον τρόπο προεπεξεργασίας αυτών, τις επιλεγμένες αρχιτεκτονικές μοντέλων και τέλος στον τρόπο αξιολόγησής τους.

- Πραγματοποιήθηκε εναλλακτικός τρόπος διεξαγωγής χαρακτηριστικών από αυτόν που προτείνεται από τους παρόχους του επιλεγμένου συνόλου δεδομένων, ώστε να ανταποκρίνεται στην φύση του προβλήματος, δηλαδή στην αυτονομία των δικτυακών συσκευών στην ανίχνευση εισβολών. Εξετάστηκαν Πολυεπίπεδα Perceptrons (Multilayer Perceptrons - MLPs), Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNNs) και Μετασχηματιστές (Transformers) όπου για την κάθε μία περίπτωση εκπαιδεύτηκαν πολλαπλά μοντέλα με διαφορετικά μεγέθη.

- Εξήχθησαν συμπεράσματα για τις βέλτιστες αρχιτεκτονικές του υπό εξέταση προβλήματος, όχι μόνο με βάση το μέγεθος αυτών και τις κλασσικές μετρικές κατηγοριοποίησης (όπως είναι η ακρίβεια και η βαθμολογία F1), αλλά και με χαρακτηριστικές μετρικές των Συστημάτων Ανίχνευσης Εισβολών (Intrusion Detection Systems - IDS).

1.3 Οργάνωση του τόμου

Η δομή της εργασίας είναι σχεδιασμένη για να παρέχει μια ολοκληρωμένη κατανόηση της εφαρμογής των σύγχρονων τεχνικών βαθιάς μάθησης στην ανίχνευση απειλών, με έμφαση στις προκλήσεις και τις λύσεις στον τομέα της κυβερνοασφάλειας. Πιο συγκεκριμένα:

- **Κεφάλαιο 2**: Καλύπτει τις βασικές αρχές των τεχνικών μηχανικής και βαθιάς μάθησης που χρησιμοποιούνται στην ανίχνευση απειλών. Περιλαμβάνει ανασκόπηση των βασικών εννοιών, όπως τα σύνολα δεδομένων, η ταξινόμηση, τα τυχαία δάση ταξινόμησης (Random Forest Classifier), και τα νευρωνικά δίκτυα. Επιπλέον, εξετάζονται τα πολυεπίπεδα Perceptrons, τα συνελικτικά νευρωνικά δίκτυα και οι μετασχηματιστές.

- **Κεφάλαιο 3**: Εστιάζει στα συστήματα ανίχνευσης εισβολών, με ανάλυση της στρατηγικής τοποθέτησης, του τρόπου εγκατάστασης, της μεθόδου ανίχνευσης και της αξιολόγησής τους. Επιπλέον, εξετάζεται το Διαδίκτυο των Πραγμάτων (Internet of Things - IoT), οι εφαρμογές του, οι προκλήσεις ασφαλείας και οι γνωστές κατηγορίες επιθέσεων.

- **Κεφάλαιο 4**: Παρουσιάζεται η πειραματική διάταξη της έρευνας. Περιγράφονται τα εργαλεία και οι πλατφόρμες που χρησιμοποιήθηκαν, όπως το NFStream, το Tensorflow και το Kaggle, καθώς και το σύνολο δεδομένων CICIoT2023. Εξηγείται η διαδικασία προετοιμασίας των δεδομένων, η εξαγωγή χαρακτηριστικών και η αντιμετώπιση της ανισορροπίας κλάσεων.

- **Κεφάλαιο 5**: Αναλύει τη διαδικασία εκπαίδευσης των μοντέλων και τα αποτελέσματα που προέκυψαν. Περιλαμβάνει την εκπαίδευση τυχαίου δάσους και νευρωνικών δικτύων. Επιπλέον, γίνεται σύγκριση των αρχιτεκτονικών και αξιολόγηση των επιδόσεων με βάση τις μετρικές ακρίβειας, την βαθμολογία F1 και τα μεγέθη των μοντέλων.

- **Κεφάλαιο 6**: Παρουσιάζονται τα συμπεράσματα της έρευνας και προτείνονται μελλοντικές κατευθύνσεις για περαιτέρω μελέτη και βελτιώσεις στην ανίχνευση απειλών σε περιβάλλοντα IoT.

Κεφάλαιο 2

Μηχανική και Βαθιά Μάθηση

Στο δεύτερο Κεφάλαιο παρουσιάζεται ένα θεωρητικό υπόβαθρο της μηχανικής και βαθιάς μάθησης. Το κεφάλαιο χωρίζεται σε τρεις ενότητες. Πρώτα, παρουσιάζονται μερικές βασικές αρχές. Έπειτα, περιγράφονται οι αρχιτεκτονικές βαθιών νευρωνικών δικτύων που χρησιμοποιούνται στην παρούσα εργασία. Η τελευταία ενότητα αφιερώνεται στην επισήμανση των περιορισμών που αντιμετωπίζει ο τομέας της βαθιάς μάθησης για δομημένα σύνολα δεδομένων, καθώς αυτοί εμφανίζονται και στο υπό εξέταση πρόβλημα.

2.1 Βασικές Αρχές

Η μηχανική μάθηση είναι ένας κλάδος της Τεχνητής Νοημοσύνης (Artificial Intelligence - AI) που εστιάζει στην αξιοποίηση δεδομένων και αλγορίθμων με σκοπό τη μίμηση του τρόπου που μαθαίνουν οι άνθρωποι. Λόγω της αυξανόμενης δημοτικότητας του τομέα της τεχνητής νοημοσύνης, καθίσταται επιτακτική η σαφής διάκριση μεταξύ βασικών εννοιών όπως αυτή της μηχανική μάθησης, της βαθιάς μάθησης και των νευρωνικών δικτύων, οι οποίες χρησιμοποιούνται συχνά στην παρούσα εργασία και συχνά συγχέονται μεταξύ τους.

Σχήμα 2.1: *Διάγραμμα Venn Εννοιών Τεχνητής Νοημοσύνης*

Η τεχνητή νοημοσύνη αποτελεί μία ευρύτερη έννοια που περιλαμβάνει οποιαδήποτε τεχνική επιτρέπει σε έναν υπολογιστή να μιμείται την ανθρώπινη ευφυΐα, ενώ η μηχανική μάθηση είναι ένα υποσύνολο της τεχνητής νοημοσύνης που περιλαμβάνει τεχνικές και αλγόριθμους που επιτρέπουν στα συστήματα να μαθαίνουν και να βελτιώνονται από δεδομένα χωρίς να είναι

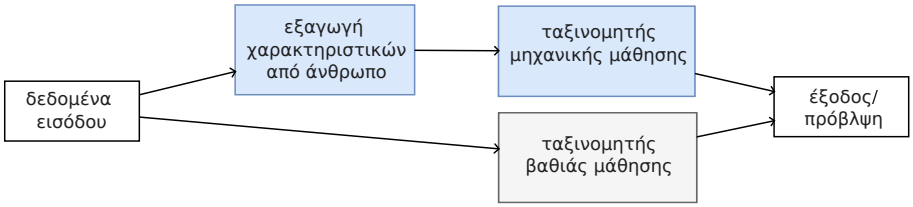
ρητά προγραμματισμένα.

Τα νευρωνικά δίκτυα είναι ένα υποσύνολο της μηχανικής μάθησης, εμπνευσμένο από τη δομή και τη λειτουργία του ανθρώπινου εγκεφάλου. Τα νευρωνικά δίκτυα αποτελούνται από στρώματα κόμβων (νευρώνων) που συνεργάζονται για την ανάλυση και την επεξεργασία δεδομένων. Τα νευρωνικά δίκτυα που έχουν πολλά κρυφά επίπεδα αποτελούν ένα εξειδικευμένο υποσύνολο αυτών και εντάσσονται στην βαθιά μάθηση και είναι ικανά να αναγνωρίζουν σύνθετα μοτίβα σε μεγάλα σύνολα δεδομένων.

Η βασική διαφορά μεταξύ βαθιάς μάθησης και μηχανικής μάθησης έγκειται στην ανάγκη της ανθρώπινης παρέμβασης για την εξαγωγή χαρακτηριστικών από τα δεδομένα:

- Η **μηχανική μάθηση** βασίζεται περισσότερο στην ανθρώπινη παρέμβαση. Οι ειδικοί καθορίζουν τα χαρακτηριστικά των δεδομένων εισόδου και για αυτό η μηχανική μάθηση είναι καταλληλότερη για δομημένα δεδομένα (δεδομένα σε μορφή πίνακα) για την εκμάθηση των μοντέλων. Μέχρι και σήμερα, η μηχανική μάθηση υπερτερεί της βαθιάς μάθησης σε δομημένα δεδομένα [7] (βλ. Ενότητα 2.3).

- Η **βαθιά μάθηση** μπορεί να αναλύσει μη δομημένα δεδομένα σε ακατέργαστη μορφή (κείμενο, εικόνες) και να εντοπίσει αυτόματα τα χαρακτηριστικά που διακρίνουν διαφορετικές κατηγορίες δεδομένων. Αυτό μειώνει την ανθρώπινη παρέμβαση και επιτρέπει τη χρήση μεγάλων όγκων δεδομένων [8].



Σχήμα 2.2: *Κύρια Διαφορά Μεταξύ Μηχανικής και Βαθιάς Μάθησης*

2.1.1 Είδη Μάθησης

Εν γένει, ο τομέας της μηχανικής μάθησης αναπτύσσει τρεις τρόπους μάθησης, ανάλογους με τους τρόπους με τους οποίους μαθαίνει ο άνθρωπος: επιβλεπόμενη μάθηση, μη επιβλεπόμενη μάθηση και ενισχυτική μάθηση. Πιο αναλυτικά:

- **Επιβλεπόμενη Μάθηση (Supervised Learning)**: Σε αυτή την περίπτωση η εκπαίδευση πραγματοποιείται χρησιμοποιώντας επισημασμένα (labeled) σύνολα δεδομένων, δηλαδή για κάθε είσοδο γνωρίζουμε την επιθυμητή έξοδο. Τα προβλήματα που μπορούν να λυθούν με αυτό το είδος είναι αυτά της ταξινόμησης (classification) και της παλινδρόμησης (regression). Αυτό είναι και το είδος Μηχανικής Μάθησης που σχετίζεται με την παρούσα εργασία.

- **Μη επιβλεπόμενη Μάθηση (Unsupervised Learning)**: Αντίθετα με την επιβλεπόμενη μάθηση, η εκπαίδευση γίνεται με μη επισημασμένα (unlabeled) σύνολα δεδομένων με σκοπό να βρεθούν κοινά χαρακτηριστικά μεταξύ τους. Το πρωταρχικό

πρόβλημα που μελετάται με αυτό το είδος είναι η συσταδοποίηση (clustering), όπου πραγματοποιείται ομαδοποίηση ενός συνόλου αντικειμένων με τέτοιο τρόπο ώστε τα αντικείμενα στην ίδια ομάδα (ή συστάδα) να είναι πιο όμοια μεταξύ τους παρά με αυτά σε άλλες ομάδες.

- **Ενισχυτική Μάθηση (Reinforcement Learning)**: Η μάθηση βασίζεται στην αλληλεπίδραση ενός πράκτορα με το περιβάλλον, όπου αυτός προσαρμόζει τη στρατηγική του με βάση το κόστος των πράξεών του με στόχο να την βελτιστοποιήσει. Χρησιμοποιείται σε εφαρμογές όπως παιχνίδια στρατηγικής ή σε έλεγχο κίνησης ρομπότ.

2.1.2 Ταξινόμηση

Η ταξινόμηση, ως πρόβλημα μηχανικής μάθησης, ανήκει στην κατηγορία μεθόδων επιβλεπόμενης μάθησης. Αυτό σημαίνει πως τα δείγματα του συνόλου δεδομένων είναι επισημασμένα, δηλαδή έχουν ετικέτα η οποία καθορίζει την κλάση του εκάστοτε δείγματος. Για παράδειγμα, αν μιλάμε για ταξινόμηση ροών ανταλλαγής δικτυακών πακέτων, η ετικέτα κάθε ροής θα όριζε αν αυτή είναι καλοήθης ή κακόβουλη.

Ο στόχος ενός προβλήματος ταξινόμησης από μαθηματική άποψη είναι να βρεθεί μια συνάρτηση f που χαρτογραφεί τα εισαγόμενα δεδομένα x από έναν χώρο εισόδου X σε μια διακριτή κατηγορία y στον χώρο εξόδου Y. Πιο συγκεκριμένα ας υποθέσουμε ότι έχουμε ένα σύνολο εκπαίδευσης $\{(x_i, y_i)\}_{i=1}^{n}$, όπου κάθε x_i είναι ένα δείγμα δεδομένων (για δομημένα σύνολα είναι μία γραμμή πίνακα) και κάθε y_i είναι η αντίστοιχη κατηγορία (ετικέτα) σε ένα σύνολο κατηγοριών $C = \{c_1, c_2, \ldots, c_k\}$.

Ο στόχος είναι να βρούμε μια συνάρτηση $f : X \to C$ (target function) που ελαχιστοποιεί την αναμενόμενη απώλεια (σφάλμα) σε νέα, άγνωστα δεδομένα, δηλαδή να ελαχιστοποιήσουμε την αναμενόμενη τιμή της απώλειας $L(f(x), y)$, όπου L είναι η συνάρτηση απώλειας (loss function) που μετράει το σφάλμα της πρόβλεψης $f(x)$ σε σχέση με την πραγματική ετικέτα y.

Στα προβλήματα επιβλεπόμενης μάθησης, το σύνολο δεδομένων χωρίζεται σε τρία σύνολα:

- Σύνολο εκπαίδευσης (training set): αποτελεί και το μεγαλύτερο μέρος του dataset καθώς με αυτό εκπαιδεύεται το μοντέλο ώστε να βρεθεί η συνάρτηση στόχου.

- Σύνολο επικύρωσης (validation set): χρησιμοποιείται για την παρακολούθηση της απόδοσης του μοντέλου κατά την εκπαίδευση, βοηθώντας στην αναγνώριση της στιγμής που το μοντέλο αρχίζει να υπερπροσαρμόζεται (overfitting) στα δεδομένα εκπαίδευσης. Όταν η απόδοση στο validation set αρχίσει να επιδεινώνεται ενώ η απόδοση στο training set συνεχίζει να βελτιώνεται, η εκπαίδευση μπορεί να σταματήσει για να αποφευχθεί το overfitting.

- Σύνολο ελέγχου (test set): δεσμεύει συνήθως το 15-20% του συνόλου και χρησιμεύει για την αξιολόγηση της επίδοσης του μοντέλου σε άγνωστα δεδομένα μετά την διαδικασία εκπαίδευσης.

Η περίπτωση κατηγοριοποίησης μίας ροής κίνησης σε δύο κλάσεις (καλοήθης ή κακόβουλη) ονομάζεται **δυαδική** (binary), ενώ αν κατηγοριοποίηση γίνεται με πάνω από δύο κλάσεις (π.χ. κατηγοριοποίηση τύπου επίθεσης) τότε χαρακτηρίζεται **πολυταξική** (multi-class)

2.1.3 Σύνολα Δεδομένων

Τα σύνολα δεδομένων (datasets), όπως υποδηλώνει και η ονομασία τους, αποτελούν συλλογές από δεδομένα. Αυτά τα δεδομένα μπορούν να είναι δομημένα είτε μη δομημένα.

Τα δομημένα δεδομένα είναι οργανωμένα σε μορφή πίνακα. Αυτά τα δεδομένα έχουν σαφή ορισμένα πεδία (στήλες) και πολλαπλούς τύπους δεδομένων, όπως αριθμοί, ημερομηνίες ή κείμενο. Ένα παράδειγμα δομημένων δεδομένων είναι ένας πίνακας σε μια βάση δεδομένων με στήλες για όνομα, ηλικία, και διεύθυνση.

Αντιθέτως, τα μη δομημένα δεδομένα δεν ακολουθούν μια προκαθορισμένη δομή. Αυτά τα δεδομένα είναι πιο «ελεύθερα», δεν χωρίζονται σε στήλες και μπορούν να περιλαμβάνουν κείμενα, εικόνες, βίντεο, αρχεία ήχου κ.ά. Η ανάλυση μη δομημένων δεδομένων είναι πιο περίπλοκη, καθώς δεν υπάρχει συγκεκριμένη διάταξη για να βασιστεί κανείς.

Τα διαθέσιμα σύνολα δεδομένων για σκοπούς έρευνας που σχετίζονται με δικτυακά προβλήματα σε σχέση με αυτά άλλων ερευνητικών τομέων είναι λιγοστά. Ο λόγος έλλειψης ανοιχτών δικτυακών συνόλων δεδομένων είναι κυρίως το απόρρητο αυτών των πληροφοριών. Οι εταιρείες και οι οργανισμοί τείνουν να προστατεύουν τα δεδομένα τους για να διασφαλίσουν την ιδιωτικότητα και την ασφάλεια των χρηστών τους, με αποτέλεσμα να περιορίζεται η διαθεσιμότητα τους για ερευνητικούς σκοπούς. Επίσης, ακόμη και όταν αυτά διατίθενται, η πρόσβαση μπορεί να είναι περιορισμένη ή να απαιτείται ειδική άδεια και διαδικασίες για την απόκτησή τους.

Ειδική περίπτωση αποτελούν τα ανοιχτά σύνολα δεδομένων του «Canadian Institute for Cybersecurity» από το πανεπιστήμιο του «New Brunswick» [9], τα οποία χωρίζονται σε διάφορες κατηγορίες ανά ερευνητικό αντικείμενο. Ένα από αυτά χρησιμοποιήθηκε και στην παρούσα εργασία.

Ανισορροπία κλάσεων

Μία δυσκολία που συχνά υπάρχει σε προβλήματα ταξινόμησης είναι η ανισορροπία των κλάσεων που εμφανίζεται σε πολλά σύνολα δεδομένων, ειδικά σε σύνολα δικτυακής κίνησης για την ασφάλεια. Δηλαδή, είναι λογικό να έχουμε πολύ περισσότερα δείγματα μιας επίθεσης, η οποία από την φύση της παράγει τεράστιο όγκο δικτυακών δεδομένων, από τα δείγματα της καλοήθους κίνησης. Αυτό πρακτικά σημαίνει πως στο σύνολο εκπαίδευσης μπορεί να υπάρχουν αρκετά περισσότερα δείγματα μίας κλάσης έναντι άλλων. Συνεπώς, αν δεν υπάρξει σωστή αντιμετώπιση, το μοντέλο κατά την εκπαίδευση ενδέχεται να παρουσιάσει **προκατάληψη (bias)** της κλάσης με τα περισσότερα δείγματα.

Ο πιο αποτελεσματικός τρόπος αντιμετώπισης αυτού του φαινομένου είναι η παραγωγή ή εύρεση αυθεντικών δεδομένων για την εξισορρόπηση των κλάσεων. Ωστόσο, αυτό τις περισσότερες περιπτώσεις υφίσταται αδύνατο. Επομένως, έχουν αναπτυχθεί διάφορες τεχνικές δειγματοληψίας (sampling techniques) που χωρίζονται σε δύο κατηγορίες μεθόδων, ονομαστικά την υπερδειγματοληψία (oversampling) και την υποδειγματοληψία (undersampling).

Η **υποδειγματοληψία** εφαρμόζεται στις μεγαλύτερες κλάσεις, δηλαδή αυτές με τα περισσότερα δείγματα, με σκοπό να αντισταθμιστούν με τις μικρότερες. Ωστόσο, αν τα δείγματα των μικρών κλάσεων είναι πολύ λίγα, τότε το σύνολο εκπαίδευσης μειώνεται σημαντικά και έτσι χάνεται πολύτιμη πληροφορία.

Η **υπερδειγματοληψία** μπορεί να γίνει με διάφορους τρόπους. Εκτός από την τυχαία δειγματοληψία, όπου τα υπάρχοντα δείγματα των μικρότερων κλάσεων απλώς αντιγράφονται τυχαία, υπάρχουν και άλλες τεχνικές που χρησιμοποιούνται για τη δημιουργία ισορροπημένων συνόλων δεδομένων. Για παράδειγμα, η συνθετική υπερδειγματοληψία (synthetic data generation), όπως η Συνθετική Τεχνική Υπερδειγματοληψίας Μειονοτήτων (Synthetic Minority Over-sampling Technique - SMOTE), επικεντρώνεται στη δημιουργία νέων δεδομένων για τις μικρότερες κλάσεις με βάση τα υπάρχοντα δείγματα, προσομοιάζοντας τα χαρακτηριστικά τους.

Μετρικές Αξιολόγησης Ακρίβειας

Οι μετρικές αξιολόγησης της ακρίβειας των προβλημάτων ταξινόμησης βασίζονται στην σύγκριση των προβλέψεων του μοντέλου και των πραγματικών ετικετών. Μερικοί από τους βασικότερους ορισμούς μετρικών παρουσιάζονται παρακάτω οι οποίοι βασίζονται στον Πίνακα Σύγχυσης (Conusion Matrix). Για λόγους απλότητας, επικεντρωνόμαστε στην περίπτωση της δυαδικής ταξινόμησης, όπου οι κλάσεις είναι δύο, οπότε τα δείγματα ανήκουν είτε στα θετικά είτε στα αρνητικά.

Πίνακας 2.1: *Πίνακας Σύγχυσης Δυαδικού Προβλήματος Ταξινόμησης*

Πραγματική Κλάση	**Προβλεπόμενη Κλάση**	
	Θετικά	**Αρνητικά**
Θετικά	Αληθώς Θετικά (True Positives - TP)	Ψευδώς Αρνητικά (False Negatives - FN)
Αρνητικά	Ψευδώς Θετικά (False Positives - FP)	Αληθώς Αρνητικά (True Negatives - TN)

$$\text{Ακρίβεια (accuracy)} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Βαθμολογία F1 (F1-score)} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

$$\text{Ευστοχία (precision)} = \frac{TP}{TP + FP}$$

$$\text{Ανάκληση (recall)} = \frac{TP}{TP + FN}$$

Σε πολυταξικά προβλήματα, η ακρίβεια αντιπροσωπεύει όλες τις κλάσεις και χρησιμεύει ως μία ολική μετρική για το μοντέλο. Ωστόσο, οι υπόλοιπες τρεις μετρικές υπολογίζονται για κάθε κλάση ξεχωριστά. Έπειτα, μπορούν να υπολογιστούν οι μέσοι όροι των μετρικών αυτών για κάθε κλάση, οπότε μία τιμή που θα αντιπροσώπευε μια γενικότερη επίδοση του μοντέλου θα ήταν η μέση τιμή (average) macro ή micro. Η macro δίνει την ίδια βαρύτητα σε κάθε

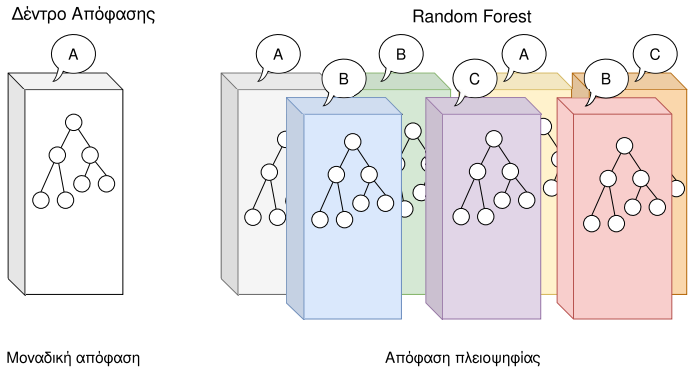
κλάση, ενώ η micro ρυθμίζει τα βάρη αναλόγως το μέγεθος της κάθε κλάσης.

Οπότε, στην περίπτωση ενός ανισόρροπου συνόλου δεδομένων, το macro-average δίνει μία πιο καθαρή εικόνα της απόδοσης του μοντέλου. Συμπερασματικά, η χρήση πολλαπλών μετρικών είναι απαραίτητη για την αξιολόγηση του μοντέλου, ειδικά όταν υπάρχουν ανισόρροπες κλάσεις.

2.1.4 Τυχαίο Δάσος Ταξινόμησης

Ο αλγόριθμος του Τυχαίου Δάσους Ταξινόμησης (Random Forest Classifier) βασίζεται στην έννοια του δέντρου αποφάσεων. Τα δέντρα αποφάσεων είναι ένας θεμελιώδης αλγόριθμος μηχανικής μάθησης για ταξινόμηση και παλινδρόμηση. Λειτουργούν δημιουργώντας μια δομή δέντρου, όπου κάθε κόμβος αντιπροσωπεύει μία ερώτηση ή μία απόφαση που βασίζεται σε ένα χαρακτηριστικό και κάθε κλάδος αντιπροσωπεύει τις πιθανές απαντήσεις σε αυτή την ερώτηση. Τα φύλλα του δέντρου αντιπροσωπεύουν την τελική πρόβλεψη.

Η εκπαίδευση ενός δέντρου απόφασης περιλαμβάνει την εύρεση των καλύτερων διαχωρισμών σε κάθε κόμβο για την ελαχιστοποίηση ενός μέτρου που ονομάζεται «gini index». Τα δέντρα αποφάσεων είναι ερμηνεύσιμα, που σημαίνει πως μπορούν εύκολα να οπτικοποιηθούν για την κατανόηση της λογικής πίσω από τις αποφάσεις τους. Για αυτό τον λόγο, τα δέντρα αποφάσεων, όπως και το τυχαίο δάσος, χρησιμοποιούνται επίσης για τον εντοπισμό σημαντικών χαρακτηριστικών στο σύνολο δεδομένων, κατατάσσοντας τα σύμφωνα με την χρησιμότητά τους στην διαδικασία της ταξινόμησης.



Σχήμα 2.3: *Κύρια Ιδέα του Τυχαίου Δάσους*

Τα τυχαία δάση λειτουργούν δημιουργώντας πολλά ανεξάρτητα δέντρα αποφάσεων χρησιμοποιώντας δειγματοληψία με αντικατάσταση (bootstrapping) από τα δεδομένα εκπαίδευσης και επιλέγοντας τυχαία υποσύνολα χαρακτηριστικών για διαχωρισμό σε κάθε κόμβο. Αυτή η τυχαιότητα μειώνει τη διακύμανση των προβλέψεων, καθώς κάθε δέντρο μπορεί να κάνει διαφορετικά λάθη. Στην περίπτωση ταξινόμησης, το τελικό αποτέλεσμα προκύπτει από την πλειοψηφία των προβλέψεων από όλα τα ξεχωριστά δέντρα, οδηγώντας σε καλύτερη γενίκευση σε νέα δεδομένα [10].

2.1.5 Νευρωνικά Δίκτυα

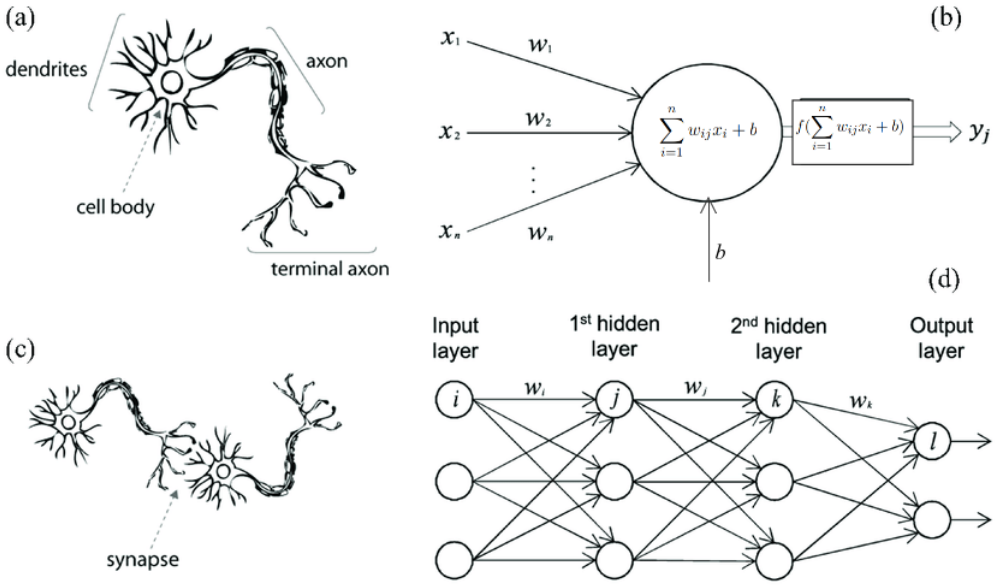
Τα Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks - ANNs) βασίστηκαν στο γεγονός ότι ο ανθρώπινος εγκέφαλος εκτελεί υπολογισμούς με εντελώς διαφορετικό τρόπο

από τον συμβατικό ψηφιακό υπολογιστή. Ο εγκέφαλος είναι ένας εξαιρετικά πολύπλοκος, μη γραμμικός, παράλληλος υπολογιστής, ο οποίος έχει τη δυνατότητα να κατασκευάζει δικούς του κανόνες συμπεριφοράς μέσω αυτού που αποκαλούμε «εμπειρία» [11].

Ένα τεχνητό νευρωνικό δίκτυο μοιάζει με τον ανθρώπινο εγκέφαλο σε δύο σημεία. Πρώτων, ότι το δίκτυο προσλαμβάνει τη γνώση από το περιβάλλον του μέσω μιας διαδικασίας μάθησης και δεύτερον ότι η ισχύς των συνδέσεων μεταξύ των νευρώνων, που αποκαλείται συναπτικό βάρος, χρησιμοποιείται για την αποθήκευση της γνώσης που αποκτιέται.

Τεχνητός Νευρώνας

Η βασική δομή ενός τεχνητού νευρωνικού δικτύου είναι ο Τεχνητός Νευρώνας (Perceptron). Ένας τεχνητός νευρώνας λαμβάνει εισροές ως ερεθίσματα από το περιβάλλον και μετά από έναν γραμμικό συνδυασμό αυτών παράγει μία έξοδο. Αντίστοιχα, ένα νευρωνικό δίκτυο αποτελείται από πολλούς τεχνητούς νευρώνες, όπου χωρίζονται σε επίπεδα στα οποία η έξοδος των νευρώνων ενός επιπέδου τροφοδοτεί τις εισόδους των νευρώνων του επόμενου επιπέδου.



Σχήμα 2.4: *Σύγκριση Βιολογικού και Τεχνητού Νευρώνα: (a) Ανθρώπινος Νευρώνας (b) Τεχνητός Νευρώνας (c) Βιολογική Σύναψη (d) Συνάψεις Τεχνητού Νευρωνικού Δικτύου [1]*

Ο τεχνητός νευρώνας αποτελείται από 4 στοιχεία:

1. **Συνάψεις**, οι οποίες χαρακτηρίζονται από την δική τους τιμή βάρους (w).

2. **Αθροιστής**, ο οποίος αθροίζει τον συνδυασμό των εισόδων με τα βάρη των συνάψεων. Πιο συγκεκριμένα, οι τιμές εισόδου (x_i) πολλαπλασιάζονται με τις τιμές βάρους.

3. **Συνάρτηση Ενεργοποίησης** (f), η οποία χρησιμεύει για τον περιορισμό του πλάτους του συστήματος εξόδου ενός νευρώνα. Υπάρχουν πολλαπλές συναρτήσεις ενεργοποίησης που μπορούν να επιλεχθούν για την ανάπτυξη ενός τεχνητού νευρώνα όπως περιγράφουμε στην συνέχεια.

4. **Πόλωση** (b), η οποία προκαλεί μία προκατάληψη στο αποτέλεσμα της συνάρτησης ενεργοποίησης.

Η έξοδος ενός τεχνητού νευρώνα δίνεται από την σχέση:

$$y = f(\sum_{i=1}^{n} w_{ij}x_i + b)$$

όπου το y συμβολίζει την έξοδο του νευρώνα, το w_i και x_i το βάρος και την είσοδο της i-στης σύναψης, το f στην συνάρτηση ενεργοποίησης και το b την πόλωση.

Συναρτήσεις Ενεργοποίησης

Οι συναρτήσεις ενεργοποίησης είναι κρίσιμες για τη λειτουργία των τεχνητών νευρωνικών δικτύων. Χρησιμοποιούνται για να εισάγουν μη γραμμικότητα στα μοντέλα, επιτρέποντάς τους να μάθουν και να εκπροσωπούν σύνθετες σχέσεις στα δεδομένα. Διαφορετικές συναρτήσεις ενεργοποίησης έχουν διαφορετικές ιδιότητες και χρησιμοποιούνται για διαφορετικούς σκοπούς ανάλογα με το πρόβλημα που προσπαθούμε να λύσουμε και τη δομή του νευρωνικού δικτύου. Υπάρχουν πολλές τέτοιες συναρτήσεις αλλά επικεντρωνόμαστε σε τρεις βασικές:

Η **σιγμοειδής (Sigmoid)** μετασχηματίζει την είσοδο σε ένα εύρος τιμών από 0 έως 1. Χρησιμοποιείται συχνά για δυαδικά προβλήματα ταξινόμησης.

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

Η **Softmax** χρησιμοποιείται συνήθως σε προβλήματα ταξινόμησης με πολλές κατηγορίες. Μετασχηματίζει τις εισόδους σε πιθανότητες που αθροίζουν στο 1.

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad \text{για } i = 1, \ldots, K$$

όπου $\mathbf{z}$ είναι το διάνυσμα εισόδων και K είναι ο αριθμός των κατηγοριών.

Η **ReLu (Rectified Linear Unit)** ενεργοποιεί τις θετικές εισόδους και απενεργοποιεί τις αρνητικές. Είναι πολύ δημοφιλής λόγω της αποτελεσματικότητάς της στην εκπαίδευση βαθιών νευρωνικών δικτύων.

$$\text{ReLU}(x) = \max(0, x)$$

Εκπαίδευση Νευρωνικών Δικτύων

Συνήθως, η εκπαίδευση των ANNs βασίζεται σε δύο κύριες μεθόδους: την οπίσθια διάδοση σφάλματος (backpropagation) και την κλίση κατάβασης (gradient descent). Αυτές οι τεχνικές συνδυάζονται για να προσαρμόζουν τα βάρη των συνάψεων, ώστε να ελαχιστοποιείται το σφάλμα μεταξύ της παραγόμενης και της επιθυμητής εξόδου.

Η διαδικασία της διάδοσης σφάλματος αποτελείται από δύο βασικά στάδια: την πρόσθια διέλευση (forward pass) στην οποία υπολογίζεται η έξοδος κάθε νευρώνα από τα δεδομένα εισόδου, καθώς και την οπίσθια διέλευση (backward pass) στην οποία υπολογίζονται οι παράγωγοι του σφάλματος ως προς τα βάρη των συνάψεων διαδίδοντας το σφάλμα προς την

αντίθετη κατεύθυνση. Οι παράγωγοι αυτές χρησιμοποιούνται για την προσαρμογή των βαρών των συνάψεων.

Μετά τον υπολογισμό των παραγώγων του σφάλματος, η κλίση κατάβασης χρησιμοποιείται για να ενημερώσει τα βάρη του δικτύου. Οι νέες τιμές των βαρών υπολογίζονται έτσι ώστε το σφάλμα να μειώνεται με έναν συγκεκριμένο ρυθμό μάθησης η (learning rate), ο οποίος καθορίζει την ταχύτητα σύγκλισης του αλγορίθμου.

Αυτές οι διαδικασίες επαναλαμβάνονται για πολλούς κύκλους εκπαίδευσης (epochs), μέχρι το δίκτυο να συγκλίνει σε μια κατάσταση όπου το σφάλμα είναι ελάχιστο ή μέχρι να εξαντληθεί ορισμένος αριθμός εποχών που έχει οριστεί από τον χρήστη. Κατά την εκπαίδευση, συνήθως παρακολουθείται η εξέλιξη του σφάλματος και η ακρίβεια του μοντέλου, και η εκπαίδευση μπορεί να σταματήσει νωρίτερα αν δεν παρατηρηθεί σημαντική βελτίωση.

2.2 Βασικά Είδη Βαθιών Νευρωνικών Δικτύων

2.2.1 Πολυεπίπεδα Perceptrons

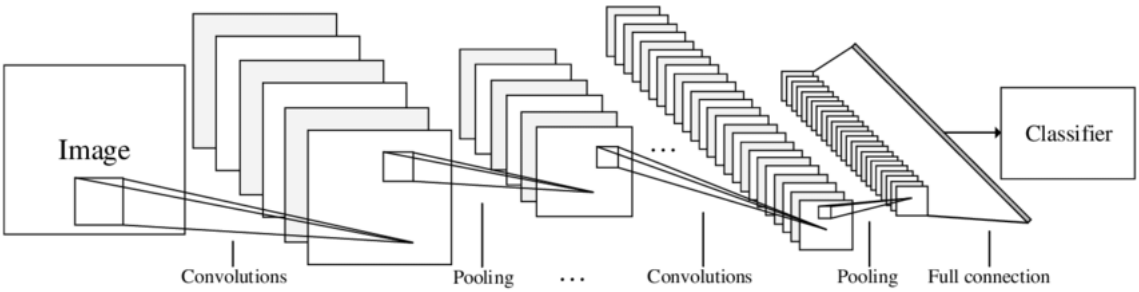
Τα Πολυεπίπεδα Perceptrons (Multilayer Perceptrons - MLPs) αποτελούν το πιο απλό είδος νευρωνικών δικτύων, τα οποία αναπτύχθηκαν μετά από μία προσπάθεια να βελτιώσουν τα perceptrons (single-layer perceptrons), τα οποία έχουν μόνο την δυνατότητα να αναγνωρίσουν γραμμικώς διαχωρίσιμα δεδομένα. Για αυτό, είναι ευέλικτα μοντέλα κατάλληλα για ένα ευρύ φάσμα προβλημάτων ταξινόμησης και παλινδρόμησης. Διαπρέπουν στην εκμάθηση πολύπλοκων, μη γραμμικών σχέσεων μεταξύ των χαρακτηριστικών εισόδου και των εξόδων.

Ένα MLP αποτελείται από πολλαπλά στρώματα, τουλάχιστον τρία (επίπεδο εισόδου, ένα ή περισσότερα κρυφά επίπεδα και επίπεδο εξόδου), διασυνδεδεμένα με τρόπο πρόσθιας τροφοδότησης (Feedforward Networks). Κάθε νευρώνας σε ένα επίπεδο συνδέεται με όλους τους νευρώνες του επόμενου επιπέδου, και αντίστοιχα οι είσοδοί είναι οι έξοδοι όλων των νευρώνων του προηγούμενου επιπέδου. Αυτά τα επίπεδα ονομάζονται Πλήρως Συνδεδεμένα Επίπεδα (Dense ή Fully Connected Layer).

2.2.2 Συνελικτικά Νευρωνικά Δίκτυα

Τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNNs) είναι ειδικά σχεδιασμένα για επεξεργασία εικόνων ή χωρικών δεδομένων. Η αρχιτεκτονική τους ενσωματώνει συνελικτικά επίπεδα που εξάγουν τοπικά χαρακτηριστικά και μοτίβα από την είσοδο, καθιστώντας τα εξαιρετικά αποτελεσματικά για εφαρμογές σχετικές με την όραση υπολογιστών, όπως η αναγνώριση εικόνων. Η ονομασία τους προέρχεται από το γεγονός ότι ένα ή περισσότερα από τα επίπεδά τους χρησιμοποιούν την πράξη της συνέλιξης. Βασίζονται σε τρία βασικά είδη επιπέδων:

1. **Συνελικτικά (Convolutional)**: Ο βασικός μηχανισμός είναι η συνέλιξη. Εφαρμόζεται στην είσοδο με ένα σύνολο μικρών φίλτρων (ή πυρήνων). Αυτά τα φίλτρα σαρώνουν τον πίνακα εισόδου κατά πλάτος και κατά μήκος και δημιουργούν νέα χαρακτηριστικά μιας πιο αφαιρετικής αναπαράστασης.



Σχήμα 2.5: *Βασική Δομή Συνελικτικού Νευρωνικού Δικτύου [2]*

2. **Υποδειγματοληψία (Pooling)**: Μετά από κάθε συνέλιξη, συνήθως ακολουθεί ένα επίπεδο υποδειγματοληψίας. Στην υποδειγματοληψία, ο πίνακας χωρίζεται σε περιοχές και εξάγεται ένα αντιπροσωπευτικό στοιχείο από κάθε περιοχή, όπως το μέγιστο (Max Pooling) στοιχείο ή ο μέσος όρος (Average Pooling), ξανά με χρήση φίλτρου. Αυτό μειώνει τη διαστατικότητα των δεδομένων, κρατώντας παράλληλα την πληροφορία των σημαντικών χαρακτηριστικών.

3. **Πλήρως συνδεδεμένα επίπεδα**: Το τελευταίο τμήμα ενός CNN αποτελείται από πλήρως συνδεδεμένα επίπεδα, όπου γίνεται η ταξινόμηση ή η πρόβλεψη. Αυτά τα επίπεδα λειτουργούν όπως τα κλασικά νευρωνικά δίκτυα (MLP) και συνδέονται με όλα τα χαρακτηριστικά που έχουν εξαχθεί προηγουμένως.

2.2.3 Μετασχηματιστές

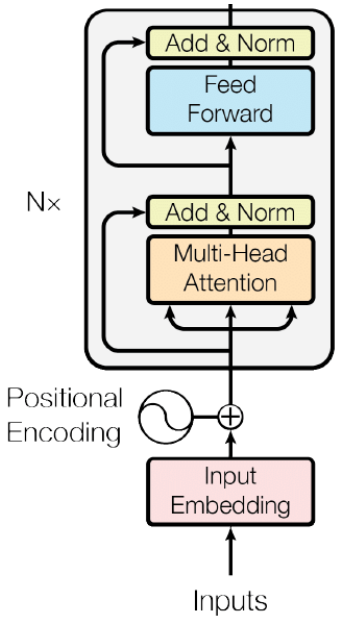
Ο Μετασχηματιστής (Transformer) είναι μία αρχιτεκτονική νευρωνικού δικτύου η οποία παρουσιάστηκε στο διάσημο άρθρο «Attention Is All You Need» από τους Vaswani et al. το 2017 [12]. Βασίζονται στον μηχανισμό της αυτο-προσοχής (self-attention) και είναι ένας τύπος νευρωνικού δικτύου που επεξεργάζεται δεδομένα εισόδου ταυτόχρονα, χωρίς να χρειάζεται να διαβάσει τα δεδομένα με τη σειρά. Αυτό τον καθιστά ιδιαίτερα αποτελεσματικό και ταχύ για εφαρμογές όπως η μετάφραση κειμένου.

Η κλασική δομή του μετασχηματιστή περιλαμβάνει κωδικοποιητές και αποκωδικοποιητές. Εμείς επικεντρωνόμαστε στη δομή του κωδικοποιητή, αναφερόμενοι σε αυτόν ως μετασχηματιστή για απλότητα, ο οποίος αποτελείται από πολλαπλά στρώματα (N), καθένα με δύο βασικά υποσυστήματα:

- **Μηχανισμός Αυτο-Προσοχής (Self-Attention Mechanism)**: Δέχεται έναν πίνακα εισόδου που περιέχει διανύσματα αναπαράστασης λέξεων (word embeddings). Κάθε λέξη στην είσοδο συγκρίνεται με κάθε άλλη λέξη μέσω μιας συνάρτησης προσοχής (attention function), η οποία υπολογίζει τη σημασία της κάθε λέξης ως προς τις άλλες. Η έξοδος είναι ένας πίνακας με τις ίδιες διαστάσεις, αλλά με βελτιωμένες αναπαραστάσεις που λαμβάνουν υπόψη τις σχέσεις μεταξύ όλων των λέξεων στην είσοδο.

- **Πλήρως Συνδεδεμένο Δίκτυο (Feedforward Neural Network)**: Δέχεται την έξοδο του μηχανισμού αυτο-προσοχής και παράγει την τελική αναπαράσταση κάθε λέξης. Κάθε λέξη περνάει μέσα από ένα μικρό νευρωνικό δίκτυο που περιλαμβάνει ένα

σύνολο πλήρως συνδεδεμένων στρωμάτων με μια μη γραμμική συνάρτηση ενεργοποίησης, όπως η ReLU.



Σχήμα 2.6: *Δομή του Μετασχηματιστή [3]*

Συνοπτικά, η συνολική διαδικασία είναι η παρακάτω:

1. Εισαγωγή Λέξεων (Word Embeddings): Ο κωδικοποιητής δέχεται έναν πίνακα εισόδου που περιέχει διανύσματα αναπαράστασης λέξεων.

2. Προσθήκη Θέσης (Positional Encoding): Προστίθεται μια πληροφορία θέσης στα διανύσματα αναπαράστασης για να διατηρηθεί η σειρά των λέξεων.

3. Εφαρμογή Αυτο-Προσοχής (Self-Attention): Ο μηχανισμός αυτο-προσοχής υπολογίζει τις βαρύτητες μεταξύ των λέξεων και βελτιώνει τις αναπαραστάσεις τους.

4. Κανονικοποίηση και Προσθήκη Υπολοίπου (Normalization and Residual Connection): Προσθέτει την αρχική είσοδο πίσω στην έξοδο της αυτο-προσοχής και εφαρμόζει κανονικοποίηση (layer normalization).

5. Πλήρως Συνδεδεμένο Νευρωνικό Δίκτυο (Feedforward Neural Network): Η έξοδος του μηχανισμού αυτο-προσοχής περνάει από ένα πλήρως συνδεδεμένο δίκτυο για την τελική επεξεργασία, προσθέτοντας μη γραμμικότητα και εμπλουτίζοντας τις αναπαραστάσεις.

6. Κανονικοποίηση και Προσθήκη Υπολοίπου: Προσθέτει την είσοδο του πλήρως συνδεδεμένου δικτύου πίσω στην έξοδο και εφαρμόζει ξανά κανονικοποίηση.

7. Επαναληπτικά Στρώματα (Repeated Layers): Η διαδικασία αυτή επαναλαμβάνεται για τα πολλαπλά στρώματα του κωδικοποιητή, βελτιώνοντας συνεχώς τις αναπαραστάσεις.

Αυτά τα βήματα επιτρέπουν στον μετασχηματιστή να μάθει πολύπλοκες σχέσεις και δομές στα δεδομένα εισόδου, καθιστώντας τον ιδιαίτερα ισχυρό και αποτελεσματικό σε σχέση με άλλες αρχιτεκτονικές.

2.3 Βαθιά Μάθηση και Δομημένα Σύνολα Δεδομένων

Τα προβλήματα με τα δομημένα σύνολα δεδομένων (ή ετερογενή δεδομένα ή δεδομένα πίνακα) αποτελούν προς το παρόν ένα από τα τελευταία άλυτα ζητήματα της έρευνας στην βαθιά μάθηση. Ενώ οι πιο πρόσφατες εξελίξεις στη φυσική γλώσσα, την όραση και την ομιλία έχουν επιτευχθεί με βαθιά μοντέλα, η επιτυχία τους στον τομέα των δομημένων δεδομένων δεν είναι ακόμα πειστική. [7][13].

2.3.1 Ιστορική Αναδρομή

Τα δομημένα σύνολα δεδομένων είναι μία από τις πιο παλιές μορφές δεδομένων που αναλύονταν στατιστικά. Πριν την ψηφιακή εποχή, η συλλογή κειμένων, εικόνων και ηχητικών αρχείων δεν ήταν δυνατή και σχεδόν όλα τα δεδομένα ήταν σε μορφή πίνακα. Παραδοσιακές τεχνικές μηχανικής μάθησης ήταν μονόδρομος για την εξαγωγή κρυφών μοτίβων στα δεδομένα. Ωστόσο, μετά την έγερση της βαθιάς μάθησης, η επιστημονική κοινότητα άρχιζε να εστιάζει στα μη δομημένα δεδομένα, όπως οι εικόνες. Έτσι, δεν υπήρξε κάποια σημαντική εξέλιξη στην ανάλυση των δεδομένων πίνακα για ένα σημαντικό διάστημα [13].

Ωστόσο, μετά την «άνθιση» του ηλεκτρονικού εμπορίου, νέες προκλήσεις ήρθαν στην επιφάνεια. Οι παραδοσιακές μέθοδοι μηχανικής μάθησης δεν είχαν ικανοποιητικά αποτελέσματα στα ετερογενή σύνολα δεδομένων που είχαν μεγάλη διαστατικότητα (πολλά χαρακτηριστικά). Αυτό πυροδότησε ένα ανανεωμένο ενδιαφέρον για την βαθιά μάθηση. Χαρακτηριστικό παράδειγμα αποτελεί το πρόβλημα της πρόβλεψης της αναλογίας «κλικ» προς αριθμό εμφανίσεων (Clickthrough rate - CTR), η οποία δείχνει πόσο συχνά οι χρήστες που βλέπουν μία διαφήμιση καταλήγουν να κάνουν «κλικ» σε αυτή [14]. Το CTR έγινε σημαντικός τομέας έρευνας που οδήγησε στην ανάπτυξη εξειδικευμένων αρχιτεκτονικών βαθιάς μάθησης που έχουν σχεδιαστεί ειδικά για την αντιμετώπιση ετερογενών δεδομένων [13].

Επίσης, η πρόσφατη επιτυχία των μοντέλων που βασίζονται στην αυτο-προσοχή, όπως οι μετασχηματιστές στην επεξεργασία κειμένου και εικόνας, οδήγησε στην εφαρμογή τους και στον τομέα των δεδομένων πίνακα. Οι ερευνητές διερευνούν ενεργά αρχιτεκτονικές μετασχηματιστών ιδιαίτερα για το χειρισμό πολύ μεγάλων συνόλων δεδομένων σε μορφή πίνακα[15][16].

2.3.2 Δυσκολίες στην εκπαίδευση

Είναι συχνά ασαφές γιατί η βαθιά μάθηση δεν μπορεί να επιτύχει το ίδιο επίπεδο προγνωστικής ικανότητας σε σχέση σε άλλους τομείς, όπως η ταξινόμηση εικόνων και η επεξεργασία φυσικής γλώσσας. Εντοπίζονται τέσσερις πιθανοί λόγοι [13].

1. **Ποιότητα δεδομένων εκπαίδευσης**: Τα δεδομένα πινάκων συχνά υποφέρουν από ελλιπείς ή ακραίες τιμές. Ακόμα, μερικές φορές αυτά τα σύνολα έχουν μικρό μέγεθος σε σχέση με την πολυπλοκότητα των χαρακτηριστικών. Ενώ αυτά τα ζητήματα επηρεάζουν όλους τους αλγόριθμους, οι παραδοσιακές μέθοδοι όπως τα δέντρα αποφάσεων μπορούν να τα χειριστούν πιο αποτελεσματικά.

2. **Έλλειψη χωρικών εξαρτήσεων**: Τα μοντέλα βαθιάς μάθησης έχουν σχεδιαστεί για δεδομένα που είναι χωρικά εξαρτώμενα. Σε αντίθεση με τις εικόνες, τα δεδομένα

πίνακα καθιστούν δύσκολο στα νευρωνικά μοντέλα να να μάθουν αυτές τις σύνθετες και μη κανονικές εξαρτήσεις από την αρχή.

3. **Εξάρτηση από προεπεξεργασία**: Η απόδοση στα δεδομένα πίνακα εξαρτάται σε μεγάλο βαθμό από τις επιλογές προεπεξεργασίας, ειδικά για κατηγορικά χαρακτηριστικά. Η προεπεξεργασία μπορεί να οδηγήσει σε απώλεια πληροφοριών και να εμποδίσει την απόδοση.

4. **Σημασία μεμονωμένων χαρακτηριστικών**: Ενώ η αλλαγή της κλάσης μίας εικόνας απαιτεί μία συντονισμένη αλλαγή σε πολλά χαρακτηριστικά, δηλαδή σε πολλά pixels, η μικρότερη αλλαγή ενός κατηγορικού, και ειδικά δυαδικού, χαρακτηριστικού μπορεί να ανατρέψει εντελώς μία πρόβλεψη. Σε αντίθεση με τα βαθιά νευρωνικά δίκτυα, τα δέντρα αποφάσεων μπορούν να χειριστούν εξαιρετικά καλά τέτοιου είδους καταστάσεις.

Ο μετασχηματισμός δομημένων δεδομένων σε μη δομημένων, με τρόπο που να μπορεί να αξιοποιηθεί από τα CNNs και Transformer, αποτελεί πρόκληση. Ωστόσο, σύμφωνα με τις μελέτες [16] και [15], έχουν προταθεί λύσεις που επιτρέπουν αυτή τη μετατροπή. Συγκεκριμένα, οι παραπάνω μελέτες αναφέρουν ότι τα δεδομένα πίνακα μπορούν να μετατραπούν σε διανύσματα (embeddings), τα οποία αντιπροσωπεύουν τις αριθμητικές και κατηγορικές τιμές των δεδομένων με τρόπους που διατηρούν τις δομές και τις σχέσεις τους. Αυτές οι ενσωματώσεις μπορούν να τροφοδοτηθούν σε Transformers, οι οποίοι με τη σειρά τους μπορούν να επεξεργαστούν τα δεδομένα με τους μηχανισμούς αυτο-προσοχής. Παρομοίως, τα CNNs μπορούν να χρησιμοποιηθούν για την επεξεργασία των δομημένων δεδομένων, μετά από μετασχηματισμό τους σε πολυδιάστατους πίνακες με διάφορες τεχνικές, δηλαδή σαν εικόνες, και εφαρμόζοντας φίλτρα για την εξαγωγή χαρακτηριστικών.

2.3.3 Αναγκαιότητα Χρήσης Βαθιών Νευρωνικών Δικτύων

Η ενασχόληση με μοντέλα βαθιάς μάθησης για την επεξεργασία δομημένων δεδομένων, παρά την ανωτερότητα των παραδοσιακών αλγορίθμων μηχανικής μάθησης σε αυτού του είδους δεδομένα, κρίνεται απαραίτητη. Η αναγκαιότητα αυτή προκύπτει κυρίως λόγω της περιορισμένης υποστήριξης των παραδοσιακών αλγορίθμων μηχανικής μάθησης από εργαλεία λογισμικού που είναι σχεδιασμένα για συσκευές περιορισμένων πόρων, όπως οι μικροελεγκτές. Ενώ μερικοί αλγόριθμοι μηχανικής μάθησης μπορούν να εκτελεστούν σε αυτές τις συσκευές, οι επιλογές είναι περιορισμένες και δεν καλύπτουν την πλήρη γκάμα των διαθέσιμων τεχνικών.

Αντιθέτως, τα βαθιά νευρωνικά δίκτυα υποστηρίζονται πλήρως από διάφορα εργαλεία σχεδιασμένα ειδικά για συσκευές περιορισμένων πόρων, όπως το TensorFlow Lite[17]. Αυτά τα εργαλεία επιτρέπουν την υλοποίηση και εκτέλεση ποικίλων δομών νευρωνικών δικτύων, προσφέροντας μεγαλύτερη ευελιξία στην έρευνα και την ανάπτυξη.

Συνεπώς, είναι ζωτικής σημασίας η έρευνα να επικεντρωθεί στο κατά πόσο τα μοντέλα βαθιάς μάθησης μπορούν να φτάσουν ή και να ξεπεράσουν τις επιδόσεις των παραδοσιακών αλγορίθμων μηχανικής μάθησης. Δεδομένης της ευελιξίας και της δυνατότητας ευρείας εφαρμογής των βαθιών νευρωνικών δικτύων σε συσκευές περιορισμένων πόρων, είναι αναγκαίο να διερευνηθούν πλήρως οι δυνατότητές τους σε διάφορα περιβάλλοντα και εφαρμογές.

Κεφάλαιο 3

Συστήματα Ανίχνευσης Εισβολών και Διαδίκτυο των Πραγμάτων

Καθώς ο κόσμος γίνεται όλο και πιο συνδεδεμένος μέσω του διαδικτύου, η ασφάλεια των πληροφοριακών συστημάτων αποκτά κεντρική σημασία. Η αυξανόμενη πολυπλοκότητα των δικτύων και των συσκευών που τα απαρτίζουν καθιστά απαραίτητη την ανάπτυξη προηγμένων συστημάτων για την προστασία τους από κακόβουλες ενέργειες. Στο πλαίσιο αυτό, τα Συστήματα Ανίχνευσης Εισβολών (Intrusion Detection Systems - IDS) αναδύονται ως κρίσιμα εργαλεία για την αναγνώριση και την αντιμετώπιση απειλών.

Παράλληλα, το Διαδίκτυο των Πραγμάτων (Internet of Things - IoT) ανοίγει νέους ορίζοντες στην τεχνολογία, επιτρέποντας τη διασύνδεση και την επικοινωνία εκατομμυρίων συσκευών και συστημάτων σε παγκόσμιο επίπεδο. Αυτή η εξέλιξη δημιουργεί νέες προκλήσεις και ανάγκες στον τομέα της ασφάλειας, καθώς οι απειλές μπορούν να επηρεάσουν όχι μόνο τα δεδομένα αλλά και τις φυσικές λειτουργίες των συνδεδεμένων συσκευών.

Το παρόν κεφάλαιο εστιάζει στην αναλυτική παρουσίαση των IDS και τη σημασία τους για την ασφάλεια των δικτύων, με ιδιαίτερη έμφαση στις προκλήσεις που παρουσιάζονται σε περιβάλλοντα IoT. Αρχικά, θα εξετάσουμε τη λειτουργία και τις κατηγορίες των IDS, καθώς και τη μέθοδο αξιολόγησής τους. Στη συνέχεια, θα αναλύσουμε τη ραγδαία εξέλιξη και τις εφαρμογές του IoT, αναδεικνύοντας τις ιδιαίτερες απαιτήσεις και τις προκλήσεις που δημιουργεί η ενσωμάτωσή του σε διάφορους τομείς της καθημερινής ζωής και της βιομηχανίας.

3.1 Συστήματα Ανίχνευσης Εισβολών

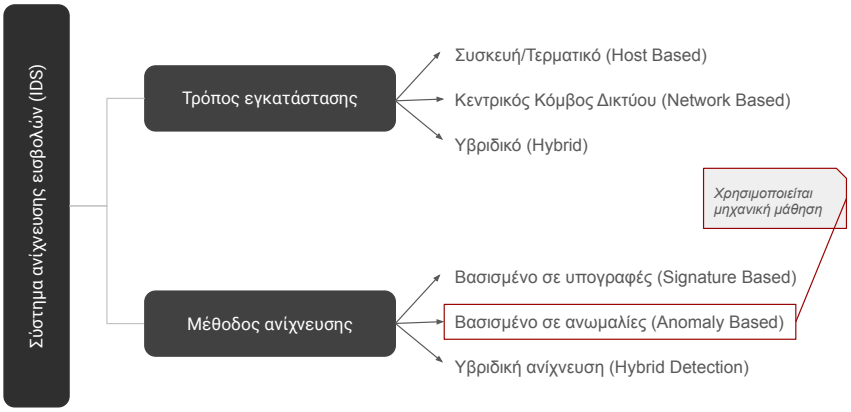
Καθώς όλο και περισσότεροι άνθρωποι χρησιμοποιούν το διαδίκτυο για προσωπικούς ή επαγγελματικούς λόγους, οι διαφορετικές κυβερνοεπιθέσεις και οι παραβιάσεις αυξάνονται καθημερινά. Το IDS αποτελεί μία από τις πιο σημαντικές παραμέτρους της κυβερνοασφάλειας, αφού χρησιμοποιείται για την αναγνώριση επιτυχημένων παραβιάσεων. Ο όρος «σύστημα ανίχνευσης εισβολών» χρησιμοποιήθηκε για πρώτη φορά από τον James Anderson στις αρχές της δεκαετίας του ΄80. Εισήγαγε την έννοια της ανίχνευσης κακόβουλης χρήσης και έθεσε τις βάσεις για τον μελλοντικό σχεδιασμό και την ανάπτυξη των IDS.

Ένα IDS είναι λογισμικό ή υλικό σχεδιασμένο να εντοπίζει οποιαδήποτε κακόβουλη δραστηριότητα ή επίθεση εναντίον του συστήματος ή του δικτύου. Συλλέγει δεδομένα από διάφορες πηγές μέσα σε έναν υπολογιστή ή δίκτυο, όπως εντολές συστήματος και αρχεία κατα-

γραφής πακέτων δικτύου. Στη συνέχεια, τα αναλύει για να εντοπίσει πιθανές παραβιάσεις του συστήματος[18].

3.1.1 Κατηγοριοποίηση

Η κατηγοριοποίηση των IDS αποτελεί σημαντικό βήμα για την κατανόηση και την επιλογή του κατάλληλου συστήματος για συγκεκριμένες ανάγκες ασφαλείας. Τα IDS μπορούν να διακριθούν με βάση δύο κύριες παραμέτρους: τον τρόπο εγκατάστασής τους (Deployment Method) και τη μέθοδο ανίχνευσης (Detection Method) που χρησιμοποιούν [19] [18].



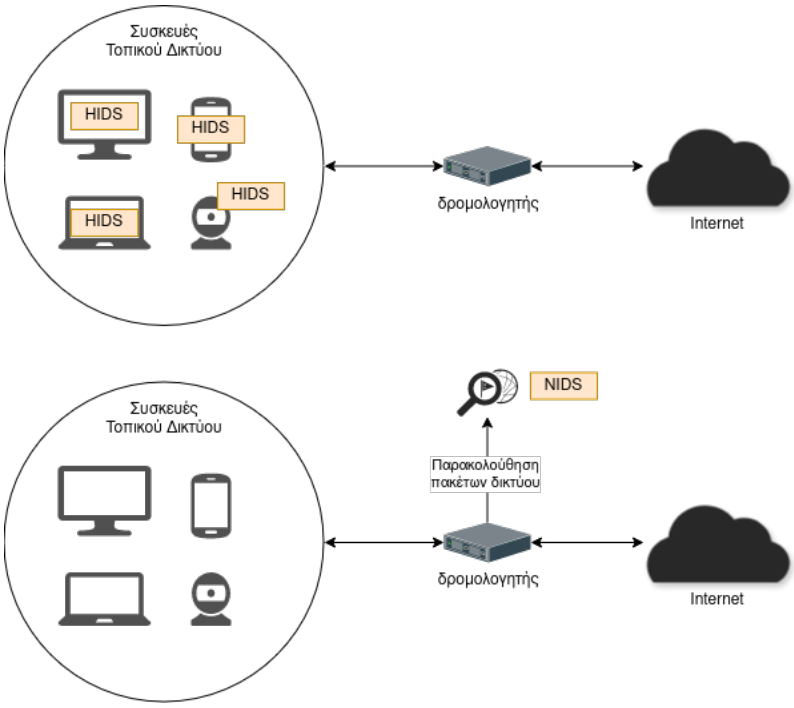
Σχήμα 3.1: *Κατηγοριοποίηση Συστημάτων Ανίχνευσης Εισβολών*

Τρόπος Εγκατάστασης

- **Host Based**: Το Host Based IDS - HIDS εγκαθίσταται ξεχωριστά σε συσκευές/τερματικά του δικύου και έχει σχεδιαστεί να παρακολουθεί, να αναλύει και να προστατεύει ένα σύστημα από εσωτερικές και εξωτερικές απειλές. Η ορατότητα ενός HIDS περιορίζεται στη συσκευή που είναι εγκατεστημένο, γεγονός που σημαίνει ότι μπορεί να παρακολουθήσει μόνο επιθέσεις που μπορεί να συμβαίνουν σε αυτή, χωρίς να γνωρίζει την κατάσταση του υπόλοιπου δικτύου στο οποίο ανήκει.

- **Network Based**: Τα Network Based IDS - NIDS τοποθετούνται σε έναν κεντρικό κόμβο ή σε πολλαπλούς στρατηγικά σε ολόκληρο το δίκτυο για να παρακολουθούν όλες τις συσκευές που συνδέονται σε αυτό και την κίνηση δεδομένων τους. Σε αντίθεση με τα HIDS, έχουν ευρύτερη ορατότητα και παρακολουθούν όλη την κίνηση του δικτύου.

Στον Πίνακα 3.1 παρουσιάζουμε τα πλεονεκτήματα και τους περιορισμούς των δύο παραπάνω στρατηγικών τοποθέτησης. Τα Hybrid συστήματα ανίχνευσης εισβολών συνδυάζουν υποσυστήματα HIDS και NIDS.



Σχήμα 3.2: *Διαφορά Μεταξύ IDS σε Επίπεδο Συσκευής και Δικτύου*

Μέθοδος Ανίχνευσης

- **Σύστημα Ανίχνευσης Εισβολών Βασισμένο σε Υπογραφές (Signature Based IDS - SIDS)**: Αναγνωρίζει γνωστές επιθέσεις συγκρίνοντας ύποπτη δραστηριότητα με υπάρχουσες υπογραφές στην βάση δεδομένων του. Είναι αποτελεσματικό για γνωστές απειλές αλλά αποτυγχάνει σε νέες επιθέσεις (zero day attacks). Πλεονεκτεί όμως σε ταχύτητα επεξεργασίας και περιορισμό ψευδών συναγερμών (false alarms).

- **Σύστημα Ανίχνευσης Εισβολών Βασισμένο σε Ανωμαλίες (Anomaly Based IDS - AIDS)**: Έχει το πλεονέκτημα να εντοπίζει και να αναφέρει ύποπτη άγνωστη συμπεριφορά. Χρησιμοποιεί μηχανική μάθηση για την μοντελοποίηση της κανονικής συμπεριφοράς του δικτύου με σκοπό να αναφέρει οποιαδήποτε σημαντική απόκλιση ως πιθανή επίθεση[20].

Συνοπτικά, τα SIDS είναι γρήγορα και αποτρέπουν τους ψευδείς συναγερμούς για γνωστές απειλές, ενώ τα AIDS ανιχνεύουν και νέες επιθέσεις αλλά χρειάζονται συνεχή εκπαίδευση.

3.1.2 Αξιολόγηση

Από την πρώτη τους εμφάνιση, τα IDS έχουν αξιολογηθεί με διάφορους τρόπους, χρησιμοποιώντας διαφορετικά σύνολα δεδομένων. Γενικά, ένα IDS μπορεί να αξιολογηθεί από δύο κύριες οπτικές γωνίες[21]:

1. **Αποδοτικότητα**: Αφορά τους πόρους που πρέπει να διατεθούν στο σύστημα, συμπεριλαμβανομένων των κύκλων CPU και της κύριας μνήμης.

	IDS Επιπέδου Συσκευής	**IDS Επιπέδου Δικτύου**
Πλεονεκτήματα	• Λειτουργεί τοπικά στην συσκευή προστατεύοντας έτσι ευαίσθητα δεδομένα. • Μπορεί αποτελεσματικά να εντοπίσει τοπικά συμβάντα. • Είναι ανεξάρτητο από τις υπόλοιπες συσκευές στο δίκτυο, χωρίς να βασίζεται η ασφάλεια της συσκευής σε έναν κεντρικό κόμβο.	• Έχουν πλήρη εικόνα για την κατάσταση του δικτύου και μπορούν να ανιχνεύσουν επιθέσεις που επηρεάζουν συνολικά το δίκτυο. • Είναι εγκατεστημένα σε υπολογιστές που έχουν αρκετούς υπολογιστικούς πόρους και έτσι μπορούν να αντέξουν πιο «βαριές» διεργασίες ανάλυσης και ανίχνευσης. • Είναι συστήματα ανεξάρτητα από το λειτουργικό περιβάλλον κάθε δικτυακής συσκευής.
Περιορισμοί	• Η ορατότητα του περιορίζεται στην συσκευή που είναι εγκατεστημένο και έτσι μπορεί να αδυνατεί να εντοπίσει επιθέσεις που επηρεάζουν συνολικά το δίκτυο. • Είναι εγκατεστημένα σε συσκευές που έχουν περιορισμένους υπολογιστικούς και ενεργειακούς πόρους. • Πρέπει να διαμορφώνεται και να διαχειρίζεται ξεχωριστά για την κάθε συσκευή.	• Ενδέχεται να εμφανιστούν καθυστερήσεις ή και διακοπές λειτουργίας στην περίπτωση υπερβολικά αυξημένης συνολικής κίνησης του δικτύου. • Μία πιθανή κατάρρευση του κεντρικού κόμβου υποβαθμίζει την ασφάλεια πολλών συσκευών στο δίκτυο. • Παρουσιάζουν το σημαντικό μειονέκτημα της παραβίασης της ιδιωτικότητας των δεδομένων σε μερικές περιπτώσεις.

Πίνακας 3.1: *Σύγκριση Μεταξύ IDS σε Επίπεδο Συσκευής και σε Επίπεδο Δικτύου*

2. **Αποτελεσματικότητα**: Ονομάζεται επίσης ακρίβεια ταξινόμησης και αντιπροσωπεύει την ικανότητα του συστήματος να διακρίνει μεταξύ καλοήθων ή κακόβουλων δραστηριοτήτων.

Για την αποδοτικότητα, στην δική μας περίπτωση εξετάζουμε μόνο το μέγεθος του μοντέλου, καθώς στις περισσότερες περιπτώσεις όσο μικρότερο το μοντέλο, τόσο λιγότερη ενέργεια αυτό απαιτεί για την εκτέλεσή του.

Για την αποτελεσματικότητα, στην βιβλιογραφία αναφέρεται πως, δυστυχώς, δεν υπάρχει ακόμα κάποιο γενικό μέτρο αποδοτικότητας για την ανίχνευση εισβολών για τα IDS[22][21]. Μάλιστα, σε πολλές δημοσιεύσεις στις οποίες εφαρμόζονται μέθοδοι ML πολύ συχνά αποκρύπτονται λεπτομέρειες για τον τρόπο διεξαγωγής των χαρακτηριστικών και έτσι η σύγκριση μεταξύ διαφορετικών μοντέλων για χρήση σε IDS γίνεται αδύνατη.

Ωστόσο, πολλοί ερευνητές χρησιμοποιούν μια ποικιλία μετρήσεων για τη ποσοτική αξιολόγηση της απόδοσης. Κλασσικές μετρικές είναι το Ποσοστό των Ψευδών Συναγερμών (False Alarm Rate - FAR) και το Ποσοστό Ανίχνευσης (Detection Rate - DR), όπου το πρώτο πρέπει να ελαχιστοποιείται και το δεύτερο να μεγιστοποιείται [23]. Επίσης, η περιοχή κάτω από την Χαρακτηριστική Καμπύλη Λειτουργίας Δέκτη (Receiver Operating Characteristic -

ROC) επίσης χρησιμοποιείται για την αξιολόγηση ενός IDS, η οποία μετράει την επίδοση ενός δυαδικού ταξινομητή.

Ποσοστό Ψευδών Συναγερμών

Το FAR ορίζεται ως ο λόγος του αριθμού των καλοήθων περιπτώσεων που εντοπίζονται ως επίθεση προς τον συνολικό αριθμό των καλοήθων περιπτώσεων και το σύνολο τιμών του είναι [0, 1]. Στην περίπτωση όπου οι True ετικέτες σημαίνουν κακόβουλη δραστηριότητα και οι False καλοήθη τότε,

$$FAR = \frac{\text{Αριθμός καλοήθων περιπτώσεων που εντοπίστηκαν ως επιθέσεις}}{\text{Συνολικός αριθμός καλοήθων περιπτώσεων}}$$

$$= \frac{FP}{FP + TN}$$

Ποσοστό Ανίχνευσης

Το DR Υπολογίζεται ως ο λόγος του αριθμού των περιπτώσεων επίθεσης που εντοπίζονται σωστά ως επιθέσεις προς τον συνολικό αριθμό των επιθέσεων και το σύνολο τιμών του είναι [0, 1]. Στην περίπτωση όπου οι True ετικέτες σημαίνουν κακόβουλη δραστηριότητα και οι False καλοήθης τότε,

$$DR = \frac{\text{Σωστά ανιχνευμένες επιθέσεις}}{\text{Συνολικός αριθμός επιθέσεων}}$$

$$= \frac{TP}{TP + FN}$$

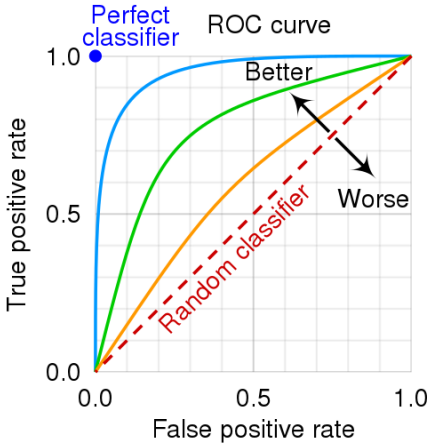
Η αντίθετη μετρική είναι το Ποσοστό Αστοχίας Miss Rate (MR), δηλαδή ο λόγος του αριθμού των περιπτώσεων επίθεσης που εντοπίζονται ως καλοήθεις προς τον συνολικό αριθμό των επιθέσεων [24]. Στην δική μας περίπτωση, εφόσον μιλάμε για επιθέσεις μπορούμε να αναφερόμαστε σε αυτό ως Attack Miss Rate - AMR. Εύκολα παρατηρεί κανείς πως ισχύει

$$AMR = 1 - DR$$

Χαρακτηριστική Καμπύλη Λειτουργίας Δέκτη

Στην προσπάθεια σχεδίασης ενός IDS, σύμφωνα με τους παραπάνω ορισμούς θα πρέπει να ελαχιστοποιηθεί η μετρική FAR και να μεγιστοποιηθεί η DR ή ισοδύναμα να ελαχιστοποιηθεί η MR.

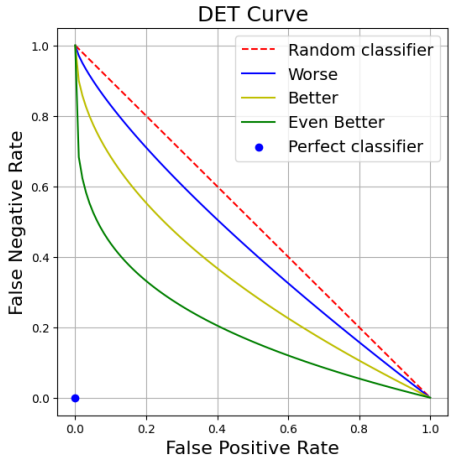
Η καμπύλη ROC είναι ένα γράφημα της FAR έναντι της DR για όλα τα πιθανά κατώφλια. Το εμβαδό κάτω από την καμπύλη ROC (Area Under the Curve - AUC) χρησιμοποιείται ως στατιστικό σύνοψης. Προερχόμενη από τη θεωρία ανίχνευσης σήματος, οι καμπύλες ROC χρησιμοποιούνται αφενός για την οπτικοποίηση της σχέσης μεταξύ του ρυθμού ανίχνευσης και του ποσοστού ψευδών θετικών ενός ταξινομητή κατά τη ρύθμισή του, και αφετέρου για τη σύγκριση της ακρίβειας διαφορετικών ταξινομητών.



Σχήμα 3.3: *Χαρακτηριστική Καμπύλη Λειτουργίας Δέκτη [4]*

Αντιστάθμιση σφάλματος ανίχνευσης

Επίσης, μπορούμε να εισάγουμε την έννοια της Αντιστάθμισης του Σφάλματος Ανίχνευσης (Detection Error Tradeoff - DET). Οι καμπύλες DET αποτελούν παραλλαγή των καμπυλών



Σχήμα 3.4: *Καμπύλη Αντιστάθμισης Σφάλματος Ανίχνευσης*

ROC, όπου στον άξονα Y απεικονίζεται το ποσοστό ψευδώς αρνητικών αποτελεσμάτων (False Negative Rate - FNR), ή ισοδύναμα του AMR στην δική μας περίπτωση, αντί του DR [25]. Σε αυτή την περίπτωση, το ιδανικό σημείο είναι η αρχή (κάτω αριστερή γωνία) του γραφήματος.

3.2 Διαδίκτυο των Πραγμάτων

Ως μία ανερχόμενη και ταχύτατα εξελισσόμενη τεχνολογία, το Διαδικτύων των Πραγμάτων (Internet of Things - IoT) έχει φέρει επανάσταση στον τρόπο ζωής των χρηστών. Για πολλά άτομα, η καθημερινότητά τους βασίζεται σε δίκτυα IoT, όπως έξυπνα περιβάλλοντα (έξυπνα σπίτια, έξυπνες πόλεις), και έξυπνα συστήματα μεταφοράς. Από την άλλη, όσον αφορά τις επιχειρήσεις και τη βιομηχανία, καινοτομίες όπως το smart manufacturing, η ανταλλαγή γνώσης (knowledge sharing) και η διαχείριση μεγάλου όγκου δεδομένων (big data management) αρχίζουν και εδραιώνονται.

Η ραγδαία εξέλιξη των συστημάτων τηλεπικοινωνιών έχει οδηγήσει σε μια νέα εποχή συνεργασίας του Διαδικτύου των Πραγμάτων (IoT) με Ασύρματα Δίκτυα Αισθητήρων (Wireless Sensor Networks - WSNs), Αναγνώριση με Ραδιοσυχνότητες (Radio Frequency Identification - RFID), και γενικότερα με οποιαδήποτε συσκευή ή δίκτυο, οποιαδήποτε στιγμή και οπουδήποτε [26].

Η ανάπτυξη του Διαδικτύου των Πραγμάτων (IoT) φέρνει μαζί της αναπόφευκτα και την πρόκληση της κυβερνοασφάλειας. Χωρίς την κατάλληλη αντιμετώπιση αυτού του ζητήματος, οι επιτιθέμενοι μπορεί να εκμεταλλευτούν τυχόν τρωτά σημεία ασφαλείας (vulnerabilities) των συσκευών και των δικτύων, αλλοιώνοντας δεδομένα ή διαταράσσοντας συστήματα. Σε ένα τέτοιο σενάριο, οι αρνητικές συνέπειες των επιθέσεων και των δυσλειτουργιών στο IoT θα μπορούσαν να εμποδίσουν τα οφέλη του.

Επιπλέον, τα παραδοσιακά πρωτόκολλα και μηχανισμοί ασφαλείας δεν επαρκούν για το IoT. Οι υφιστάμενες συσκευές περιορίζονται από χαμηλά επίπεδα επεκτασιμότητας, ακεραιότητας και διαλειτουργικότητας. Συνεπώς, απαιτείται η ανάπτυξη νέων μεθοδολογιών και τεχνολογιών που θα μπορούν να ανταποκριθούν στις αυστηρές απαιτήσεις ασφαλείας, απορρήτου και αξιοπιστίας του IoT [26].

3.2.1 Εφαρμογές και Αρχιτεκτονική

Το Διαδίκτυο των Πραγμάτων φέρνει επανάσταση σε διάφορους τομείς, από τα έξυπνα σπίτια και τις πόλεις έως την ρομποτική, την υγειονομική περίθαλψη και την ενέργεια. Η τεχνολογία αυτή, σε συνδυασμό με τη μηχανική μάθηση, ανοίγει αμέτρητες δυνατότητες για την αυτοματοποίηση και την βελτιστοποίηση λειτουργιών.

Στον βιομηχανικό τομέα, το Βιομηχανικό Διαδίκτυο των Πραγμάτων (Industrial IoT IIoT) φέρνει τη Βιομηχανία 4.0 (Industry 4.0), αυξάνοντας την παραγωγικότητα και μειώνοντας το κόστος μέσω της ανάλυσης δεδομένων και του έξυπνου ελέγχου. Ένα παράδειγμα του υπάρχοντος (IIoT) είναι τα μη επανδρωμένα εναέρια οχήματα (unmanned aerial vehicles - UAVs) που επιθεωρούν αγωγούς πετρελαίου και παρακολουθούν την ασφάλεια των τροφίμων χρησιμοποιώντας αισθητήρες [27].

Η τυπική αρχιτεκτονική ενός συστήματος IoT αποτελείται από τρία κύρια επίπεδα: συλλογής δεδομένων, δικτύου και εφαρμογών. Το επίπεδο συλλογής, συνδέει τις φυσικές συσκευές με το δίκτυο, παρακολουθεί το περιβάλλον και μεταδίδει δεδομένα. Το επίπεδο δικτύου λαμβάνει αυτά τα δεδομένα και τα μεταφέρει σε άλλες συσκευές. Τέλος, το επίπεδο εφαρμογής λειτουργεί ως διασύνδεση χρήστη, παρέχοντας πρόσβαση στα δεδομένα και επιτρέποντας την παροχή ουσιαστικών υπηρεσιών. Σε αυτό το σχήμα αρχιτεκτονικής, το επίπεδο συλλογής περιλαμβάνει αισθητήρες, ενεργοποιητές και τερματικές συσκευές IoT. Οι αισθητήρες ανιχνεύουν αλλαγές στο περιβάλλον και μεταδίδουν αυτές τις πληροφορίες. Οι ενεργοποιητές, που εμφανίζονται συχνά π.χ. σε εφαρμογές έξυπνων σπιτιών, ελέγχουν μηχανήματα ή συστήματα.

Η σύνδεση των αισθητήρων μπορεί να γίνει μέσω ενός τοπικού δικτύου (Local Area Network - LAN) ή ενός δικτύου προσωπικής περιοχής (Personal Area Network - PAN). Για ανάγκες χαμηλού ρυθμού δεδομένων και χαμηλής κατανάλωσης ενέργειας, οι αισθητήρες μπορούν επίσης να συνδεθούν σε δίκτυα ευρείας περιοχής (Wide Area Newotk - WAN) [27].

Η επιλογή της κατάλληλης τεχνολογίας δικτύου για κάθε εφαρμογή IoT εξαρτάται από

Σχήμα 3.5: *Επίπεδα Αρχιτεκτονικής Διαδικτύου των Πραγμάτων*

πολλές παραμέτρους, όπως η εμβέλεια, η κατανάλωση ενέργειας, το κόστος και η πυκνότητα των συσκευών. Παρακάτω, θα εξετάσουμε τρεις κύριες κατηγορίες αγορών IoT: LAN, Low Power WAN (LPWAN) και Κυψελωτών Δικτύων (Cellular), και θα αναλύσουμε τα χαρακτηριστικά και τις χρήσεις τους:

LAN

Τα δίκτυα LAN περιλαμβάνουν τεχνολογίες όπως WiFi, Zigbee και Bluetooth, οι οποίες χρησιμοποιούνται για τη σύνδεση συσκευών σε μικρές αποστάσεις. Αυτές οι τεχνολογίες είναι κατάλληλες για κινητές συσκευές, οικιακές εφαρμογές, και δίκτυα πλέγματος (mesh). Τα δίκτυα LAN είναι ιδανικά για εφαρμογές που απαιτούν χαμηλή κατανάλωση ενέργειας και μικρή εμβέλεια, αλλά δεν είναι κατάλληλα για εφαρμογές που απαιτούν μεγάλη διάρκεια ζωής μπαταρίας και μακρινές αποστάσεις. Οι τεχνολογίες αυτές κατέχουν το 30% της αγοράς IoT. [28].

LPWAN

Οι λύσεις LPWAN εξυπηρετούν τις ανάγκες εφαρμογών που απαιτούν μεγάλες αποστάσεις και χαμηλή κατανάλωση ενέργειας. Αυτά τα δίκτυα είναι κατάλληλα για την σύνδεση μεγάλου αριθμού συσκευών και αισθητήρων, προσφέροντας κάλυψη σε βάθος εσωτερικών χώρων με χαμηλό κόστος. Παρόλο που τα LPWAN είναι εξαιρετικά για την επίτευξη μεγάλης εμβέλειας και χαμηλού κόστους, δεν είναι κατάλληλα για εφαρμογές που απαιτούν υψηλό ρυθμό μετάδοσης δεδομένων, ακριβή τοποθέτηση και υψηλή αξιοπιστία (Ultra-Reliable Low Latency Communications - URLLC). Οι λύσεις αυτές κυριαρχούν στην αγορά IoT με 60% μερίδιο αγοράς, αποτελώντας μια από τις πιο ταχέως αναπτυσσόμενες κατηγορίες [28].

Κυψελωτά Δίκτυα

Τα δίκτυα Cellular, που περιλαμβάνουν τεχνολογίες όπως LTE και 5G προσφέρουν αξιόπιστες λύσεις για εφαρμογές IoT που απαιτούν μεγάλες αποστάσεις και υψηλό ρυθμό μετάδοσης δεδομένων. Τα κυψελωτά δίκτυα είναι κατάλληλα για εφαρμογές που χρειάζονται εκτεταμένη κάλυψη, ακριβή τοποθεσία και υψηλή διαθεσιμότητα, όπως η αυτόνομη οδήγηση σε περιβάλλοντα έξυπνων πόλεων. Ωστόσο, παρουσιάζουν μειονεκτήματα όσον αφορά την κατανάλωση ενέργειας, το κόστος και τον αριθμό των συνδεδεμένων συσκευών. Αυτά τα δίκτυα κατέχουν το 10% της αγοράς IoT και βασίζονται σε καλά καθιερωμένα πρότυπα που διασφαλίζουν την αξιοπιστία και την απόδοση των εφαρμογών [28].

Σε αυτή την εργασία, η εφαρμογή που μελετάται είναι της πρώτης κατηγορίας (LAN), και πιο συγκεκριμένα χρησιμοποιείται dataset που περιέχει δικτυακά δεδομένα μίας τοπολογίας smart home όπως και θα δούμε στην Ενότητα 4.2.

3.2.2 Ασφάλεια στο IoT και Προκλήσεις

Η ασφάλεια των IoT συσκευών αποτελεί ένα από τα πιο κρίσιμα ζητήματα στην εποχή της ψηφιακής συνδεσιμότητας. Μια πρόσφατη έρευνα της Bitdefender [29] εξέτασε περίπου 120 εκατομμύρια IoT συσκευές, οι οποίες δημιούργησαν 3,6 δισεκατομμύρια συμβάντα ασφαλείας παγκοσμίως, αποκαλύπτοντας τρωτά σημεία ασφαλείας (vulnerabilities) για ένα έξυπνο σπίτι. Οι σύγχρονοι οικιακοί χώροι περιλαμβάνουν κατά μέσο όρο 46 συσκευές στις ΗΠΑ και 25 συσκευές στην Ευρώπη, προσφέροντας στους επιτιθέμενους πολλές πιθανότητες για εκμετάλλευση. Μάλιστα, τα οικιακά δίκτυα δέχονται κατά μέσο όρο 8 επιθέσεις κάθε 24 ώρες σύμφωνα με την εταιρεία.

Επομένως, η αυξημένη συνδεσιμότητα και η ποικιλία των συσκευών εισάγουν σημαντικές προκλήσεις ασφάλειας. Αυτές οι συσκευές είναι ευάλωτες σε κυβερνοεπιθέσεις για διάφορους λόγους, όπως η περιορισμένη υπολογιστική ισχύς, το ευρύ «πρόσφορο έδαφος» για επιθέσεις, η ποικιλομορφία των συσκευών και η απομακρυσμένη εγκατάσταση:

1. Οι IoT συσκευές συχνά διαθέτουν περιορισμένη υπολογιστική ισχύ, μνήμη και ενεργειακούς πόρους. Αυτό περιορίζει την ικανότητά τους να εφαρμόζουν ισχυρά μέτρα ασφαλείας όπως ισχυρή κρυπτογράφηση και προχωρημένα συστήματα ανίχνευσης εισβολών [29].

2. Ο μεγάλος αριθμός διασυνδεδεμένων συσκευών δημιουργεί περισσότερες ευκαιρίες για τους επιτιθέμενους. Κάθε διασυνδεδεμένη συσκευή μπορεί να αποτελέσει πιθανό σημείο εισόδου για εισβολές, καθιστώντας το δίκτυο πιο δύσκολο να προστατευτεί στο σύνολό του [30].

3. Οι IoT συσκευές προέρχονται από διάφορους κατασκευαστές με διαφορετικά πρότυπα και πρωτόκολλα ασφαλείας. Αυτή η ποικιλομορφία δυσκολεύει την εφαρμογή ομοιόμορφων πολιτικών ασφαλείας σε όλες τις συσκευές [29] [30].

4. Πολλές IoT συσκευές λειτουργούν σε απομακρυσμένα και μη επιβλεπόμενα περιβάλλοντα, κάνοντας τη φυσική ασφάλεια και τις τακτικές ενημερώσεις δύσκολες. Αυτό αυξάνει

τον κίνδυνο φυσικής παραβίασης και καθιστά τις συσκευές παρατεταμένα ευάλωτες σε γνωστές επιθέσεις [30].

3.2.3 Γνωστές Κατηγορίες Επιθέσεων στο IoT

Οι κατηγορίες επιθέσεων που εξετάζουμε συμβαίνουν σχεδόν σε όλα τα επίπεδα του μοντέλου OSI [31]. Δεν εξετάζουμε επιθέσεις στο φυσικό επίπεδο ωστόσο, όπως frequency jamming (υποκλοπή συχνότητας) ή sniffing (υποκλοπή πακέτων δικτύου), καθώς αυτές οι επιθέσεις δεν παράγουν δικτυακά πακέτα και τα δεδομένα εκπαίδευσης των μοντέλων βασίζονται στις ροές ανταλλαγής πακέτων.

Παρακάτω περιγράφονται συνοπτικά οι βασικές κατηγορίες επιθέσεων σε IoT δίκτυα [32]. Αυτές οι κατηγορίες αποτελούν και τις κλάσεις του προβλήματος ταξινόμησης που εξετάζουμε. Η κάθε κατηγορία περιλαμβάνει πολλαπλούς τύπους επιθέσεων που μπορεί να διεξάγονται σε διαφορετικό επίπεδο του μοντέλου OSI όπως απεικονίζεται στο Σχήμα 3.6.

Dos & DDoS

Οι επιθέσεις Άρνησης Υπηρεσίας (Denial of Service - DoS) και Μεγάλης Κλίμακας Άρνησης Υπηρεσίας (Distributed Denial of Service - DDoS) έχουν στόχο να καταστήσουν μια συσκευή ή υπηρεσία μη διαθέσιμη στους χρήστες της. Η κύρια διαφορά μεταξύ των δύο είναι ότι στην DDoS χρησιμοποιούνται πολλοί υπολογιστές (ή πολλές συσκευές) ως επιτιθέμενοι, οι οποίοι μπορεί να είναι μέρος ενός botnet, δηλαδή ενός δικτύου υπολογιστών που έχουν μολυνθεί και ελέγχονται απομακρυσμένα από κάποια κεντρική πηγή. Στην DoS ο επιτιθέμενος υπολογιστής είναι ένας.

Αυτές οι επιθέσεις έχουν σκοπό την κατάρρευση του διακομιστή (στόχος) μέσω της υπερφόρτωσής του με τεράστιο όγκο αιτήσεων ή δεδομένων. Κάποιες ενδεικτικές ειδικές επιθέσεις σε αυτή την κατηγορία είναι:

- ICMP/HTTP/UDP/TCP Flood: Επίθεση με μεγάλη ποσότητα αιτημάτων ICMP, HTTP, UDP ή TCP για να υπερφορτωθεί ο στόχος.
- SYN Flood: Αποστολή μεγάλου αριθμού πακέτων SYN χωρίς ολοκλήρωση της χειραψίας TCP.
- Slowloris: Χρήση μερικών αιτημάτων HTTP για να διατηρηθούν πολλές ανοιχτές συνδέσεις προς τον στόχο.

Recon

Οι επιθέσεις Reconnaissance (Recon) αποσκοπούν στη συλλογή πληροφοριών για το δίκτυο και τις συσκευές που το απαρτίζουν, προετοιμάζοντας το έδαφος για άλλες επιθέσεις:

- Ping Sweep: Αποστολή αιτημάτων ICMP Echo σε πληθώρα IP διευθύνσεων για να εντοπιστούν ενεργοί κόμβοι.
- Port Scan: Έλεγχος για ανοικτές θύρες σε μια συσκευή.

- OS Scan: Αναγνώριση του λειτουργικού συστήματος μιας συσκευής μέσω ανάλυσης των δικτυακών απαντήσεων.

Web-based

Οι επιθέσεις αυτές στοχεύουν τις υπηρεσίες δικτυακών εφαρμογών των IoT συσκευών και περιλαμβάνουν:

- SQL Injection: Εισαγωγή κακόβουλου SQL κώδικα σε εφαρμογές ιστού για πρόσβαση σε βάσεις δεδομένων.
- Command Injection: Εισαγωγή κακόβουλων εντολών σε εφαρμογές ιστού για εκτέλεση εντολών στο σύστημα.
- Cross-Site Scripting (XSS): Εισαγωγή κακόβουλου κώδικα σε ιστοσελίδες που εκτελείται από τους περιηγητές (Web Browsers) των χρηστών.

Spoofing

Οι επιθέσεις Spoofing επιτρέπουν στους επιτιθέμενους να λειτουργούν υπό την ταυτότητα μίας νόμιμης οντότητας:

- ARP Spoofing: Αποστολή παραποιημένων ARP μηνυμάτων με σκοπό την αντιστοίχηση της MAC διεύθυνσης του επιτιθέμενου με την IP ενός νόμιμου συστήματος.
- DNS Spoofing: Αλλοίωση των καταχωρήσεων DNS για να ανακατευθυνθούν οι χρήστες σε παραποιημένους ή κακόβουλους ιστότοπους.

Brute Force

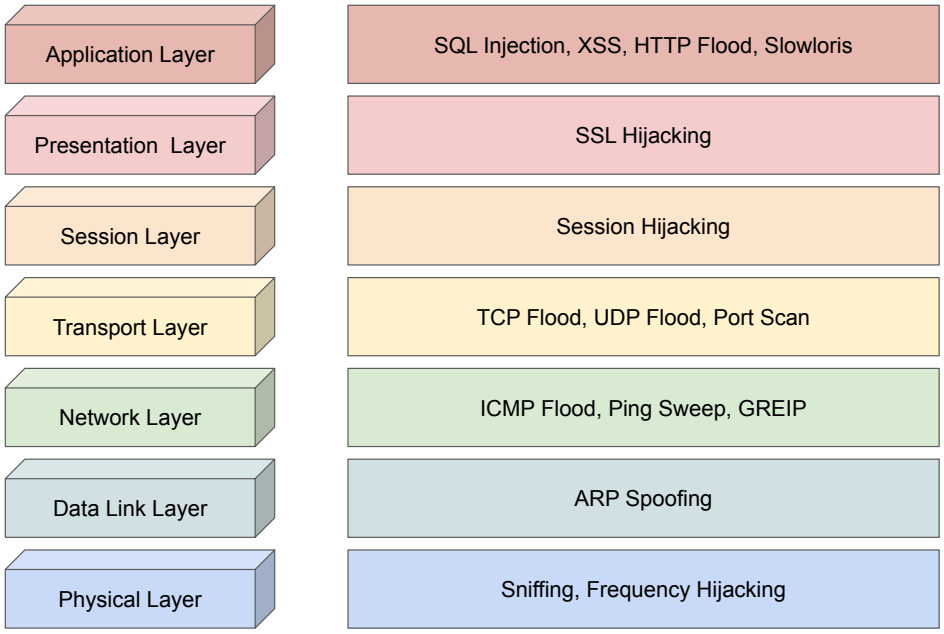
Οι επιθέσεις Brute Force επιχειρούν να αποκτήσουν πρόσβαση σε συστήματα χρησιμοποιώντας μεγάλο αριθμό προσπαθειών με πολλούς διαφορετικούς συνδυασμούς στοιχείων ταυτοποίησης, όπως π.χ. η Dictionary Attack, στην οποία πραγματοποιείται δοκιμή λέξεων και αριθμών από προκαθορισμένη λίστα για την εύρεση σωστού κωδικού πρόσβασης.

Mirai

Το Mirai είναι ένα γνωστό botnet που στόχευσε IoT συσκευές και χρησιμοποιήθηκε για να πραγματοποιήσει μεγάλης κλίμακας επιθέσεις DDoS. Ανακαλύφθηκε για πρώτη φορά το 2016 και είχε σημαντική επίδραση στο διαδίκτυο, προκαλώντας διακοπές υπηρεσιών σε μεγάλα websites και υπηρεσίες.

Το Mirai botnet δημιουργήθηκε για να σαρώνει το διαδίκτυο για IoT συσκευές που χρησιμοποιούν προεπιλεγμένα ή αδύναμα διαπιστευτήρια σύνδεσης (credentials) (όπως κωδικοί ίδιοι με όνομα χρήστη «admin»). Μόλις εντοπιστεί μια ευάλωτη συσκευή, το Mirai τη μολύνει και τη μετατρέπει σε bot που μπορεί να χρησιμοποιηθεί για να συμμετάσχει σε DDoS επιθέσεις.

- GREIP: Επίθεση με GRE πακέτα που περιέχουν τυχαίες IPs και θύρες.
- GREETH: Παρόμοια με το GREIP αλλά με επικέντρωση στην επικεφαλίδα Ethernet .



Σχήμα 3.6: *Διαχωρισμός Ενδεικτικών Επιθέσεων στα Επίπεδα του Μοντέλου OSI*

Κεφάλαιο 4

Πειραματική Διάταξη

Σε αυτό το κεφάλαιο παρουσιάζουμε το σύνολο δεδομένων που χρησιμοποιήθηκε για την εκπαίδευση και αξιολόγηση των μοντέλων. Συγκεκριμένα, πρώτα περιγράφουμε τα εργαλεία που χρησιμοποιήσαμε για την προεπεξεργασία των δεδομένων, την ανάπτυξη και την εκπαίδευση των μοντέλων. Εξετάζουμε τον τρόπο συλλογής των δεδομένων και οπτικοποιούμε μερικά χαρακτηριστικά και στατιστικά στοιχεία του συνόλου. Τέλος, περιγράφουμε τις αρχιτεκτονικές που επιλέξαμε να αναπτύξουμε.

4.1 Εργαλεία

4.1.1 NFStream

Το NFStream [33] αποτελεί ένα εργαλείο που εξάγει χαρακτηριστικά ροών δικτύου, είτε σε πραγματικό χρόνο είτε μέσω αρχείων καταγραφής πακέτων (packet capture - pcap). Ο τρόπος λειτουργίας του συμπεριλαμβάνει την συλλογή και ομαδοποίηση πακέτων δικτύου που έχουν μερικά κοινά χαρακτηριστικά, σχηματίζοντας έτσι ροές (flows). Τα χαρακτηριστικά αυτά είναι:

- Διεύθυνση IP πηγής: Η διεύθυνση IP από την οποία αποστέλλονται τα πακέτα.
- Διεύθυνση IP προορισμού: Η διεύθυνση IP από την οποία παραλαμβάνονται τα πακέτα.
- Αριθμός θύρας πηγής: Ο αριθμός θύρας TCP ή UDP που χρησιμοποιείται από την πηγή.
- Αριθμός θύρας προορισμού: Ο αριθμός θύρας TCP ή UDP που χρησιμοποιείται από τον προορισμό.
- Πρωτόκολλο: Το πρωτόκολλο που χρησιμοποιείται για τη μεταφορά των πακέτων (π.χ., TCP, UDP, ICMP).

Η ομαδοποίηση των πακέτων σε ροές επιτρέπει στο NFStream να εξάγει χρήσιμα χαρακτηριστικά για κάθε ροή. Τα εξαγόμενα χαρακτηριστικά παρουσιάζονται στον Πίνακα 4.1.

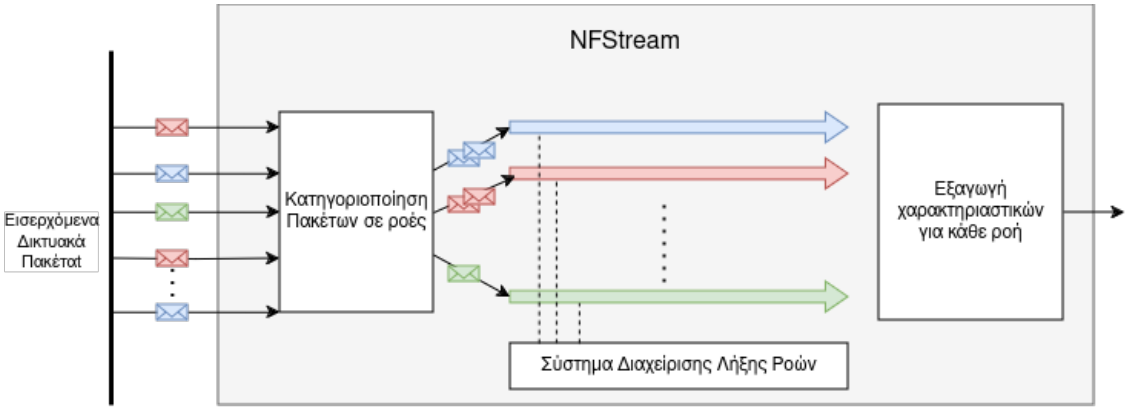
Επιπλέον, το NFStream αξιοποιεί τη βιβλιοθήκη nDPI για να αναγνωρίσει τους τύπους των εφαρμογών που δημιουργούν τις ροές παράγοντας έτσι κατηγορικά χαρακτηριστικά για την κάθε ροή.

Όνομα	**S2D**	**D2S**	**BD**	**Περιγραφή**
id			✓	Αναγνωριστικό ροής
expiration_id			✓	Αναγνωριστικό λήξης ροής (0 για λήξη αδρανούς, 1 για λήξη ενεργού και αρνητική για προσαρμοσμένη λήξη)
src_ip			✓	Διεύθυνση IP πηγής σε μορφή συμβολοσειράς
src_mac			✓	Διεύθυνση MAC πηγής σε μορφή συμβολοσειράς
src_port			✓	Θύρα πηγής επιπέδου μεταφοράς
dst_ip			✓	Διεύθυνση IP προορισμού σε μορφή συμβολοσειράς
dst_mac			✓	Διεύθυνση MAC προορισμού σε μορφή συμβολοσειράς
dst_port			✓	Θύρα προορισμού επιπέδου μεταφοράς
protocol			✓	Αναγνωριστικό πρωτοκόλλου επιπέδου μεταφοράς
ip_version			✓	Έκδοση IP ροής (4 ή 6)
first_seen_ms	✓	✓	✓	Χρονοσφραγίδα σε χιλιοστά του δευτερολέπτου στο πρώτο πακέτο ροής
last_seen_ms	✓	✓	✓	Χρονοσφραγίδα σε χιλιοστά του δευτερολέπτου στο τελευταίο πακέτο ροής
duration_ms	✓	✓	✓	Διάρκεια ροής σε χιλιοστά του δευτερολέπτου
packets	✓	✓	✓	Αριθμός πακέτων ροής
bytes	✓	✓	✓	Αριθμός των bytes ροής
min_ps	✓	✓	✓	Ελάχιστο μέγεθος πακέτου ροής
mean_ps	✓	✓	✓	Μέσο μέγεθος πακέτου ροής
stdev_ps	✓	✓	✓	Τυπική απόκλιση μεγέθους πακέτου ροής
maximum_ps	✓	✓	✓	Μέγιστο μέγεθος πακέτου ροής
min_piat_ms	✓	✓	✓	Ελάχιστος χρόνος μεταξύ πακέτων ροής σε χιλιοστά του δευτερολέπτου
mean_piat_ms	✓	✓	✓	Μέσος χρόνος άφιξης μεταξύ πακέτων ροής σε χιλιοστά του δευτερολέπτου
stdev_piat_ms	✓	✓	✓	Τυπική απόκλιση χρόνου άφιξης μεταξύ πακέτων ροής σε χιλιοστά του δευτερολέπτου
maximum_piat_ms	✓	✓	✓	Μέγιστος χρόνος άφιξης μεταξύ πακέτων ροής σε χιλιοστά του δευτερολέπτου
syn_packets	✓	✓	✓	Αριθμός πακέτων με ενεργοποιημένη τη σημαία TCP SYN
cwr_packets	✓	✓	✓	Αριθμός πακέτων με ενεργοποιημένη τη σημαία TCP CWR
ece_packets	✓	✓	✓	Αριθμός πακέτων με ενεργοποιημένη τη σημαία TCP ECE
urg_packets	✓	✓	✓	Αριθμός πακέτων με ενεργοποιημένη τη σημαία TCP URG
ack_packets	✓	✓	✓	Αριθμός πακέτων με ενεργοποιημένη τη σημαία TCP ACK
psh_packets	✓	✓	✓	Αριθμός πακέτων με ενεργοποιημένη τη σημαία TCP PSH
rst_packets	✓	✓	✓	Αριθμός πακέτων με ενεργοποιημένη τη σημαία TCP RST
fin_packets	✓	✓	✓	Αριθμός πακέτων με ενεργοποιημένη τη σημαία TCP FIN
split_direction			✓	Λίστα των N πρώτων κατευθύνσεων πακέτων ροής (0: S2D, 1: D2C, -1: κανένα πακέτο)
split_ps			✓	Λίστα των N πρώτων μεγεθών πακέτων ροής (-1 όταν δεν υπάρχει πακέτο)
split_piat_ms			✓	Λίστα των N πρώτων χρόνων άφιξης μεταξύ πακέτων ροής (πάντα 0 για το πρώτο πακέτο, -1 όταν δεν υπάρχει πακέτο)
application_name			✓	Όνομα εφαρμογής nDPI
application_category_name			✓	Όνομα κατηγορίας εφαρμογής nDPI
application_is_guessed			✓	Υποδεικνύει αν η ανίχνευση βασίζεται σε καθαρή ανάλυση ή σε εικασία βάσει θύρας

Πίνακας 4.1: *Εξαγόμενα Χαρακτηριστικά Δικτυακής Ροής από το NFStream.*
S2D: Source to Destination (Από Πηγή σε Προορισμό) ¸ *D2S: Destination to Source (Από Προορισμό σε Πηγή)* ¸ *BD: Bidirectional (Αμφίδρομα)*

Ακόμα, το NFStream ενσωματώνει ένα σύστημα διαχείρισης λήξης των ροών στον τρόπο λειτουργίας του ελέγχοντας τον χρόνο ζωής των ροών δικτύου, διασφαλίζοντας ότι οι άχρηστες ή παλιές ροές τερματίζονται αυτομάτως. Η διαχείριση λήξης ροής βασίζεται σε τρεις λογικές λήξης:

1. Λήξη Ενεργού Ροής: Τερματίζει μια ροή που παραμένει ενεργή για ένα προκαθορισμένο χρονικό διάστημα.
2. Λήξη Αδρανούς Ροής: Τερματίζει μια ροή που παραμένει αδρανής για ένα προκαθορισμένο χρονικό διάστημα.
3. Προσαρμοσμένη Λήξη: Επιτρέπει στον χρήστη να ορίσει μια προσαρμοσμένη λύση λήξης, όπως ένα όριο πακέτων ροής.



Σχήμα 4.1: *Λειτουργία του NFStream σε υψηλό επίπεδο*

Φαίνεται λογικό πως η επιλογή του τρόπου λήξης μιας ροής μπορεί να έχει μεγάλο αντίκτυπο στην παραγωγή των χαρακτηριστικών. Σε αυτή την εργασία, όπως θα περιγράψουμε και παρακάτω, δοκιμάσαμε μόνο την 2η επιλογή, όπου οι ροές λήγουν μετά από αδράνεια. Η λόγος αυτής της επιλογής έγκειται στην διαίσθηση πως έτσι αποτυπώνεται η συμπεριφορά της ροής «ολιστικά» και δεν υπάρχει απώλεια πληροφορίας. Βέβαια αυτή είναι μία επιλογή που χρειάζεται περισσότερη διερεύνηση. Δηλαδή, σε ένα σενάριο όπου θα θέλαμε να ελέγχουμε τις ροές ανά πολύ μικρά τακτικά χρονικά διαστήματα, ίσως θα έπρεπε να επιλέξουμε την 1η επιλογή με ένα πολύ μικρό προκαθορισμένο χρονικό διάστημα.

4.1.2 TensorFlow

Το TensorFlow [34] είναι ένα εργαλείο για την δημιουργία και την εκτέλεση αλγορίθμων μηχανικής μάθησης που αναπτύχθηκε από την Google. Ένα μοντέλο μηχανικής μάθησης που δημιουργείται με το TensorFlow μπορεί να εκτελεστεί σε διάφορα συστήματα χωρίς να χρειάζεται σχεδόν καμία τροποποίηση, από κινητές συσκευές όπως τηλέφωνα και tablet μέχρι μεγάλης κλίμακας κατανεμημένα συστήματα με εκαντοτάδες υπολογιστές και χιλιάδες Κάρτες Γραφικών (Graphics Processing Unit - GPU). Το σύστημα είναι ευέλικτο και μπορεί να χρησιμοποιηθεί για την ανάπτυξη πολλών διαφορετικών αλγορίθμων, ειδικά για την εκπαίδευση βαθιών νευρωνικών δικτύων. Έχει εφαρμοστεί σε αμέτρητους τομείς της επιστήμης των υπολογιστών και άλλων επιστημών, όπως αναγνώριση ομιλίας, όραση υπολογιστών, ρομποτική,

ανάκτηση πληροφοριών, επεξεργασία φυσικής γλώσσας, εξαγωγή γεωγραφικών πληροφοριών και ανακάλυψη φαρμάκων.

Η δυνατότητα προσαρμογής σε πολλαπλές συσκευές αλλά και η δυνατότητα ανάπτυξης διαφορετικών αρχιτεκτονικών νευρωνικών δικτύων αποτελούν τους δύο κύριους λόγους επιλογής του TensorFlow ως πλαίσιο ανάπτυξης (framework) στην παρούσα εργασία.

4.1.3 Kaggle

Στο πλαίσιο της διπλωματικής εργασίας, χρησιμοποιήθηκε η πλατφόρμα Kaggle [35] για την εκπαίδευση των βαθιών νευρωνικών δικτύων. Το Kaggle αποτελεί μια διαδικτυακή κοινότητα και πλατφόρμα διαγωνισμών για χρήστες που ασχολούνται με την επιστήμη δεδομένων και την μηχανική μάθηση. Η πλατφόρμα προσφέρει σημαντικά πλεονεκτήματα για την εκπαίδευση νευρωνικών δικτύων:

1. **Δωρεάν πρόσβαση σε υπολογιστικούς πόρους**: Παρέχει δωρεάν πρόσβαση σε GPU και μνήμη τυχαίας προσπέλασης (Random Access Memory - RAM), απαραίτητους πόρους για την εκπαίδευση νευρωνικών δικτύων, ιδιαίτερα μεγάλων μοντέλων.

2. **Ποικιλία συνόλων δεδομένων**: Η πλατφόρμα διαθέτει μια πλούσια συλλογή από ανοιχτά σύνολα δεδομένων σε διάφορους τομείς, προσφέροντας υλικό για πειραματισμούς και ανάπτυξη μοντέλων.

3. **Συνεργασία και ανταλλαγή γνώσεων**: Φιλοξενεί μια ενεργή κοινότητα χρηστών, όπου σπουδαστές, ερευνητές και επαγγελματίες μπορούν να μοιραστούν κώδικα, ιδέες και λύσεις, συμβάλλοντας σε μία συλλογική πρόοδο.

4. **Διαγωνισμοί**: Η πλατφόρμα διοργανώνει τακτικά διαγωνισμούς με πραγματικά δεδομένα και προβλήματα, προσφέροντας μια ευκαιρία για δοκιμή δεξιοτήτων, σύγκριση με άλλους ερευνητές και διεκδίκηση βραβείων.

Η χρήση του Kaggle αποδείχθηκε ιδιαίτερα ωφέλιμη για την εκπαίδευση των νευρωνικών δικτύων στο πλαίσιο της διπλωματικής εργασίας. Η πρόσβαση σε ισχυρούς υπολογιστικούς πόρους συνέβαλαν σημαντικά στην ταχύτερη εκπαίδευση των μοντέλων.

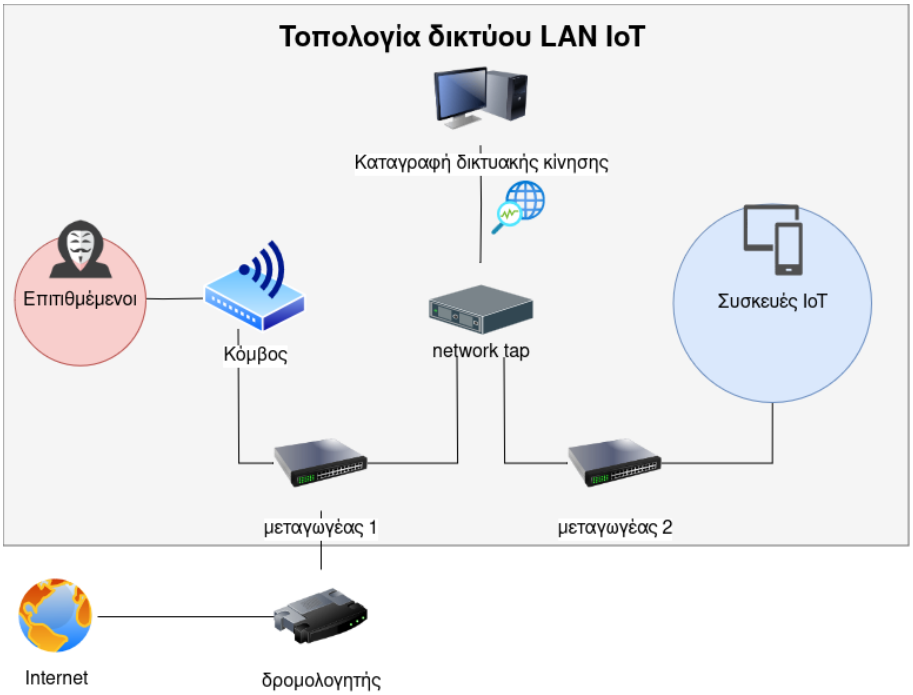
4.2 Σύνολο Δεδομένων CICIoT2023

Σύμφωνα με τους δημιουργούς του συνόλου δεδομένων CICIoT2023 [32], από το Καναδικό Ινστιτούτο Κυβερνοασφάλειας (Canadian Institute for Cybersecurity) του Πανεπιστημίου του New Brunswick, έχουν γίνει προσπάθειες στο παρελθόν για τη δημιουργία συνόλων δεδομένων που αποτελούνται από επιθέσεις σε συσκευές IoT. Ωστόσο, σε αυτές τις προσπάθειες αρκετές πιθανές επιθέσεις δεν λαμβάνονται υπόψη, αλλά επίσης δεν χρησιμοποιείται μία εκτεταμένη τοπολογία δικτύου με αρκετές και πραγματικές συσκευές IoT για την παραγωγή των δεδομένων. Το CICIoT2023 δημιουργήθηκε λύνοντας αυτές τις ελλείψεις για να προωθήσει την ανάπτυξη εφαρμογών ασφάλειας σε πραγματικά σενάρια λειτουργίας του IoT. Επιλέξαμε το συγκεκριμένο σύνολο δεδομένων γιατί ξεπερνά τους περιορισμούς των υπαρχόντων συνόλων, καθώς περιλαμβάνει 33 επιθέσεις που εκτελέστηκαν σε μία LAN τοπολογία IoT 105

συσκευών και ταξινομούνται σε επτά κατηγορίες (DDOS, Dos, Recon, Web-based, brute force, spoofing και Mirai).

4.2.1 Τοπολογία Δικτύου

Η τοπολογία δικτύου που χρησιμοποιήθηκε για την παραγωγή του συγκεκριμένου συνόλου δεδομένων σχεδιάστηκε έτσι ώστε να μιμείται σε μεγάλο βαθμό μία πραγματική εφαρμογή από IoT συσκευές και υπηρεσίες σε ένα σενάριο έξυπνου σπιτιού. Το δίκτυο αποτελείται από συνολικά 105 συσκευές εκ των οποίων οι 67 συμπεριλαμβάνονται άμεσα στις επιθέσεις που προσημειώθηκαν, ενώ υπάρχουν άλλες 38 συσκευές τύπου Zigbee και Z-Wave που είναι συνδεδεμένες σε πέντε σταθμούς βάσης (hubs). Το δίκτυο χωρίζεται σε δύο νοητά μέρη, τα οποία θα περιγράψουμε κάνοντας αναφορές στο Σχήμα 4.2.



Σχήμα 4.2: *Τοπολογία Τοπικού Δικτύου CICIoT2023*

Το πρώτο μέρος του δικτύου βασίζεται στον «μεταγωγέα 1» ο οποίος συνδέεται με έναν δρομολογητή και έναν κόμβο (hub). Ο δρομολογητής παρέχει πρόσβαση στο διαδίκτυο, ενώ το hub συνδέεται με 7 συσκευές τύπου Rasberry Pi [36]. Αυτές οι συσκευές χρησιμοποιούνται για την πραγματοποίηση των επιθέσεων και άλλων κακόβουλων δραστηριοτήτων στις προσομοιώσεις. Αυτό είναι ένα μοναδικό χαρακτηριστικό της τοπολογίας του CICIoT2023, δηλαδή ότι οι επιτιθέμενοι αποτελούν εσωτερική απειλή στο δίκτυο, και όχι εξωτερική.

Το δεύτερο μέρος του δικτύου βασίζεται στον «μεταγωγέα 2», ο οποίος είναι συνδεδεμένος με το υποδίκτυο της τοπολογίας το οποίο περιέχει όλες τις συσκευές IoT, οι οποίες μπορεί να είναι συνδεδεμένες είτε σε κάποιο smart hub ή σε κάποια βάση Zigbee/Z-wave.

Τέλος, στη τοπολογία του εργαστηρίου του Καναδικού Ινστιτούτου, έχει προστεθεί ένα «Network Tap», το οποίο αποτελεί μία δικτυακή συσκευή που συλλέγει όλη την κίνηση του IoT και την στέλνει σε υπολογιστές οι οποίοι την αποθηκεύουν σε μορφή packet capture (pcap) αρχείων μέσω εργαλείων όπως το Wireshark. Τα network taps είναι σχεδιασμένα

έτσι ώστε να συνδέονται στο δίκτυο και να δημιουργούν αντίγραφα της δικτυακής κίνησης με τρόπο που δεν επηρεάζει την κανονική λειτουργία του δικτύου, καθώς διαθέτουν έναν μη παρεμβατικό και παθητικό τρόπο σύνδεσης, χωρίς να εισάγουν καμία καθυστέρηση ή να επηρεάζουν την απόδοση. Αυτές οι συσκευές διαθέτουν θύρες δικτύου και θύρες παρακολούθησης, συνδέοντας τους επιτιθέμενους, τα θύματα και τον υπολογιστή παρακολούθησης μεταξύ τους.

Περισσότερες λεπτομέρειες σχετικά με την τοπολογία και τις συσκευές που χρησιμοποιούνται μπορούν να βρεθούν στην επίσημη ιστοσελίδα του συνόλου δεδομένων[32].

4.2.2 Συσκευές

Είναι σημαντικό να αναλύσουμε μερικές από τις IoT συσκευές που χρησιμοποιούνται ώστε να αποφανθούμε για το μέγεθος της μνήμης RAM που διαθέτουν. Με αυτόν τον τρόπο, αποσκοπούμε να έχουμε μία πιο καθαρή αντίληψη για ένα ικανοποιητικό μέγεθος που μπορούν να έχουν τα μοντέλα των νευρωνικών δικτύων που πρόκειται να «τρέχουν» σε αυτές τις συσκευές. Δυστυχώς, η εύρεση των τεχνικών χαρακτηριστικών των εμπορικών συσκευών που χρησιμοποιήθηκαν στο σύνολο δεδομένων είναι αδύνατη, καθώς πολλοί από τους επίσημους κατασκευαστές δεν παρέχουν αυτές τις πληροφορίες δημοσίως. Ωστόσο, οι συσκευές του IoT μπορούν να κατηγοριοποιηθούν σε τρεις κλάσεις: Τις «χαμηλού επιπέδου (Low-End)», τις «μεσαίου επιπέδου (Middle-end)» και τις «υψηλού επιπέδου (High-end)»[37].

1. **Χαμηλού επιπέδου IoT συσκευές**: Περιορισμένες σε πόρους μνήμης RAM και κύριας μνήμης (flash memory) σε δεκάδες ή εκατοντάδες kilobytes. Χρησιμοποιούνται κυρίως για βασικές εφαρμογές αισθητήρων και ενεργοποιητών.

2. **Μεσαίου επιπέδου IoT συσκευές**: Με περισσότερους πόρους από τις χαμηλού επιπέδου, μπορούν να εκτελέσουν βασικούς αλγόριθμους επεξεργασίας εικόνας και διαθέτουν πολλαπλές τεχνολογίες επικοινωνίας.

3. **Υψηλού επιπέδου IoT συσκευές**: Διαθέτουν ισχυρούς επεξεργαστές, αρκετή RAM, και δυνατότητα εκτέλεσης παραδοσιακών λειτουργικών συστημάτων όπως Linux ή Windows 10 IoT Core. Είναι ικανές για σύνθετους υπολογισμούς και μεγάλα μοντέλα μηχανικής μάθησης. Οι περισσότερες από αυτές έχουν την δυνατότητα εισαγωγής πρόσθετης κάρτας μνήμης SD.

Τα εύρη των μεγεθών της κύριας μνήμης και της μνήμης RAM αυτών των κατηγοριών, με βάση την σχετική έρευνα, παρουσιάζονται στον Πίνακα 4.2.

		Χαμηλού επιπέδου	**Μεσαίου επιπέδου**	**Υψηλού επιπέδου**
Μνήμη RAM	Ελάχιστο	∼10KB	96KB	64MB
	Μέγιστο	∼50KB	64MB	2GB
Κύρια Μνήμη	Ελάχιστο	∼100KB	32KB	8MB
	Μέγιστο	∼250KB	32MB	8GB

Πίνακας 4.2: *Κύρια μνήμη και μνήμη RAM για τρεις κατηγορίες συσκευών IoT*

Ωστόσο, σε αυτό το σημείο πρέπει να αναφέρουμε πως δεν διαθέτουν όλες οι συσκευές κάρτα δικτύου με φυσική διεύθυνση MAC. Αυτές οι συσκευές είναι εκείνες που χρησιμοποιούν μόνο τα πρωτόκολλα επικοινωνίας Z-wave και Zigbee για να επικοινωνούν με τον σταθμό βάσης τους. Ο κάθε σταθμός βάσης ωστόσο κατέχει διεύθυνση MAC. Εμείς εξετάζουμε τις συσκευές που επικοινωνούν με δικτυακά πρωτόκολλα τύπου IP και άρα μας ενδιαφέρουν μόνο οι συσκευές που διαθέτουν διεύθυνση MAC.

4.2.3 Τρόπος Παραγωγής Δεδομένων

Στη συνέχεια, περιγράφουμε πώς πραγματοποιήθηκε η παραγωγή, η εξαγωγή και η επισήμανση δεδομένων για κάθε σενάριο (επίθεσης ή μη). Σε πρώτη φάση, χρησιμοποιήθηκαν διαφορετικά εργαλεία για τη διεξαγωγή κάθε επίθεσης από τους ερευνητές για την εκτέλεση επιθέσεων κατά των IoT συσκευών[32]. Στη συνέχεια, στη διάρκεια κάθε ξεχωριστής επίθεσης, καταγράφηκε η κίνηση του δικτύου σε μορφή pcap χρησιμοποιώντας το Wireshark. Τέλος, για κάθε εκτελεσμένη επίθεση, ολόκληρο το αρχείο καταγραφής επισημάνθηκε αντίστοιχα με την εν λόγω επίθεση. Το συνολικό μέγεθος των αρχείων pcap που παράχθηκαν ανέρχεται στα 548GB.

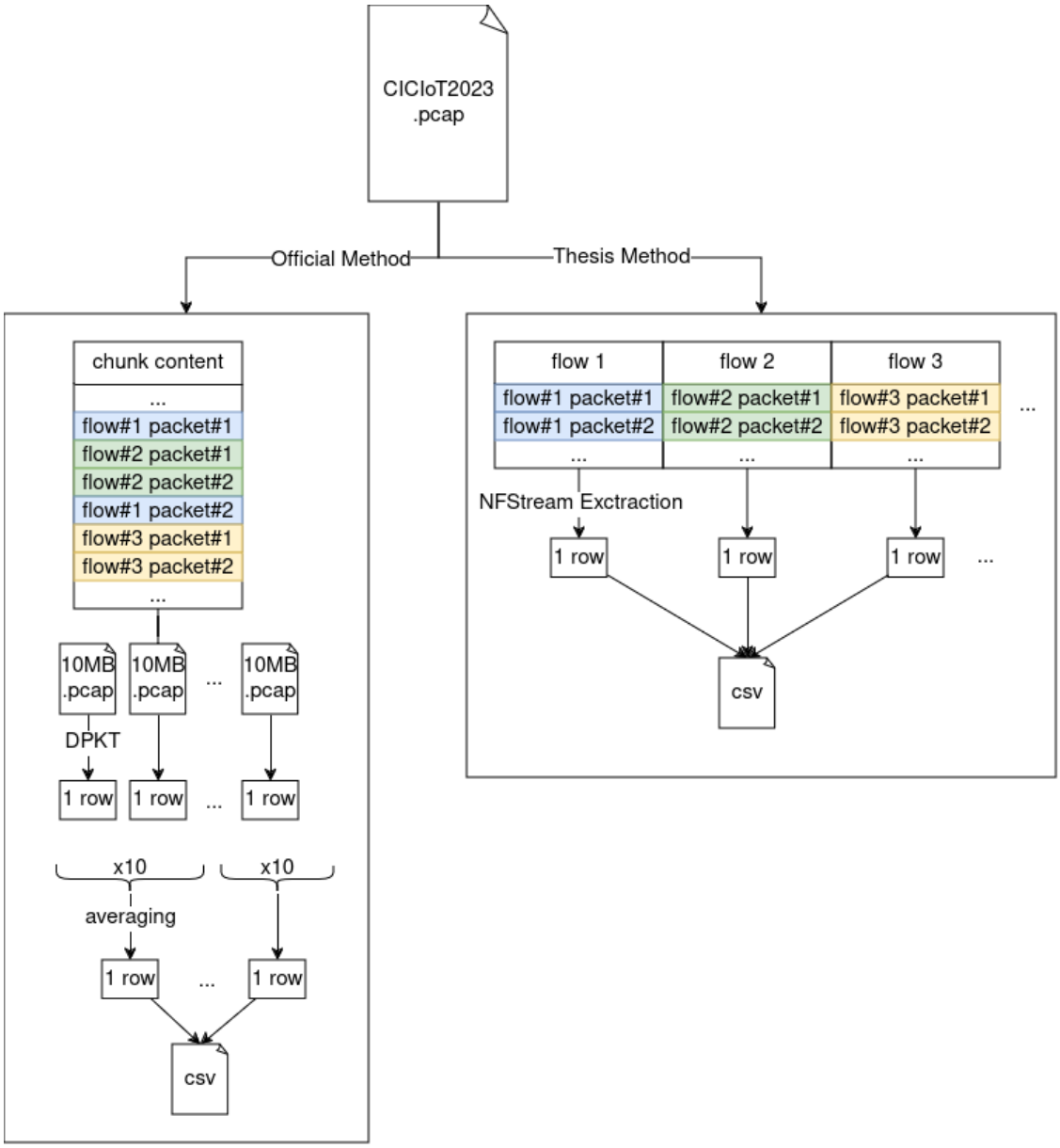
Στην παρούσα εργασία, χρησιμοποιήθηκαν μόνο τα αρχεία pcap για την εξαγωγή των χαρακτηριστικών όπως θα δούμε και στην Υποενότητα 4.3.1. Ωστόσο, το CICIoT2023 διατίθεται και σε μία εκδοχή csv, εκτός από pcap, η οποία διαθέτει έτοιμα χαρακτηριστικά για άμεση χρήση σε εφαρμογές μηχανικής μάθησης. Ο τρόπος με τον οποίο πραγματοποιήθηκε η εξαγωγή της εκδοχής csv από τους ερευνητές περιγράφεται ακολούθως:

1. Πρώτα, τα αρχεία pcap χωρίζονται σε μικρά κομμάτια αρχείων μεγέθους 10MB.

2. Για κάθε κομμάτι αρχείου χρησιμοποιείται το πακέτο DPKT [38] για την εξαγωγή χαρακτηριστικών της καταγεγραμμένης κίνησης.

3. Έπειτα, για την κάθε επίθεση, συνδυάζονται 10 ή 100 κομμάτια αρχείων υπολογίζοντας τους μέσους όρους των χαρακτηριστικών. Έτσι, η κάθε «γραμμή» στο csv αποτελεί τον μέσο όρο χαρακτηριστικών αυτά τα κομμάτια αρχείων pcap.

Ο λόγος που επιλέξαμε να χρησιμοποιήσουμε το NFStream για την εξαγωγή των χαρακτηριστικών από τα pcap, αντί να χρησιμοποιήσουμε απευθείας τα έτοιμα χαρακτηριστικά που παρέχει το CICIoT2023 σε csv, είναι το γεγονός ότι θέλουμε να εστιάσουμε στην εφαρμογή ενός HIDS και όχι ενός NIDS. Όπως περιγράφηκε προηγουμένως, το έτοιμο csv αντιπροσωπεύει χαρακτηριστικά της κίνησης του δικτύου από μία ολιστική ή γενική παρακολούθηση του δικτύου, καθώς η εξαγωγή πραγματοποιήθηκε με τεμαχισμό ολόκληρης της κίνησης. Ωστόσο, όπως είδαμε και στον τρόπο λειτουργίας του NFStream στην Υποενότητα 4.1.1, η εξαγωγή των χαρακτηριστικών πραγματοποιείται για κάθε ξεχωριστή ροή του δικτύου, όπου η κάθε ροή αντιστοιχεί σε μία αμφίδρομη επικοινωνία μεταξύ μόνο δύο διευθύνσεων (μεταξύ συσκευής IoT και επιτιθέμενου ή διακομιστή). Έτσι, γίνεται σαφές πως ο δεύτερος τρόπος εξαγωγής χαρακτηριστικών ταιριάζει καλύτερα σε σενάρια HIDS.



Σχήμα 4.3: *Σύγκριση Τρόπων Διεξαγωγής Χαρακτηριστικών*

4.3 Προετοιμασία Δεδομένων

Πριν προχωρήσουμε στον σχεδιασμό και την εκπαίδευση των βαθιών νευρωνικών δικτύων, όπως σε κάθε πρόβλημα μηχανικής μάθησης, πρέπει να προετοιμάσουμε πρώτα τα δεδομένα. Πρώτα, εξάγουμε τα χαρακτηριστικά από τα pcap αρχεία του CICIoT2023 με το NFStream.

4.3.1 Εξαγωγή Χαρακτηριστικών

Στην υποενότητα 4.2.3 περιγράψαμε πώς παράγονται τα pcap αρχεία. Σύμφωνα με αυτά που έχουν αναφερθεί έως τώρα, ένα αρχείο pcap που είναι επισημασμένο με μία συγκεκριμένη επίθεση περιέχει την κίνηση όλου του δικτύου που καταγράφηκε κατά την διάρκεια αυτής. Αυτό σημαίνει πως είναι πιθανό στο ίδιο αρχείο να περιέχονται μαζί με τις ροές της επίθεσης (δηλαδή μεταξύ των συσκευών-θυμάτων και συσκευών-θυτών) άλλες δευτερεύουσες ροές που

μπορούν να θεωρηθούν ως «θόρυβος» (π.χ. επικοινωνία συσκευών μεταξύ τους στη LAN τοπολογία ή ακόμα και επικοινωνία συσκευών με εξωτερικούς διακομιστές). Επομένως, στις περιπτώσεις επιθέσεων η κίνηση θα πρέπει να φιλτραριστεί έτσι ώστε να περιέχονται μόνο οι ροές μεταξύ συσκευών-θυμάτων και συσκευών-θυτών. Αυτό γίνεται να πραγματοποιηθεί μέσω του NFStream εισάγοντας ένα φίλτρο BPF [39] της μορφής:

$$
\begin{gathered}
((ether\ src\ MAC_{attacker_1} \text{ or } ether\ src\ MAC_{attacker_2} \text{ or } \cdots ether\ src\ MAC_{attacker_n}) \\
and \\
(ether\ dst\ MAC_{victim_1} \text{ or } ether\ dst\ MAC_{victim_2} \text{ or } \cdots ether\ dst\ MAC_{victim_m})) \\
or \\
((ether\ dst\ MAC_{attacker_1} \text{ or } ether\ dst\ MAC_{attacker_2} \text{ or } \cdots ether\ dst\ MAC_{attacker_n}) \\
and \\
(ether\ src\ MAC_{victim_1} \text{ or } ether\ src\ MAC_{victim_2} \text{ or } \cdots ether\ src\ MAC_{victim_m}))
\end{gathered}
$$

όπου n ο αριθμός των συσκευών των επιτιθεμένων, m ο αριθμός των IoT συσκευών θυμάτων. Μέσω των φυσικών διευθύνσεων MAC που αναφέρονται στις λεπτομέρειες του CICIoT2023 μπορούμε να απομονώσουμε τις ροές που μας ενδιαφέρουν, δηλαδή την κίνηση από τους θύτες προς τα θύματα και αντιστρόφως. Αντιθέτως, για τα αρχεία pcap που αντιπροσωπεύουν την καλοήθη κίνηση, δεν εφαρμόζουμε κανένα φίλτρο.

Έπειτα, μετά το φιλτράρισμα, εξάγονται τα χαρακτηριστικά κάθε ροής που αναγράφονται στον πίνακα 4.1. Από αυτά τα χαρακτηριστικά μερικά αποφασίστηκε να μην συμπεριληφθούν στο τελικό csv:

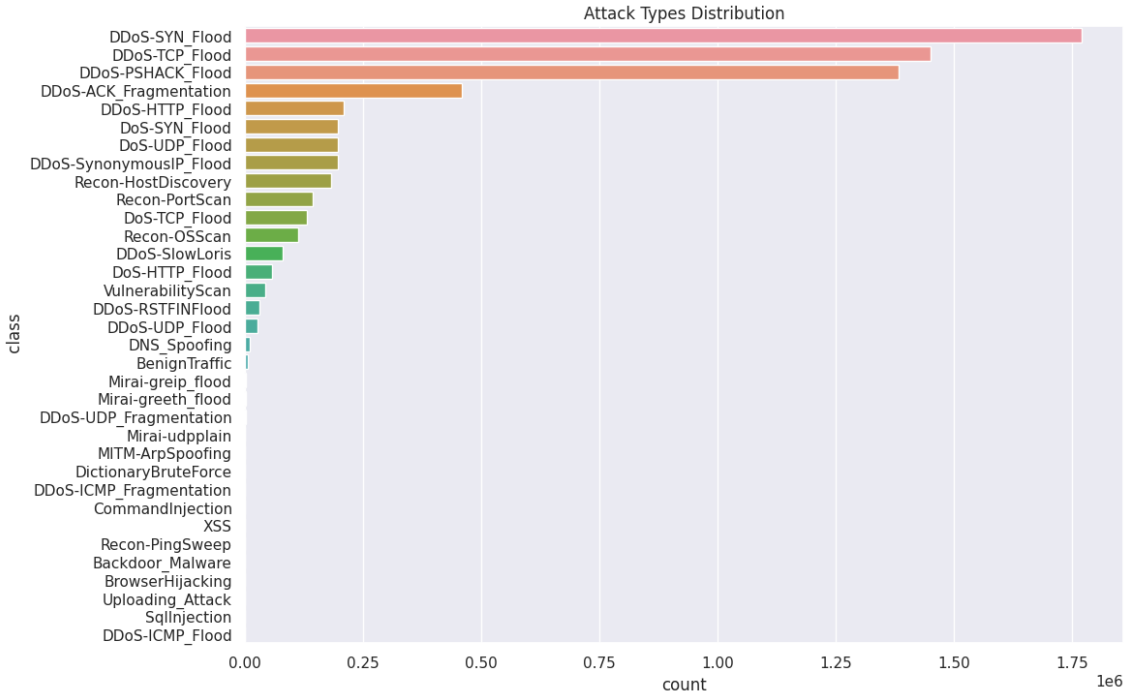
- Τα id, src_ip, src_port, src_mac, dst_ip, dst_port, dst_mac γιατί είναι μοναδικά για την κάθε ροή ή συσκευή και για αυτό υπονομεύουν τη «γενίκευση» των μοντέλων.
- Τα first_seen_ms, last_seen_ms, καθώς δεν προσφέρουν κάποια πρόσθετη πληροφορία από το χαρακτηριστικό του συνδυασμού αυτών (το duration) το οποίο και συμπεριλαμβάνουμε.

Αφού γίνει η εξαγή των csv από όλα τα pcap, ενώνονται σε ένα ενιαίο csv. Συνολικά προκύπτουν 6.691.873 datapoints (δηλαδή ροές) και 66 χαρακτηριστικά, εκ των οποίων τα 3 είναι κατηγορικά και χρειάζονται κωδικοποίηση (το protocol, το application_name και το application_category_name). Το application_guessed και το expiration_id είναι επίσης κατηγορικά, ωστόσο επειδή παίρνουν δυαδική αριθμητική τιμή (0 ή 1) δεν χρειάζονται κωδικοποίηση. Τέλος, για τα χαρακτηριστικά «split» που είναι σε μορφή λίστας, υπολογίζουμε τον μέσο όρο των τιμών και τα αντικαθιστούμε με αυτόν.

4.3.2 Ανισορροπία Κλάσεων

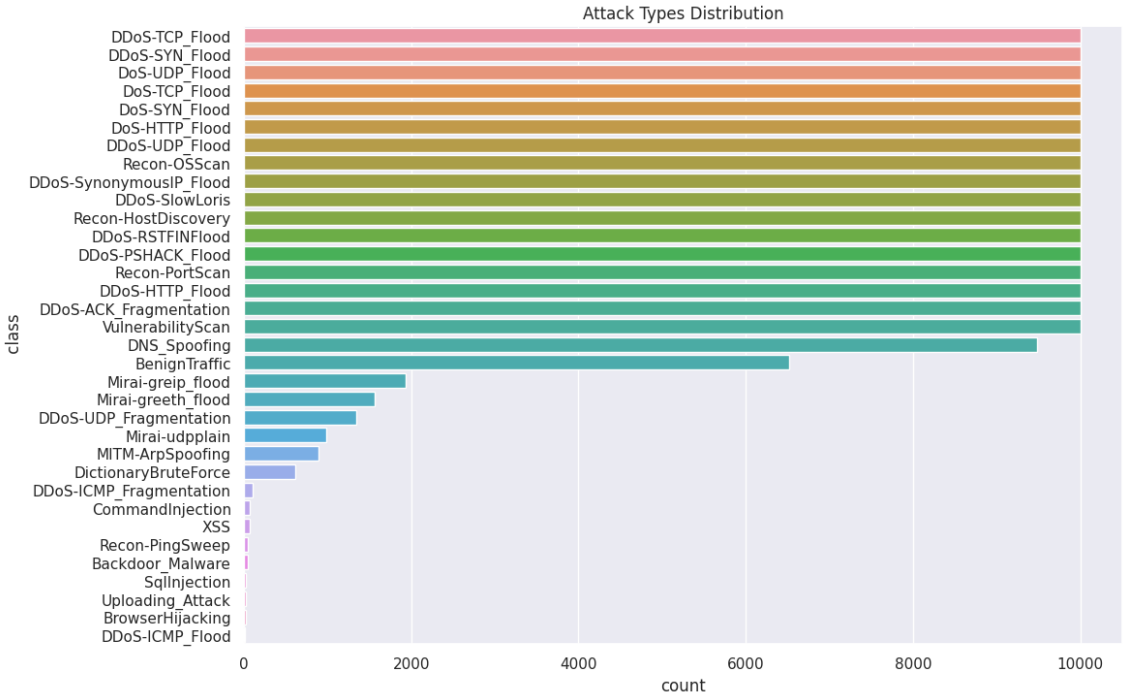
Το πρόβλημα είναι η ταξινόμηση των ροών σε 8 κλάσεις: DDoS, DoS, Recon, Spoofing, Mirai, BruteForce, Web-based και Benign. Ωστόσο, οι ροές είναι επίσης κατηγοριοποιημένες σε 34 πιο ειδικές (π.χ. DDoS-TCP-Flood, XSS, DNS Spoofing κτλ). Η κατανομή αυτών των 34 τύπων επιθέσεων που παρουσιάζεται παρακάτω αποκαλύπτει την τεράστια ανισορροπία που

υπάρχει, όχι μόνο μεταξύ των 8 γενικότερων κατηγοριών, αλλά ακόμα και σε υποκατηγορίες. Για παράδειγμα, η DDoS-SYN_Flood έχει σχεδόν 1.75 εκατομμύρια δείγματα, ενώ η DDoS-ICMP_Flood σχεδόν κανένα. Είναι εμφανές λοιπόν πως η ποιότητα του συνόλου δεν είναι ικανοποιητική, καθώς υπάρχουν και πολλές κατηγορίες με πολύ λίγα (δεκάδες ή εκατοντάδες) δείγματα.

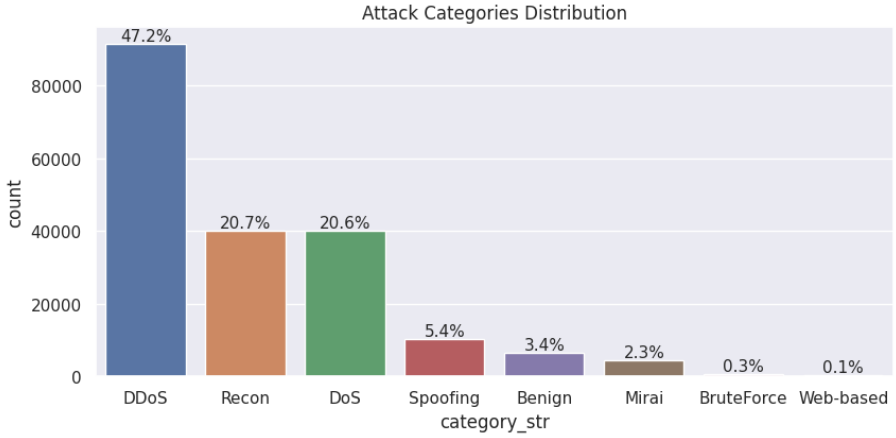


Σχήμα 4.4: *Κατανομή των 34 Κλάσεων πριν την Υποδειγματοληψία*

Προσπαθούμε να εξομαλύνουμε αυτό το πρόβλημα με υποδειγματοληψία των «μεγαλύτερων» κλάσεων. Βέβαια, τα προβλήματα βαθιάς μάθησης χρειάζονται πάρα πολλά δεδομένα για να κατακτήσουν πολύ υψηλή ακρίβεια και το ιδανικό θα ήταν να πραγματοποιούσαμε κάποια μέθοδο υπερδειγματοληψίας. Ωστόσο, εδώ η ανισορροπία είναι τεράστιας κλίμακας και σε μια τέτοια προσέγγιση θα χρειαζόταν εκτεταμένη παραγωγή τεχνητών δεδομένων η οποία θα καθιστούσε σχεδόν ολόκληρο το σύνολο δεδομένων συνθετικό. Έτσι, αποφασίσαμε να βάλουμε ένα άνω όριο στον αριθμό δειγμάτων για όλες τις 34 υποκλάσεις. Πιο συγκεκριμένα, θέσαμε το μέγιστο σε 10.000 δείγματα και έτσι τουλάχιστον οι μισές υποκλάσεις είναι ισορροπημένες.



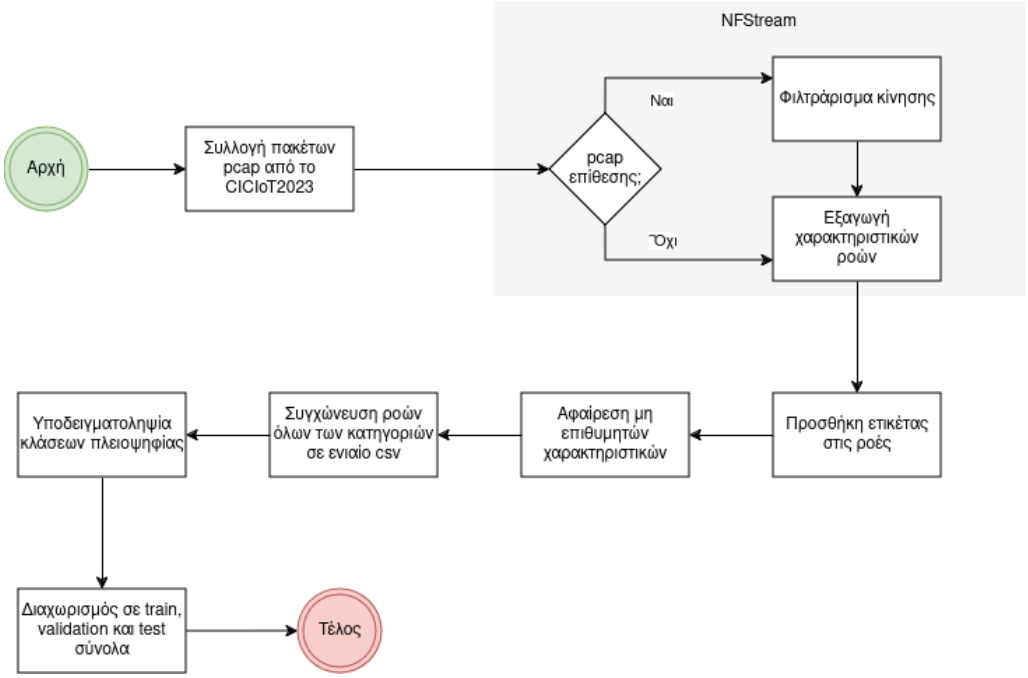
Σχήμα 4.5: *Κατανομή των 34 Κλάσεων μετά την Υποδειγματοληψία*



Σχήμα 4.6: *Κατανομή των 8 κλάσεων μετά την Υποδειγματοληψία*

Παρατηρώντας την κατανομή των δειγμάτων στις 8 γενικότερες κατηγορίες, φαίνεται πως υπάρχει ακόμη αισθητή ανισορροπία. Η κλάση DDoS αποτελεί τουλάχιστον το μισό σύνολο δεδομένων, οι Recon και DoS το 1/5 αντίστοιχα, ενώ οι υπόλοιπες κλάσεις μαζί δεν ξεπερνούν το 12% του συνόλου. Ειδικά οι κλάσεις BruteForce και Web-based βρίσκονται κάτω από το 0.5% και άρα δεν περιμένουμε καλή επίδοση ταξινόμησης για αυτές από τα μοντέλα. Συνολικά, μετά από την υποδειγματοληψία, τα δείγματα δεδομένων μειώθηκαν από 6.691.873 σε **193.754**, δηλαδή κατά 97.1%.

Έπειτα, χωρίσαμε το επεξεργασμένο σύνολο δεδομένων στα τρία υποσύνολα εκπαίδευσης (68%), επικύρωσης (12%) και ελέγχου (20%) τα οποία και χρησιμοποιούμε ίδια για όλα τα μοντέλα. Στο Σχήμα 4.7 παρουσιάζεται το διάγραμμα ροής ενεργειών για την παραγωγή αυτών των υποσυνόλων.



Σχήμα 4.7: *Παραγωγή Συνόλου Εκπαίδευσης, Επικύρωσης και Ελέγχου*

Κεφάλαιο 5

Εκπαίδευση και Αποτελέσματα

Σε αυτή την ενότητα θα παρουσιάσουμε και θα αναλύσουμε τη δομή των μοντέλων καθώς και τα αποτελέσματα από την εκπαίδευση και την αξιολόγησή τους. Στόχος είναι η σύγκριση των αποτελεσμάτων για την εύρεση του βέλτιστου μοντέλου.

5.1 Τυχαίο Δάσος

Γνωρίζοντας πως για δεδομένα πίνακα το τυχαίο δάσος (ως αλγόριθμος μηχανικής μάθησης) πετυχαίνει πολύ καλά αποτελέσματα, χρησιμοποιούμε την απόδοσή του ως επίπεδο αναφοράς. Για αυτό, δοκιμάζουμε το σύνολο εκπαίδευσης και ελέγχου με τον αλγόριθμο του τυχαίου δάσους με σκοπό να αποφανθούμε για τα σκορ που επιτυγχάνει για κάθε κλάση. Επίσης, συγκρίνουμε το τελικό σύνολο δεδομένων που προτείνουμε (εξαγωγή με NFStream και υποδειγματοληψία) με αυτό που προτείνουν οι δημιουργοί του CICIoT2023 (βλ. Υποενότητα 4.2.3).

	Precision	Recall	F1	Δείγματα
Benign	0.91	0.91	0.91	1309
DDoS	1.00	0.99	0.99	18211
DoS	0.99	0.99	0.99	7982
BruteForce	0.96	0.90	0.93	132
Spoofing	0.97	0.96	0.96	2056
Recon	1.00	0.99	0.99	8104
Web-based	0.17	0.38	0.23	45
Mirai	0.94	0.95	0.94	912
Accuracy			0.99	
Macro Avg	0.86	0.88	0.87	38751
Weighted Avg	0.99	0.99	0.99	38751

Πίνακας 5.1: *Αποτελέσματα Ταξινόμησης Τυχαίου Δάσους με τον Τρόπο Εξαγωγής Χαρακτηριστικών της Παρούσας Εργασίας*

Precision	Recall	F1	Δείγματα
0.89	0.97	0.93	243,322
1.00	1.00	1.00	7,526,151
1.00	1.00	1.00	1,792,167
0.93	0.06	0.11	2,982
0.84	0.82	0.83	107,780
0.85	0.78	0.82	77,212
0.80	0.06	0.11	4,165
1.00	1.00	1.00	583,102
		1.00	
0.92	0.84	0.85	10,336,881
1.00	1.00	1.00	10,336,881

Πίνακας 5.2: *Αποτελέσματα Ταξινόμησης Τυχαίου Δάσους με τον Τρόπο Εξαγωγής Χαρακτηριστικών που Προτείνεται στο CICIoT2023*

Ο συνδυασμός του τρόπου εξαγωγής των χαρακτηριστικών που επιλέξαμε μαζί με την προσπάθεια εξισορρόπησής των κλάσεων μέσω υποδειγματοληψίας φαίνεται να είχε θετικό αποτέλεσμα συγκρίνοντας τα αποτελέσματα του ίδιου αλγορίθμου για το προτεινόμενο τρόπο

εξαγωγής του συνόλου CICIoT2023, καθώς οι βαθμολογίες F1 των κλάσεων με τα πολύ λιγότερα δείγματα (BruteForce, Spoofing, Recon, Web-based που πετύχαμε είναι υψηλότερες. Αξιοσημείωτη είναι, επίσης, η διαφορά του συνολικού αριθμού δειγμάτων (38.751 έναντι 10.336.881), η οποία, σε συνδυασμό με τα υψηλά σκορ και στις δύο περιπτώσεις, αποτελεί απόδειξη της σημαντικότητας της ισορροπίας των κλάσεων.

Εστιάζοντας μόνο στον Πίνακα 5.1, παρατηρούμε επίσης πως οι κλάσεις DDoS, DoS και Recon είναι αυτές που αναγνωρίζονται καλύτερα από το μοντέλο, όπως άλλωστε ήταν αναμενόμενο, αφού κατέχουν τα μεγαλύτερα ποσοστά δεδομένων του συνόλου. Αντιθέτως, μεγάλη εντύπωση δημιουργεί η κλάση BruteForce, η οποία, παρά το γεγονός ότι έχει πολύ λίγα δείγματα σημειώνει εξαιρετικές επιδόσεις. Μάλιστα, αν συγκριθεί με την αντίστοιχη επίδοση του ομολόγου συνόλου φαίνεται πως ο τρόπος εξαγωγής χαρακτηριστικών ή και η υποδειγματοληψία που προτείνουμε κατάφερε να βελτιώσει σημαντικά την ποιότητα των δεδομένων της κλάσης. Τέλος, τη μεγαλύτερη δυσκολία παρουσιάζει η Web-based κατηγορία, ακόμα και αν υπάρχει τουλάχιστον διπλάσια βελτίωση από το προτεινόμενο σύνολο δεδομένων ($F1$ $0.11 \rightarrow 0.23$).

Πίνακας 5.3: *Πίνακας Σύγχυσης Τυχαίου Δάσους του Δικού μας Συνόλου Δεδομένων*

	Benign	DDoS	DoS	BruteForce	Spoofing	Recon	Web-based	Mirai
Benign	1.195	5	1	3	27	21	48	9
DDoS	15	18.062	94	0	13	0	0	27
DoS	1	48	7.928	0	2	0	0	3
BruteForce	3	0	0	119	7	1	6	2
Spoofing	39	8	3	0	1.969	6	18	13
Recon	38	3	0	5	5	8.047	1	0
Web-based	16	1	0	0	10	1	17	0
Mirai	13	11	0	2	6	1	11	868

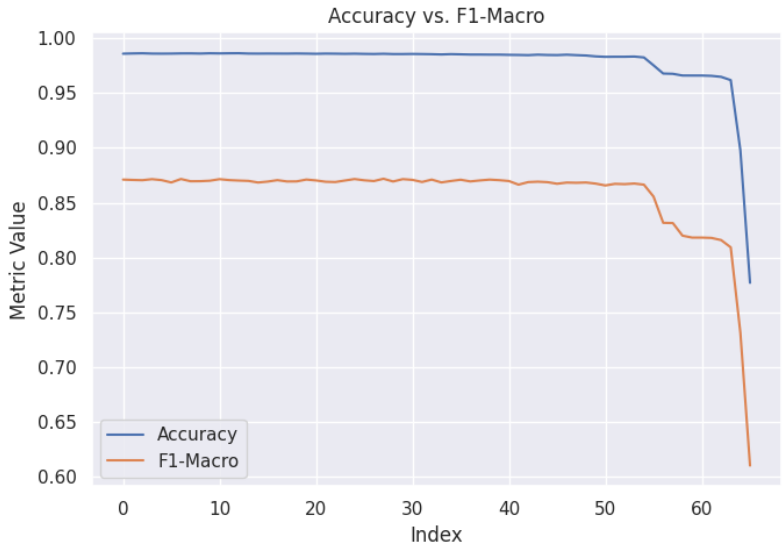
Πίνακας 5.4: *Πίνακας Σύγχυσης Τυχαίου Δάσους του Συνόλου Δεδομένων που Προτείνεται στο CICIoT2023*

	Benign	DDoS	DoS	BruteForce	Spoofing	Recon	Web-based	Mirai
Benign	234.929	24	2	4	5.159	3.192	8	4
DDoS	15	7,525,049	557	0	173	339	0	18
DoS	7	1.088	1.790.979	0	47	12	0	34
BruteForce	1,342	1	0	169	626	844	1	0
Spoofing	14.618	18	6	1	88.371	4.743	30	11
Recon	11.565	1.418	11	6	5.591	60.006	17	16
Web-based	1,140	3	1	1	2.792	1.265	230	1
Mirai	5	603	18	0	30	100	0	582.921

Ακόμα, εξετάζουμε τη σημαντικότητα των χαρακτηριστικών που εξάγει το τυχαίο δάσος με βάση το Mean Decrease in Impurity (MDI). Το MDI είναι μια μέθοδος υπολογισμού της σημαντικότητας των χαρακτηριστικών σε ένα δέντρο απόφασης ή σε ένα σύνολο δέντρων, όπως το τυχαίο δάσος. Συγκεκριμένα, υπολογίζει τη μείωση του impurity που προκαλεί κάθε χαρακτηριστικό σε όλες τις αποφάσεις των δέντρων. Όσο μεγαλύτερη είναι η μείωση

αυτής της τιμής, τόσο πιο σημαντικό θεωρείται το χαρακτηριστικό. Η σημαντικότητα των χαρακτηριστικών είναι χρήσιμη γιατί μας επιτρέπει να εντοπίσουμε μια πιθανή ιεραρχία στα χαρακτηριστικά.

Αφού υπολογίσουμε τη σειρά σημαντικότητας βάσει MDI εξετάζουμε την απόδοση του τυχαίου δάσους καθώς αφαιρούμε χαρακτηριστικά από το λιγότερο προς το περισσότερο σημαντικό. Με αυτόν τον τρόπο αποσκοπούμε στην εύρεση των κρίσιμων σημείων όπου η απόδοση μειώνεται αισθητά και άρα στην εύρεση των ελάχιστων διαστάσεων χωρίς απώλεια της απόδοσης. Στο Σχήμα 5.2 απεικονίζεται η επίδραση στην ακρίβεια και στην βαθμολογία F1 που προκαλεί η μείωση διαστατικότητας στον αλγόριθμο τυχαίου δάσους. Ο άξονας x αντιστοιχεί στον αριθμό αφαιρούμενων χαρακτηριστικών με την αντίστροφη σειρά σημαντικότητας, δηλαδή $x = 10$ σημαίνει πως αφαιρέθηκαν τα 10 λιγότερο σημαντικά χαρακτηριστικά.



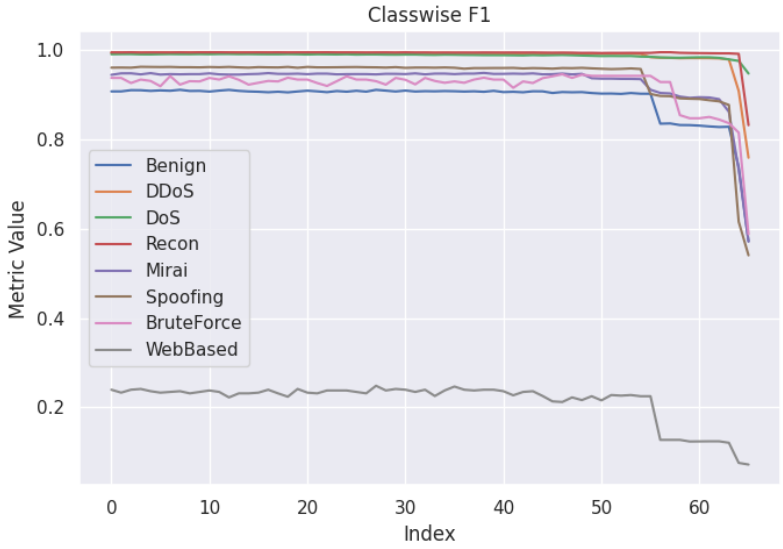
Σχήμα 5.1: *Επίδραση Μείωσης Διαστατικότητας στο Τυχαίο Δάσος (Ακρίβεια και στην Βαθμολογία F1)*

Η μείωση της διαστατικότητας είναι κρίσιμη για τα επόμενα βήματα, όπου θα χρησιμοποιήσουμε μοντέλα βαθιάς μάθησης και θα δοκιμάσουμε διαφορετικές διαστάσεις δεδομένων, συμπεριλαμβανομένων και πολύ μικρών υποσυνόλων χαρακτηριστικών. Με αυτόν τον τρόπο, ελπίζουμε να βελτιστοποιήσουμε την απόδοση των μοντέλων μας και να μειώσουμε σημαντικά την υπολογιστική πολυπλοκότητα.

Παρατηρούμε πως η κρίσιμη περιοχή βρίσκεται για Index από περίπου 52 έως 60, δηλαδή για διαστατικότητα από 14 έως 6 (αφού συνολικά έχουμε 66 χαρακτηριστικά). Επομένως, οι δοκιμές του αριθμού των διαστάσεων που πρόκειται να πραγματοποιήσουμε κατά την εκπαίδευση των νευρωνικών δικτύων θα είναι πιο πυκνές στο διάστημα 6-14. Αποφασίσαμε, λοιπόν, να πραγματοποιήσουμε τις εξής δοκιμές πλήθους διαστάσεων: [7, 9, 11, 13, 15, 30, 66].

5.2 Νευρωνικά Δίκτυα

Όπως έχει ήδη αναφερθεί, σε αυτή την εργασία χρησιμοποιούνται τρεις κύριες αρχιτεκτονικές νευρωνικών δικτύων: πολυεπίπεδα Perceptrons, συνελικτικά νευρωνικά δίκτυα, και



Σχήμα 5.2: *Επίδραση Μείωσης Διαστατικότητας στο Τυχαίο Δάσος (Βαθμολογία F1 Κλάσεων)*

μετασχηματιστές. Κάθε μια από αυτές τις αρχιτεκτονικές δοκιμάζεται με διαφορετικά μοντέλα που διαφοροποιούνται ως προς τη διάσταση εισόδου και τους συνδυασμούς επιπέδων που χρησιμοποιούνται. Σκοπός είναι να καθοριστεί ποια αρχιτεκτονική ή διαμόρφωση αποδίδει καλύτερα στα δεδομένα μας.

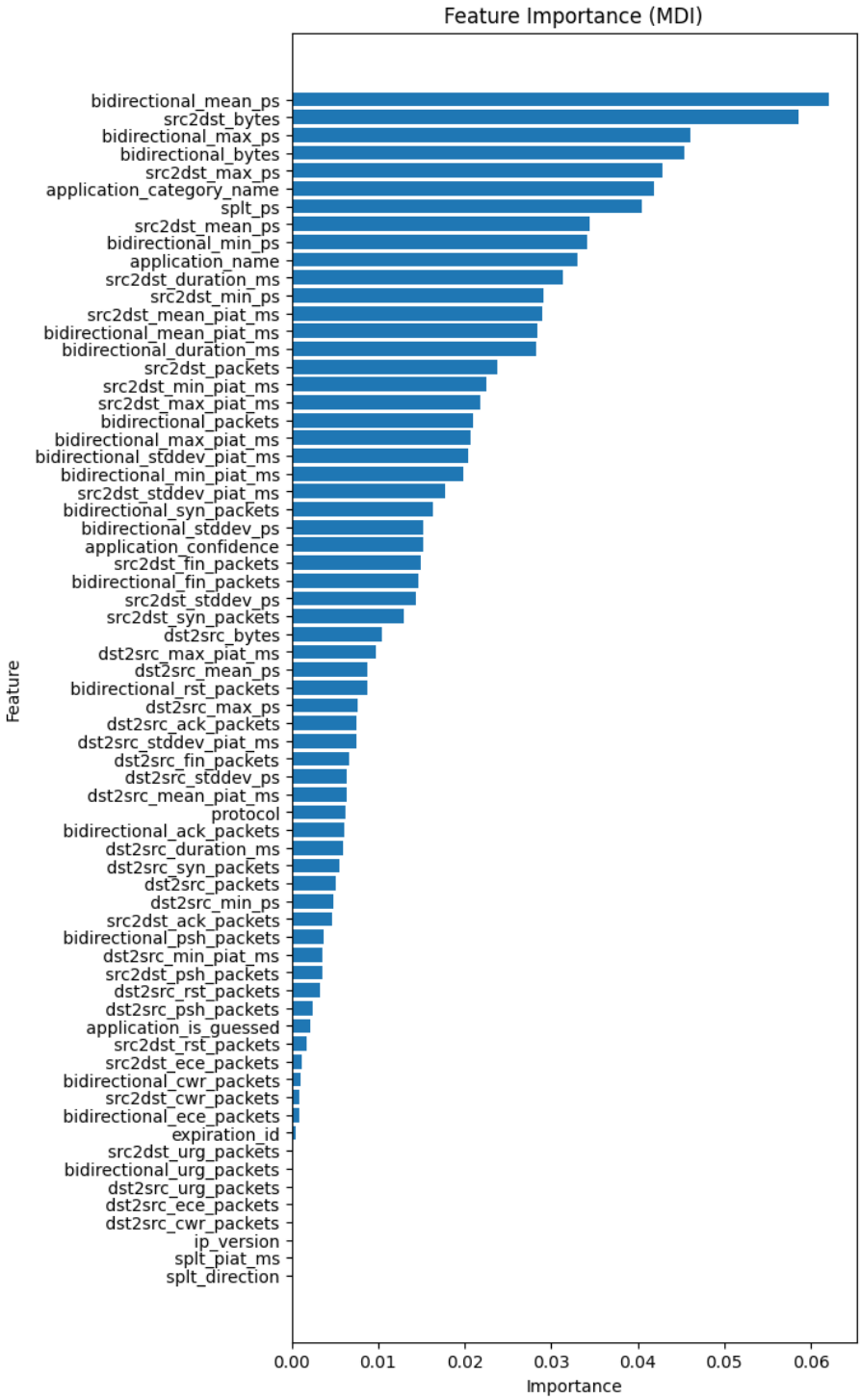
5.2.1 Κωδικοποίηση Κατηγορικών Χαρακτηριστικών

Σε αυτό το σημείο, περιγράφουμε τον τρόπο κωδικοποίησης των τριών κατηγορικών χαρακτηριστικών «protocol», «application_name» και «application_category_name» (βλ. υποενότητα 4.3.1).

1. Το «protocol» έχει 5 μοναδικές τιμές σε αριθμητική. Εφόσον οι μοναδικές τιμές είναι λίγες χρησιμοποιούμε την κωδικοποίηση one-hot μέσω της συνάρτησης IntegerLookup.

2. Το «application_name» έχει 153 μοναδικές τιμές σε μορφή συμβολοσειράς. Αρχικά, μετατρέπουμε κάθε τιμή σε μορφή αριθμού μέσω της συνάρτησης string_lookup. Έπειτα, εφόσον οι μοναδικές τιμές είναι αρκετές δεν συμφέρει να χρησιμοποιήσουμε την κωδικοποίηση one-hot. Για αυτό χρησιμοποιούμε ένα embedding_layer 10 διαστάσεων. Έτσι, μετατρέπουμε κάθε όνομα εφαρμογής ως ένα διάνυσμα 10 διαστάσεων το οποίο βελτιώνει τη δυνατότητα αναπαράστασης του κατά την διάρκεια της εκπαίδευσης. Στο τέλος, μόνο στις περιπτώσεις των CNNs και MLPs χρησιμοποιείται ένα απλό reshape_layer για λόγους συμβατότητας με το concatenate επίπεδο του δικτύου, όπως θα εξηγήσουμε και στην συνέχεια.

3. Το «application_category_name» έχει 24 μοναδικές τιμές σε μορφή συμβολοσειράς. Ακολουθούμε ακριβώς την ίδια μέθοδο με την κωδικοποίηση του «application_name» και το μετατρέπουμε σε ένα διάνυσμα 5 διαστάσεων.

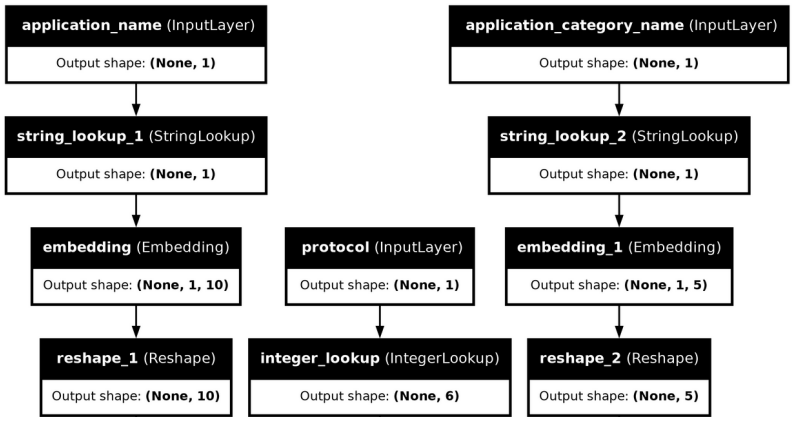


Σχήμα 5.3: *Σημαντικότητα Χαρακτηριστικών Βάσει του Τυχαίου Δάσους*

Αυτός ο τρόπος κωδικοποίησης των παραπάνω χαρακτηριστικών χρησιμοποιείται για όλα τα μοντέλα όλων των αρχιτεκτονικών που εξετάζουμε.

5.2.2 Εκπαίδευση

Η εκπαίδευση των μοντέλων για όλες τις αρχιτεκτονικές πραγματοποιήθηκε με τον Adam optimizer για 100 εποχές και περιλάμβανε τη χρήση early stopping για την αποτροπή της



Σχήμα 5.4: *Κωδικοποίηση Χαρακτηριστικών στο TensorFlow*

υπερπροσαρμογής, διακόπτοντας την εκπαίδευση αν η απόδοση στο σύνολο επικύρωσης δεν βελτιωνόταν για 3 συνεχόμενες εποχές.

5.2.3 Πολυεπίπεδα Perceptron

Επίπεδο Εισόδου

Στο επίπεδο εισόδου ενός MLP όλες οι τιμές των χαρακτηριστικών συνενώνονται σε ένα επίπεδο (concatenate layer) το οποίο τροφοδοτεί το νευρωνικό δίκτυο. Δηλαδή, δεν πραγματοποιείται κάποια μετατροπή των χαρακτηριστικών σε διανύσματα εικόνων ή διανύσματα αναπαράστασης (embeddings).

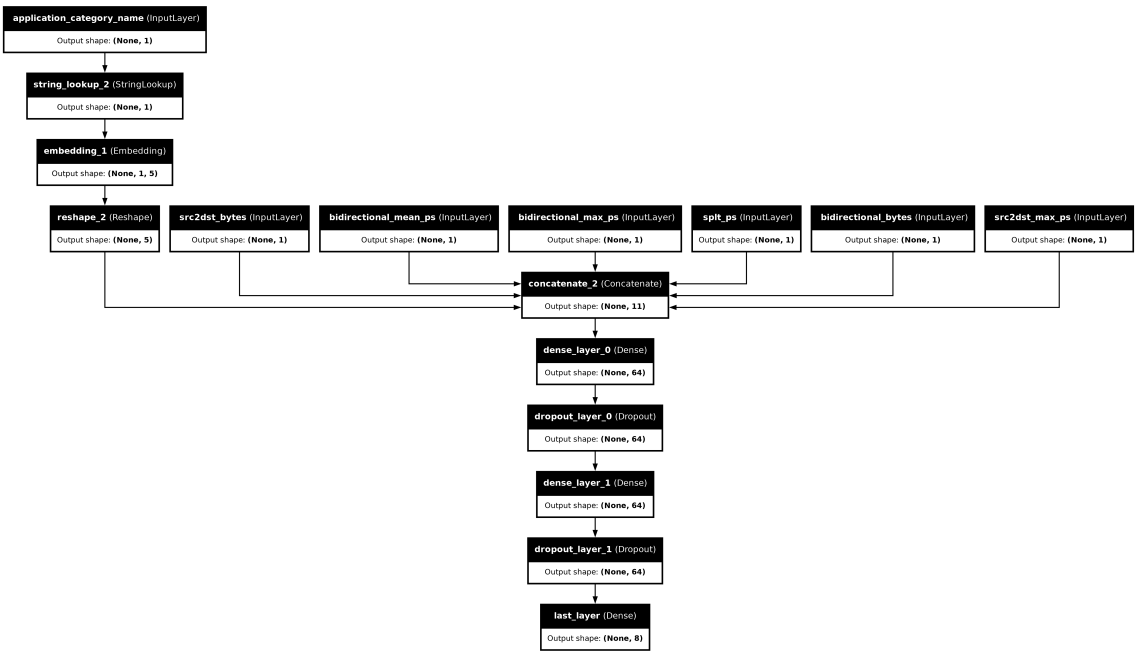
Συνδυασμοί Επιπέδων & Αρχιτεκτονική

Συνολικά σχεδιάσαμε 588 MLPs με διαφορετικούς συνδυασμούς επιπέδων. Συγκεκριμένα, για κάθε διάσταση εισόδου, κατασκευάσαμε αρχιτεκτονικές με:

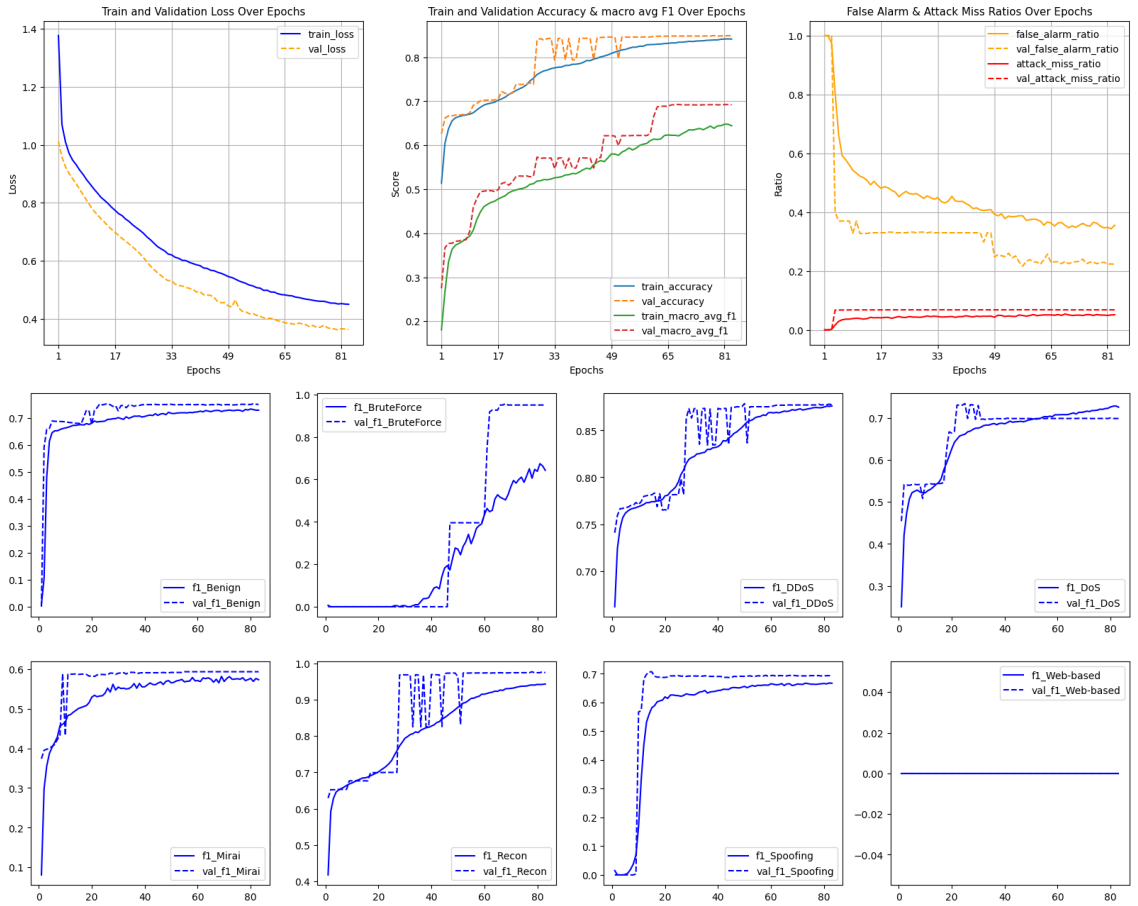
1. Ένα επίπεδο με 8, 16, 32 ή 64 νευρώνες.
2. Δύο επίπεδα με συνδυασμούς των 8, 16, 32 και 64 νευρώνων.
3. Τρία επίπεδα με συνδυασμούς των 8, 16, 32 και 64 νευρώνων.

Κάθε μοντέλο MLP κατασκευάστηκε με βάση την ακόλουθη γενική αρχιτεκτονική:

1. Είσοδος: Τα κωδικοποιημένα χαρακτηριστικά εισόδου, συνδυασμένα σε ένα ενιαίο επίπεδο.
2. Κρυφά Επίπεδα: Τα κρυφά επίπεδα με ενεργοποίηση ReLU και 50% dropout για τη μείωση της υπερπροσαρμογής.
3. Έξοδος: Ένα επίπεδο με 8 νευρώνες και ενεργοποίηση softmax για την πρόβλεψη των κατηγοριών.



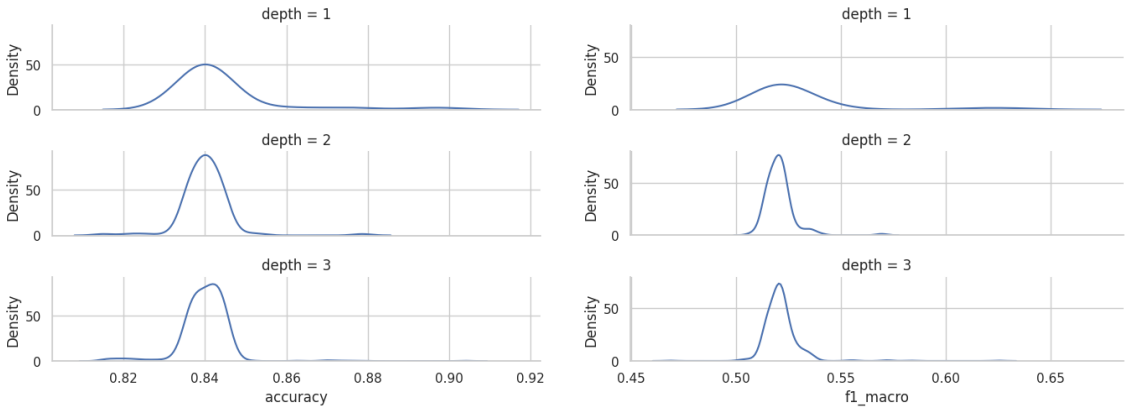
(α′) *Σχεδιασμός*



(β′) *Εκπαίδευση*

Σχήμα 5.5: *Παράδειγμα MLP Δύο Κρυφών Επιπέδων*

Αποτελέσματα



(α′) *Κατανομή Ακρίβειας* (β′) *Κατανομή F1*

Σχήμα 5.6: *Κατανομές Ακρίβειας και F1 για MLP (Διαφορετικά Βάθη)*

Αρχικά, μελετάμε την επίδραση που έχει το βάθος στα MLP μέσω των σχημάτων 5.6 και 5.7 όπου μπορούμε να κάνουμε τις εξής παρατηρήσεις:

1. Βάθος 1:

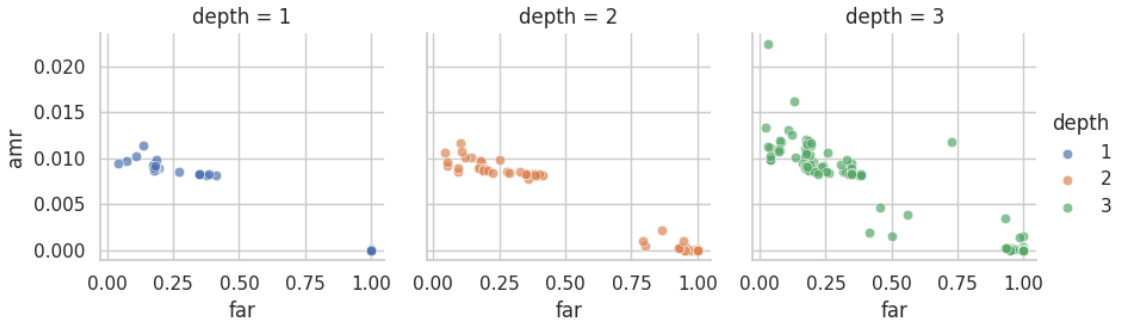
   - Ακρίβεια: Η κατανομή της ακρίβειας δείχνει μια συγκέντρωση γύρω από την τιμή 0.84, υποδεικνύοντας ότι τα μοντέλα με ένα κρυφό επίπεδο έχουν σχετικά καλή ακρίβεια σε σχέση με άλλα βάθη.
   - F1: Η κατανομή της F1 βαθμολογίας είναι πιο διασπαρμένη, με μια μικρή αιχμή γύρω από την τιμή 0.52, γεγονός που δείχνει μια ποικιλία στην απόδοση των μοντέλων ως προς τη διακριτική ικανότητα μεταξύ των κατηγοριών.
   - DET: Τα μοντέλα με βάθος 1 παρουσιάζουν μεγάλη διακύμανση στις τιμές FAR και AMR. Αυτό δείχνει ότι υπάρχουν μοντέλα που έχουν καλή ανίχνευση επιθέσεων (χαμηλό AMR) αλλά με υψηλό ρυθμό ψευδών συναγερμών (υψηλό FAR) και αντίστροφα.

2. Βάθος 2:

   - Ακρίβεια: Η κατανομή της ακρίβειας είναι πιο συγκεντρωμένη και πλησιάζει περισσότερο την τιμή 0.84, υποδηλώνοντας μια σταθερή απόδοση σε σχέση με τα μοντέλα με ένα κρυφό επίπεδο.
   - F1: Η κατανομή της F1 βαθμολογίας είναι ελαφρώς βελτιωμένη με περισσότερα μοντέλα να συγκεντρώνονται γύρω από την μέση τιμή.
   - DET: Τα μοντέλα με βάθος 2 έχουν γενικά ελαφρώς καλύτερη απόδοση με χαμηλότερα FAR, υποδηλώνοντας καλύτερη ισορροπία μεταξύ ανίχνευσης και ψευδών συναγερμών σε σχέση με τα μοντέλα με βάθος 1.

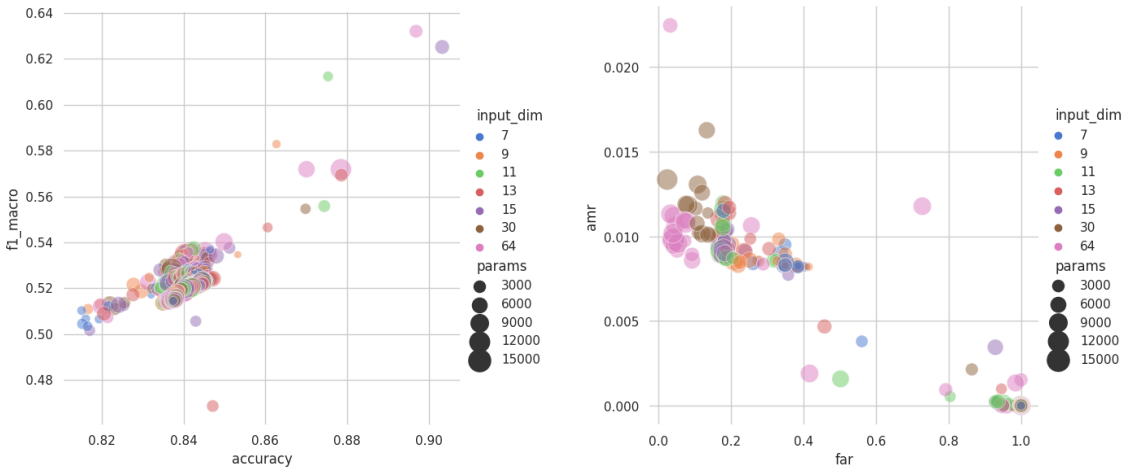
3. Βάθος 3:

- Ακρίβεια: Η κατανομή της ακρίβειας παραμένει περίπου στα ίδια επίπεδα με τα προηγούμενα βάθη, δείχνοντας μικρή βελτίωση ή σταθερότητα.
- F1: Η κατανομή της F1 βαθμολογίας δεν παρουσιάζει ιδιαίτερη βελτίωση από την αντίστοιχη του βάθους «2».
- DET: Τα μοντέλα με βάθος 3 παρουσιάζουν επίσης διασπορά στις τιμές FAR και AMR αλλά με καλύτερη συνολική απόδοση από τα μοντέλα με βάθος 1 και 2. Υπάρχουν μοντέλα που επιτυγχάνουν χαμηλότερα επίπεδα AMR και FAR, δείχνοντας βελτιωμένη ακρίβεια και αποδοτικότητα.



Σχήμα 5.7: *Απεικόνιση DET για MLP (Διαφορετικά Βάθη)*

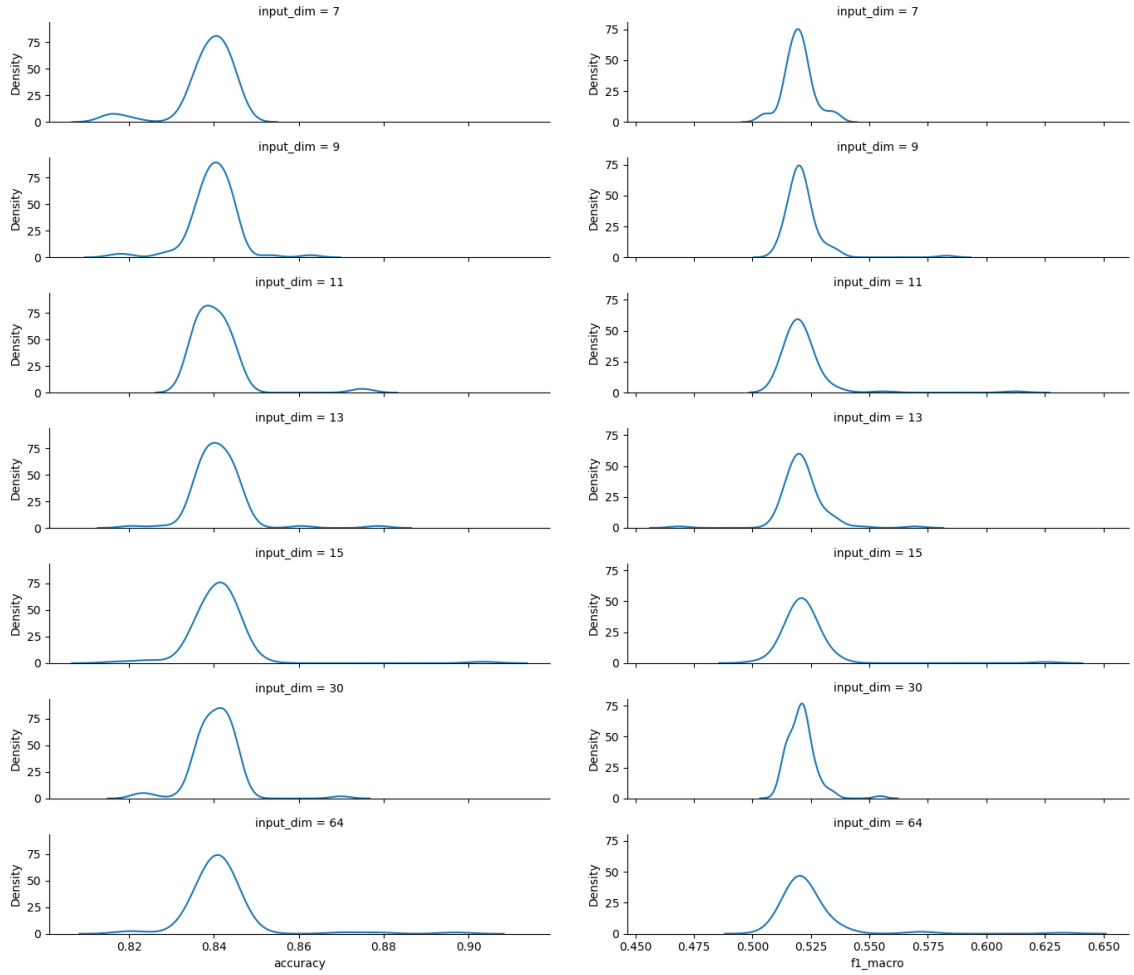
Γενικά, παρατηρούμε ότι η αύξηση του βάθους των κρυφών επιπέδων στα MLP μοντέλα τείνει να βελτιώνει την απόδοση, ειδικά όσον αφορά τη διακριτική ικανότητα μεταξύ κατηγοριών και την ισορροπία μεταξύ ανίχνευσης επιθέσεων και ψευδών συναγερμών. Ωστόσο, η βελτίωση δεν είναι γραμμική και υπάρχουν σημεία κορεσμού όπου η αύξηση του βάθους δεν οδηγεί απαραίτητα σε σημαντική βελτίωση της απόδοσης.



(α′) *Ακρίβεια - F1* (β′) *DET*

Σχήμα 5.8: *Διάγραμμα Διασποράς Ακρίβειας-F1 και DET για MLP*

Επίσης είναι αξιοσημείωτο πως στο το Σχήμα 5.9 οι διαφορές στις διαστάσεις εισόδου δεν φαίνεται να επηρεάζουν σημαντικά την απόδοση των μοντέλων σε ό,τι αφορά την ακρίβεια και τη F1 βαθμολογία. Όλες οι κατανομές δείχνουν ότι τα μοντέλα έχουν σταθερή απόδοση με μικρές αποκλίσεις.



(α′) *Κατανομή ακρίβειας* (β′) *Κατανομή F1*

Σχήμα 5.9: *Κατανομές Ακρίβειας και F1 για MLP (Διαφορετικές Διαστάσεις Εισόδου)*

Συμπερασματικά, τα MLP μοντέλα παρουσιάζουν σταθερή απόδοση ανεξαρτήτως των διαστάσεων εισόδου και του αριθμού των παραμέτρων. Η ακρίβεια και η F1 βαθμολογία είναι σταθερές, ενώ οι μετρικές ανίχνευσης (FAR και AMR) δεν φαίνεται να επηρεάζονται σημαντικά από αυτές τις παραμέτρους. Αυτό υποδηλώνει ότι τα MLP μοντέλα αποδίδουν σταθερά μέτρια σε ένα εύρος τιμών για τις διαστάσεις εισόδου και τον αριθμό των παραμέτρων.

5.2.4 Συνελικτικά Νευρωνικά Δίκτυα

Επίπεδο Εισόδου

Στα CNNs, τα δεδομένα εισόδου μετατρέπονται σε εικόνες για να μπορέσουν να περάσουν από τα συνελικτικά επίπεδα. Η διαδικασία αυτή περιλαμβάνει τα εξής βήματα:

1. Είσοδος Δεδομένων: Τα δεδομένα εισόδου περιλαμβάνουν αριθμητικά και κωδικοποιημένα κατηγορικά χαρακτηριστικά.

2. Συνένωση Δεδομένων: Όλα τα αριθμητικά και τα κατηγορικά χαρακτηριστικά συνενώνονται σε ένα ενιαίο διάνυσμα χρησιμοποιώντας το επίπεδο concatenate.

3. Μετατροπή σε Εικόνες: Το συνενωμένο διάνυσμα περνάει από τόσα πυκνά επίπεδα όσα έχουμε ορίσει για κανάλια. Αυτά δημιουργούν νέα διανύσματα με διαστάσεις ίσες με το μέγεθος των συνολικών pixels, που έχουμε ορίσει σαν μέγεθος εικόνας, το καθένα. Κάθε ένα από αυτά τα διανύσματα αναδιαμορφώνεται (reshape) σε μια δισδιάστατη εικόνα.

4. Δημιουργία Εικόνας για το CNN: Τα αναδιαμορφωμένα διανύσματα συνενώνονται ξανά σε ένα ενιαίο επίπεδο concatenate, δημιουργώντας έτσι μια εικόνα ενός ή πολλαπλών καναλιών.

Συνδυασμοί Επιπέδων & Αρχιτεκτονική

Συνολικά σχεδιάσαμε 1.512 CNNs με διαφορετικούς συνδυασμούς επιπέδων. Συγκεκριμένα, για κάθε διάσταση εισόδου, κατασκευάσαμε αρχιτεκτονικές με:

1. Αριθμός καναλιών 1, 2 ή 3.

2. Μέγεθος εικόνας $3x3$, $4x4$, $5x5$, $6x6$, $7x7$, $8x8$, $9x9$, $10x10$.

3. Πρώτο συνελικτικό επίπεδο με ενεργοποίηση ReLU και φίλτρα 2 ή 4.

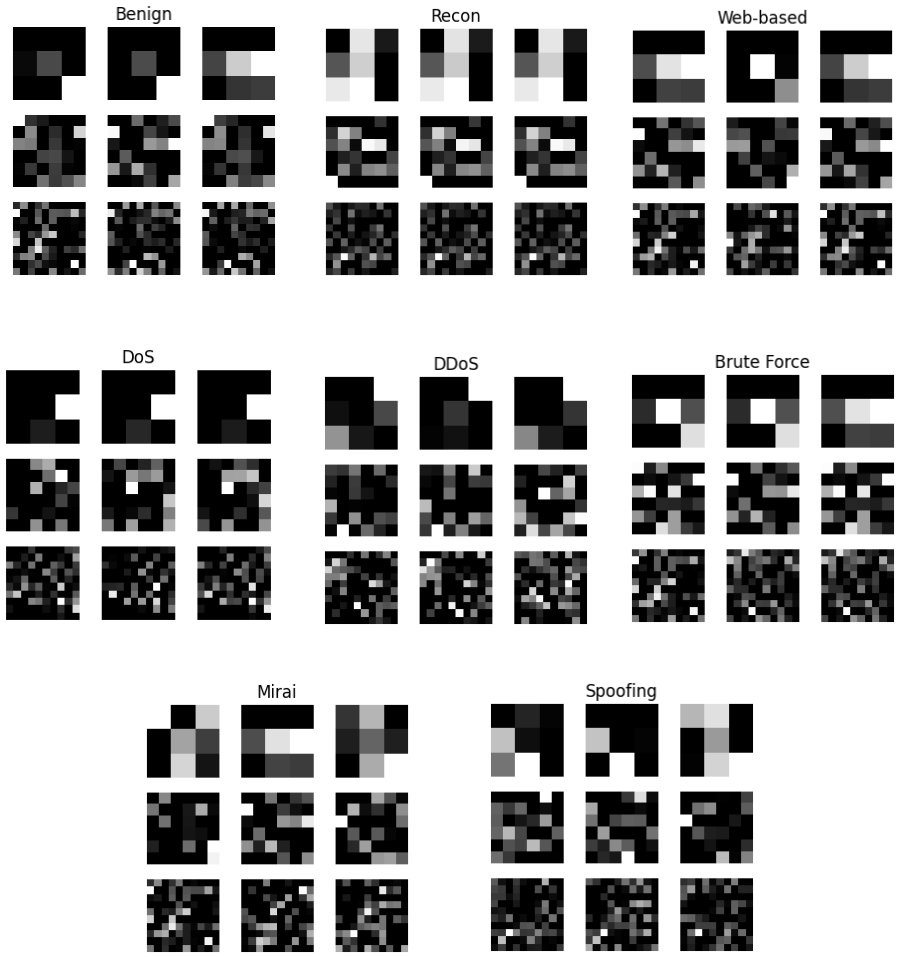
4. Δεύτερο συνελικτικό επίπεδο (αν υπάρχει) με ενεργοποίηση ReLU και φίλτρα 2, 4, 8 ή 16. Φροντίζουμε το πλήθος των φίλτρων να είναι μεγαλύτερο ή ίσο σε σχέση με αυτό του πρώτου συνελικτικού επιπέδου.

Κάθε μοντέλο CNN κατασκευάστηκε με βάση την ακόλουθη γενική αρχιτεκτονική:

1. Είσοδος: Τα κωδικοποιημένα χαρακτηριστικά εισόδου, συνδυασμένα σε ένα ενιαίο επίπεδο.

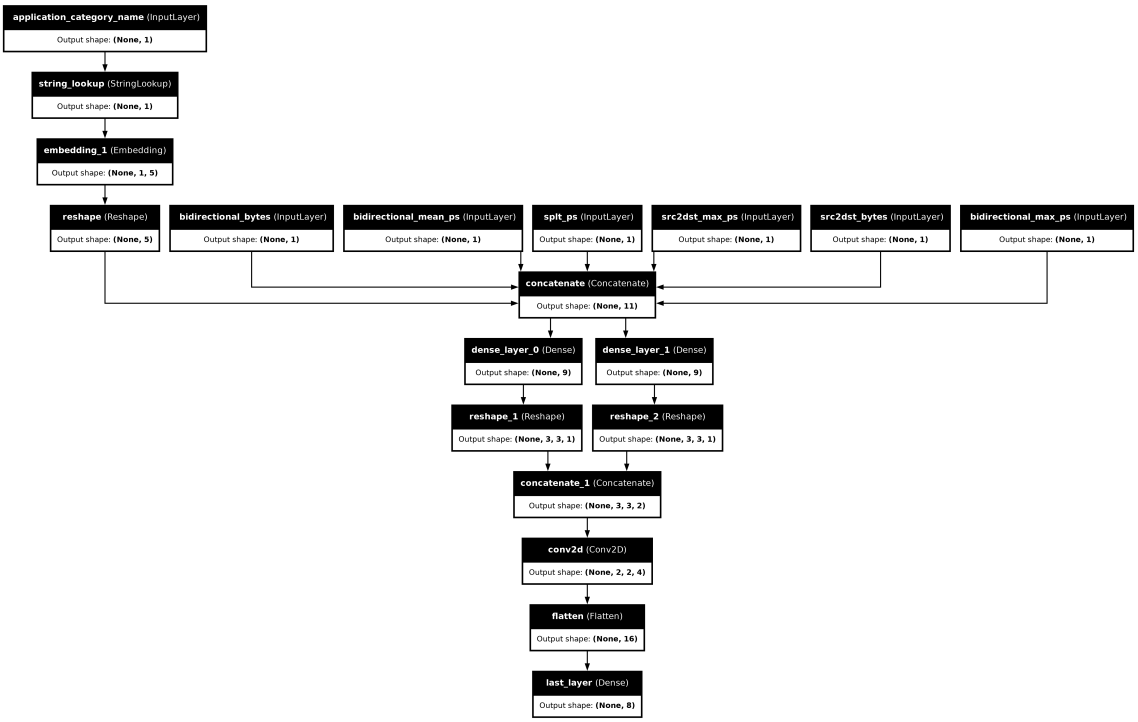
2. Επαναδιάταξη: Τα χαρακτηριστικά εισόδου μετασχηματίζονται σε εικόνες.

3. Συνελικτικά Επίπεδα: Το αποτέλεσμα περνάει από ένα ή δύο συνελικτικά επίπεδα Conv2D με φίλτρα, ενεργοποίηση ReLU, και batch normalization. Αν η διάσταση της εικόνας είναι μικρότερη από $7x7$ τότε χρησιμοποιούμε $kernel\_size = (2, 2)$ και $strides = (1, 1)$, αλλιώς $kernel\_size = (3, 3)$ και $strides = (2, 2)$.

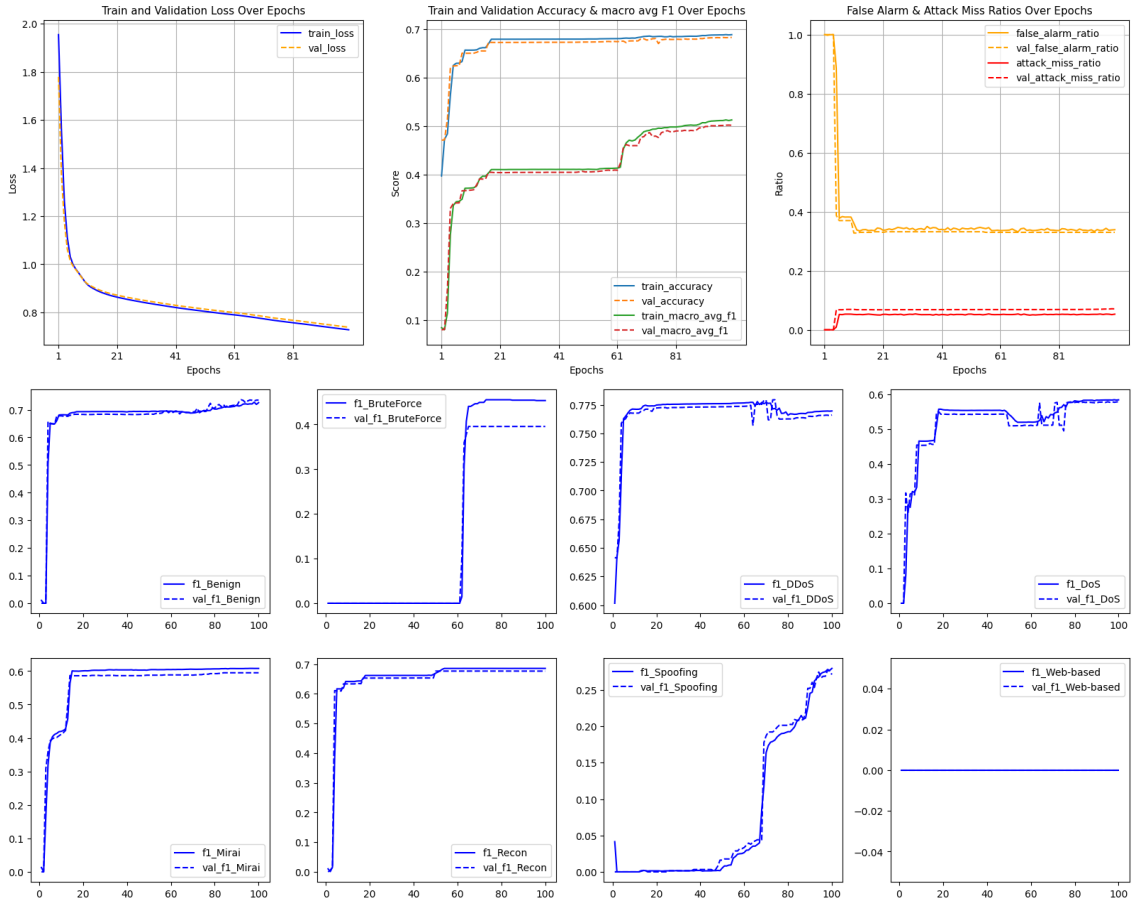


Σχήμα 5.10: *Αναπαράσταση Δεδομένων Εισόδου για Εκπαιδευμένα CNN Μοναδικού Καναλιού*

4. Επιπέδωση και Τελικό Επίπεδο: Το αποτέλεσμα από τα συνελικτικά επίπεδα εισάγεται σε ειδικό επίπεδο για να μετατραπεί σε μονοδιάστατο διάνυσμα. Το τελικό επίπεδο είναι ένα πυκνό επίπεδο με ενεργοποίηση softmax για την ταξινόμηση των δεδομένων σε 8 κατηγορίες.



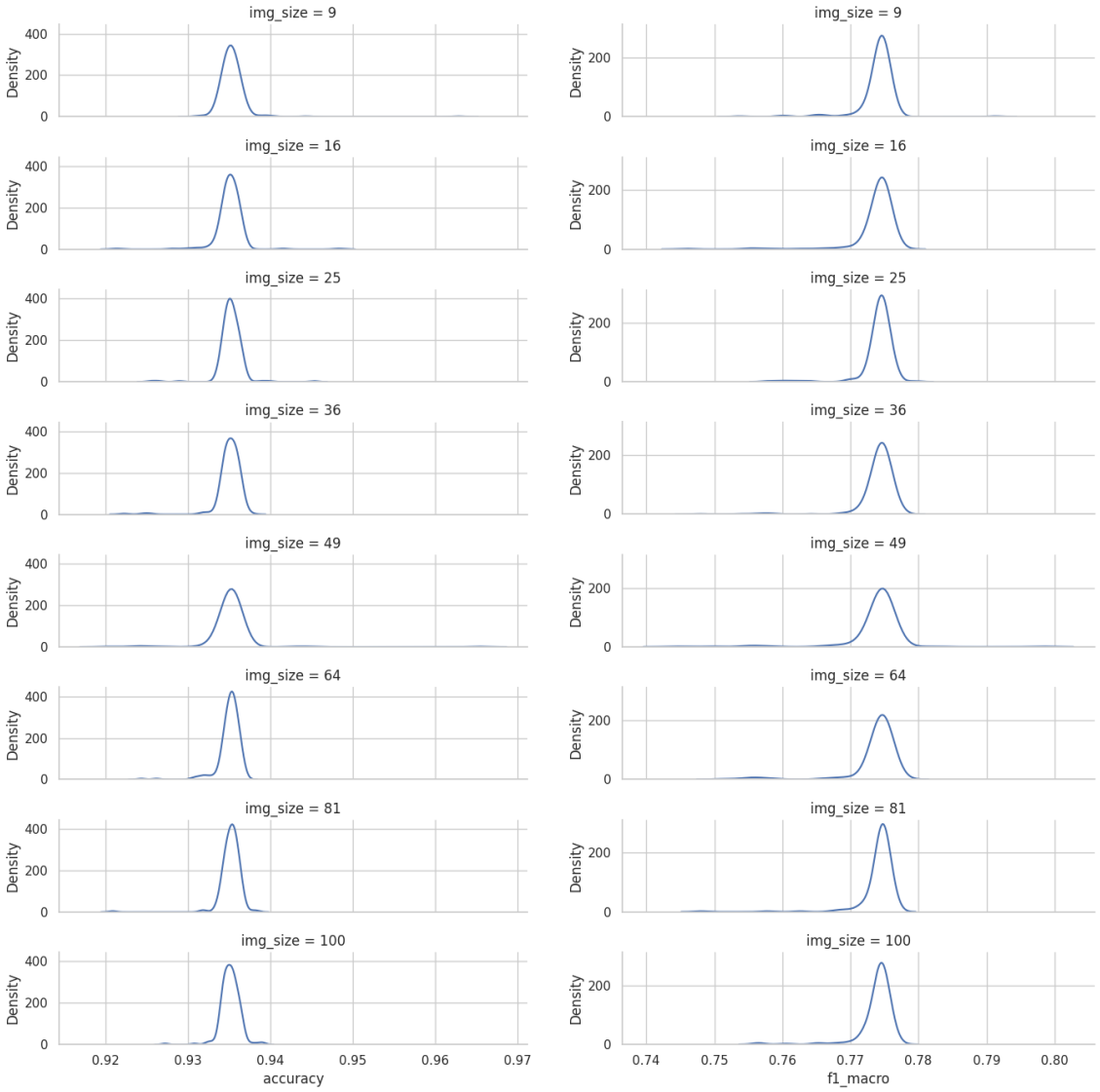
(α′) *Σχεδιασμός*



(β′) *Εκπαίδευση*

Σχήμα 5.11: *Παράδειγμα CNN Δύο Καναλιών*

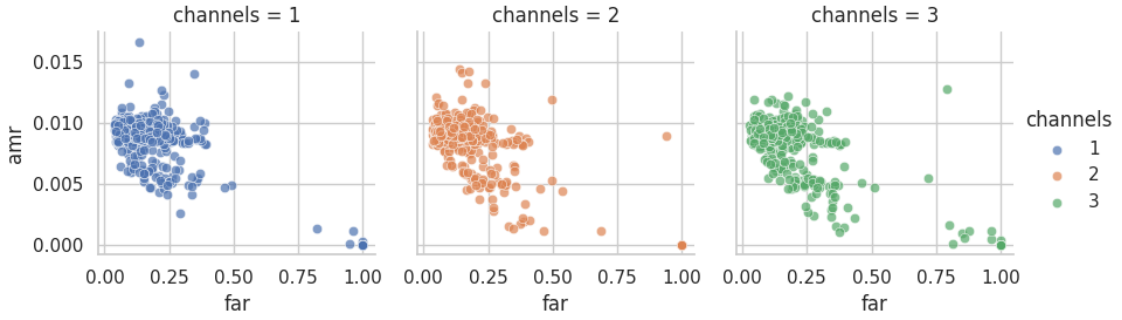
Αποτελέσματα



(α′) *Κατανομή σκορ ακρίβειας* (β′) *Κατανομή σκορ F1*

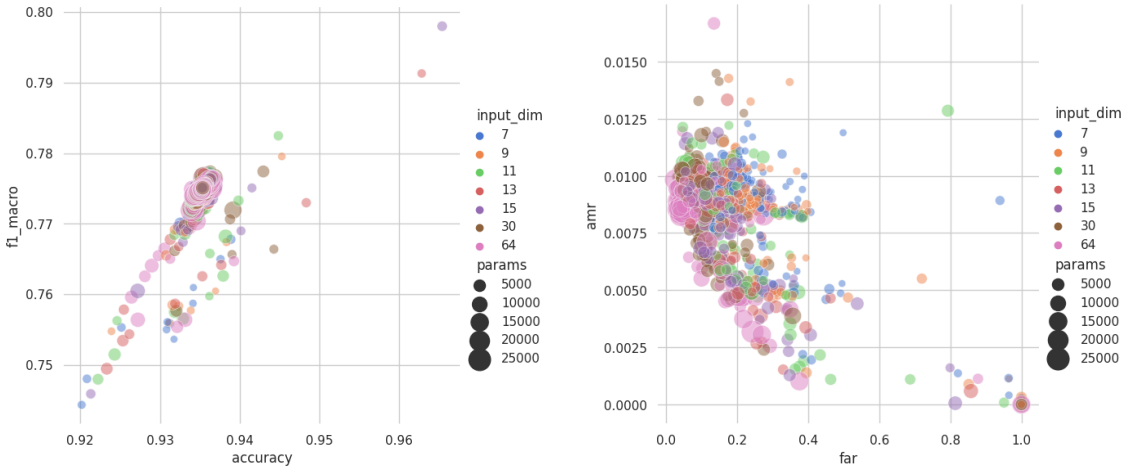
Σχήμα 5.12: *Κατανομές Ακρίβειας και F1 για CNN (Διαφορετικές Διαστάσεις Εικόνας)*

Οι κατανομές του Σχήματος 5.12 δείχνουν ότι ένα μεγαλύτερο μέγεθος εικόνας δεν σημαίνει απαραίτητα ότι είναι πιθανότερο να υπάρχουν καλύτερες τιμές F1 και ακρίβειας.



Σχήμα 5.13: *Απεικόνιση DET για CNN (Διαφορετικός Αριθμός Καναλιών)*

Τα διαγράμματα του Σχήματος 5.13 δείχνουν ότι τα μοντέλα με δύο ή τρία κανάλια έχουν καλύτερη απόδοση, καθώς βρίσκονται πιο κοντά στην κάτω αριστερή γωνία του διαγράμματος (σημείο (0,0)), υποδηλώνοντας χαμηλότερα ποσοστά AMR και FAR.



(α΄) *Ακρίβεια - F1* (β΄) *DET*

Σχήμα 5.14: *Διάγραμμα Διασποράς Ακρίβειας-F1 και DET για CNN*

Τέλος, στο διάγραμμα του Σχήματος 5.14β΄ παρατηρούμε πως τα μοντέλα που βρίσκονται πιο κοντά στο ιδανικό σημείο (0,0) έχουν μέγιστη διαστατικότητα εισόδου, και επομένως έχουν περισσότερες παραμέτρους.

5.2.5 Μετασχηματιστές

Επίπεδο Εισόδου

Στα μοντέλα κωδικοποιητών μετασχηματιστών τα δεδομένα εισόδου περνούν μέσα από πολυεπίπεδες διαδικασίες για να μπορέσουν να προσαρμοστούν στην αρχιτεκτονική. Η διαδικασία αυτή, μεταξύ άλλων, περιλαμβάνει δύο τρόπους επεξεργασίας των αριθμητικών χαρακτηριστικών εισόδου, τους οποίους ονομάζουμε «τρόπος 1» και «τρόπος 2» για την διευκόλυνση αναφοράς σε αυτούς. Γενικά, περιλαμβάνονται τα εξής βήματα:

1. Μετατροπή κατηγορικών δεδομένων σε διανύσματα αναπαράστασης: Τα κατηγορικά δεδομένα εισόδου έχουμε δει ήδη πως μετατρέπονται σε διανύσματα χρησιμοποιώντας την συνάρτηση embedding. Όλα τα διανύσματα των κατηγορικών δεδομένων συνενώνονται σε έναν πίνακα από tokens.

2. Μετατροπή αριθμητικών δεδομένων σε embeddings: Όπως είπαμε, δοκιμάζουμε δύο τρόπους επεξεργασίας των αριθμητικών δεδομένων

   - «Τρόπος 1»: Κάθε αριθμητική είσοδος τροφοδοτεί ένα ξεχωριστό πυκνό επίπεδο με αριθμό νευρώνων ίσο με τον προκαθορισμένο μέγεθος των embeddings. Έπειτα, η έξοδος αυτών εισέρχεται από ένα επίπεδο αναδιαμόρφωσης (reshape) όπου προστίθεται ακόμα μία διάσταση ώστε τελικά να συνενωθούν όλα σε έναν πίνακα.
   - «Τρόπος 2»: Όλες οι αριθμητικές είσοδοι συνενώνονται και τροφοδοτούν το ίδιο πυκνό επίπεδο με αριθμό νευρώνων ίσο με τον προκαθορισμένο μέγεθος των embeddings επί τον πλήθος των αριθμητικών χαρακτηριστικών. Έπειτα, η έξοδος εισέρχεται σε ένα επίπεδο αναδιαμόρφωσης (reshape) όπου ο μονοδιάστατος πίνακας που περιέχει τα embeddings όλων των αριθμητικών χαρακτηριστικών μετατρέπεται σε δισδιάστατος (αριθμός αριθμητικών χαρακτηριστικών x προκαθορισμένη διάσταση embeddings). Επί της ουσίας, η διαφορά έγκειται στο γεγονός πως κατά την εκπαίδευση, όλα τα διανύσματα των tokents που τροφοδοτούν τον κωδικοποιητή περιέχουν πληροφορία του συνδυασμού όλων των αριθμητικών χαρακτηριστικών, και όχι μόνο ενός.

3. Οι πίνακες των αριθμητικών και των κατηγορικών χαρακτηριστικών που έχουν προκύψει συνενώνονται ώστε να τροφοδοτήσουν τον κωδικοποιητή.

Συνδυασμοί Επιπέδων & Αρχιτεκτονική

Συνολικά σχεδιάσαμε 420 μοντέλα κωδικοποιητών μετασχηματιστών με διαφορετικούς συνδυασμούς επιπέδων. Συγκεκριμένα, για κάθε διάσταση εισόδου, κατασκευάσαμε αρχιτεκτονικές με:

1. Διαστάσεις Διανυσμάτων Αναπαράστασης (embedding dimensions) 5, 10, 20, 40 και 80.

2. Αριθμός Κεφαλών Προσοχής (attention heads) 1 ή 3.

3. Αριθμός Κωδικοποιητών (encoder layers) 1, 2 ή 3.

4. Χρήση ή μη χρήσης ξεχωριστών πυκνών επιπέδων για κάθε χαρακτηριστικό, δηλαδή «τρόπος 1» ή «τρόπος 2» παραγωγής embeddings.

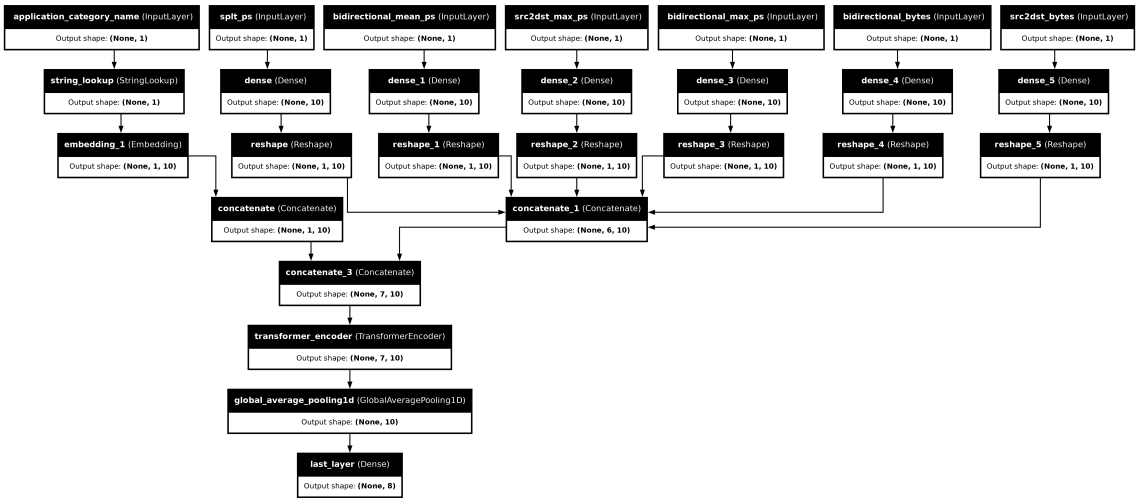
Κάθε μοντέλο κατασκευάστηκε με βάση την ακόλουθη γενική αρχιτεκτονική:

1. Είσοδος: Τα κωδικοποιημένα χαρακτηριστικά εισόδου, συνδυασμένα σε ένα ενιαίο επίπεδο.

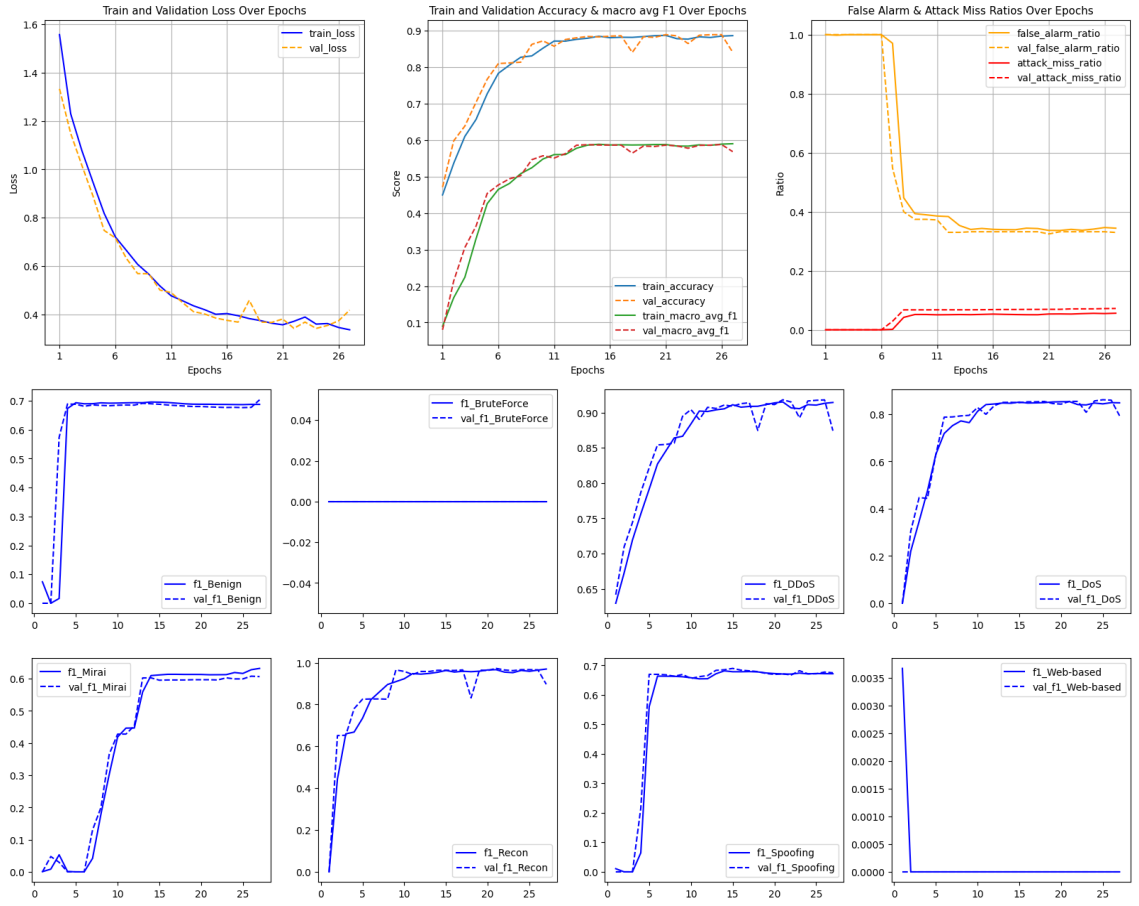
2. Κωδικοποιητές: Το αποτέλεσμα περνάει από έναν αριθμό κωδικοποιητών με πολλαπλές κεφαλές προσοχής και ενεργοποίηση ReLU.

3. Επίπεδο Συγκέντρωσης: Χρησιμοποιείται ένα επίπεδο μέσης συγκέντρωσης (GlobalAveragePooling1D) για να μειώσει τις διαστάσεις του αποτελέσματος.

4. Τελικό Επίπεδο: Το αποτέλεσμα περνάει από ένα τελικό πυκνό επίπεδο (Dense) με ενεργοποίηση softmax για την ταξινόμηση των δεδομένων σε 8 κατηγορίες.

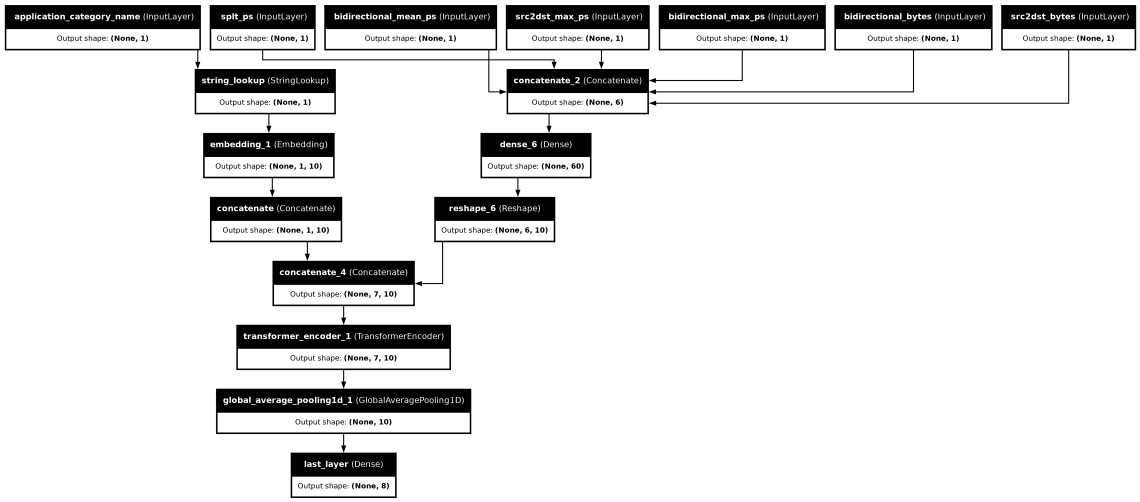


(α′) *Σχεδιασμός*

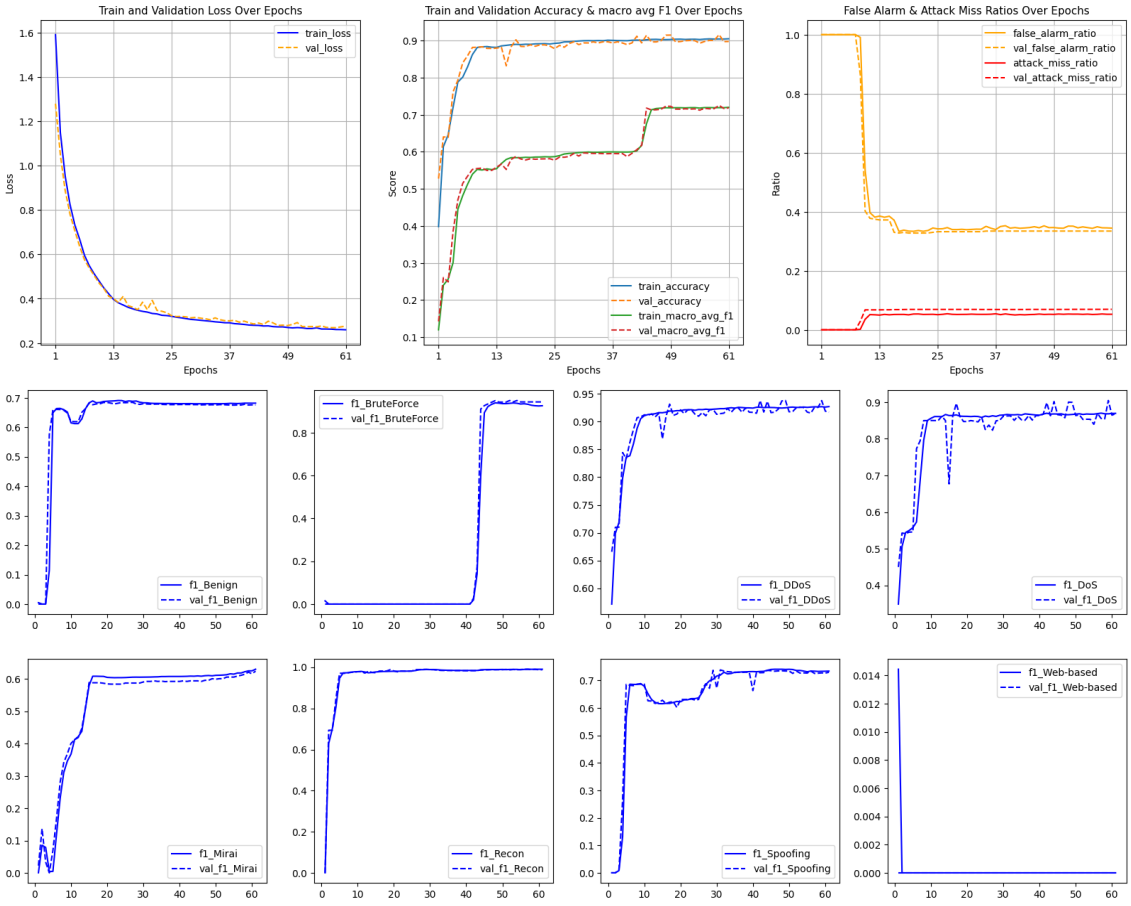


(β′) *Εκπαίδευση*

Σχήμα 5.15: *Παράδειγμα Transformer «Τρόπος 1»*



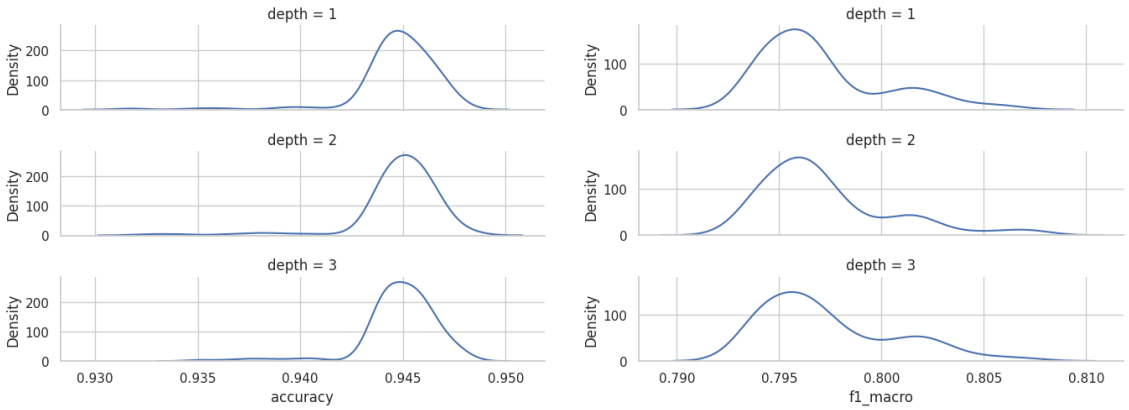
(α′) *Σχεδιασμός*



(β′) *Εκπαίδευση*

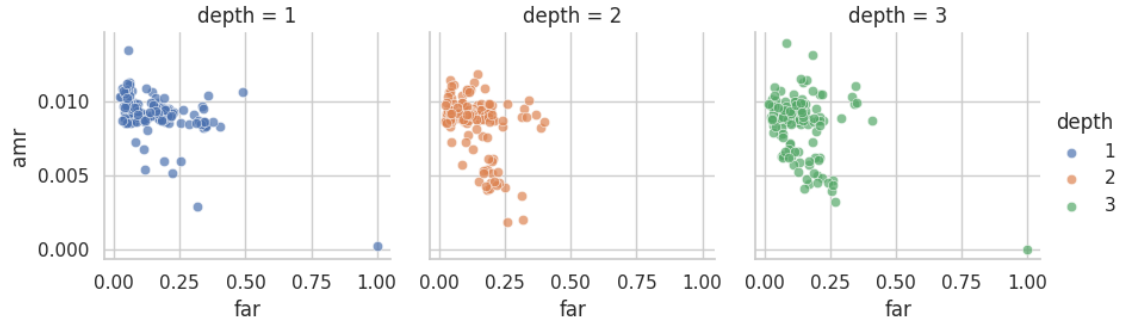
Σχήμα 5.16: *Παράδειγμα Transformer «Τρόπος 2»*

Αποτελέσματα



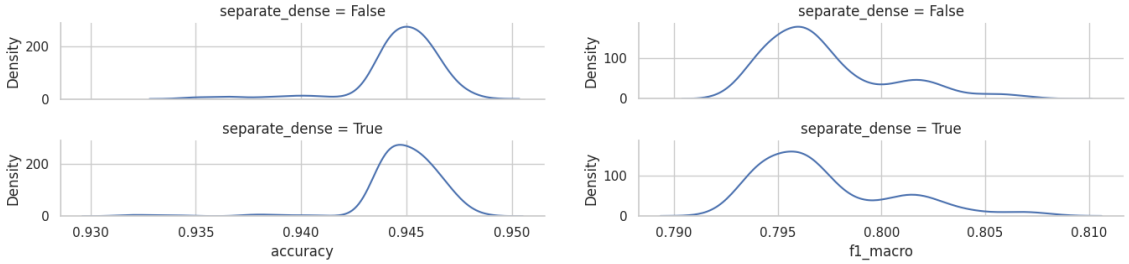
(α′) *Κατανομή Ακρίβειας* (β′) *Κατανομή F1*

Σχήμα 5.17: *Κατανομές Ακρίβειας και F1 για Transformer (Διαφορετικός Αριθμός Κωδικοποιητών)*



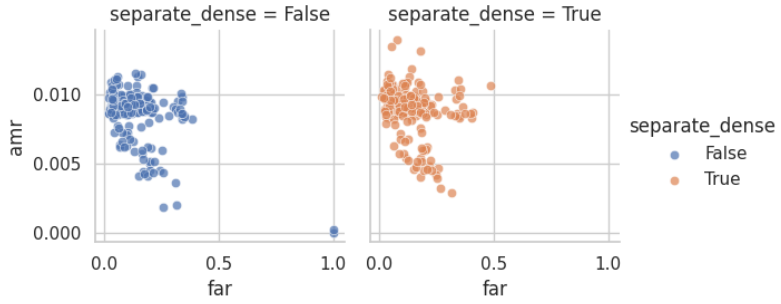
Σχήμα 5.18: *Απεικόνιση DET για Transforer (Διαφορετικός Αριθμός Κωδικοποιητών)*

Από τις κατανομές σκορ ακρίβειας και F1 (Σχήμα 5.17) παρατηρούμε πως για το βάθος του δικτύου δεν φαίνεται να επηρεάζει ιδιαίτερα τις κατανομές. Ωστόσο, τα διαγράμματα DET για τα διαφορετικά βάθη (Σχήμα 5.18) δείχνουν ότι τα μοντέλα με μεγαλύτερο βάθος έχουν χαμηλότερες τιμές AMR και FAR. Η διαφορά στην απόδοση είναι εμφανής με περισσότερα σημεία να συγκεντρώνονται κοντά στο ιδανικό σημείο (κάτω αριστερή γωνία) για βάθος 2 και 3.



(α′) *Κατανομή Ακρίβειας* (β′) *Κατανομή F1*

Σχήμα 5.19: *Κατανομές Ακρίβειας και F1 για Transformer (Διαφορετικοί Τρόποι Δημιουργίας Embeddings)*

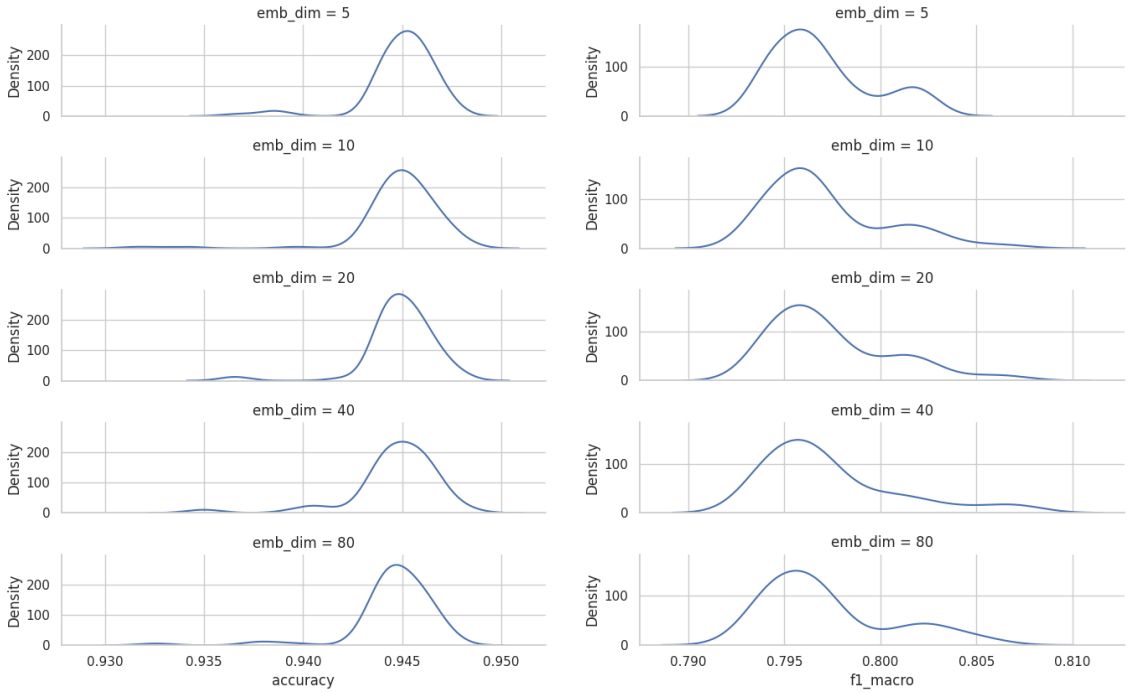


Σχήμα 5.20: *Απεικόνιση DET για Transforer (Διαφορετικοί Τρόποι Δημιουργίας Embeddings)*

Από τις κατανομές σκορ ακρίβειας και F1 για τους διαφορετικούς τρόπους δημιουργίας embeddings (separate_dense = «τρόπος 1») (Σχήμα 5.19), παρατηρούμε τα εξής:

- Ακρίβεια : Η χρήση του «τρόπου 1» ή του «τρόπου 2» δεν φαίνεται να επηρεάζει σημαντικά την ακρίβεια, με τις κορυφές των κατανομών να είναι κοντά στο 0.94.
- F1: Οι κατανομές του F1-score πάλι δείχνουν ότι η χρήση του «τρόπου 1» ή το «τρόπου 2» έναντι του άλλου δεν βελτιώνει αισθητά την απόδοση. Ωστόσο, η κατανομή του «τρόπου 2» φτάνει έως την ακραία τιμή του (0.81) σε αντίθεση με του «τρόπου 1».

Τα διαγράμματα DET για τους διαφορετικούς τρόπους μετατροπής σε embeddings (σχήμα 5.20) δείχνουν ότι κάποια λίγα μοντέλα με χρήση του «τρόπου 2» έχουν καταφέρει να συγκεντρώσουν πολύ πιο χαμηλά AMR. Ωστόσο δεν φαίνεται να μπορεί να γίνει κάποια σημαντική σύγκριση.

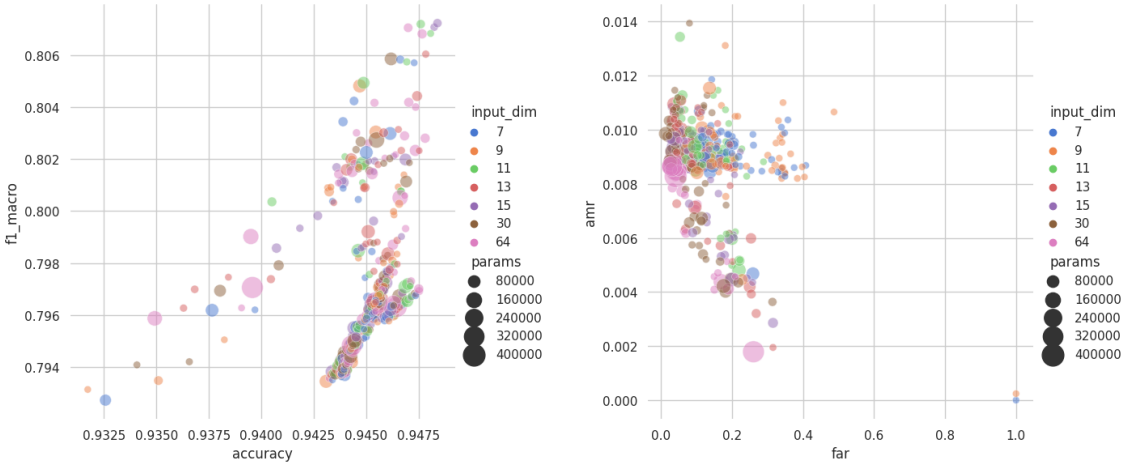


(α′) *Κατανομή Ακρίβειας* (β′) *Κατανομή F1*

Σχήμα 5.21: *Κατανομές Ακρίβειας και F1 για Transformer (Διαφορετικές Διαστάσεις Embeddings)*

Το Σχήμα 5.21 παρουσιάζει τις κατανομές των σκορ ακρίβειας και F1 για διαφορετικές διαστάσεις embeddings (emb_dim στο διάγραμμα). Παρατηρούμε πως όσο αυξάνεται η διάσταση των embeddings από 5 σε 80, δεν παρατηρείται κάποια γενική αύξηση της μέσης ακρίβειας. Υπάρχει σταθερή κατανομή με μικρή απόκλιση στις υψηλότερες διαστάσεις, δείχνοντας σταθερή απόδοση.

Αντιθέτως, στις κατανομές της F1 παρατηρείτε πως καθώς αυξάνεται ο αριθμός των διαστάσεων των διανυσμάτων, παρουσιάζονται θετικές μετατοπίσεις, καθώς περισσότερη μάζα μετατοπίζεται προς τις μεγαλύτερες τιμές. Το ίδιο παρατηρούμε και στο Σχήμα 5.22β′ που αφορά τις διαστάσεις εισόδου.

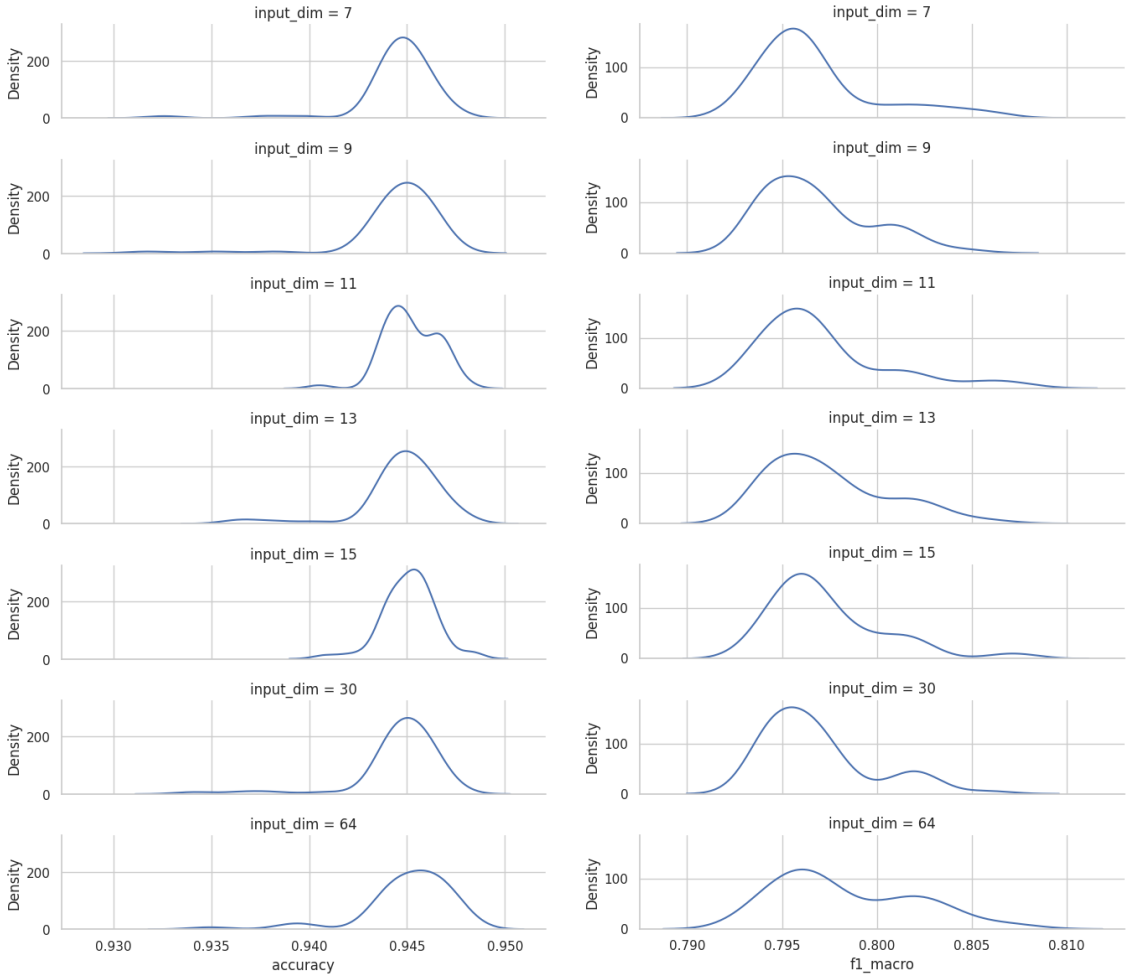


(α′) *Ακρίβεια - F1* (β′) *DET*

Σχήμα 5.23: *Διάγραμμα Διασποράς Ακρίβειας-F1 και DET για Transformer*

Τέλος, παρατηρούμε στο DET του Σχήματος 5.23β′ πως τα σημεία δείχνουν ότι οι διαφορετικές διαστάσεις εισόδου επηρεάζουν την κατανομή της απόδοσης, με τις υψηλότερες διαστάσεις να παρέχουν καλύτερη και πιο σταθερή απόδοση (πιο κοντά στο σημείο (0,0)).

Γενικά, παρατηρούμε πως η αύξηση της διάστασης embeddings και εισόδου βελτιώνει τα F1 scores. Οι υψηλότερες διαστάσεις οδηγούν σε πιο συγκεντρωμένες κατανομές, υποδεικνύοντας μεγαλύτερη σταθερότητα και αξιοπιστία στην απόδοση των μοντέλων.



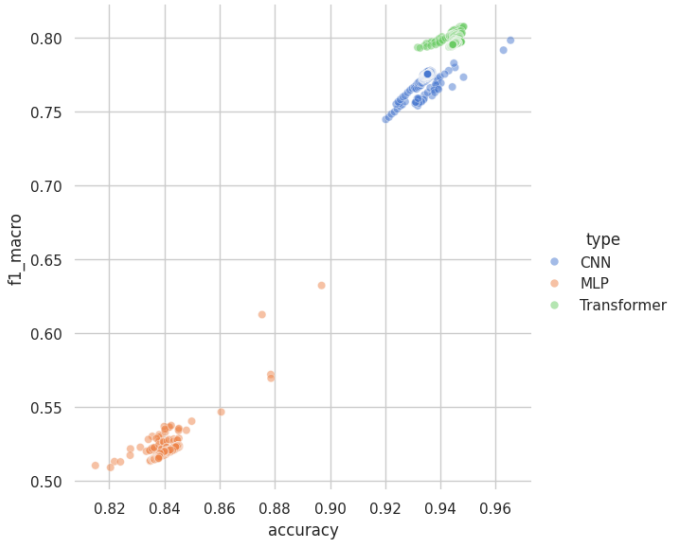
(α′) *Κατανομή Ακρίβειας* (β′) *Κατανομή F1*

Σχήμα 5.22: *Κατανομές Ακρίβειας F1 για Transformer (Διαφορετικές Διαστάσεις Εισόδου)*

5.3 Γενικές Παρατηρήσεις Αποτελεσμάτων

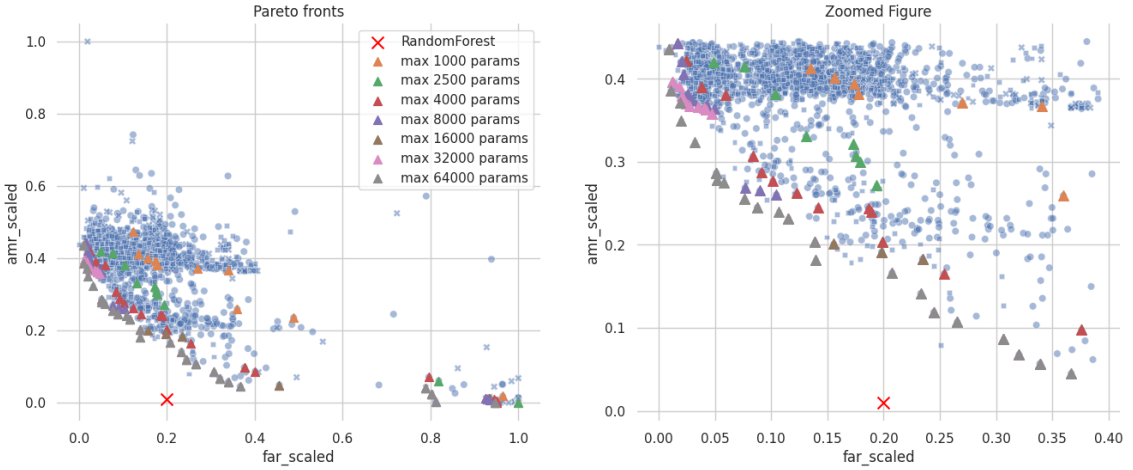
Αρχικά, στο Σχήμα 5.24 παρουσιάζεται η απόδοση των μοντέλων MLPs, CNNs, και Transformers με βάση τις μετρικές F1 και ακρίβειας. Από την ανάλυση του διαγράμματος παρατηρούμε τα εξής:

- **Transformers**: Αυτά τα μοντέλα έχουν την καλύτερη απόδοση ως προς την μετρική F1, καθώς συγκεντρώνονται στα υψηλότερα σημεία του διαγράμματος.
- **CNNs**: Γενικά παρουσιάζουν ελαφρώς χαμηλότερες τιμές F1 σε σχέση με τα Transformers και περίπου ίδιες τιμές ακρίβειας. Ωστόσο, υπάρχουν λίγες περιπτώσεις μοντέλων που παρουσιάζουν ελαφρώς καλύτερη ακρίβεια.
- **MLPs**: Έχουν τη χαμηλότερη απόδοση σε σύγκριση με τα άλλα δύο είδη μοντέλων, με τις τιμές F1 και ακρίβειας να είναι αισθητά χαμηλότερες.



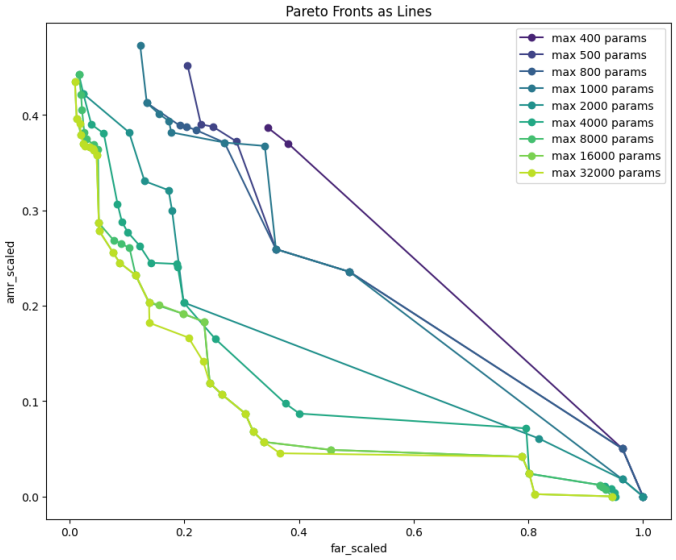
Σχήμα 5.24: *Απόδοση MLPs, CNNs, και Transformers (Ακρίβεια και F1)*

Το Σχήμα 5.25 απεικονίζει τα Pareto Fronts για τα μοντέλα με διαφορετικό μέγιστο αριθμό παραμέτρων στο διάγραμμα DET. Στο διάγραμμα αυτό, τα μοντέλα απεικονίζονται ως προς τις μετρικές FAR και AMR, που είναι οι μετρικές που θέλουμε να ελαχιστοποιήσουμε. Η βελτίωση της μίας μετρικής συνήθως επιδεινώνει την άλλη, και το Pareto Front μας βοηθά να βρούμε τα καλύτερα δυνατά μοντέλα που εξισορροπούν αυτές τις δύο αντικρουόμενες μετρικές. Πιο συγκεκριμένα, ένα σημείο στο Pareto Front σημαίνει ότι δεν υπάρχει άλλο μοντέλο που να είναι καλύτερο και στις δύο μετρικές ταυτόχρονα. Αν ένα μοντέλο έχει χαμηλότερο FAR από ένα άλλο χωρίς να έχει χειρότερο AMR, τότε το πρώτο μοντέλο θεωρείται καλύτερο. Έτσι, τα μοντέλα που βρίσκονται στο Pareto Front είναι τα βέλτιστα, καθώς αντιπροσωπεύουν την καλύτερη δυνατή ισορροπία μεταξύ FAR και AMR.



Σχήμα 5.25: *DET Pareto Fronts για Μοντέλα Διαφορετικού Μέγιστου Αριθμού Παραμέτρων*

Στο Σχήμα 5.26 παρατηρούμε πιο καθαρά ότι όσο αυξάνεται ο μέγιστος αριθμός παραμέτρων, τόσο πιο κοντά σχηματίζεται η καμπύλη του Pareto Front στο ιδανικό σημείο. Παρατηρούμε ότι μετά τις 16.000 παραμέτρους, τα Pareto Fronts αρχίζουν να ταυτίζονται, ενώ για μικρότερους αριθμούς παραμέτρων υπάρχει εμφανής διαφορά. Αυτό υποδηλώνει ότι η αύξηση των παραμέτρων βελτιώνει την απόδοση των μοντέλων μέχρι ενός σημείου, μετά το οποίο η απόδοση φαίνεται να σταθεροποιείται.



Σχήμα 5.26: *Επίδραση Αριθμού Παραμέτρων στο Pareto Front*

5.4 Σύγκριση Μοντέλων

Συγκρίνουμε τα μοντέλα σε κατηγορίες για μέγιστο αριθμό παραμέτρων 400, 500, 800, 1.000, 2.000, 4.000, 8.000, 16.000 και 32.000. Για κάθε μοντέλο υπολογίζονται τα KBs απαιτούμενης RAM για να εκτελεστούν τα μοντέλα, με δεδομένο πως η κάθε παράμετρος αντιστοιχεί σε 4 bytes. Για την κάθε κατηγορία μπορούμε να πάρουμε 3 βέλτιστα μοντέλα,

ένα για κάθε μέγιστο των μετρικών accuracy, macro average F1-score και τέλος καλύτερου DET.

5.4.1 Κριτήριο Βέλτιστης Ακρίβειας

Πίνακας 5.5: *Τα Αποτελέσματα των Βέλτιστων Μοντέλων Βάσει Accuracy στις Διαφορετικές Κατηγορίες Μέγιστου Μεγέθους*

Βέλτιστα Μοντέλα Βάσει Accuracy										
Μέγιστο Μέγεθος	Τύπος	Πλήθος Εισόδων	Παράμετροι	accuracy	F1 macro	precision micro	recall micro	AMR	FAR	KB
400	Transformer	7	382	0.947	0.796	0.959	0.937	0.0104	0.3575	1.49
500	Transformer	9	402	0.947	0.801	0.959	0.937	0.0083	0.4041	1.57
800	Transformer	9	402	0.947	0.801	0.959	0.937	0.0083	0.4041	1.57
1000	Transformer	9	402	0.947	0.801	0.959	0.937	0.0083	0.4041	1.57
2000	CNN	13	1983	0.963	0.791	0.966	0.962	0.0096	0.1749	7.75
4000	CNN	15	3355	0.965	0.798	0.969	0.964	0.0089	0.1383	13.11
8000	CNN	15	3355	0.965	0.798	0.969	0.964	0.0089	0.1383	13.11
16000	CNN	15	3355	0.965	0.798	0.969	0.964	0.0089	0.1383	13.11
32000	CNN	15	3355	0.965	0.798	0.969	0.964	0.0089	0.1383	13.11
∞	CNN	15	3355	0.965	0.798	0.969	0.964	0.0089	0.1383	13.11

5.4.2 Κριτήριο Βέλτιστης Βαθμολογίας F1

Πίνακας 5.6: *Τα Αποτελέσματα των Βέλτιστων Μοντέλων Βάσει Βαθμολογία F1 στις Διαφορετικές Κατηγορίες Μέγιστου Μεγέθους*

Βέλτιστα Μοντέλα Βάσει Βαθμολογία F1										
Μέγιστο Μέγεθος	Τύπος	Πλήθος Εισόδων	Παράμετροι	accuracy	F1 macro	precision micro	recall micro	AMR	FAR	KB
400	Transformer	7	382	0.947	0.796	0.959	0.937	0.0104	0.3575	1.49
500	Transformer	9	402	0.947	0.801	0.959	0.937	0.0083	0.4041	1.57
800	Transformer	9	402	0.947	0.801	0.959	0.937	0.0083	0.4041	1.57
1000	Transformer	9	402	0.947	0.801	0.959	0.937	0.0083	0.4041	1.57
2000	Transformer	13	1793	0.948	0.802	0.959	0.938	0.0088	0.1749	7.00
4000	Transformer	7	2548	0.947	0.806	0.958	0.936	0.0087	0.2238	9.95
8000	Transformer	11	4009	0.948	0.807	0.958	0.937	0.0062	0.2093	15.66
16000	Transformer	15	12548	0.948	0.807	0.958	0.937	0.0107	0.0351	49.02
32000	Transformer	15	28448	0.948	0.807	0.958	0.938	0.0068	0.1283	111.12
∞	Transformer	15	28448	0.948	0.807	0.958	0.938	0.0068	0.1283	111.12

5.4.3 Κριτήριο Βέλτιστης Αντιστάθμισης Σφάλματος Ανίχνευσης

Πίνακας 5.7: *Τα Αποτελέσματα των Βέλτιστων Μοντέλων Βάσει DET στις Διαφορετικές Κατηγορίες Μέγιστου Μεγέθους*

Βέλτιστα Μοντέλα Βάσει DET										
Μέγιστο Μέγεθος	Τύπος	Πλήθος Εισόδων	Παράμετροι	accuracy	F1 macro	precision micro	recall micro	AMR	FAR	KB
400	CNN	9	336	0.936	0.775	0.951	0.922	0.0087	0.3545	1.31
500	CNN	7	422	0.935	0.774	0.950	0.920	0.0058	0.3682	1.65
800	CNN	9	752	0.935	0.774	0.950	0.919	0.0087	0.2024	2.94
1000	CNN	9	886	0.935	0.776	0.952	0.920	0.0086	0.1879	3.46
2000	Transformer	64	1876	0.944	0.794	0.956	0.933	0.0046	0.2093	7.33
4000	CNN	30	3538	0.936	0.775	0.951	0.922	0.0055	0.1528	13.82
8000	Transformer	64	4528	0.943	0.794	0.956	0.933	0.0046	0.1497	17.69
16000	Transformer	64	4528	0.943	0.794	0.956	0.933	0.0046	0.1497	17.69
32000	Transformer	64	20330	0.948	0.797	0.959	0.938	0.0041	0.1505	79.41
∞	Transformer	64	20330	0.948	0.797	0.959	0.938	0.0041	0.1505	79.41

Περιγραφή Διαδικασίας Επιλογής Βέλτιστου Μοντέλου με Αντιστάθμιση Σφάλματος Ανίχνευσης

Η διαδικασία εύρεσης του βέλτιστου μοντέλου βάσει DET πραγματοποιείται μέσω της ανάλυσης των Pareto fronts τα οποία βοηθούν στην αναγνώριση των μοντέλων που προσφέρουν την καλύτερη ισορροπία μεταξύ των αντικρουόμενων μετρικών, δηλαδή του FAR και AMR. Ακολουθεί η περιγραφή της διαδικασίας για δύο και τρεις διαστάσεις.

Στην δισδιάστατη περίπτωση, για κάθε μοντέλο στο Pareto front, υπολογίζουμε την Ευκλείδεια απόσταση αυτού από ένα ιδανικό σημείο που αντιπροσωπεύει ένα μοντέλο με ιδανικές επιδόσεις. Το Ουτοπικό Σημείο (Utopia Point), το οποίο ορίζεται ως ένα θεωρητικό σημείο στο Pareto front, όπου κάθε παράμετρος έχει την καλύτερη δυνατή απόδοση θα μπορούσε να αποτελεί ένα τέτοιο σημείο. Δημιουργείτε συνδυάζοντας τα δύο σημεία με τις βέλτιστες επιδόσεις για κάθε παράμετρο, δημιουργώντας ένα σημείο αναφοράς που, αν και δεν είναι πάντα επιτεύξιμο, καθοδηγεί τη βελτιστοποίηση των συστημάτων[40]. Ωστόσο, μπορούμε, επίσης, να χρησιμοποιήσουμε ως ιδανικό σημείο το μοντέλο του RandomForest το οποίο παρουσιάζει $\text{FAR}_{RF} = 0.08$ και $\text{AMR}_{RF} = 0.003$, δηλαδή αρκετά καλύτερες επιδόσεις από τα υπόλοιπα μοντέλα. Επομένως, χρησιμοποιώντας τον γνωστό μαθηματικός τύπος για την απόσταση d λαμβάνουμε τις αποστάσεις των σημείων από το RF:

$$d = \sqrt{(\text{FAR}_{model} - \text{FAR}_{RF})^2 + (\text{AMR}_{model} - \text{AMR}_{RF})^2}$$

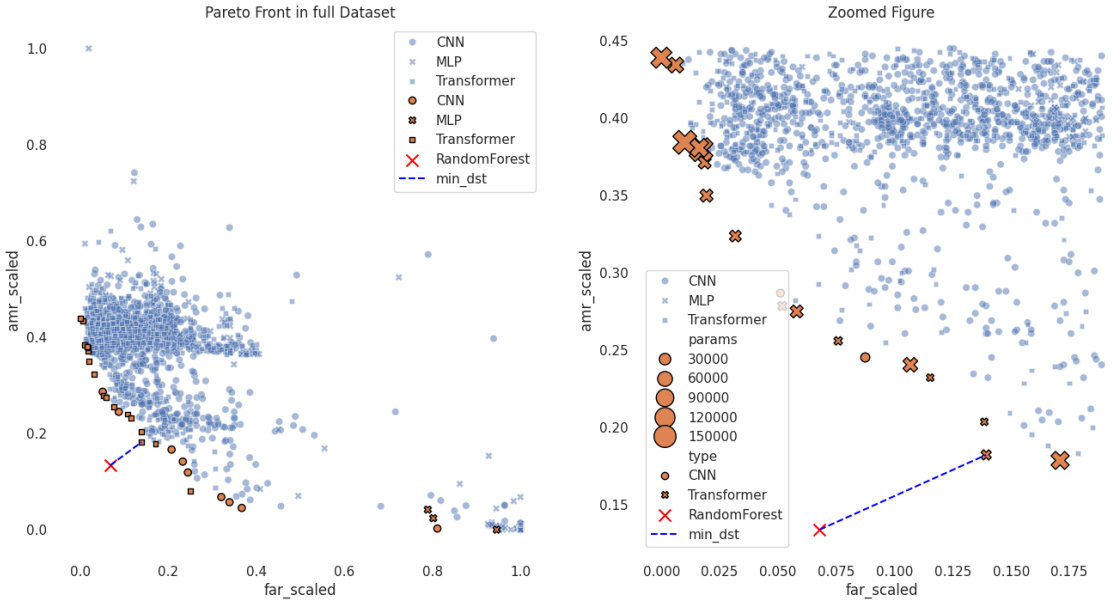
Είναι σημαντικό να προσθέσουμε πως πριν τον υπολογισμό της απόστασης εφαρμόζουμε min-max scaling στις μετρικές FAR και AMR για να διασφαλίσουμε ότι η απόσταση υπολογίζεται με τέτοιο τρόπο έτσι ώστε οι μετρικές να λαμβάνουν το ίδιο βάρος. Η φόρμουλα του

min-max scaling για μια μεταβλητή x είναι:

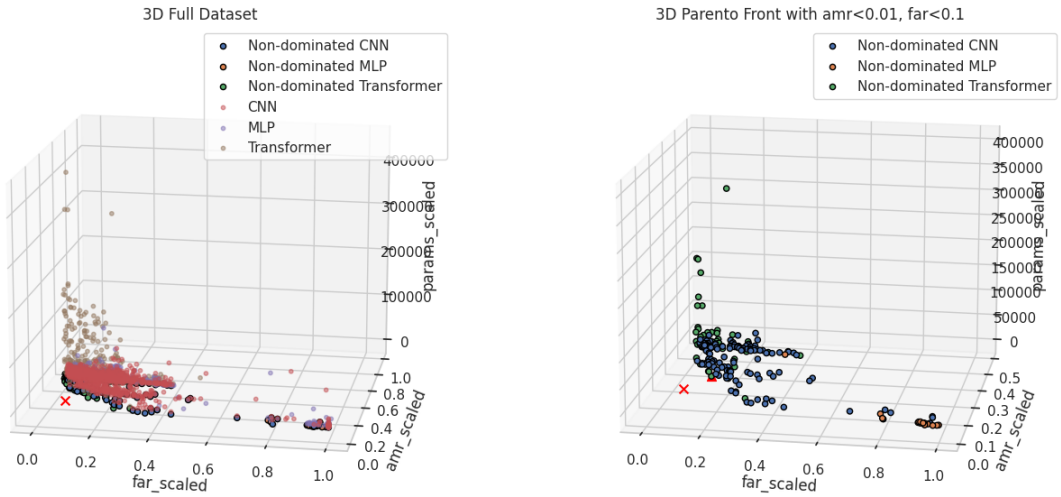
$$x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

Αυτό εξασφαλίζει ότι οι τιμές είναι μεταξύ 0 και 1.

Το μοντέλο με τη μικρότερη απόσταση από το σημείο $(0, 0)$ θεωρείται το βέλτιστο, καθώς προσφέρει την καλύτερη ισορροπία μεταξύ των δύο μετρικών. Στο παράδειγμα του Σχήματος 5.27 ως ιδανικό σημείο χρησιμοποιείται το μοντέλο του RandomForest το οποίο παρουσιάζει βέλτιστες επιδόσεις και χρησιμοποιείται ως σημείο αναφοράς.



Σχήμα 5.27: *Οπτικοποίηση Pareto Front σε 2 Διαστάσεις*



Σχήμα 5.28: *Οπτικοποίηση Pareto Front σε 3 Διαστάσεις*

Στην τρισδιάστατη ανάλυση, προσθέτουμε και το μέγεθος των μοντέλων ως τρίτη διάσταση. Ακολουθούμε την ίδια διαδικασία, δηλαδή εφαρμογή min-max scaling στη διάσταση

του μεγέθους, εφόσον θέλουμε να δίνεται το ίδιο βάρος μεταξύ των τριών διαστάσεων στην Ευκλείδεια απόσταση.

Αυτή η μεθοδολογία εξασφαλίζει ότι τα μοντέλα αξιολογούνται με ίσο βάρος στις κρίσιμες μετρικές, παρέχοντας έτσι μια αντικειμενική επιλογή του βέλτιστου μοντέλου με βάση το DET.

Κεφάλαιο 6

Επίλογος

Στο πλαίσιο της παρούσας διπλωματικής εργασίας, διερευνήθηκαν σύγχρονες τεχνικές μηχανικής και βαθιάς μάθησης για την ανίχνευση εισβολών σε περιβάλλοντα IoT. Η εργασία αυτή ανέδειξε τις προκλήσεις που συνοδεύουν την ανάπτυξη και την εφαρμογή αποδοτικών μοντέλων υψηλής ακρίβειας για την ασφάλεια των δικτύων, εστιάζοντας στο κριτήριο της επιβάρυνσης των συσκευών.

Συνδυάζοντας τα αποτελέσματα των αναλύσεων και των πειραμάτων με τα θέματα που καλύφθηκαν στα προηγούμενα κεφάλαια, γίνεται σαφές ότι η ανάπτυξη αποτελεσματικών μοντέλων ανίχνευσης εισβολών απαιτεί μια ολοκληρωμένη προσέγγιση. Στο Κεφάλαιο 2, παρουσιάστηκαν οι βασικές αρχές της μηχανικής και βαθιάς μάθησης, καθώς και οι προκλήσεις που αντιμετωπίζουν τα δομημένα σύνολα δεδομένων και τα προβλήματα ταξινόμησης με ανισόρροπες κλάσεις, κάτι πολύ συχνό στα προβλήματα ασφάλειας δικτύων. Στο Κεφάλαιο 3 αναλύθηκαν τα συστήματα ανίχνευσης εισβολών, ο τρόπος αξιολόγησής τους και οι ιδιαίτερες ανάγκες ασφάλειας των συσκευών του IoT. Το Κεφάλαιο 4 περιέγραψε την πειραματική διάταξη και την προετοιμασία των δεδομένων, ενώ το Κεφάλαιο 5 επικεντρώθηκε στην εκπαίδευση και αξιολόγηση των μοντέλων.

Η ανάπτυξη μοντέλων βαθιάς μάθησης που μπορούν να λειτουργήσουν αποτελεσματικά σε περιβάλλοντα IoT με περιορισμένους πόρους αποτελεί μια κρίσιμη ανάγκη στον τομέα της κυβερνοασφάλειας. Με αυτήν την εργασία, έγινε ένα βήμα προς την κατεύθυνση της αυτονομίας των δικτυακών συσκευών στην ανίχνευση εισβολών, προτείνοντας λύσεις που εξισορροπούν την απόδοση και την αποδοτικότητα.

Στην συνέχεια, παρουσιάζονται τα συμπεράσματα που προέκυψαν από την ανάλυση και τα πειράματα που πραγματοποιήθηκαν, καθώς και οι μελλοντικές κατευθύνσεις για περαιτέρω βελτιώσεις και έρευνα στον τομέα αυτό.

6.1 Συμπεράσματα

Επιχειρούμε να συνοψίσουμε τα κύρια ευρήματα και συμπεράσματα που προέκυψαν από την ανάλυση και τις δοκιμές που πραγματοποιήθηκαν στη μελέτη αυτή. Μέσα από τη διερεύνηση διαφόρων τεχνικών και εργαλείων, καθώς και την αξιολόγηση της απόδοσης πολλαπλών μοντέλων νευρωνικών δικτύων, αναδείχθηκαν σημαντικά σημεία που αφορούν την αποδοτικότητα και την αποτελεσματικότητα στην ανίχνευση εισβολών σε περιβάλλοντα IoT.

1. Η ανισορροπία των κλάσεων αποτελεί ένα από τα σημαντικότερα εμπόδια για την α-

νάπτυξη αποτελεσματικών μοντέλων βαθιάς μάθησης. Οι τεχνικές υπερδειγματοληψίας, αν και θεωρητικά επαρκείς, αποδεικνύονται μη ρεαλιστικές λόγω της έντονης ανισορροπίας που παρατηρείται στα δεδομένα. Αυτό το ζήτημα καθιστά την ανάπτυξη ακριβών μοντέλων ιδιαίτερη πρόκληση και αναγκάζει την επιστημονική κοινότητα να βρίσκεται σε συνεχή αναζήτηση νέων λύσεων. Η προστασία των προσωπικών δεδομένων, καθώς και η φύση των επιθέσεων, δυσχεραίνουν την παραγωγή ενός υψηλής ποιότητας συνόλου δεδομένων, επιτείνοντας το πρόβλημα της ανισορροπίας.

2. Στις εφαρμογές μηχανικής μάθησης, η διαδικασία εξαγωγής χαρακτηριστικών παρουσιάζει ποικιλομορφία, με πολλούς διαφορετικούς τρόπους να εφαρμόζονται ανάλογα με τις ανάγκες της εκάστοτε έρευνας. Η υιοθέτηση του εργαλείου NFStream σε ευρεία κλίμακα θα μπορούσε να προσφέρει μια ενοποιημένη και αξιόπιστη προσέγγιση στην εξαγωγή χαρακτηριστικών, καθιστώντας τα αποτελέσματα των ερευνών συγκρίσιμα. Η απόδοση του αλγορίθμου τυχαίου δάσους σε συνδυασμό με το σύνολο δεδομένων CICIoT2023 επιβεβαίωσε την αξιοπιστία του NFStream, αποδεικνύοντας ότι προσφέρει μια σταθερή και αποδοτική λύση σε αυτά τα προβλήματα.

3. Από την αξιολόγηση των μοντέλων, διαπιστώθηκε ότι τα μοντέλα μετασχηματιστών προσφέρουν την υψηλότερη απόδοση, ακολουθούμενα από τα συνελικτικά δίκτυα, ενώ τα πολυεπίπεδα Perceptrons παρουσίασαν σημαντικά χαμηλότερη απόδοση. Η διαδικασία μετατροπής των δομημένων δεδομένων σε μη δομημένες μορφές απέδωσε θετικά αποτελέσματα, επιτρέποντας την αποτελεσματικότερη χρήση των μοντέλων. Ωστόσο, κανένα από τα δοκιμασμένα μοντέλα δεν κατάφερε να ξεπεράσει ή να φτάσει τις επιδόσεις του τυχαίου δάσους, κάτι που υποδηλώνει την ανωτερότητα αυτού του αλγορίθμου στις συγκεκριμένες συνθήκες.

4. Η αξιολόγηση των μοντέλων δεν μπορεί να βασίζεται σε ένα μόνο κριτήριο, καθώς οι απαιτήσεις και οι στόχοι της κάθε εφαρμογής διαφέρουν. Ανάλογα με την εκάστοτε εφαρμογή, ορισμένα μοντέλα μπορεί να είναι πιο κατάλληλα από άλλα. Για παράδειγμα, εάν το κύριο ενδιαφέρον εστιάζεται στην ακριβή ταξινόμηση των επιθέσεων χωρίς να λαμβάνεται υπόψη ο αριθμός των χαμένων επιθέσεων ή των ψευδών συναγερμών, τότε το κριτήριο F1 αποτελεί την καταλληλότερη επιλογή. Αντίθετα, για την εξισορρόπηση του FAR και του AMR, το κριτήριο DET είναι πιο ενδεδειγμένο. Αξιοσημείωτο είναι ότι υπάρχουν μοντέλα με πολύ μικρό μέγεθος, σχεδόν 1 KB, τα οποία επιτυγχάνουν ικανοποιητικά αποτελέσματα, υποδεικνύοντας ότι οι συσκευές IoT χαμηλού επιπέδου μπορούν να επωφεληθούν από τη χρήση αυτών των μοντέλων.

5. Η ανάλυση των μοντέλων έδειξε ότι η αντιστάθμιση μεταξύ FAR και AMR βελτιώνεται όταν τα μοντέλα διαθέτουν μεγαλύτερο μέγεθος παραμέτρων. Ωστόσο, αυτή η παρατήρηση δεν αποτελεί απόλυτο κανόνα, αλλά μάλλον μια γενική τάση. Φαίνεται ότι υπάρχει ένα κατώφλι απόδοσης, το οποίο δεν μπορεί να ξεπεραστεί με τις αρχιτεκτονικές που εξετάστηκαν, επισημαίνοντας την ανάγκη για περαιτέρω έρευνα και ανάπτυξη πιο αποδοτικών μοντέλων.

Τα παραπάνω συμπεράσματα αναδεικνύουν τις σημαντικές προκλήσεις και τις δυνατότητες

βελτίωσης που υπάρχουν στον τομέα της ανίχνευσης εισβολών με τη χρήση τεχνικών μηχανικής και βαθιάς μάθησης. Παρά τα εμπόδια, η παρούσα εργασία συμβάλλει σημαντικά στην κατανόηση και την ανάπτυξη πιο αποδοτικών και αποτελεσματικών μοντέλων για την ασφάλεια των δικτύων, ανοίγοντας τον δρόμο για περαιτέρω βελτιώσεις και καινοτομίες.

6.2 Μελλοντικές Κατευθύνσεις

Η παρούσα εργασία ανέδειξε πολλές σημαντικές προκλήσεις και δυνατότητες για την ανίχνευση εισβολών, χρησιμοποιώντας σύγχρονες τεχνικές μηχανικής και βαθιάς μάθησης και απέδειξε πως υπάρχουν πολλά περιθώρια για περαιτέρω βελτιώσεις και έρευνα. Οι μελλοντικές κατευθύνσεις που προτείνονται επικεντρώνονται στη βελτίωση της ακρίβειας, της αποδοτικότητας και της γενίκευσης των μοντέλων ανίχνευσης, λαμβάνοντας υπόψη τις διαφορετικές ανάγκες και προκλήσεις που προκύπτουν από την ανάπτυξη και εφαρμογή τους σε πραγματικά περιβάλλοντα. Στη συνέχεια, παρουσιάζονται μερικές προτάσεις για περαιτέρω έρευνα.

1. **Δοκιμές για Δυαδική και Πολυταξική Ταξινόμηση**: Μια άμεση κατεύθυνση της παρούσας εργασίας είναι η διερεύνηση και αξιολόγηση των μοντέλων ανίχνευσης εισβολών τόσο για δυαδική ταξινόμηση όσο και για πολυταξική ταξινόμηση, άνω των 8 κλάσεων. Η δυαδική ταξινόμηση θα επιτρέψει την αναγνώριση της παρουσίας κακόβουλης δραστηριότητας με υψηλή ακρίβεια και χαμηλό κόστος υπολογιστικών πόρων, ενώ η πολυταξική ταξινόμηση, που περιλαμβάνει 34 κλάσεις στην περίπτωση του συνόλου δεδομένων που χρησιμοποιείται στην εργασία, θα προσφέρει ακριβή προσδιορισμό του τύπου της εισβολής, ενισχύοντας την ασφάλεια των δικτύων με στοχευμένη απόκριση σε κάθε απειλή. Παράλληλα, παρουσιάζει ενδιαφέρον να εξεταστεί το νέο μέγεθος των μοντέλων που θα προκύψουν από αυτές τις δοκιμές, σε σύγκριση με το μέγεθος των μοντέλων που χρησιμοποιήθηκαν στην παρούσα εργασία. Αυτή η σύγκριση θα βοηθήσει στην κατανόηση της επίδρασης του αριθμού των κλάσεων στο μέγεθος και την αποδοτικότητα των μοντέλων, παρέχοντας περισσότερες πληροφορίες για τη βελτιστοποίηση σε περιβάλλοντα με περιορισμένους πόρους.

2. **Μελέτη της Καθυστέρησης Ανίχνευσης σε Ολοκληρωμένο Σύστημα Προσομοίωσης**: Σημαντική θα ήταν επίσης η διερεύνηση της καθυστέρησης ανίχνευσης εισβολών σε ένα ολοκληρωμένο σύστημα προσομοίωσης. Συγκεκριμένα, θα πρέπει να μελετηθεί ο χρόνος λήξης των ροών από το NFStream, ώστε να διαπιστωθεί αν η χρήση μικρών χρονικών παραθύρων μπορεί να επιτευχθεί χωρίς απώλεια κρίσιμων πληροφοριών της ροής, ενώ παράλληλα να διατηρείται η υψηλή απόδοση των μοντέλων. Επιπροσθέτως, είναι απαραίτητο να εξεταστεί ο χρόνος καθυστέρησης των διαδικασιών συμπερασματολογίας των μοντέλων σε πραγματικές συσκευές, με σκοπό την μέτρηση καθυστέρησης ενός ολοκληρωμένου συστήματος ανίχνευσης που ενσωματώνει τις λειτουργίες του NFStream και τη συμπερασματολογία των μοντέλων, χρησιμοποιώντας ένα ελαφρύ framework όπως το TFLite.

3. **Διερεύνηση Διαφορετικών Συνόλων Δεδομένων Ασφάλειας Δικτύων**: Στην παρούσα εργασία χρησιμοποιήθηκε μοναδικό σύνολο δεδομένων συγκεκριμένης δι-

κτυακής τοπολογίας. Ωστόσο, η γενίκευση του μοντέλου σε διαφορετικές τοπολογίες δεν έχει εξεταστεί. Επομένως, η διερεύνηση πρόσθετων συνόλων δεδομένων ασφάλειας δικτύων, πέρα από αυτό που χρησιμοποιήθηκε στην παρούσα μελέτη θα μπορούσε να οδηγήσει στο να επιτευχθεί μία γενικευμένη και αξιόπιστη λύση για την ανίχνευση απειλών σε γενικευμένα δικτυακά περιβάλλοντα. Η δυσκολία αυτής της προσέγγισης, ωστόσο, έγκειται στο γεγονός πως τα ανοικτά σύνολα δικτυακών δεδομένων είναι περιορισμένα. Μία πιθανή λύση θα ήταν η παραγωγή νέων δικτυακών δεδομένων σε εργαστηριακό περιβάλλον για τις απαιτήσεις της έρευνας.

4. **Χρήση Τεχνικών Υπερδειγματοληψίας για Βελτίωση της Ανισορροπίας Δεδομένων**: Η ανισορροπία των δεδομένων παραμένει μία από τις μεγαλύτερες προκλήσεις στην ανάπτυξη αποδοτικών μοντέλων μηχανικής μάθησης. Συμπερασματικά, η διερεύνηση της εφαρμογής διαφόρων τεχνικών υπερδειγματοληψίας, όπως το SMOTE, για την αντιμετώπιση αυτής της πρόκλησης αποτελούν μία σημαντική μελλοντική κατεύθυνση. Συγκεκριμένα, θα πρέπει να μελετηθεί κατά πόσο κατάλληλες και αποδοτικές είναι τέτοιες τεχνικές σε δικτυακά σύνολα δεδομένων. Η έρευνα αυτή θα μπορούσε να εξετάσει τις διαφορετικές παραλλαγές του SMOTE και άλλων τεχνικών υπερδειγματοληψίας, προκειμένου να προσδιοριστεί ποια προσέγγιση είναι η πιο αποτελεσματική για τη βελτίωση της ανισορροπίας των κλάσεων και τη βελτίωση της απόδοσης των μοντέλων ανίχνευσης εισβολών.

5. **Διερεύνηση Τροποποίησης Δεδομένων Πακέτων σε Επίπεδο Bytes**: Η δυσκολία εισαγωγής δομημένων δεδομένων σε συνελικτικά δίκτυα και transformers καθιστά σημαντική την εξερεύνηση της τροποποίησης των συνόλων δεδομένων που περιέχουν αρχεία καταγραφής πακέτων. Μία πιθανή λύση στο πρόβλημα είναι η αναπαράσταση των δεδομένων σε σειρές από bytes των πακέτων ολόκληρων ροών, αντί σε μορφή πίνακα. Σε αυτή την περίπτωση, η χρήση Transformers, τα οποία είναι εξειδικευμένα μοντέλα για την επεξεργασία μεγάλων συμβολοσειρών, θα μπορούσε να αποδειχθεί εξαιρετικά αποδοτική, επιτρέποντας περισσότερο αποτελεσματική κατηγοριοποίηση.

Συνολικά, η παρούσα εργασία συνέβαλε στην κατανόηση και βελτίωση των τεχνικών ανίχνευσης εισβολών σε περιβάλλοντα IoT, αναδεικνύοντας τις προκλήσεις και τις δυνατότητες των σύγχρονων νευρωνικών δικτύων. Τα συμπεράσματα και οι προτάσεις για μελλοντική έρευνα θέτουν τα θεμέλια για περαιτέρω βελτιώσεις, με στόχο την ανάπτυξη πιο αποδοτικών και αποτελεσματικών συστημάτων ανίχνευσης. Με την προσέγγιση αυτή, ευελπιστούμε να συμβάλλουμε στην ενίσχυση της ασφάλειας των δικτύων και των συσκευών στο συνεχώς μεταβαλλόμενο ψηφιακό τοπίο.

Βιβλιογραφία

[1] Zhenzhu Meng, Yating Hu και Christophe Ancey. *Using a Data Driven Approach to Predict Waves Generated by Gravity Driven Mass Flows. Water*, 12, 2020.

[2] Jieyuan Wang, Ying Qian, Qingqing Ye και Biao Wang. *Image retrieval method based on metric learning for convolutional neural network. IOP Conference Series: Materials Science and Engineering*, 231:012002, 2017.

[3] Giuseppe Castellucci, Valentina Bellomaria, Andrea Favalli και Raniero Romagnoli. *Multi-lingual Intent Detection and Slot Filling in a Joint BERT-based Model*, 2019.

[4] *Receiver operation characteristic.* `https://en.wikipedia.org/wiki/Receiver_operating_characteristic`. Ημερομηνία πρόσβασης: 12-6-2024.

[5] Bengio Y. Hinton G. LeCun, Y. *Deep Learning. Nature*, 521:436–444, 2015.

[6] Yukhe Lavinia, Ramakrishnan Durairajan, Reza Rejaie και Walter Willinger. *Challenges in Using ML for Networking Research: How to Label If You Must. Proceedings of the Workshop on Network Meets AI & ML*, NetAI '20, σελίδα 21–27, New York, NY, USA, 2020. Association for Computing Machinery.

[7] Léo Grinsztajn, Edouard Oyallon και Gaël Varoquaux. *Why do tree-based models still outperform deep learning on tabular data?*, 2022.

[8] *What is machine learning (ML)?* `https://www.ibm.com/topics/machine-learning`. Ημερομηνία πρόσβασης: 6-6-2024.

[9] *Canadian Institute for Cybersecurity, Datasets.* `https://www.unb.ca/cic/datasets/`. Ημερομηνία πρόσβασης: 6-6-2024.

[10] *Random Forests, scikit learn.* `https://scikit-learn.org/stable/modules/ensemble.html#forest`. Ημερομηνία πρόσβασης: 7-6-2024.

[11] Simon S. Haykin. *Neural networks and learning machines.* Pearson Education, Upper Saddle River, NJ, 3η έκδοση, 2009.

[12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser και Illia Polosukhin. *Attention Is All You Need*, 2023.

[13] Vadim Borisov, Tobias Leemann, Kathrin Seßler, Johannes Haug, Martin Pawelczyk και Gjergji Kasneci. *Deep Neural Networks and Tabular Data: A Survey. CoRR*, αβς/2110.01889, 2021.

[14] *Clickthrough rate (CTR): Definition.* `https://support.google.com/google-ads/answer/2615875?hl=en`. Ημερομηνία πρόσβασης: 12-6-2024.

[15] Gilbert Badaro, Mohammed Saeed και Paolo Papotti. *Transformers for Tabular Data Representation: A Survey of Models and Applications. Transactions of the Association for Computational Linguistics*, 11:227–249, 2023.

[16] Yury Gorishniy, Ivan Rubachev και Artem Babenko. *On Embeddings for Numerical Features in Tabular Deep Learning*, 2023.

[17] *TensorFlow Lite.* `https://www.tensorflow.org/lite`. Ημερομηνία πρόσβασης: 19-6-2024.

[18] Suad Mohammed Othman και Fadl Mutaher Ba Alwi Ammar Thabit Zahary Nabeel T.Alsohybe. *Survey on Intrusion Detection System Types. International Journal of Cyber-Security and Digital Forensics*, 7:444–462, 2018.

[19] Asmaa Shaker Ashoor και Sharad Gore. *Importance of intrusion detection system (IDS). International Journal of Scientific and Engineering Research*, 2(1):1–4, 2011.

[20] Addison Shaver, Zhipeng Liu, Niraj Thapa, Kaushik Roy, Balakrishna Gokaraju και Xiaohon Yuan. *Anomaly Based Intrusion Detection for IoT with Machine Learning. 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, σελίδες 1–6, 2020.

[21] Dr. Gulshan Kumar Ahuja. *Evaluation Metrics for Intrusion Detection Systems-A Study. International Journal of Computer Science and Mobile Applications*, 11, 2015.

[22] Roberto Magán-Carrión, Daniel Urda, Ignacio Díaz-Cano και Bernabé Dorronsoro. *Towards a Reliable Comparison and Evaluation of Network Intrusion Detection Systems Based on Machine Learning Approaches. Applied Sciences*, 10(5), 2020.

[23] Merve Ozkan-Okay, Refik Samet, Ömer Aslan και Deepti Gupta. *A Comprehensive Systematic Literature Review on Intrusion Detection Systems. IEEE Access*, 9:157727–157760, 2021.

[24] *Confusion Matrix.* `https://en.wikipedia.org/wiki/Confusion_matrix`. Ημερομηνία πρόσβασης: 12-6-2024.

[25] *Detection error tradeoff (DET) curve.* `https://scikit-learn.org/stable/auto_examples/model_selection/plot_det.html#sphx-glr-auto-examples-model-selection-plot-det-py`. Ημερομηνία πρόσβασης: 12-6-2024.

[26] Yang Lu και Li Da Xu. *Internet of Things (IoT) Cybersecurity Research: A Review of Current Research Topics. IEEE Internet of Things Journal*, 6(2):2103–2115, 2019.

[27] Prathyusha M R και Biswajit Bhowmik. *IoT Evolution and Recent Advancements. 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, τόμος 1, σελίδες 1725–1730, 2023.

[28] Ayman Elnashar. *IoT evolution towards a super-connected world*, 2019.

[29] BitDefender. *The 2023 IoT Security Landscape Report.* `https://www.bitdefender.com/files/News/CaseStudies/study/429/2023-IoT-Security-Landscape-Report.pdf`, 2023. Ημερομηνία πρόσβασης: 16-6-2024.

[30] THALES. *IoT security issues in 2022: A business perspective.* `https://www.thalesgroup.com/en/markets/digital-identity-and-security/iot/magazine/internet-threats`, 2022. Ημερομηνία πρόσβασης: 16-6-2024.

[31] KUKUTLA TEJONATH REDDY. *OSI Layers and Their Impact on Network Security. networks*, 1:2.

[32] Euclides Carlos Pinto Neto, Sajjad Dadkhah, Raphael Ferreira, Alireza Zohourian, Rongxing Lu και Ali A. Ghorbani. *CICIoT2023: A Real-Time Dataset and Benchmark for Large-Scale Attacks in IoT Environment. Sensors*, 23(13), 2023.

[33] Zied Aouini και Adrian Pekar. *NFStream: A flexible network data analysis framework. Computer Networks*, 204:108719, 2022.

[34] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu και Xiaoqiang Zheng. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015. Software available from tensorflow.org.

[35] *Kaggle Community.* `https://www.kaggle.com/`. Ημερομηνία πρόσβασης: 16-6-2024.

[36] *Raspberry Pi.* `https://www.raspberrypi.com/`. Ημερομηνία πρόσβασης: 19-6-2024.

[37] Mike O. Ojo, Stefano Giordano, Gregorio Procissi και Ilias N. Seitanidis. *A Review of Low-End, Middle-End, and High-End Iot Devices. IEEE Access*, 6:70528–70554, 2018.

[38] *DPKT. Dpkt Documentation.* `https://https://dpkt.readthedocs.io/en/latest/`. Ημερομηνία πρόσβασης: 19-6-2024.

[39] *Berkeley packet filters.* `https://www.ibm.com/docs/en/qsip/7.4?topic=queries-berkeley-packet-filters`. Ημερομηνία πρόσβασης: 19-6-2024.

[40] Lu Lu, Christine Anderson-Cook και Timothy Robinson. *Optimization of Designed Experiments Based on Multiple Criteria Utilizing a Pareto Frontier. Technometrics*, 53:353–365, 2012.

Συντομογραφίες - Αρκτικόλεξα - Ακρωνύμια

βλ.	βλέπε
π.χ.	παραδείγματος χάρη
κτλ	και τα λοιπά
κ.ά.	και άλλα
ML	Machine Learning
DL	Deep Learning
QoS	Quality of Service
QoE	Quality of Experience
IoT	Internet of Things
FLOPS	Floating Point Operations Per Second
DNN	Deep Neural Network
MLP	Multilayer Perceptron
CNN	Convolutional Neural Network
IDS	Intrusion Detection System
AI	Artificial Intelligence
SMOTE	Synthetic Minority Over-sampling Technique
TP	True Positives
FP	False Positives
TN	True Negatives
FN	False Negatives
ANN	Artificial Neural Network
IDS	Intrusion Detection System
CTR	Clickthrough Rate
HIDS	Host Based Intrusion Detection System
NIDS	Network Based Intrusion Detection System
SIDS	Signature Based Intrusion Detection System
AIDS	Anomaly Based Intrusion Detection System
FAR	False Alarm Rate
DR	Detection Rate
AMR	Attack Miss Rate
ROC	Receiver Operating Characteristic
DET	Detection Error Tradeoff
WSN	Wireless Sensor Network
RFID	Radio Frequency Identification
IIoT	Industrial Internet of Things

LAN	Local Area Network
LPWAN	Low Power Wide Area Network
URLLC	Ultra-Reliable Low Latency Communications
OSI	Open Systems Interconnection
DoS	Denial of Service
DDoS	Distributed Denial of Service
ICMP	Internet Control Message Protocol
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
HTTP	HyperText Transfer Protocol
SYN	Synchronize
ARP	Address Resolution Protocol
DNS	Domain Name System
GRE	Generic Routing Encapsulation
GREIP	Generic Routing Encapsulation Internet Protocol
GREETH	Generic Routing Encapsulation Ethernet
XSS	Cross-Site Scripting
OS	Operating System
SQL	Structured Query Language
Recon	Reconnaisance
IP	Internet Protocol
GPU	Graphics Processing Unit
RAM	Random Access Memory
KB	Kilobyte
SD	Secure Digital
MAC	Media Access Control
DPKT	Data Packet
BPF	Berkeley Packet Filter
MDI	Mean Decrease in Impurity
PCAP	Packet Capture
CONV2D	Two-Dimensional Convolution

Απόδοση ξενόγλωσσων όρων

Απόδοση	Ξενόγλωσσος όρος
δεδομένα κίνησης	traffic data
Μηχανική Μάθηση	Machine Learning
Βαθιά Μάθηση	Deep Learning
Ποιότητα Υπηρεσίας	Quality of Service
ποιότητα εμπειρίας	Quality of Experience
συμπερασματολογία	inference
καλοήθης	benign
κακόβουλη	malicious
Διαδίκτυο των Πραγμάτων	Internet of Things
συσκευές του άκρου	edge devices
παράμετροι	weights
Πράξεις Κινητής Υποδιαστολής Ανά Δευτερόλεπτο	Floating Point Operations Per Second
Καθυστέρηση	Latency
Επιβάρυνση του Συστήματος	Workload
ευκολία εγκατάστασης	ease of deployment
Βαθύ Νευρωνικό Δίκτυο	Deep Neural Network
μέθοδοι συμπίεσης	compression techniques
Πολυεπίπεδο Perceptron	Multilayer Perceptron
Συνελικτικά Νευρωνικά Δίκτυα	Convolutional Neural Network
Μετασχηματιστές	Transformers
Σύστημα Ανίχνευσης Απειλών	Intrusion Detection System
Τυχαίο Δάσος	Random Forest
Τεχνητή Νοημοσύνη	Artificial Intelligence
Επιβλεπόμενη Μάθηση	Supervised Learning
επισημασμένα δεδομένα	labeled data
ταξινόμηση	classification
παλινδρόμηση	regression
Μη επιβλεπόμενη Μάθηση	Unsupervised Learning
μη επισημασμένα δεδομένα	unlabeled data
συσταδοποίηση	clustering
Ενισχυτική Μάθηση	Reinforcement Learning
Συνάρτηση Στόχου	Target Function
Συνάρτηση Απώλειας	Loss Function
σύνολο εκπαίδευσης	training set

σύνολο επικύρωσης	validation set
σύνολο ελέγχου	test set
δυαδική ταξινόμηση	binary classification
πολυταξική ταξινόμηση	multi-class classification
σύνολο δεδομένων	dataset
εργαλείο λογισμικού	framework
Καναδικό Ινστιτούτο Κυβερνοασφάλειας	Canadian Institute for Cybersecurity
προκατάληψη	bias
τεχνικές δειγματοληψίας	sampling techniques
υπερδειγματοληψία	over-sampling
επέκταση δεδομένων	data augmentation
υποδειγματοληψία	under-sampling
Πίνακας Σύγχυσης	Confusion Matrix
ορθότητα	accuracy
βαθμολογία F1	F1-score
ευστοχία	precision
ανάκληση	recall
Τεχνητά Νευρωνικά Δίκτυα	Artificial Neural Networks
Τεχνητός Νευρώνας	Perceptron
διανύσματα αναπαράστασης	embeddings
Φίλτρο Πακέτων Berkeley	Berkeley Packet Filter
Σύλληψη Πακέτων	Packet Capture
Μονάδα Επεξεργασίας Γραφικών	Graphics Processing Unit
Μνήμη Τυχαίας Προσπέλασης	Random Access Memory
Λειτουργικό Σύστημα	Operating System
Βιομηχανικό Διαδίκτυο των Πραγμάτων	Industrial Internet of Things
Τοπικό Δίκτυο	Local Area Network
Δίκτυο Χαμηλής Ισχύος Μεγάλης Εμβέλειας	Low Power Wide Area Network
Αξιόπιστες Επικοινωνίες Χαμηλής Καθυστέρησης	Ultra-Reliable Low Latency Communications
Ποσοστό Κλικ	Clickthrough Rate
Σύστημα Ανίχνευσης Εισβολών	Intrusion Detection System
Σύστημα Ανίχνευσης Εισβολών Επιπέδου Συσκευής	Host Based Intrusion Detection System
Σύστημα Ανίχνευσης Εισβολών Επιπέδου Δικτύου	Network Based Intrusion Detection System
Σύστημα Ανίχνευσης Εισβολών Βάσει Υπογραφών	Signature Based Intrusion Detection System
Σύστημα Ανίχνευσης Εισβολών Βάσει Ανωμαλιών	Anomaly Based Intrusion Detection System
Ποσοστό Ψευδούς Συναγερμού	False Alarm Rate
Ποσοστό Ανίχνευσης	Detection Rate
Ποσοστό Αποτυχίας Ανίχνευσης Επίθεσης	Attack Miss Rate
Χαρακτηριστική Καμπύλη Λειτουργίας Δέκτη	Receiver Operating Characteristic
Αντιστάθμιση Σφάλματος Ανίχνευσης	Detection Error Tradeoff