



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ**

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

**“ΘΕΩΡΗΤΙΚΗ ΜΕΛΕΤΗ ΤΕΧΝΙΚΩΝ ΑΝΑΛΥΣΗΣ
ΧΡΟΝΟΣΕΙΡΩΝ ΓΙΑ ΤΟΝ ΕΝΤΟΠΙΣΜΟ ΑΝΩΜΑΛΙΩΝ
ΣΕ ΕΦΑΡΜΟΓΕΣ ΣΤΟ ΔΙΑΔΙΚΤΥΟ ΤΩΝ ΠΡΑΓΜΑΤΩΝ”**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της/του

Μανωλτσίδου Άννας

Επιβλέπων : Μεσαριτάκης Χάρης

Μέλη εξεταστικής επιτροπής:

Σάμος, Σεπτέμβριος 2023

Η σελίδα αυτή είναι σκόπιμα λευκή.

Πρόλογος και ευχαριστίες

Θα επιθυμούσα να αποδώσω τις ευχαριστίες μου στα πρόσωπα που με βοήθησαν και με υποστήριξαν σε όλο το πέρασ της διάρκειας της συγγραφής αυτής της διπλωματικής εργασίας.

Αρχικά θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα Καθηγητή του Τμήματος Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων, κύριο Χάρη Μεσαριτάκη για την εμπιστοσύνη που μου έδειξε αναθέτοντάς μου αυτή την εργασία, την υπομονή και την πολύτιμη καθοδήγησή καθόλη την διάρκεια έως την εκπόνηση αυτής.

Το μεγαλύτερο όμως ευχαριστώ, το χρωστάω κυρίως στους γονείς μου Ελένη και Χρήστο, που μου έδωσαν τη δυνατότητα να σπουδάσω και να εξελιχθώ, που πάντα με στηρίζουν στις επιλογές που παίρνω και που είναι δίπλα μου σε κάθε μου βήμα.

Τέλος, θα ήθελα να ευχαριστήσω και τους φίλους μου που ήταν εκεί όταν χρειάστηκε, με υποστήριξαν και μου υπενθύμιζαν να μην παραιτούμαι από τους στόχους μου.

© 2023

του/της

Μανωλτσίδου Άννας

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Η σελίδα αυτή είναι σκόπιμα λευκή.

Πίνακας περιεχομένων

1 Εισαγωγή	6
1.1 Διαδίκτυο των Πραγμάτων (IoT)	6
1.2 Ανίχνευση ανωμαλιών (Anomaly Detection) στο IoT	8
1.2.1 Ανίχνευση Ανωμαλιών στο IoT	8
1.2.2 Σημασία της Ανίχνευσης Ανωμαλιών στο IoT	9
1.2.3 Τεχνικές Ανίχνευσης Ανωμαλιών στο IoT	10
1.3 Ταξινόμηση και Κατηγοριοποίηση Ανωμαλιών	10
1.3.1 Τύποι Ανωμαλιών	11
1.3.2 Περιοχές Εμφάνισης Ανωμαλιών	11
1.4 Δομή της διπλωματικής	12
2 Αντιμετώπιση Ανωμαλιών	13
2.1 Τρόποι Μελέτης Ανωμαλιών	13
2.1.1 Επιβλεπόμενη Μάθηση	13
2.1.2 Μη Επιβλεπόμενη Μάθηση	15
2.1.3 Ημι-Επιβλεπόμενη Μάθηση	16
2.2 Προσεγγίσεις και Τεχνικές	18
2.2.1 Ανίχνευση Ατυχημάτων	18
2.2.2 Ανίχνευση Συστημάτων	19
2.2.3 Ανίχνευση Ανωμαλιών με βάση τη Γνώση	19
2.2.4 Ανίχνευση Ασφαλείας	20
2.2.5 Στατιστικές Μέθοδοι	20
2.3 Τεχνικές Ανάλυσης Ανωμαλιών	21
2.3.1 Στατιστική Ανάλυση	21
2.3.2 Μηχανική Μάθηση	22
2.3.3 Ανάλυση Φάσματος	23
2.3.4 Ανίχνευση Συμπεριφορικών Ανωμαλιών	24
2.3.5 Ανάλυση Αποτύπωσης	24
2.3.6 Ανίχνευση Παραμέτρων	25
2.3.7 Ανάλυση Συμπεριφοράς	25
2.3.8 Ανάλυση Δικτύου	26
2.4 Εισαγωγή στα Νευρωνικά Δίκτυα	27
2.4.1 Δομή των Νευρωνικών Δικτύων	27
2.4.2 Τύποι Νευρωνικών Δικτύων	27
2.4.3 Εφαρμογές των Νευρωνικών Δικτύων	31
3 Deep Neural Networks (DNNs)	34
3.1 Γενική εικόνα	34
3.2 Βασικές Αρχές των Deep Neural Networks	35
3.2.1 Βασικά Στοιχεία των DNNs	35
3.3 Εκπαίδευση των DNNs	44
3.3.1 Διαδικασία Εκπαίδευσης	44

3.3.2 Προκλήσεις στην Εκπαίδευση των DNNs	49
3.4 Εφαρμογές των DNNs	50
3.4.1 Εικόνα και Βίντεο	50
3.4.2 Φυσική Γλώσσα	54
3.4.3 Υγεία	54
3.4.4 Αυτόνομα Οχήματα	60
3.4.5 Κοινωνικά Δίκτυα και Διαδίκτυο	64
4 Recurrent Neural Networks (RNNs)	67
4.1 Εισαγωγικά	67
4.2 Θεμελιώδης Αρχή των RNNs	68
4.2.1 Αναδρομική Λειτουργία	68
4.2.2 Χρήση στη Μοντελοποίηση Χρονοσειρών	70
4.2.3 Προκλήσεις και Εξελίξεις	75
4.3 Μοντέλα-Αρχιτεκτονικές RNN	76
4.3.1 Γραμμικές Αναδρομικές Μονάδες (LRUs)	76
4.3.2 Μονάδες με Μοναδική Μνήμη (UMUs)	80
4.3.3 Αναδρομικές Μονάδες Μακράς Προθεσμίας (LSTM)	81
4.4 Εφαρμογές των RNNs	85
4.4.1 Επεξεργασία Φυσικής Γλώσσας	85
4.4.2 Πρόβλεψη Χρονοσειρών	93
4.4.3 Επεξεργασία Ήχου και Μουσικής	96
5 Reservoir Computing	102
5.1 Εισαγωγή	102
5.2 Αρχές του Υπολογισμού Δεξαμενής	102
5.2.1 Αρχές Λειτουργίας	103
5.2.2 Επιλεγμένες Παραλλαγές του Υπολογισμού Δεξαμενής	104
5.2.3 Εφαρμογές στην Επιστήμη και τη Τεχνολογία	110
5.2.4 Προκλήσεις και Μελλοντικές Εξελίξεις	111
5.3 Η Έννοια του Υπολογισμού Δεξαμενής	111
5.3.1 Βασικές Έννοιες του Υπολογισμού Δεξαμενής	112
5.4 Εκπαίδευση και Λειτουργία της Δεξαμενής	114
5.4.1 Η Δεξαμενή ως Μοντέλο Υπολογιστικής Μάθησης	115
5.4.2 Εκπαίδευση της Δεξαμενής	115
5.5 Πλεονεκτήματα και Εφαρμογές	117
5.5.1 Πλεονεκτήματα της Τεχνικής	118
5.5.2 Εφαρμογές της Δεξαμενής	120
6 Παραδείγματα Ανάλυσης Ανωμαλιών με Reservoir Computing	141
6.1 Reservoir Computing Meets Smart Grids: Attack Detection Using Delayed Feedback Networks	141
6.2 Detecting Dynamic Attacks in Smart Grids Using Reservoir Computing: A Spiking Delayed Feedback Reservoir Based Approach	153
7 Σύνοψη	171
7.1 Επιγραμματική περίληψη	171
7.2 Κριτική Αξιολόγηση των ειδών Νευρωνικών Δικτύων	172
7.3 Σύγκριση των ειδών	173

7.3.1 Τύπος Νευρωνικού Δικτύου: DNNs	173
7.3.2 Τύπος Νευρωνικού Δικτύου: RNNs	174
7.3.3 Τύπος Νευρωνικού Δικτύου: Reservoir Computing	174
7.4 Πλεονεκτήματα και Μειονεκτήματα	175
7.4.1 Τύπος Νευρωνικού Δικτύου: Deep Neural Networks (DNNs)	175
7.4.2 Τύπος Νευρωνικού Δικτύου: Recurrent Neural Networks (RNNs)	176
7.4.3 Τύπος Νευρωνικού Δικτύου: Reservoir Computing	177
Βιβλιογραφία	179
References	184

Περίληψη

Με τη ραγδαία επέκταση του Διαδικτύου των Πραγμάτων (IoT) και την αυξανόμενη ενσωμάτωσή του σε διάφορους τομείς, η διασφάλιση της ασφάλειας και της αξιοπιστίας των συστημάτων IoT καθίσταται υψίστης σημασίας. Η ανίχνευση ανωμαλιών είναι μια κρίσιμη τεχνική για τον εντοπισμό και τον μετριασμό πιθανών απειλών και ανωμαλιών σε περιβάλλοντα IoT. Το παρόν έγγραφο παρέχει μια εκτενή διερεύνηση της ανίχνευσης ανωμαλιών στο πλαίσιο του IoT, εστιάζοντας σε διάφορες μεθοδολογίες και τεχνικές για την αποτελεσματική ανίχνευση.

Το έγγραφο ξεκινά με μια εισαγωγή στο IoT και τη σημασία του, ακολουθούμενη από μια εμπειρισταωμένη συζήτηση σχετικά με τη σημασία της ανίχνευσης ανωμαλιών σε περιβάλλοντα IoT. Παρουσιάζονται διάφορες τεχνικές ανίχνευσης ανωμαλιών ειδικά για το IoT, επισημαίνοντας τα πλεονεκτήματα και τους περιορισμούς τους. Επιπλέον, το έγγραφο εμβαθύνει στην ταξινόμηση και την κατηγοριοποίηση των ανωμαλιών, εντοπίζοντας τις περιοχές όπου εμφανίζονται συνήθως ανωμαλίες στα δίκτυα IoT.

Στην επιδίωξη της αντιμετώπισης των ανωμαλιών, το έγγραφο διερευνά διάφορες προσεγγίσεις και μεθοδολογίες για τη μελέτη των ανωμαλιών, συμπεριλαμβανομένων τεχνικών μάθησης με επίβλεψη, χωρίς επίβλεψη και με ημι-πίεση. Επιπλέον, διερευνά διάφορες τεχνικές ανίχνευσης, όπως η μηχανική μάθηση, η ανίχνευση ατυχημάτων, η παρακολούθηση συστημάτων, η ανίχνευση ανωμαλιών βάσει γνώσης, οι προσεγγίσεις βάσει ασφάλειας και οι στατιστικές μέθοδοι.

Στη συνέχεια, το έγγραφο εστιάζει στην ανάλυση των τεχνικών ανίχνευσης ανωμαλιών, συμπεριλαμβανομένης της στατιστικής ανάλυσης, των μεθόδων μηχανικής μάθησης, της φασματικής ανάλυσης, της ανίχνευσης ανωμαλιών συμπεριφοράς, της αναγνώρισης προτύπων, της ανάλυσης παραμέτρων, της ανάλυσης συμπεριφοράς και της ανάλυσης δικτύου. Επιπλέον, παρέχεται μια εισαγωγή στα νευρωνικά δίκτυα, εξηγώντας τις δομές, τους τύπους και τις εφαρμογές τους στην ανίχνευση ανωμαλιών.

Περαιτέρω ενότητες εμβαθύνουν στις βασικές αρχές των βαθιών νευρωνικών δικτύων (DNN), στις βασικές αρχές τους και στη διαδικασία εκπαίδευσης. Διερευνάται επίσης η εφαρμογή των DNN στην ανίχνευση ανωμαλιών, υπογραμμίζοντας τα δυνητικά οφέλη τους στην ανίχνευση σύνθετων και λεπτών ανωμαλιών.

Μετά τη συζήτηση σχετικά με τα DNN, παρουσιάζονται τα αναδρομικά νευρωνικά δίκτυα (RNN) και η υπολογιστική δεξαμενή (Reservoir Computing), παρουσιάζοντας τη σημασία και τις δυνατότητές τους στην ανίχνευση ανωμαλιών IoT.

Εν κατακλείδι, το παρόν έγγραφο προσφέρει μια ολοκληρωμένη επισκόπηση της ανίχνευσης ανωμαλιών σε περιβάλλοντα IoT, παρουσιάζοντας ένα ευρύ φάσμα τεχνικών και μεθοδολογιών που οι ερευνητές και οι επαγγελματίες μπορούν να χρησιμοποιήσουν για να ενισχύσουν την ασφάλεια και την αξιοπιστία των συστημάτων IoT. Οι γνώσεις που παρέχονται σε αυτό το έγγραφο χρησιμεύουν ως πολύτιμος πόρος για την κατανόηση και την εφαρμογή αποτελεσματικών στρατηγικών ανίχνευσης ανωμαλιών σε εφαρμογές IoT.

Λέξεις Κλειδιά: Τεχνικές ανάλυσης χρονοσειρών, Ανίχνευση ανωμαλιών, Εφαρμογές στο διαδίκτυο των πραγμάτων, Αλγόριθμοι μηχανικής μάθησης, Αισθητήρες, Συστήματα επιτήρησης, Ασφάλεια

Abstract

"Theoretical Study of Time Series Analysis Techniques for Anomaly Detection in Internet of Things (IoT) Applications"

The theoretical study focuses on the development of methods and techniques for detecting and analyzing uncommon observations or anomalies in time series data derived from Internet of Things (IoT) applications.

With the rapid expansion of the Internet of Things (IoT) and its increasing integration into various domains, ensuring the security and reliability of IoT systems becomes paramount. Anomaly detection is a crucial technique for identifying and mitigating potential threats and abnormalities in IoT environments. This paper provides an extensive exploration of anomaly detection in the IoT context, focusing on various methodologies and techniques for effective detection.

The paper begins with an introduction to IoT and its significance, followed by an in-depth discussion on the importance of anomaly detection in IoT environments. Various anomaly detection techniques specific to IoT are presented, highlighting their strengths and limitations. Moreover, the paper delves into the classification and categorization of anomalies, pinpointing the areas where anomalies commonly occur in IoT networks.

In the pursuit of addressing anomalies, the paper explores different approaches and methodologies for studying anomalies, including supervised, unsupervised, and semi-supervised learning techniques. Furthermore, it investigates various detection techniques, such as machine learning, accident detection, system monitoring, knowledge-based anomaly detection, security-based approaches, and statistical methods.

The paper then shifts its focus to the analysis of anomaly detection techniques, including statistical analysis, machine learning methods, spectral analysis, behavioral anomaly detection, pattern recognition, parameter analysis, behavioral analysis, and network analysis. Additionally, an introduction to neural networks is provided, explaining their structures, types, and applications in anomaly detection.

Further sections delve into the fundamentals of Deep Neural Networks (DNNs), their basic principles, and the training process. The application of DNNs in anomaly detection is also explored, highlighting their potential benefits in detecting complex and subtle anomalies.

Following the discussion on DNNs, Recurrent Neural Networks (RNNs) and Reservoir Computing are introduced, showcasing their relevance and capabilities in IoT anomaly detection.

In conclusion, this paper offers a comprehensive overview of anomaly detection in IoT environments, presenting a diverse array of techniques and methodologies that researchers and practitioners can employ to enhance the security and reliability of IoT systems. The insights provided in this paper serve as a valuable resource for understanding and implementing effective anomaly detection strategies in IoT deployments.

Keywords: *Time series analysis techniques, Anomaly detection, Internet of Things (IoT) applications, Machine learning algorithms, Sensors, Monitoring systems, Security*

1

Εισαγωγή

1.1 Διαδίκτυο των Πραγμάτων (IoT)

Το Διαδίκτυο των πραγμάτων, ή εν συντομία IoT, είναι ένας ταχέως αναπτυσσόμενος τομέας που αναφέρεται στη διασύνδεση συσκευών και μηχανημάτων μέσω του διαδικτύου. Με το IoT, τα καθημερινά αντικείμενα μπορούν να επικοινωνούν και να αλληλεπιδρούν μεταξύ τους, επιτρέποντας ένα επίπεδο αυτοματοποίησης, συλλογής δεδομένων και ανάλυσης που ήταν προηγουμένως αδιανόητο.[1] Το IoT περιλαμβάνει ένα ευρύ φάσμα συσκευών, όπως αισθητήρες, έξυπνες συσκευές και φορητές συσκευές, οι οποίες συλλέγουν και μοιράζονται δεδομένα μεταξύ τους και με άλλα συστήματα.[2] Τα δεδομένα που παράγονται από τις συσκευές IoT είναι τεράστια και ποικίλα και απαιτούν καινοτόμες μεθόδους συλλογής, επεξεργασίας και ανάλυσης.[3] Παρακάτω παρέχεται μια γενική επισκόπηση του IoT, τονίζοντας τις εφαρμογές, τα οφέλη και τις προκλήσεις που προκύπτουν από αυτή την τεχνολογία.

Στον πυρήνα του, το IoT περιστρέφεται γύρω από τη σύνδεση συσκευών μέσω ενός δικτύου.[4] Αυτό το δίκτυο μπορεί να είναι οτιδήποτε, από ένα τοπικό δίκτυο (LAN) έως το ίδιο το διαδίκτυο. Οι συσκευές IoT κυμαίνονται από τις απλές, όπως οι αισθητήρες θερμοκρασίας και λαμπτήρες, έως τις πιο πολύπλοκες, όπως τα αυτόνομα αυτοκίνητα και τα μη επανδρωμένα αεροσκάφη.[5] Ωστόσο, ανεξάρτητα από την πολυπλοκότητά τους, ο απώτερος στόχος αυτών των συνδεδεμένων συσκευών είναι να φέρουν αποτελεσματικότητα στον κόσμο με την αυτοματοποίηση των διαδικασιών, τη μείωση της σπατάλης και τη βελτίωση της ποιότητας ζωής.

Ένα από τα σημαντικότερα πλεονεκτήματα του IoT είναι ο όγκος των δεδομένων που μπορούν να συλλεχθούν από τις συνδεδεμένες συσκευές. Τα δεδομένα αυτά μπορούν να χρησιμοποιηθούν για διάφορους σκοπούς, όπως η προγνωστική συντήρηση και η βελτιστοποίηση διαδικασιών στη μεταποίηση ή παρακολούθηση της υγιεινομικής περίθαλψης μέσω φορητών συσκευών και η ανάλυση της ροής της κυκλοφορίας σε έξυπνες πόλεις, μεταξύ πολλών άλλων.

Δεδομένων των τεράστιων δυνατοτήτων του IoT, ένα ευρύ φάσμα βιομηχανιών το εκμεταλλεύεται. Για παράδειγμα, ο τομέας της γεωργίας μπορεί να χρησιμοποιήσει αισθητήρες στα χωράφια για τη βελτιστοποίηση της ανάπτυξης των καλλιεργειών,[6] ενώ ο κλάδος της υγιεινομικής περίθαλψης μπορεί να χρησιμοποιήσει συσκευές IoT για την παρακολούθηση ασθενειών και τον εξ' αποστάσεως εντοπισμό τυχόν πιθανών κινδύνων για την υγεία σε πραγματικό χρόνο.[7] Ο κλάδος των μεταφορών μπορεί να χρησιμοποιήσει συσκευές IoT για την παρακολούθηση της κυκλοφορίας και τη βελτιστοποίηση των διαδρομών, μειώνοντας τον χρόνο ταξιδιού, ενώ ο τομέας της ενέργειας μπορεί να χρησιμοποιήσει δεδομένα σε πραγματικό χρόνο για την πρόβλεψη και την προληπτική αντιμετώπιση τυχόν προβλημάτων που προκύπτουν.[8]

Παρ' όλα αυτά, το IoT δεν αποτελεί πανάκεια για όλα τα προβλήματα. Υπάρχουν ορισμένες εγγενείς προκλήσεις που συνοδεύουν την εφαρμογή αυτής της τεχνολογίας. Η πρωταρχική ανησυχία αφορά την ασφάλεια και την προστασία της ιδιωτικής ζωής, καθώς εκατομμύρια συσκευές συνδεδεμένες στο διαδίκτυο αποτελούν ελκυστικό στόχο για τους εγκληματίες του κυβερνοχώρου.[\[9\]](#) Οι συσκευές αντιμετωπίζουν επίσης προκλήσεις όπως ο κατακερματισμός, όπου διαφορετικές συσκευές από διαφορετικούς κατασκευαστές και κλάδους ενδέχεται να μην διασυνδέονται ή να μην μοιράζονται δεδομένα, ενώ τα παλαιά συστήματα ενδέχεται να μην λειτουργούν με τις νέες τεχνολογίες. Επίσης, υπάρχει κόστος υλοποίησης που μπορεί να είναι σημαντικό ανάλογα με το εύρος της ανάπτυξης του IoT και τις εμπλεκόμενες συσκευές.

Οι συσκευές IoT είναι εξοπλισμένες με αισθητήρες που μπορούν να μετρήσουν διάφορες περιβαλλοντικές και λειτουργικές παραμέτρους, όπως η θερμοκρασία, η υγρασία, η πίεση, η δόνηση και η κίνηση. Αυτοί οι αισθητήρες παράγουν δεδομένα χρονοσειράς, τα οποία είναι μια ακολουθία σημείων δεδομένων που μετρώνται σε τακτά χρονικά διαστήματα με την πάροδο του χρόνου. Τα δεδομένα χρονοσειρών μπορούν να αναλυθούν με τη χρήση διαφόρων τεχνικών, όπως η ανάλυση χρονοσειρών (Time Series Analysis), η μηχανική μάθηση (Machine Learning) και η εξόρυξη δεδομένων (Data Mining).[\[10\]](#)

Η ανάλυση χρονοσειρών είναι μια στατιστική τεχνική που αποσκοπεί στη μοντελοποίηση των υποκείμενων τάσεων, προτύπων και εποχικότητας στα δεδομένα χρονολογικών σειρών. Περιλαμβάνει μεθόδους όπως ο αυτοπαλίνδρομος ολοκληρωμένος κινητός μέσος όρος (Autoregressive Integrated Moving Average - ARIMA), η εκθετική εξομάλυνση (Exponential Smoothing - ETS), η εποχική αποσύνθεση χρονοσειρών (Seasonal-Trend decomposition using Loess - STL) και η ανάλυση Fourier. Αυτές οι τεχνικές μπορούν να χρησιμοποιηθούν για την πρόβλεψη και την ανίχνευση ανωμαλιών σε δεδομένα IoT.[\[9\]\[10\]](#)

Η μηχανική μάθηση είναι μια άλλη τεχνική που έχει κερδίσει δημοτικότητα στις εφαρμογές IoT. Τα μοντέλα μηχανικής μάθησης μπορούν να μαθαίνουν από τα δεδομένα και να εντοπίζουν σύνθετα μοτίβα που μπορεί να μην ανιχνεύονται με τη χρήση παραδοσιακών στατιστικών μεθόδων. Οι αλγόριθμοι μηχανικής μάθησης περιλαμβάνουν τα νευρωνικά δίκτυα (Neural Networks), τα δέντρα αποφάσεων (Decision Trees), τις μηχανές διανυσμάτων υποστήριξης (Support Vector Machines) και τους αλγορίθμους ομαδοποίησης (Clustering Algorithms). Αυτοί οι αλγόριθμοι μπορούν να χρησιμοποιηθούν για προγνωστική μοντελοποίηση, ανίχνευση ανωμαλιών και ταξινόμηση σε εφαρμογές IoT.[\[9\]](#)

Η εξόρυξη δεδομένων είναι μια τεχνική που αποσκοπεί στην εξαγωγή χρήσιμων πληροφοριών και συμπερασμάτων από μεγάλα σύνολα δεδομένων. Περιλαμβάνει τεχνικές όπως η εξόρυξη κανόνων συσχέτισης (Association rule mining), η ομαδοποίηση (Clustering) και η ταξινόμηση (Classification). Η εξόρυξη δεδομένων μπορεί να χρησιμοποιηθεί για διάφορες εφαρμογές IoT, όπως η προγνωστική συντήρηση, η ανίχνευση απάτης και η ανίχνευση ανωμαλιών.[\[10\]](#)

Εκτός από αυτές τις τεχνικές, οι εφαρμογές IoT απαιτούν επίσης μια ισχυρή υποδομή επικοινωνίας για τη μετάδοση και την αποθήκευση δεδομένων. Η υποδομή IoT περιλαμβάνει διάφορες τεχνολογίες όπως είναι τα ασύρματα δίκτυα αισθητήρων (Wireless Sensor Networks), το υπολογιστικό νέφος (Cloud Computing) και η υπολογιστική ακμής (Edge Computing). Αυτές οι τεχνολογίες επιτρέπουν την επεξεργασία, αποθήκευση και ανάλυση δεδομένων σε πραγματικό χρόνο και είναι απαραίτητες για την επιτυχία των εφαρμογών IoT.[\[1\]\[6\]](#)

Η προβολή ανόδου του Internet of Things (IoT) είναι εντυπωσιακή, καθώς η τεχνολογία συνεχίζει να εξελίσσεται και νέες εφαρμογές αναπτύσσονται σε διάφορους τομείς. Οι προβλέψεις δείχνουν ότι ο αριθμός των συνδεδεμένων συσκευών IoT θα αυξηθεί δραματικά τα επόμενα χρόνια.

Ο λόγος πίσω από την άνοδο του IoT είναι η συνεχής εξέλιξη της τεχνολογίας, η αύξηση της χωρητικότητας των δικτύων και η πτώση του κόστους των αισθητήρων και των συσκευών. Αυτό έχει οδηγήσει στην αποδοχή και την υιοθέτηση της τεχνολογίας IoT σε πολλούς τομείς, όπως έξυπνα σπίτια, έξυπνες πόλεις, βιομηχανία, υγεία, γεωργία, αυτοκινητοβιομηχανία και πολλούς άλλους. Οι εφαρμογές του IoT είναι ευρέως ποικίλες και περιλαμβάνουν την παρακολούθηση και τον έλεγχο των συσκευών απομακρυσμένα, την ανίχνευση και την πρόβλεψη ανωμαλιών, την αυτόματη διαχείριση της ενέργειας, τη βελτιστοποίηση των διαδικασιών παραγωγής και πολλές άλλες εφαρμογές που συμβάλλουν στην ευκολία, την αποδοτικότητα και τη βελτίωση της ποιότητας ζωής.

Με την άνοδο του IoT, αναμένεται επίσης μια αύξηση στον όγκο των δεδομένων που παράγονται από τις συσκευές, προκαλώντας μια πρόκληση για την αποθήκευση, την επεξεργασία και την ανάλυση αυτών των δεδομένων. Η ανάπτυξη τεχνολογιών που διευκολύνουν τη διαχείριση των μεγάλων δεδομένων (big data) και την ανάλυσή τους αναμένεται να παίζει σημαντικό ρόλο στην αξιοποίηση του δυναμικού του IoT. Παράλληλα, πρέπει να δοθεί έμφαση στην ασφάλεια και την προστασία των δεδομένων που συλλέγονται και μεταδίδονται από τις συνδεδεμένες συσκευές. Η αυξανόμενη συνδεσιμότητα μπορεί να δημιουργήσει νέους κινδύνους, και είναι σημαντικό να ληφθούν μέτρα για να διασφαλιστεί η ιδιωτικότητα και η ασφάλεια των δεδομένων που συνδέονται με το IoT.

Συνολικά, το IoT παρέχει ένα ευρύ φάσμα ευκαιριών και προκλήσεων για τους ερευνητές και τους επαγγελματίες. Το θεωρητικό υπόβαθρο του περιλαμβάνει διάφορες τεχνικές, όπως η ανάλυση χρονοσειρών, η μηχανική μάθηση και η εξόρυξη δεδομένων, καθώς και την υποδομή επικοινωνίας που απαιτείται για τις εφαρμογές του IoT. Η διεξοδική κατανόηση αυτών των τεχνικών και τεχνολογιών είναι απαραίτητη για την επιτυχή υλοποίηση εφαρμογών IoT σε διάφορους τομείς, όπως η υγειονομική περίθαλψη, οι μεταφορές και η μεταποίηση.

Συμπερασματικά, το IoT αντιπροσωπεύει μια σημαντική τεχνολογική πρόοδο που έχει τεράστιες δυνατότητες για τη βελτίωση της ποιότητας ζωής και την αύξηση της αποδοτικότητας. Ενώ τα οφέλη του IoT είναι τεράστια, πρέπει να αντιμετωπιστούν προκλήσεις όπως το απόρρητο των δεδομένων και η ασφάλεια για να εξασφαλιστεί η επιτυχής εφαρμογή της τεχνολογίας IoT. Καθώς το IoT συνεχίζει να αναπτύσσεται και να επεκτείνεται, είναι πιθανό να μετασχηματίσει τις βιομηχανίες και να συνδέσει περισσότερες συσκευές για να διασφαλίσει τη βέλτιστη απόδοση και την ποιοτική ανάλυση δεδομένων.

1.2 Ανίχνευση ανωμαλιών (Anomaly Detection) στο IoT

Το Internet of Things (IoT) έχει εκτοξευθεί ως μια αναπτυσσόμενη τεχνολογία που επηρεάζει θετικά πολλούς τομείς, από την έξυπνη πόλη και τη βιομηχανία έως την υγεία και την ασφάλεια. Ωστόσο, με την αύξηση του αριθμού των συνδεδεμένων συσκευών IoT, προκύπτει η ανάγκη για ανίχνευση ανωμαλιών.

Η ανίχνευση ανωμαλιών στο IoT αποτελεί κρίσιμη πτυχή για την ασφάλεια, την προστασία και την ομαλή λειτουργία των συστημάτων. Σε αυτήν την ενότητα, θα εξετάσουμε τη σημασία της ανίχνευσης ανωμαλιών στο IoT και τον ρόλο που διαδραματίζει στην εξασφάλιση της αξιοπιστίας και της αποτελεσματικότητας του συστήματος.

1.2.1 Ανίχνευση Ανωμαλιών στο IoT

Η ανίχνευση ανωμαλιών στο Internet of Things (IoT) αποτελεί κρίσιμη πτυχή για την ασφάλεια και την αποτελεσματική λειτουργία των συστημάτων. Καθώς οι συσκευές IoT

συνδέονται και αλληλεπιδρούν μεταξύ τους, παράγοντας μεγάλους όγκους δεδομένων, είναι απαραίτητο να αναγνωρίζονται ανωμαλίες που μπορεί να υπονομεύουν την ασφάλεια ή να επηρεάζουν την επίδοση του συστήματος.

Οι ανωμαλίες μπορεί να προκύψουν από πολλούς παράγοντες, όπως κακόβουλες επιθέσεις, σφάλματα λογισμικού ή υλικού, εσφαλμένες ρυθμίσεις συστήματος ή μη συνήθεις συμπεριφορές συσκευών.[\[11\]](#) Η ανίχνευση ανωμαλιών αποσκοπεί στον εντοπισμό τέτοιων ανωμαλιών και στην προστασία του συστήματος από πιθανά προβλήματα.

Για την ανίχνευση ανωμαλιών στο IoT, χρησιμοποιούνται διάφορες τεχνικές και μεθοδολογίες. Αυτές περιλαμβάνουν την ανάλυση των δεδομένων για τον εντοπισμό μη τυπικών προτύπων και ανωμαλιών, τη χρήση μοντέλων μηχανικής μάθησης για την αναγνώριση μη συνήθων συμπεριφορών και την εφαρμογή αλγορίθμων ανίχνευσης ανωμαλιών που βασίζονται σε κανόνες εμπειρογνομώνων.[\[12\]](#)

Η ανίχνευση ανωμαλιών στο IoT είναι κρίσιμη για την πρόληψη απωλειών δεδομένων, τη διασφάλιση της ιδιωτικότητας των χρηστών και την αποτροπή επιθέσεων. Επιπλέον, βοηθά στη βελτίωση της αποτελεσματικότητας και της αξιοπιστίας των συστημάτων IoT, επιτρέποντας την αντίδραση σε προβλήματα και ανωμαλίες πριν αυτά επηρεάσουν την κανονική λειτουργία του συστήματος.

Με την αναπτυξιακή τάση του IoT και την αυξημένη συνδεσιμότητα των συσκευών, η ανίχνευση ανωμαλιών αποκτά ακόμη μεγαλύτερη σημασία, καθώς οι πιθανότητες για ανεπιθύμητα γεγονότα και επιθέσεις αυξάνονται.[\[13\]](#) Οι προηγμένες τεχνικές ανίχνευσης ανωμαλιών θα συνεχίσουν να αναπτύσσονται και να εξελίσσονται προκειμένου να παρέχουν ακόμη πιο αποτελεσματική προστασία και ασφάλεια στα συστήματα IoT.

1.2.2 Σημασία της Ανίχνευσης Ανωμαλιών στο IoT

Η ανίχνευση ανωμαλιών στο IoT έχει κρίσιμη σημασία για τη διασφάλιση της αξιοπιστίας και της αποτελεσματικότητας του συστήματος. Ένα από τα βασικά χαρακτηριστικά του IoT είναι η συνεχής συλλογή, ανταλλαγή και ανάλυση δεδομένων από συνδεδεμένες συσκευές. Ωστόσο, αυτή η πληθώρα δεδομένων και η συνεχής ροή πληροφοριών μπορεί να καθιστά δύσκολο τον εντοπισμό ανωμαλιών ή παρατηρήσεων που υποδηλώνουν πιθανά προβλήματα.

Η έγκαιρη ανίχνευση ανωμαλιών επιτρέπει την αποτελεσματική αντίδραση σε επικίνδυνες καταστάσεις. Μπορεί να ανιχνεύσει ανεπιθύμητες ενέργειες, κακόβουλες επιθέσεις, αλλαγές στα μοτίβα λειτουργίας των συσκευών ή την υπερβολική χρήση πόρων. Με την πρόληψη αυτών των προβλημάτων, μπορούν να αποφευχθούν ζημιές, απώλεια δεδομένων, αποτυχίες συστημάτων ή ακόμη και απώλειες ανθρώπινων ζώων σε περιπτώσεις που σχετίζονται με την ασφάλεια.

Επιπλέον, η ανίχνευση ανωμαλιών στο IoT βοηθά στην πρόληψη και την αντιμετώπιση απειλών κυβερνοασφάλειας. Οι κακόβουλες επιθέσεις μπορούν να προέλθουν από εσωτερικές ή εξωτερικές πηγές, και οι ανιχνευτές ανωμαλιών μπορούν να εντοπίσουν ύποπτες δραστηριότητες ή παρεκκλίσεις από την κανονικότητα, προειδοποιώντας τους υπεύθυνους να λάβουν απαραίτητα μέτρα ασφαλείας.[\[13\]](#)

Εν Κατακλείδι, η ανίχνευση ανωμαλιών στο IoT είναι απαραίτητη για την αξιοπιστία, την ασφάλεια και την αποτελεσματικότητα των συστημάτων. Η έγκαιρη ανίχνευση ανωμαλιών επιτρέπει την πρόληψη προβλημάτων, την προστασία από απειλές και τη βελτίωση της απόδοσης. Με την συνεχή εξέλιξη του IoT, η ανίχνευση ανωμαλιών θα παραμείνει ένας ζωτικός παράγοντας για την αποτελεσματική και ασφαλή λειτουργία των συστημάτων IoT.

1.2.3 Τεχνικές Ανίχνευσης Ανωμαλιών στο IoT

Υπάρχουν διάφορες τεχνικές ανίχνευσης ανωμαλιών που μπορούν να εφαρμοστούν στο IoT. Η χρήση μηχανικής μάθησης και αλγορίθμων αναλυτικής στατιστικής επιτρέπει τον εντοπισμό μη συνήθων προτύπων και ανωμαλιών στα δεδομένα. Αυτή η προσέγγιση στηρίζεται στην εκπαίδευση μοντέλων μηχανικής μάθησης με χρήση ιστορικών δεδομένων, έτσι ώστε το σύστημα να μπορεί να αναγνωρίζει τα συνήθη πρότυπα και να εντοπίζει τυχόν ανωμαλίες που αποκλίνουν από αυτά.[\[14\]\[11\]](#)

Μερικές από τις πιο δημοφιλείς τεχνικές ανίχνευσης ανωμαλιών περιλαμβάνουν τη χρήση μοντέλων αυτόματης εξόρυξης δεδομένων, όπως οι αλγόριθμοι clustering και association rules mining.[\[15\]](#) Αυτές οι τεχνικές επιτρέπουν τον εντοπισμό ανατρεπτικών προτύπων ή την εύρεση συσχετίσεων μεταξύ διαφόρων δεδομένων. Με βάση αυτές τις πληροφορίες, μπορεί να γίνει ανίχνευση ανωμαλιών ή αποκάλυψη προβληματικών περιπτώσεων.[\[16\]](#)

Επιπλέον, η χρήση τεχνικών όπως η ανάλυση συστημάτων και χρονοσειρών μπορεί να βοηθήσει στον εντοπισμό ανωμαλιών με βάση τη συμπεριφορά των δεδομένων σε συγκεκριμένες χρονικές στιγμές. Αυτές οι τεχνικές μπορούν να ανιχνεύσουν ανωμαλίες που σχετίζονται με την απόκλιση των μετρήσεων από τα αναμενόμενα μοντέλα ή την εμφάνιση ανεπιθύμητων προτύπων.

Οι προηγμένες τεχνικές ανίχνευσης ανωμαλιών προσφέρουν μια πολύτιμη εργαλειοθήκη για την εξέταση και τον εντοπισμό τυχόν παρατηρούμενων ανωμαλιών στα συστήματα IoT. Η σωστή εφαρμογή τεχνικών ανίχνευσης ανωμαλιών είναι ζωτικής σημασίας για την εξασφάλιση της αξιοπιστίας, της αποτελεσματικότητας και της ασφάλειας των συστημάτων IoT και η επιτυχημένη ανίχνευση ανωμαλιών μπορεί να οδηγήσει σε πρόληψη απωλειών, βελτιωμένη ασφάλεια και αποτελεσματικότερη λειτουργία των συστημάτων IoT.[\[17\]](#)

Συμπερασματικά, η ανίχνευση ανωμαλιών στο Internet of Things (IoT) είναι ζωτικής σημασίας για την ασφάλεια, την προστασία και την ομαλή λειτουργία των συστημάτων. Η έγκαιρη ανίχνευση ανωμαλιών επιτρέπει την πρόληψη πιθανών προβλημάτων και την προστασία από απειλές. Οι τεχνικές ανίχνευσης ανωμαλιών, όπως η χρήση μηχανικής μάθησης, παίζουν έναν σημαντικό ρόλο στον εντοπισμό και την αντιμετώπιση των ανωμαλιών. Με τη συνεχή εξέλιξη του IoT, η ανίχνευση ανωμαλιών θα συνεχίσει να αποτελεί έναν κρίσιμο παράγοντα για την αποτελεσματική και ασφαλή λειτουργία των συστημάτων IoT.

1.3 Ταξινόμηση και Κατηγοριοποίηση Ανωμαλιών

Το Internet of Things (IoT) έχει επανασχεδιάσει τον τρόπο με τον οποίο αλληλεπιδρούμε και λειτουργούν οι συσκευές γύρω μας. Μέσω της σύνδεσης και της ανταλλαγής δεδομένων μεταξύ φυσικών και εικονικών συσκευών, το IoT μας προσφέρει νέες δυνατότητες και λειτουργίες. Ωστόσο, με την αύξηση του αριθμού των συνδεδεμένων συσκευών και της πολυπλοκότητας του συστήματος, ανακύπτουν νέες προκλήσεις σχετικά με την ασφάλεια, την απόδοση και τη διαχείριση του IoT. Η ανίχνευση ανωμαλιών αποκτά ολοένα και μεγαλύτερη σημασία, καθώς μας επιτρέπει να εντοπίσουμε και να αντιμετωπίσουμε δυνητικά προβλήματα και παραβιάσεις στο σύστημα IoT. Στην παρούσα ενότητα, θα εξετάσουμε την ταξινόμηση των ανωμαλιών στο IoT, τους διάφορους τύπους ανωμαλιών που μπορούν να εμφανιστούν και τις περιοχές όπου συχνά εμφανίζονται.

1.3.1 Τύποι Ανωμαλιών

1.3.1.1 Ανωμαλίες Δεδομένων

Οι ανωμαλίες αυτές συμβαίνουν όταν τα δεδομένα που λαμβάνονται από τις συσκευές IoT διαφέρουν από τα αναμενόμενα ή από τα συνήθη πρότυπα λειτουργίας. Μπορεί να προκαλείται από μετρήσεις εκτός φυσιολογικών ορίων, αλλαγές στις προτιμώμενες συνθήκες λειτουργίας, απόκλιση από το αναμενόμενο πρότυπο ή ακόμα και από την απώλεια δεδομένων. Αυτές οι ανωμαλίες μπορούν να οφείλονται σε τεχνικές αποτυχίες, μηχανικά προβλήματα ή ακόμα και κακόβουλες επιθέσεις. Η ανίχνευση αυτών των ανωμαλιών αποτελεί σημαντικό εργαλείο για τον εντοπισμό πιθανών προβλημάτων και απειλών που μπορεί να επηρεάσουν την απόδοση και την ασφάλεια του συστήματος IoT.

1.3.1.2 Ανωμαλίες Συμπεριφοράς

Οι ανωμαλίες αυτές αφορούν τις συνήθειες και τη συμπεριφορά των συσκευών IoT. Όταν μια συσκευή αποκλίνει από το αναμενόμενο μοτίβο συμπεριφοράς της, μπορεί να υποδηλώνει πρόβλημα, όπως πιθανή κακόβουλη επιθετική δραστηριότητα ή παραβίαση της ασφάλειας. Αυτές οι ανωμαλίες μπορούν να σχετίζονται με την αλλαγή των προτιμήσεων του χρήστη, την εκτέλεση μη αναμενόμενων ενεργειών ή ακόμα και με την παρουσία επιθετικού λογισμικού. Η ανίχνευση αυτών των ανωμαλιών μπορεί να βοηθήσει στον εντοπισμό πιθανών κακόβουλων επιθέσεων και στην πρόληψη ασφαλειικών παραβιάσεων.

1.3.1.3 Ανωμαλίες Απόκρισης Δικτύου

Οι ανωμαλίες αυτές συμβαίνουν σε επίπεδο δικτύου, όταν παρατηρούνται απρόσμενες αποκρίσεις, καθυστερήσεις ή απώλεια πακέτων μεταξύ των συσκευών IoT. Αυτό μπορεί να επηρεάσει την αξιοπιστία και την απόδοση του συστήματος. Η ανίχνευση αυτών των ανωμαλιών μπορεί να βοηθήσει στον εντοπισμό προβλημάτων στο δίκτυο και στη βελτίωση της απόδοσης των συσκευών IoT.

1.3.2 Περιοχές Εμφάνισης Ανωμαλιών

1.3.2.1 Ανωμαλίες στην Ενέργεια

Η ενέργεια αποτελεί κρίσιμο παράγοντα για τη λειτουργία των συσκευών IoT. Ανωμαλίες μπορούν να εμφανιστούν στην κατανάλωση ενέργειας ή στην απόδοση της μπαταρίας. Για παράδειγμα, μια συσκευή IoT μπορεί να καταναλώνει απροσδόκητα μεγάλη ποσότητα ενέργειας ή να απενεργοποιείται πρόωρα λόγω αδυναμίας της μπαταρίας να διατηρήσει τη φόρτίσή της. Αυτές οι ανωμαλίες μπορεί να οφείλονται σε τεχνικές αποτυχίες, μηχανικά προβλήματα ή ακόμα και σε κακόβουλες επιθέσεις.[\[15\]](#)

1.3.2.2 Ανωμαλίες στην Ασφάλεια

Οι συσκευές IoT είναι ευάλωτες σε ποικίλες απειλές ασφαλείας. Ανωμαλίες μπορεί να περιλαμβάνουν μη εξουσιοδοτημένη πρόσβαση στη συσκευή, παραβίαση της ακεραιότητας των δεδομένων, κλοπή προσωπικών πληροφοριών ή ακόμα και έγχυση κακόβουλου λογισμικού. Οι ανωμαλίες αυτές μπορούν να απειλήσουν την ιδιωτικότητα, την ασφάλεια και την εμπιστοσύνη στο σύστημα IoT.[\[15\]](#)

1.3.2.3 Ανωμαλίες στην Επικοινωνία

Η επικοινωνία ανάμεσα στις συσκευές IoT είναι απαραίτητη για τη λειτουργία του συστήματος. Ανωμαλίες μπορούν να παρατηρηθούν όταν υπάρχουν προβλήματα στην ασύρματη επικοινωνία, αστάθειες στο δίκτυο ή ακόμα και απώλεια πακέτων δικτύου. Αυτό μπορεί να οδηγήσει σε αποτυχία της επικοινωνίας ανάμεσα στις συσκευές, μειωμένη απόδοση ή καθυστερήσεις στην ανταλλαγή δεδομένων.[\[15\]](#)

Ο προσδιορισμός και η κατανόηση αυτών των περιοχών εμφάνισης ανωμαλιών στο IoT μας επιτρέπει να εστιάσουμε την προσοχή μας στους κρίσιμους τομείς και να αναπτύξουμε αποτελεσματικές τεχνικές ανίχνευσης και πρόληψης. Μέσω της εφαρμογής προηγμένων αλγορίθμων αναγνώρισης προτύπων και μηχανικής μάθησης, μπορούμε να αναγνωρίσουμε τα σημάδια αυτών των ανωμαλιών και να αντιμετωπίσουμε τα προβλήματα εγκαίρως.

Συμπερασματικά, η ταξινόμια των ανωμαλιών στο IoT μας βοηθά να κατανοήσουμε και να αντιμετωπίσουμε τις πιθανές απειλές και προβλήματα που μπορούν να επηρεάσουν την απόδοση και την ασφάλεια του συστήματος IoT. Με την ανάλυση των διαφόρων τύπων ανωμαλιών και των περιοχών όπου συχνά εμφανίζονται, μπορούμε να αναπτύξουμε αποτελεσματικές τεχνικές ανίχνευσης και πρόληψης για τη διατήρηση της ασφάλειας και της απόδοσης του συστήματος IoT. Η συνεχής έρευνα και ανάπτυξη σε αυτόν τον τομέα είναι αναγκαία για τη διασφάλιση της ασφάλειας και της αξιοπιστίας του IoT καθώς εξελίσσεται και επεκτείνεται σε νέες εφαρμογές και πεδία.

1.4 Δομή της διπλωματικής

Στο Κεφάλαιο 1, θα γίνει μία γενική ανασκόπηση του ευρύτερου θέματος της εργασίας, το Διαδίκτυο των Πραγμάτων (Internet of Things - IoT). Θα αναλυθεί ακόμα η έννοια της ανίχνευσης ανωμαλιών (Anomaly Detection) στον χώρο αυτό, η σημαντικότητά της και ακολούθως μία ταξινόμια των ανωμαλιών που μπορεί να εντοπίζονται.

Το Κεφάλαιο 2, αφορά την αντιμετώπιση ανωμαλιών στο IoT. Σε αυτό, γίνεται αναφορά στους βασικούς τρόπους μελέτης ανωμαλιών, στα Νευρωνικά Δίκτυα και τα είδη τους, πιο συγκεκριμένα στα Deep Neural Networks, Recurrent Neural Networks και Reservoir Computing, καθώς και σε τεχνικές ανάλυσης των ανωμαλιών.

Στα Κεφάλαια 3,4 και 5 γίνεται λεπτομερέστερη αναφορά και ανάλυση των θεμάτων Deep Neural Networks, Recurrent Neural Networks και Reservoir Computing αντίστοιχα.

Το Κεφάλαιο 6, περιλαμβάνει μία κριτική αξιολόγηση των παραπάνω κεφαλαίων καθώς και συγκρίσεις μεταξύ των ειδών, πλεονεκτήματα και μειονεκτήματα και τέλος τα συμπεράσματα της εργασίας συνολικά.

2

Αντιμετώπιση Ανωμαλιών

2.1 Τρόποι Μελέτης Ανωμαλιών

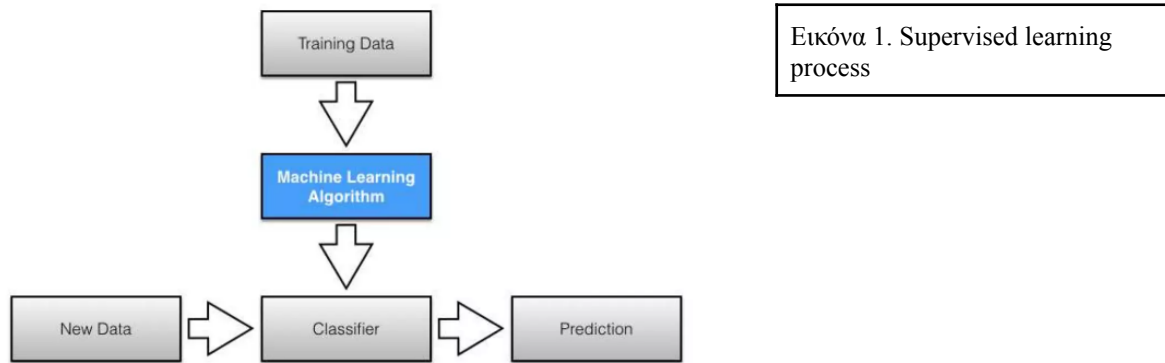
Οι τρόποι μελέτης ανωμαλιών στο Διαδίκτυο των Πραγμάτων μπορούν να ποικίλουν ανάλογα με τις απαιτήσεις και τους στόχους του συστήματος. Οι βασικοί τρόποι μελέτης ανωμαλιών στο IoT περιλαμβάνουν την επιβλεπόμενη μάθηση, τη μη επιβλεπόμενη μάθηση και την ημι-επιβλεπόμενη μάθηση. Ας εξετάσουμε αναλυτικά κάθε προσέγγιση:

2.1.1 Επιβλεπόμενη Μάθηση

Σε αυτήν την προσέγγιση, χρησιμοποιούνται ετικετοποιημένα δεδομένα για την εκπαίδευση των μοντέλων. Τα ετικετοποιημένα δεδομένα περιέχουν πληροφορίες για τις ανωμαλίες που έχουν ήδη γίνει γνωστές, όπως συμπεριφορές που αντιστοιχούν σε προηγούμενα ατυχήματα ή προβλήματα. Η μηχανή μάθησης μπορεί να αναγνωρίσει αυτά τα μοτίβα και να εφαρμόσει αυτήν τη γνώση για τον εντοπισμό νέων ανωμαλιών. Η επιβλεπόμενη μάθηση απαιτεί την ύπαρξη ετικετών ανωμαλίας στα δεδομένα εκπαίδευσης, οι οποίες πρέπει να παρέχονται από εμπειρογνώμονες ή να προκύπτουν από προηγούμενες εμπειρίες.

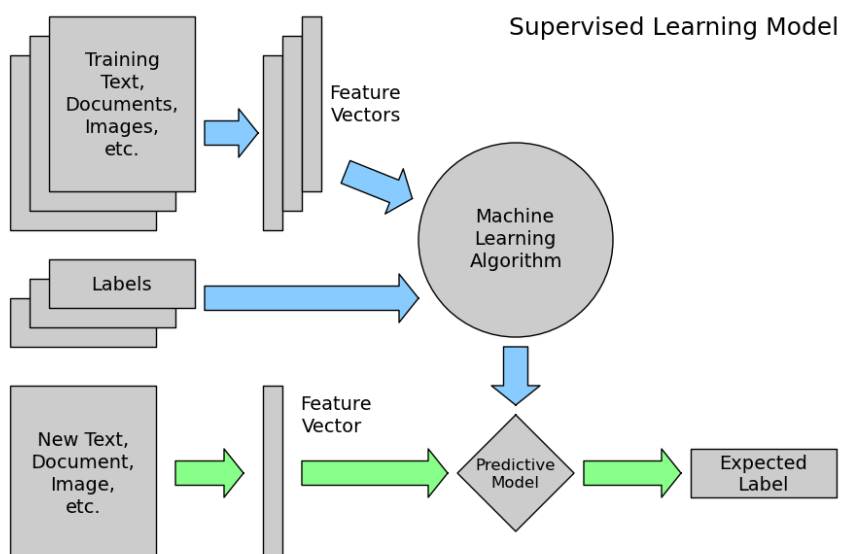
Με βάση τη μελέτη “[An overview of the supervised machine learning methods](#)” του Vladimir Nasteski εμβαθύνουμε περαιτέρω στην προσέγγιση αυτή. Η διαδικασία εκμάθησης σε ένα απλό μοντέλο μηχανικής μάθησης χωρίζεται σε δύο βήματα: εκπαίδευση και δοκιμή. Στη διαδικασία εκπαίδευσης, δείγματα σε δεδομένα εκπαίδευσης λαμβάνονται ως είσοδος στα οποία τα χαρακτηριστικά μαθαίνονται από τον αλγόριθμο μάθησης ή τον εκπαιδευόμενο και χτίζουν το μοντέλο μάθησης. Στη διαδικασία δοκιμής, το μοντέλο εκμάθησης χρησιμοποιεί τη μηχανή εκτέλεσης για να κάνει την πρόβλεψη για τη δοκιμή ή τα δεδομένα παραγωγής. Τα δεδομένα με ετικέτα είναι η έξοδος του μοντέλου εκμάθησης που δίνει την τελική πρόβλεψη ή ταξινομημένα δεδομένα.

Η εποπτευόμενη μάθηση (Εικόνα 1) είναι η πιο κοινή τεχνική στα προβλήματα ταξινόμησης, καθώς ο στόχος είναι συχνά να κάνουμε τη μηχανή να μάθει ένα σύστημα ταξινόμησης που έχουμε δημιουργήσει.



Συνηθέστερα, η εποπτευόμενη μάθηση αφήνει απροσδιόριστη την πιθανότητα για είσοδο, όπως μια είσοδο όπου είναι γνωστή η αναμενόμενη έξοδος. Αυτή η διαδικασία παρέχει σύνολο δεδομένων που αποτελείται από χαρακτηριστικά και ετικέτες. Το κύριο καθήκον είναι η κατασκευή ενός εκτιμητή ικανού να προβλέψει την ετικέτα ενός αντικειμένου που δίνεται από το σύνολο των χαρακτηριστικών. Στη συνέχεια, ο αλγόριθμος εκμάθησης λαμβάνει ένα σύνολο χαρακτηριστικών ως εισόδων μαζί με τις σωστές εξόδους και μαθαίνει συγκρίνοντας την πραγματική του έξοδο με τις διορθωμένες εξόδους για να βρει σφάλματα. Στη συνέχεια, τροποποιεί το μοντέλο ανάλογα. Το μοντέλο που δημιουργείται δεν χρειάζεται εφόσον είναι διαθέσιμες οι εισοδοί, αλλά εάν λείπουν κάποιες από τις τιμές εισόδου, δεν είναι δυνατό να συναχθεί τίποτα για τις εξόδους.

Η εποπτευόμενη μάθηση είναι η πιο κοινή τεχνική εκπαίδευσης για ουδέτερα δίκτυα και δέντρα αποφάσεων. Και τα δύο εξαρτώνται από τις πληροφορίες που δίνονται από την προκαθορισμένη ταξινόμηση. Επίσης, αυτή η εκμάθηση χρησιμοποιείται σε εφαρμογές όπου τα ιστορικά δεδομένα προβλέπουν πιθανά συμβάντα χαρακτηριστικών. Υπάρχουν πολλά πρακτικά παραδείγματα αυτής της μάθησης, για παράδειγμα μια εφαρμογή που προβλέπει το είδος της ίριδας, δεδομένου ενός συνόλου μετρήσεων του άνθους της. Όπως αναφέρθηκε προηγουμένως, οι εποπτευόμενες μαθησιακές εργασίες χωρίζονται σε δύο κατηγορίες: ταξινόμηση και παλινδρόμηση. Στην ταξινόμηση, η ετικέτα είναι διακριτή, ενώ στην παλινδρόμηση, η ετικέτα είναι συνεχής.



Όπως φαίνεται στην παραπάνω εικόνα, ο αλγόριθμος κάνει τη διάκριση μεταξύ των παρατηρούμενων δεδομένων X δηλαδή των δεδομένων εκπαίδευσης, στις περισσότερες περιπτώσεις δομημένων δεδομένων που δίνονται στο μοντέλο κατά τη διάρκεια της εκπαιδευτικής διαδικασίας. Σε αυτή τη διαδικασία, ο αλγόριθμος εποπτευόμενης μάθησης χτίζει το μοντέλο πρόβλεψης. Μετά την εκπαίδευσή του, το προσαρμοσμένο μοντέλο θα προσπαθήσει να προβλέψει τις πιο πιθανές ετικέτες για ένα νέο σύνολο δεδομένων X στο σετ δοκιμών. Ανάλογα με τη φύση του στόχου y , η εποπτευόμενη μάθηση μπορεί να ταξινομηθεί:

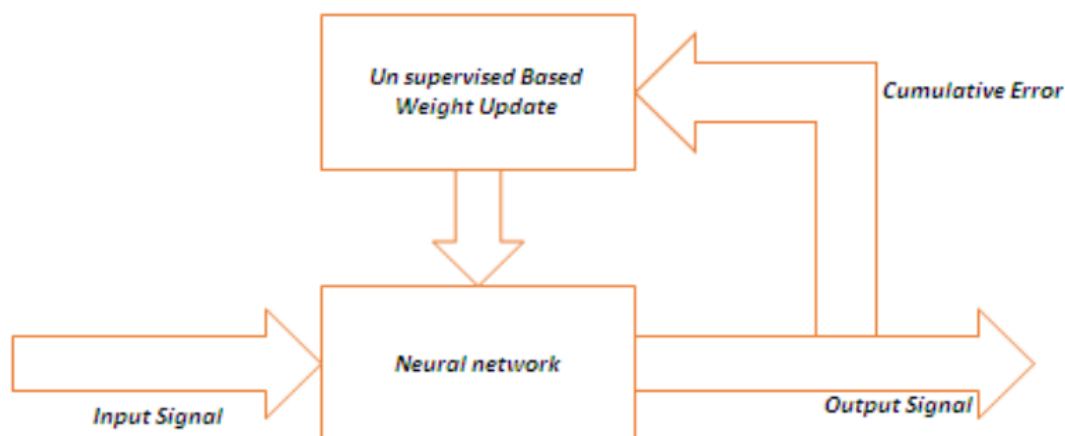
- Εάν y έχει τιμές σε ένα σταθερό σύνολο κατηγορικών αποτελεσμάτων (integers), η εργασία για την πρόβλεψη του y ονομάζεται ταξινόμηση

- Αν y έχει τιμές κινητής υποδιαστολής, η εργασία πρόβλεψης y ονομάζεται παλινδρόμηση

2.1.2 Μη Επιβλεπόμενη Μάθηση

Σε αυτήν την προσέγγιση, χρησιμοποιούνται μη ετικετοποιημένα δεδομένα. Οι αλγόριθμοι μηχανικής μάθησης ανιχνεύουν ανωμαλίες βασιζόμενοι στην απόκλιση από τα αναμενόμενα μοτίβα των δεδομένων. Αυτή η προσέγγιση είναι χρήσιμη όταν δεν υπάρχουν δεδομένα με ετικέτες για τις ανωμαλίες ή όταν οι ανωμαλίες είναι άγνωστες. Οι αλγόριθμοι μηχανικής μάθησης αναλύουν την δομή των δεδομένων και αναγνωρίζουν αταίριαστα μοτίβα, αποκαλύπτοντας έτσι περιοχές των δεδομένων που ενδεχομένως περιέχουν ανωμαλίες.

Με βάση τη μελέτη “[Unsupervised Learning Based On Artificial Neural Network: A Review](#)” των HU Dike, Y Zhou και KK Deveerasetty, εμβαθύνουμε περαιτέρω στην προσέγγιση αυτή. Η μάθηση χωρίς επίβλεψη υποδηλώνει πώς ένα δίκτυο μπορεί να μελετήσει για να υποδηλώσει ορισμένα σχέδια εισόδου σε μια μέθοδο που αναπαράγει την αριθμητική διάταξη της συνολικής συγκέντρωσης σχεδίων ή μοτίβων εισόδου. Είναι μια χρέωση μηχανικής μάθησης για την υπόθεση μιας συνάρτησης για τον ορισμό της κρυφής δομής από τα δεδομένα χωρίς ετικέτα. Αυτό δεν είναι παρά οι αλγόριθμοι εκμάθησης που δεν περιέχουν ετικέτες για την επίβλεψη της μάθησης/εκπαίδευσης. Στον αλγόριθμο, ένας μεγάλος αριθμός δεδομένων και χαρακτηριστικών κάθε παρατήρησης παρέχονται με εισόδους αλλά όχι με την επιθυμητή έξοδο. Η μάθηση χωρίς επίβλεψη χρησιμοποιείται συνήθως (π.χ. ομαδοποίηση - clustering) για τον διαχωρισμό των εικόνων σε δύο σύνολα ή συμπλέγματα με βάση ορισμένα εγγενή χαρακτηριστικά των εικόνων όπως χρώμα, μέγεθος, σχήμα κ.λπ. Το Σχήμα 1 περιγράφει τη σύνδεση δικτύου εκπαίδευσης.



Σχήμα 1. Unsupervised Learning.

Υποδηλώνεται εξίσου ως αλγόριθμος αυτοοργάνωσης (self-organizing) ή προσαρμοστικής μάθησης (adaptive learning) λόγω του γεγονότος ότι δεν υπάρχει εξωτερική βάση για την παροχή πληροφοριών στο δίκτυο. Αυτό βασίζεται στα τοπικά δεδομένα και τον εσωτερικό μηχανισμό (δηλαδή, το σύστημα αποκτά γνώση από μόνο του μαθαίνοντας και εξοικειώνοντας τις ζωτικές πληροφορίες στα δεδομένα εισόδου). Τα σχήματα εκπαίδευσης και τα μοτίβα εισαγωγής προσφέρονται στο σύστημα τα οποία το σύστημα τακτοποιεί ως δεδομένα σε κατηγορίες ή ομάδες. Μια ομάδα προτύπων εκπαίδευσης ή δεδομένων δίνεται στο σύστημα στο στάδιο του επιπέδου εισόδου. τα βάρη συσχέτισης δικτύου συντονίζονται για κάποιο είδος ανταγωνισμού μεταξύ των κόμβων του επιπέδου εξόδου όπου ο κόμβος με την υψηλότερη τιμή θα είναι ο επιτυχής υποψήφιος.

Η μάθηση χωρίς επίβλεψη είναι κυρίως χρήσιμη για αλγορίθμους ομαδοποίησης και συσχέτισης. Πρόσφατα ευρήματα υποδηλώνουν ότι η ομαδοποίηση σε μη εποπτευόμενη μάθηση για νευρωνικά δίκτυα τροφοδοσίας παρεμποδίζεται από τη χαμηλή ταχύτητα και τη χαμηλή ακρίβεια με ζητήματα υψηλής πολυπλοκότητας μνήμης.[18] Η ομαδοποίηση παίζει σημαντικό ρόλο στη μηχανική μάθηση και σε πεδία πραγματικών εφαρμογών και στην ιατρική μελέτη. Εν τω μεταξύ, η κύρια πρόκληση που σχετίζεται με την μάθηση χωρίς επίβλεψη είναι η επιβάρυνση της ανακάλυψης κρυφών δομών σε δεδομένα χωρίς ετικέτα.[19] Ο βασικός στόχος της μάθησης χωρίς επίβλεψη είναι απλώς να βρεθεί η πιθανή δομή που είναι εγγενής στα δεδομένα εισόδου.

2.1.3 Ημι-Επιβλεπόμενη Μάθηση

Αυτή η προσέγγιση συνδυάζει τα πλεονεκτήματα της επιβλεπόμενης και της μη επιβλεπόμενης μάθησης. Αρχικά, γίνεται εκπαίδευση με ετικετοποιημένα δεδομένα, τα οποία περιέχουν πληροφορίες για γνωστές ανωμαλίες. Κατόπιν, το εκπαιδευμένο μοντέλο χρησιμοποιείται για τον εντοπισμό ανωμαλιών σε μη ετικετοποιημένα δεδομένα, όπου πιθανώς να υπάρχουν ανωμαλίες που δεν έχουν εκπαιδευτεί ακόμα. Αυτή η προσέγγιση επιτρέπει τον εντοπισμό νέων ανωμαλιών ενώ εκμεταλλεύεται την υπάρχουσα γνώση για τις γνωστές ανωμαλίες.

Με βάση τη μελέτη “[Semi-supervised learning by disagreement](#)” των Zhi-Hua Zhou και Ming Li, εμβαθύνουμε περεταίρω στην προσέγγιση αυτή.

Στην ημι-εποπτευόμενη μάθηση, ένα σύνολο δεδομένων εκπαίδευσης με ετικέτα $L = \{(x_1, y_1), (x_2, y_2), \dots, (x_{|L|}, y_{|L|})\}$ και ένα σύνολο δεδομένων εκπαίδευσης χωρίς ετικέτα $U = \{x'_1, x'_2, \dots, x'_{|U|}\}$ παρουσιάζονται στον αλγόριθμο εκμάθησης για την κατασκευή μιας συνάρτησης $f: X \rightarrow Y$ για την πρόβλεψη των ετικετών των αόρατων περιπτώσεων, όπου X και Y είναι, αντίστοιχα, ο χώρος εισόδου και ο χώρος εξόδου, $x_i, x'_j \in X$ ($i = 1, 2, \dots, |L|, j = 1, 2, \dots, |U|$) είναι d -dimensional διανύσματα χαρακτηριστικών που προέρχονται από το X , και το $y_i \in Y$ είναι η ετικέτα του x_i , συνήθως $|L| \ll |U|$.

Στην πραγματικότητα, κάποια απλή χρήση παραδειγμάτων χωρίς ετικέτα εμφανίστηκε ακόμη νωρίτερα από την ημι-εποπτευόμενη μάθηση. Λόγω των δυσκολιών στην ενσωμάτωση μη επισημασμένων δεδομένων απευθείας σε συμβατικές μεθόδους εποπτευόμενης μάθησης (π.χ.

νευρωνικά δίκτυα BP) και της έλλειψης ξεκάθαρης κατανόησης της αξίας των δεδομένων χωρίς ετικέτα στη διαδικασία μάθησης, η μελέτη της ημι-εποπτευόμενης μάθησης τράβηξε την προσοχή μόνο μετά τα μέσα της δεκαετίας του 1990. Καθώς η ζήτηση για αυτόματη εκμετάλλευση δεδομένων χωρίς ετικέτα αυξάνεται και η αξία των δεδομένων χωρίς ετικέτα αποκαλύφθηκε από κάποιες πρώτες αναλύσεις, η ημι-εποπτευόμενη μάθηση έχει γίνει ένα καυτό θέμα.

Οι περισσότερες πρώτες μελέτες δεν παρείχαν πληροφορίες ή εξήγηση για τον λόγο για τον οποίο τα μη επισημασμένα δεδομένα μπορεί να είναι ωφέλιμα. Οι Miller και Uyar παρείχαν πιθανώς την πρώτη εξήγηση για τη χρησιμότητα των μη επισημασμένων δεδομένων από την άποψη της εκτίμησης της διανομής δεδομένων. Υπέθεσαν ότι τα δεδομένα προέρχονταν από ένα μοντέλο Gaussian μείγματος με συστατικά μείγματα L , δηλαδή:

$$f(x|\theta) = \sum_{l=1}^L \alpha_l f(x|\theta_l)$$

όπου α_l είναι ο συντελεστής μείγματος που ικανοποιεί $\sum_{l=1}^L \alpha_l = 1$, ενώ $\theta = \{\theta_l\}$ είναι οι παράμετροι του μοντέλου. Σε αυτήν την περίπτωση, η ετικέτα c_i μπορεί να θεωρηθεί μια τυχαία μεταβλητή C της οποίας η κατανομή $P(c_i|x_i, m_i)$ προσδιορίζεται από το συστατικό μίγμα m_i και το διάνυσμα χαρακτηριστικών x_i . Ο βέλτιστος κανόνας ταξινόμησης για αυτό το μοντέλο είναι το κριτήριο MAP (maximum a posterior), δηλαδή:

$$h(x) = \underset{k}{\operatorname{argmax}} \sum_j P(c_i = k|x_i, m_i = j)P(m_i = j|x_i)$$

$$\text{όπου, } P(m_i = j|x_i) = \frac{\alpha_j f(x_i|\theta_j)}{\sum_{l=1}^L \alpha_l f(x_i|\theta_l)}$$

Έτσι, ο στόχος της μάθησης επιτυγχάνεται με την εκτίμηση των όρων $P(c_i = k|x_i, m_i = j)$ και $P(m_i = j|x_i)$ από τα δεδομένα εκπαίδευσης. Μπορεί να βρεθεί ότι μόνο η εκτίμηση της πρώτης πιθανότητας περιλαμβάνει την ετικέτα κλάσης. Έτσι, μη επισημασμένα παραδείγματα μπορούν να χρησιμοποιηθούν για τη βελτίωση της εκτίμησης της δεύτερης πιθανότητας και, ως εκ τούτου, τη βελτίωση της απόδοσης της μαθημένης υπόθεσης. Αργότερα, αναλύθηκε η τιμή των μη επισημασμένων δεδομένων για παραμετρικά μοντέλα. Πρότειναν ότι εάν ένα παραμετρικό μοντέλο μπορεί να αποσυντεθεί ως $P(x, y|\theta) = P(y|x, \theta)P(x|\theta)$, η χρήση παραδειγμάτων χωρίς ετικέτα μπορεί να βοηθήσει στην επίτευξη καλύτερης εκτίμησης του μοντέλου parameters

Υπάρχουν δύο βασικές παραδοχές στην ημι-εποπτευόμενη μάθηση, η cluster assumption και η manifold assumption. Η πρώτη υποθέτει ότι τα δεδομένα με παρόμοιες εισόδους πρέπει να έχουν παρόμοιες ετικέτες κλάσης, η δεύτερη υποθέτει ότι τα δεδομένα με παρόμοιες εισόδους πρέπει να έχουν παρόμοιες εξόδους. Η cluster assumption αφορά την ταξινόμηση, ενώ η πολλαπλή υπόθεση μπορεί επίσης να εφαρμοστεί σε εργασίες εκτός της ταξινόμησης. Κατά κάποιο τρόπο, η manifold assumption είναι μια γενίκευση της υπόθεσης της συστάδας. Αυτές οι υποθέσεις συνδέονται στενά με την ιδέα του διαχωρισμού χαμηλής πυκνότητας (low density separation), η οποία έχει ληφθεί από πολλούς ημι-εποπτευόμενους αλγόριθμους μάθησης. Ανεξάρτητα από το ποια υπόθεση λαμβάνεται, η κοινή υποκείμενη πεποίθηση είναι ότι τα μη

επισημασμένα δεδομένα παρέχουν κάποιες χρήσιμες πληροφορίες σχετικά με τη διανομή δεδομένων βασικής αλήθειας. Έτσι, ένα κλειδί της ημι-εποπτευόμενης μάθησης είναι η εκμετάλλευση των πληροφοριών διανομής που αποκαλύπτονται από παραδείγματα χωρίς ετικέτα.[\[17\]](#)

Πολλοί ημι-εποπτευόμενοι αλγόριθμοι μάθησης έχουν αναπτυχθεί. Σε γενικές γραμμές, μπορούν να κατηγοριοποιηθούν σε τέσσερις κατηγορίες, μεθόδους παραγωγής (generative methods), S3VM (Semi Supervised Support Vector Machines), μεθόδους που βασίζονται σε γραφήματα (graph-based methods) και μεθόδους που βασίζονται σε διαφωνίες (disagreement-based methods).[\[19\]](#)

Καθένας από τους παραπάνω τρόπους μελέτης έχει τα πλεονεκτήματα και τις αστοχίες του και η επιλογή τους εξαρτάται από την συγκεκριμένη περίπτωση και τα χαρακτηριστικά του εκάστοτε προβλήματος. Σε κάθε περίπτωση, ο συνδυασμός διαφορετικών προσεγγίσεων μπορεί να βελτιώσει την απόδοση και την ακρίβεια της ανίχνευσης ανωμαλιών στο IoT.

2.2 Προσεγγίσεις και Τεχνικές

Οι τρόποι μελέτης ανωμαλιών στο Διαδίκτυο των Πραγμάτων περιλαμβάνουν ακόμα, μια σειρά προσεγγίσεων και τεχνικών που επιτρέπουν την αναγνώριση, την ανίχνευση και την αντιμετώπιση ανωμαλιών. Αυτοί οι τρόποι μελέτης προσφέρουν τη δυνατότητα προληπτικής δράσης, αυξάνοντας την ασφάλεια, την αξιοπιστία και την αποδοτικότητα των συστημάτων IoT. Ορισμένοι από τους κύριους τρόπους μελέτης ανωμαλιών στο IoT περιλαμβάνουν τα εξής:

2.2.1 Ανίχνευση Ατυχημάτων

Οι τεχνικές ανίχνευσης ατυχημάτων στον τομέα του IoT επικεντρώνονται στον εντοπισμό ανωμαλιών που μπορεί να οφείλονται σε βλάβες ή αποτυχίες των συστημάτων IoT. Η ανίχνευση αυτών των ατυχημάτων είναι σημαντική για την ασφάλεια, την αξιοπιστία και την αποτελεσματική λειτουργία των συσκευών και των δικτύων του IoT.

Η διαδικασία ανίχνευσης ατυχημάτων συνήθως ξεκινά με τη συλλογή δεδομένων που σχετίζονται με την κατάσταση και την απόδοση των συσκευών του IoT. Αυτά τα δεδομένα μπορεί να περιλαμβάνουν παραμέτρους όπως η θερμοκρασία, η πίεση, η ταχύτητα, η τάση και άλλες μετρήσεις που αντιπροσωπεύουν την κατάσταση των συσκευών.

Στη συνέχεια, τα δεδομένα αυτά υπόκεινται σε ανάλυση και επεξεργασία για τον εντοπισμό αποκλίσεων από την κανονική λειτουργία. Εδώ είναι όπου οι τεχνικές μηχανικής μάθησης μπορούν να εφαρμοστούν. Με τη χρήση αλγορίθμων μηχανικής μάθησης, μπορούν να εκπαιδευτούν μοντέλα που μπορούν να αναγνωρίσουν τα αναμενόμενα μοτίβα και να εντοπίσουν τυχόν αποκλίσεις από αυτά.[\[20\]](#)

Η ανίχνευση ατυχημάτων μέσω τεχνικών μηχανικής μάθησης μπορεί να βοηθήσει στην πρόληψη ατυχημάτων, στην εξοικονόμηση κόστους συντήρησης και στη βελτίωση της ασφάλειας και της αποδοτικότητας των συστημάτων IoT.

2.2.2 Ανίχνευση Συστημάτων

Η ανίχνευση συστημάτων αποτελεί μια προσέγγιση που επικεντρώνεται στον παρατηρητή και τον εντοπισμό συγκεκριμένων μοτίβων και συμπεριφορών στα δεδομένα του IoT που υποδηλώνουν ανωμαλίες.[\[21\]](#) Ο στόχος είναι να εντοπιστούν αποκλίσεις από την αναμενόμενη λειτουργία του συστήματος.

Η ανίχνευση συστημάτων αναλύει την κατάσταση και την απόδοση των συστημάτων του IoT, παρακολουθώντας τα δεδομένα που παράγονται από αισθητήρες, συσκευές και δίκτυα. Μελετά τις χρονοσειρές μετρήσεων, τις παραμέτρους λειτουργίας και τις αλληλεπιδράσεις μεταξύ των συστατικών του συστήματος. Οι τρόποι ανίχνευσης συστημάτων μπορεί να περιλαμβάνουν:

- **Ανάλυση Χρονοσειρών**

Η ανάλυση των χρονοσειρών μετρήσεων από αισθητήρες και συσκευές μπορεί να αποκαλύψει ανωμαλίες και αποκλίσεις από τα αναμενόμενα μοτίβα. Μέθοδοι όπως η απλή κυλιόμενη μέση των δεδομένων, η αυτοσυσχέτιση και η σπεκτραλή ανάλυση μπορούν να αναδείξουν παρατηρήσεις που υποδηλώνουν ανωμαλίες.

- **Συναρτησιακή Ανάλυση**

Η ανάλυση των συναρτήσεων πιθανότητας και των στατιστικών μετρικών μπορεί να αποκαλύψει παρατηρήσεις που υποδηλώνουν αποκλίσεις από τα αναμενόμενα μοτίβα. Αυτή η προσέγγιση επικεντρώνεται στην ανάλυση των κατανομών, των πυκνοτήτων και των μέσων τιμών των δεδομένων για τον εντοπισμό ανωμαλιών.

- **Μοντέλα Μηχανικής Μάθησης**

Οι αλγόριθμοι μηχανικής μάθησης μπορούν να εκπαιδευτούν για να αναγνωρίσουν τα αναμενόμενα μοτίβα των δεδομένων του IoT και να εντοπίσουν αποκλίσεις από αυτά. Μπορούν να ανιχνεύσουν ανωμαλίες σε πραγματικό χρόνο με βάση την εκπαίδευση πάνω σε ιστορικά δεδομένα ή με αξιοποίηση τεχνικών όπως οι αυτομάτως προσαρμοσίμες μεταβλητές (καταστάσεις) και τα νευρωνικά δίκτυα.

- **Επιβλεπόμενη Μάθηση**

Η επιβλεπόμενη μάθηση εμπλέκει την εκπαίδευση μοντέλων με χρήση ετικετών ανωμαλίας, δηλαδή δεδομένων που είναι γνωστό ότι περιέχουν ανωμαλίες. Αυτά τα μοντέλα μπορούν να αναγνωρίσουν ανωμαλίες με βάση την εκπαίδευσή τους στα επισημασμένα δεδομένα ανωμαλίας.

2.2.3 Ανίχνευση Ανωμαλιών με βάση τη Γνώση

Η προσέγγιση της ανίχνευσης ανωμαλιών με βάση τη γνώση βασίζεται στην εμπειρογνωμοσύνη των ειδικών και των εμπειρογνομόνων για τα συστήματα IoT. Σε αυτήν, αναλύονται οι γνωστές περιπτώσεις ανωμαλιών και δημιουργούνται κανόνες ή μοντέλα για τον εντοπισμό αυτών των ανωμαλιών στα δεδομένα.

Κατά τη διάρκεια της λειτουργίας του συστήματος IoT, τα δεδομένα συγκρίνονται με τους κανόνες ή τα μοντέλα γνώσης που έχουν οριστεί. Εάν παρατηρηθεί μια ανωμαλία που αποκλίνει από την αναμενόμενη συμπεριφορά, γίνεται η ανίχνευση και η αντίδραση αναλόγως. Αυτό μπορεί να περιλαμβάνει την εκτέλεση ειδοποιήσεων, τη λήψη μέτρων ασφαλείας ή την προαγωγή περαιτέρω ανάλυσης και έρευνας.

Η προσέγγιση αυτή βασίζεται στην προηγούμενη γνώση των εμπειρογνομόνων και μπορεί να είναι ευέλικτη στην αντιμετώπιση νέων μοτίβων ανωμαλιών που δεν έχουν προηγουμένως καταγραφεί. Ωστόσο, η αποτελεσματικότητα αυτής της προσέγγισης εξαρτάται από την ποιότητα και την καλή κάλυψη της γνώσης των εμπειρογνομόνων.

2.2.4 Ανίχνευση Ασφαλείας

Η ανίχνευση ασφαλείας αποτελεί σημαντική πτυχή στον τομέα του IoT και ασχολείται με τον εντοπισμό και την αντιμετώπιση ανωμαλιών που σχετίζονται με την ασφάλεια των συσκευών και των δικτύων του IoT.[\[21\]](#) Για την ανίχνευση ασφαλείας, πραγματοποιείται παρακολούθηση της δραστηριότητας στο δίκτυο και στις συσκευές IoT. Αυτό συμπεριλαμβάνει τη συλλογή δεδομένων σχετικά με την κίνηση των δεδομένων, τις επικοινωνίες μεταξύ συσκευών, την αλληλεπίδραση με τα πρωτόκολλα και τις διεπαφές, καθώς και την ανίχνευση ενδεχόμενων προσπαθειών εισβολής ή κακόβουλων ενεργειών.

Ανάμεσα στις ανωμαλίες που μπορούν να ανιχνευθούν, συμπεριλαμβάνονται η ανίχνευση κακόβουλου λογισμικού (malware), η μη εξουσιοδοτημένη πρόσβαση σε συσκευές IoT, η παραβίαση ασφαλείας των δικτύων και η διακίνηση παραποιημένων δεδομένων. Με την ανίχνευση ασφαλείας, μπορεί να επιτευχθεί η πρόληψη απειλών και η αντίδραση σε επιθέσεις στον τομέα του IoT.

2.2.5 Στατιστικές Μέθοδοι

Οι στατιστικές μέθοδοι αποτελούν ένα αναλυτικό πλαίσιο που χρησιμοποιείται για την ανίχνευση ανωμαλιών στα δεδομένα του IoT. Αυτές οι μέθοδοι εστιάζουν στην εξέταση των κατανομών και των συναρτήσεων πιθανότητας των δεδομένων για τον εντοπισμό αποκλίσεων από τα αναμενόμενα μοτίβα.[\[17\]](#)

Χρησιμοποιούνται για την ανάλυση και την εξαγωγή στατιστικών χαρακτηριστικών από τα δεδομένα του IoT. Μελετούν τις κατανομές των δεδομένων, την αλληλεπίδραση μεταξύ διαφορετικών μεταβλητών και τις συναρτήσεις πιθανότητας για να κατανοήσουν το πώς αναπτύσσονται τα δεδομένα στον χώρο του IoT.

Μια συνηθισμένη προσέγγιση είναι η χρήση στατιστικών μοντέλων, όπως οι κατανομές πιθανότητας, για την περιγραφή των αναμενόμενων μοτίβων στα δεδομένα του IoT. Κατά την εκπαίδευση αυτών των μοντέλων, μπορούν να αποκτηθούν γνώσεις σχετικά με τον τρόπο που τα δεδομένα ακολουθούν κανονικά μοντέλα κατανομής και πιθανότητας.

Κατόπιν, η ανίχνευση ανωμαλιών μπορεί να πραγματοποιηθεί με τον συγκριτικό έλεγχο των παρατηρούμενων δεδομένων με τις κατανομές που περιγράφουν τα αναμενόμενα μοτίβα. Εάν τα παρατηρούμενα δεδομένα αποκλίνουν σημαντικά από τις κατανομές που προβλέπονται από τα μοντέλα, τότε μπορεί να εντοπιστεί μια ανωμαλία.

Οι στατιστικές μέθοδοι μπορούν να εφαρμοστούν σε διάφορους τομείς του IoT, όπως η ανίχνευση αποκλίσεων σε μετρήσεις αισθητήρων, η αναγνώριση παρατηρούμενων συμπεριφορών που δεν συμβαδίζουν με τα αναμενόμενα πρότυπα, και η ανίχνευση σχέσεων και εξαρτήσεων μεταξύ διαφόρων μεταβλητών.

Οι παραπάνω τρόποι μελέτης ανωμαλιών στο IoT συνήθως συνδυάζονται και χρησιμοποιούνται από κοινού για την αποτελεσματική ανίχνευση και αντιμετώπιση ανωμαλιών στο περιβάλλον του IoT, επιτρέποντας μας να αναγνωρίζουμε και να αντιμετωπίζουμε τις απειλές

και τα προβλήματα που επηρεάζουν την απόδοση, την ασφάλεια και την αξιοπιστία των συστημάτων IoT.

Η συνδυασμένη χρήση διαφορετικών μεθόδων και προσεγγίσεων μπορεί να ενισχύσει την ακρίβεια και την αξιοπιστία των αποτελεσμάτων και να βοηθήσει στην εξαγωγή σημαντικών πληροφοριών από τα δεδομένα του IoT.

Η συνεχής έρευνα και ανάπτυξη σε αυτόν τον τομέα είναι κρίσιμη για την προαγωγή της ασφάλειας και της αξιοπιστίας των συστημάτων IoT και για την προστασία των χρηστών και των δεδομένων τους.

2.3 Τεχνικές Ανάλυσης Ανωμαλιών

Ο βασικός τρόπος μελέτης ανωμαλιών στο IoT περιλαμβάνει τη χρήση διαφόρων τεχνικών ανάλυσης και προσεγγίσεων που μας επιτρέπουν να αναγνωρίσουμε, να ανιχνεύσουμε και να αντιμετωπίσουμε τις ανωμαλίες στα συστήματα IoT.

Οι τεχνικές ανάλυσης ανωμαλιών είναι απαραίτητες για τον εντοπισμό και την αντιμετώπιση των ανωμαλιών στα συστήματα IoT. Αυτές οι τεχνικές μας επιτρέπουν να αναγνωρίσουμε τις μη συνήθεις συμπεριφορές και τα μοτίβα που αποκλίνουν από τα αναμενόμενα, προσφέροντας προληπτική δράση και αύξηση της ασφάλειας και της αξιοπιστίας των συστημάτων IoT. Ορισμένες από τις συνηθέστερες τεχνικές ανάλυσης ανωμαλιών περιλαμβάνουν:

2.3.1 Στατιστική Ανάλυση

Η στατιστική ανάλυση αποτελεί μια ισχυρή και ευρέως χρησιμοποιούμενη τεχνική για την εξαγωγή γνώσης και πληροφορίας από τα δεδομένα που συλλέγονται από το Διαδίκτυο των Πραγμάτων.[22] Η διαρκής επέκταση των συνδεδεμένων συσκευών στο περιβάλλον μας έχει οδηγήσει στη συγκέντρωση μεγάλου όγκου δεδομένων, και η στατιστική ανάλυση αποτελεί απαραίτητο εργαλείο για την εξαγωγή χρήσιμης πληροφορίας από αυτά.

Μια από τις βασικές αρχές της στατιστικής ανάλυσης είναι η περιγραφή και η κατανόηση των δεδομένων μέσω μέτρων και μετρικών. Αυτό περιλαμβάνει τον υπολογισμό του μέσου όρου, της διακύμανσης, του εύρους, των κατανομών πιθανότητας, και άλλων στατιστικών δεικτών που βοηθούν στην κατανόηση των παρατηρούμενων προτύπων και των ανωμαλιών. Με τη χρήση των στατιστικών αυτών δεικτών,[23] είναι δυνατή η ανίχνευση τυχόν αποκλίσεων ή ανωμαλιών που μπορεί να υπάρχουν στα δεδομένα IoT.

Μια συνηθισμένη πρακτική στη στατιστική ανάλυση είναι η χρήση διαγραμμάτων, γραφημάτων και διανυσματικών πεδίων για την απεικόνιση των δεδομένων και την οπτική ανίχνευση πιθανών προτύπων.[24] Αυτή η οπτικοποίηση μπορεί να διευκολύνει την εντοπισμό ανωμαλιών που ενδεχομένως να μην είναι εύκολα αντιληπτές σε υψηλές διαστάσεις δεδομένων.

Επιπλέον, η στατιστική ανάλυση μπορεί να χρησιμοποιηθεί για την εκτίμηση της αξιοπιστίας και της εγκυρότητας των μετρήσεων που πραγματοποιούνται από τις συνδεδεμένες συσκευές IoT.[25] Αυτό είναι ιδιαίτερα σημαντικό όταν τα δεδομένα χρησιμοποιούνται για τη λήψη αποφάσεων και την ανάπτυξη νέων τεχνολογιών που εξαρτώνται από την ακρίβεια των μετρήσεων.

Μια αποτελεσματική στατιστική ανάλυση απαιτεί επίσης την επιλογή της κατάλληλης μεθόδου ανάλυσης, καθώς και την επεξεργασία των δεδομένων πριν από την εφαρμογή των στατιστικών τεχνικών.[26] Οι μέθοδοι αυτές μπορεί να περιλαμβάνουν την κανονικοποίηση, την

αφαίρεση θορύβου, την αντιμετώπιση των απουσιάζουσων τιμών (missing data), και την ανάλυση συσχέτισης μεταξύ διαφορετικών μεταβλητών.

Κάποιες από τις κύριες τεχνικές στατιστικής ανάλυσης που χρησιμοποιούνται για την επεξεργασία δεδομένων IoT περιλαμβάνουν:

- **Ανάλυση Διακύμανσης (Variance Analysis)**

Χρησιμοποιείται για την αξιολόγηση των διαφορετικών μετρήσεων και την ανίχνευση ενδεχόμενων ανωμαλιών μεταξύ των παρατηρήσεων. Η διακύμανση αποτελεί μια μέτρηση της απόκλισης ή της διαφοροποίησης των παρατηρήσεων από τον μέσο όρο. Μικρή διακύμανση υποδεικνύει ότι οι τιμές είναι συγκεντρωμένες κοντά στο μέσο όρο, ενώ μεγάλη διακύμανση υποδεικνύει ότι οι τιμές διασκορπίζονται περισσότερο.

- **Παλινδρόμηση (Regression Analysis)**

Η παλινδρόμηση αποτελεί μια ισχυρή στατιστική τεχνική που χρησιμοποιείται για την εξέταση των συσχετίσεων μεταξύ δύο ή περισσότερων μεταβλητών. Η κύρια χρήση της παλινδρόμησης είναι η πρόβλεψη των τιμών μιας εξαρτημένης μεταβλητής βάσει των τιμών μιας ή περισσότερων ανεξάρτητων μεταβλητών. Στην παλινδρόμηση, προσπαθούμε να κατανοήσουμε τον τρόπο με τον οποίο μια ή περισσότερες ανεξάρτητες μεταβλητές επηρεάζουν τη μεταβλητή που θέλουμε να προβλέψουμε. Μέσω της παλινδρόμησης, μπορούμε να εκτιμήσουμε την καμπύλη γραμμής ή το μοντέλο που περιγράφει τη συσχέτιση ανάμεσα στις μεταβλητές και να χρησιμοποιήσουμε αυτό το μοντέλο για την πρόβλεψη των τιμών της εξαρτημένης μεταβλητής για νέες τιμές των ανεξάρτητων μεταβλητών. Είναι μια δυνατή εργαλειοθήκη για τον εντοπισμό της επιρροής μεταξύ δεδομένων IoT και την κατανόηση της πιθανής απόκλισης ή συμπεριφοράς τους.

- **Κατανομή Πιθανότητας (Probability Distribution)**

Η ανάλυση των κατανομών πιθανότητας μπορεί να βοηθήσει στον προσδιορισμό της κατανομής των δεδομένων και την ανίχνευση ανωμαλιών που ενδέχεται να προκύπτουν από την απόκλιση των δεδομένων από την κανονική κατανομή. Οι κατανομές πιθανότητας περιγράφουν την πιθανότητα ενός συμβάντος να συμβεί. Αυτό είναι χρήσιμο για να κατανοήσουμε την πιθανότητα να συμβούν συγκεκριμένα γεγονότα στα δεδομένα IoT.

Μπορεί επίσης να χρησιμοποιηθούν και άλλες στατιστικές τεχνικές, ανάλογα με τον συγκεκριμένο σκοπό και τα χαρακτηριστικά των δεδομένων που αναλύονται.

2.3.2 Μηχανική Μάθηση

Η μηχανική μάθηση αποτελεί έναν δημοφιλή τρόπο μελέτης και ανίχνευσης ανωμαλιών στον τομέα του IoT. Η εφαρμογή τεχνικών μηχανικής μάθησης, περιλαμβάνει τη χρήση αλγορίθμων μηχανικής μάθησης για την εκπαίδευση μοντέλων που μπορούν να αναγνωρίσουν ανωμαλίες στα δεδομένα. Αυτά τα μοντέλα μπορούν να εκπαιδευτούν να αναγνωρίζουν τα αναμενόμενα μοτίβα και να εντοπίζουν τυχόν αποκλίσεις από αυτά.

Αντιπροσωπεύει μια σημαντική προσέγγιση στην ανάλυση ανωμαλιών στα δεδομένα IoT, καθώς παρέχει ευέλικτες μεθόδους για την ανίχνευση μοτίβων και τον εντοπισμό ανωμαλιών. Οι

αλγόριθμοι μηχανικής μάθησης μπορούν να εκπαιδευτούν για να αναγνωρίσουν και να ανταποκριθούν σε διάφορες ανωμαλίες, από απλές έως πολύπλοκες, ανάλογα με την ποικιλία και τον όγκο των δεδομένων που συλλέγονται από τις συσκευές IoT.[21]

Τα νευρωνικά δίκτυα αποτελούν έναν τύπο μηχανικής μάθησης που είναι ιδιαίτερα αποτελεσματικός στον εντοπισμό και την ανάλυση πολύπλοκων μοτίβων στα δεδομένα.[27] Η χρήση νευρωνικών δικτύων μπορεί να βοηθήσει στην αναγνώριση ανωμαλιών που μπορεί να είναι δύσκολο να ανιχνευθούν με τις παραδοσιακές στατιστικές τεχνικές.

Οι αλγόριθμοι ομαδοποίησης (clustering) αποτελούν μια άλλη κατηγορία μηχανικής μάθησης που χρησιμοποιείται για τον εντοπισμό μοτίβων στα δεδομένα IoT. Οι αλγόριθμοι αυτοί μπορούν να ομαδοποιήσουν τα δεδομένα σε συστάδες βάσει των παρόμοιων χαρακτηριστικών τους, κάτι που μπορεί να αποκαλύψει διάφορα μοτίβα και ανωμαλίες.[28]

Οι αλγόριθμοι ανίχνευσης ανωμαλιών (anomaly detection algorithms) είναι ιδιαίτερα σημαντικοί για την επισήμανση και τον εντοπισμό ανωμαλιών στα δεδομένα IoT. Αυτοί οι αλγόριθμοι είναι σε θέση να εντοπίσουν ανεπανόρθωτες αποκλίσεις και παράξενα πρότυπα που μπορούν να υποδείξουν προβλήματα, βλάβες ή ασυνήθιστη συμπεριφορά στις συσκευές ή το περιβάλλον παρατηρήσεων.[29]

Κάποιες συγκεκριμένες μέθοδοι μηχανικής μάθησης που έχουν χρησιμοποιηθεί με επιτυχία για την ανάλυση ανωμαλιών στα δεδομένα IoT περιλαμβάνουν τους αλγόριθμους υποστήριξης διανυσματικών μηχανών (Support Vector Machines - SVM), τα αναδρομικά νευρωνικά δίκτυα (Recurrent Neural Networks - RNNs), τις μεθόδους ομαδοποίησης, όπως οι αλγόριθμοι K-Means και DBSCAN, καθώς και μέθοδοι βαθιάς ενισχυτικής μάθησης για την αναγνώριση ανωμαλιών.[19]

2.3.3 Ανάλυση Φάσματος

Η ανάλυση φάσματος είναι μια ισχυρή τεχνική που χρησιμοποιείται για τον εντοπισμό και την ανάλυση των συχνοτήτων και των σημάτων που περιλαμβάνονται στα δεδομένα. Στο πεδίο των δεδομένων IoT, αυτή η τεχνική μπορεί να εφαρμοστεί για την ανίχνευση μη συνήθων μοτίβων, παραμορφώσεων ή παρεμβολών που μπορεί να υπάρχουν στα σήματα που παράγονται από τις συνδεδεμένες συσκευές.

Κατά τη διάρκεια της ανάλυσης φάσματος, τα δεδομένα μελετούνται στο πεδίο των συχνοτήτων, αποκαλύπτοντας έτσι τις διάφορες συχνότητες που απαρτίζουν το σήμα.[24] Αυτό μπορεί να επιτρέψει τον εντοπισμό μοτίβων ή ανωμαλιών που μπορεί να είναι δύσκολο να ανιχνευθούν στο χρονικό πεδίο. Ειδικά όταν ασχολούμαστε με σήματα που περιέχουν θόρυβο ή παρασιτικές επιδράσεις, η ανάλυση φάσματος μπορεί να αποτελέσει αποτελεσματικό εργαλείο για τον εντοπισμό και την επισήμανση των ανωμαλιών.

Οι μέθοδοι ανάλυσης φάσματος μπορούν να εφαρμοστούν σε διάφορα είδη δεδομένων IoT, όπως αισθητήρες περιβάλλοντος, ηλεκτρονικά σήματα, ροές δεδομένων και πολλά άλλα.[30] Μέσω της ανάλυσης φάσματος, μπορούμε να αποκτήσουμε ενδείξεις για τις ενδεχόμενες ανωμαλίες που ενδέχεται να προκύπτουν από αυτά τα δεδομένα και να αναλάβουμε δράση για την αντιμετώπιση τους.

Σε συνδυασμό με άλλες τεχνικές όπως η μηχανική μάθηση, η ανάλυση φάσματος μπορεί να βοηθήσει στη βελτίωση της ακρίβειας και της αποτελεσματικότητας των συστημάτων ανίχνευσης ανωμαλιών στα δεδομένα IoT. Συνεπώς, αποτελεί σημαντική εργαλειοθήκη για την προστασία και τη βελτίωση της λειτουργίας των συστημάτων IoT.

2.3.4 Ανίχνευση Συμπεριφορικών Ανωμαλιών

Η ανίχνευση συμπεριφορικών ανωμαλιών αποτελεί μια σημαντική τεχνική για την ασφάλεια και την προστασία των συστημάτων IoT. Καθώς τα συστήματα IoT αποτελούνται από πληθώρα συνδεδεμένων συσκευών και αισθητήρων, είναι σημαντικό να αναγνωρίζουμε τυχόν ανωμαλίες στη συμπεριφορά τους που θα μπορούσαν να υποδεικνύουν επιθέσεις ή απρόσμενες καταστάσεις. [\[22\]\[30\]](#)

Η ανίχνευση συμπεριφορικών ανωμαλιών εστιάζει στην παρακολούθηση και ανάλυση των μοτίβων συμπεριφοράς των συσκευών και των χρηστών στο οικοσύστημα IoT. Αρχικά, καταγράφονται τα συνήθη μοντέλα συμπεριφοράς των διαφόρων συσκευών, και η ανάλυση επικεντρώνεται στον εντοπισμό αποκλίσεων από αυτά τα μοντέλα. Εάν μια συσκευή ή ένας χρήστης εμφανίσει συμπεριφορά που δεν συμφωνεί με τα προαποθεμένα μοντέλα ή παρουσιάσει ανεπιθύμητες δραστηριότητες, τότε μπορεί να υποπτευθεί ότι υπάρχει ανωμαλία ή πιθανή επίθεση. [\[29\]](#)

Για την ανίχνευση συμπεριφορικών ανωμαλιών, χρησιμοποιούνται ποικίλοι αλγόριθμοι μηχανικής μάθησης, όπως τα νευρωνικά δίκτυα, οι αλγόριθμοι clustering, οι αλγόριθμοι ανίχνευσης ανωμαλιών κ.ά. Επιπλέον, μπορούν να χρησιμοποιηθούν στατιστικές τεχνικές, όπως η ανάλυση φάσματος που αναφέρθηκε προηγουμένως, για την ανίχνευση συμπεριφορικών αποκλίσεων.

Η ανίχνευση συμπεριφορικών ανωμαλιών είναι κρίσιμη για τη διατήρηση της ασφάλειας και της ομαλής λειτουργίας των συστημάτων IoT. Με τη βοήθεια προηγμένων τεχνικών ανίχνευσης, μπορούμε να προλάβουμε πιθανούς κινδύνους, να αναλάβουμε έγκαιρες δράσεις και να εξασφαλίσουμε την αξιοπιστία και ασφάλεια των συστημάτων μας. [\[32\]\[33\]](#)

2.3.5 Ανάλυση Αποτύπωσης

Η ανάλυση αποτύπωσης (data fingerprinting) είναι μια τεχνική που επικεντρώνεται στη συλλογή και ανάλυση των δεδομένων που προέρχονται από τους αισθητήρες του IoT. [\[22\]](#) Καθώς οι αισθητήρες συλλέγουν δεδομένα από το περιβάλλον, όπως θερμοκρασία, υγρασία, φωτεινότητα και άλλες παραμέτρους, τα δεδομένα αποτυπώνουν τον τρόπο λειτουργίας και την κατάσταση του περιβάλλοντος.

Ο σκοπός της ανάλυσης αποτύπωσης είναι να αναγνωρίσει απρόβλεπτες μεταβολές και ανωμαλίες στα δεδομένα που συλλέγονται από τους αισθητήρες. Με αυτόν τον τρόπο, μπορεί να εντοπιστούν προβλήματα, ατυχήματα, αλλαγές στο περιβάλλον, ή ακόμα και επιθέσεις ασφαλείας. Η ανάλυση αποτύπωσης είναι χρήσιμη στον τομέα του IoT, καθώς οι αισθητήρες μπορεί να είναι ευάλωτοι σε διάφορα ζητήματα, όπως η εξάντληση της μπαταρίας, η κακή λειτουργία ή η παραβίαση.

Για να επιτευχθεί η ανάλυση αποτύπωσης, απαιτείται η χρήση αλγορίθμων και τεχνικών ανίχνευσης ανωμαλιών. Οι αλγόριθμοι αυτοί συγκρίνουν τις τρέχουσες μετρήσεις με τις προηγούμενες, εντοπίζοντας έτσι τις απότομες μεταβολές, τις ασυνήθιστες τιμές ή άλλα σημάδια ανωμαλίας. Η ανάλυση αποτύπωσης μπορεί να χρησιμοποιηθεί σε πολλούς τομείς, όπως η βιομηχανία, οι έξυπνες πόλεις, η γεωργία, η υγειονομική περίθαλψη, και άλλοι τομείς που χρησιμοποιούν συσκευές IoT και αισθητήρες. [\[30\]](#)

Συνοψίζοντας, η ανάλυση αποτύπωσης αποτελεί μια πολύτιμη τεχνική για την εντοπισμό ανωμαλιών στα δεδομένα IoT και την αντίδραση σε επιθέσεις, προβλήματα ή αλλαγές που

απαιτούν προσοχή ή παρέμβαση. Με τη χρήση αυτής της τεχνικής, μπορούν να εντοπιστούν άμεσα και να αντιμετωπιστούν τυχόν ζητήματα που επηρεάζουν τη λειτουργία και την ασφάλεια των συσκευών IoT και του περιβάλλοντος τους.

2.3.6 Ανίχνευση Παραμέτρων

Η ανίχνευση παραμέτρων αποτελεί μια προσέγγιση που επικεντρώνεται στον εντοπισμό ανωμαλιών με την ανάλυση των παραμέτρων που σχετίζονται με τα συστήματα Internet of Things (IoT). Στα συστήματα IoT, υπάρχουν πολλές παράμετροι που παρακολουθούνται, όπως η θερμοκρασία, η υγρασία, η πίεση, η ταχύτητα, κ.λπ., και οι τιμές τους συλλέγονται και αποθηκεύονται σε μεγάλους όγκους δεδομένων.

Η ανίχνευση παραμέτρων στοχεύει στο να μελετήσει τις αναμενόμενες τιμές των παραμέτρων, που μπορούν να προκύψουν από προηγούμενη έρευνα, συμπεριφορά ή πρότυπα. Κατόπιν, αναζητά αποκλίσεις από αυτές τις αναμενόμενες τιμές, που μπορεί να υποδεικνύουν ανωμαλίες ή ανεπιθύμητες καταστάσεις. Οι ανωμαλίες μπορεί να αντιστοιχούν σε μη αναμενόμενες διακυμάνσεις, εκτόνωση τιμών, αποκλίσεις από συνήθεις περιόδους, ή ανεπιθύμητες τιμές που υποδηλώνουν δυσλειτουργίες ή προβλήματα.

Για να υλοποιηθεί η ανίχνευση παραμέτρων, χρησιμοποιούνται συχνά στατιστικές τεχνικές, μηχανική μάθηση, και αλγόριθμοι επεξεργασίας σημάτων.[\[20\]](#) Επιπλέον, η ανίχνευση παραμέτρων μπορεί να εφαρμοστεί σε ποικίλους τομείς, όπως η βιομηχανία, η υγειονομική περίθαλψη, η ασφάλεια, και το περιβάλλον.

Η συνδυασμένη χρήση ανάλυσης φάσματος, μηχανικής μάθησης και ανίχνευσης παραμέτρων αποτελεί ισχυρή προσέγγιση για τον εντοπισμό ανωμαλιών στα δεδομένα IoT και την προστασία του συστήματος από ανεπιθύμητες καταστάσεις ή προβλήματα.[\[19\]](#) Σε έναν τόσο πολύπλοκο και δυναμικό κόσμο όπως το IoT, η ανίχνευση ανωμαλιών και η πρόληψή τους αποτελούν κρίσιμη πρόκληση, και η χρήση προηγμένων τεχνικών όπως η ανάλυση φάσματος αποδεικνύεται απαραίτητη για τη διατήρηση της ασφάλειας και της αποτελεσματικότητας των συστημάτων IoT.

2.3.7 Ανάλυση Συμπεριφοράς

Η ανάλυση συμπεριφοράς αποτελεί μια σημαντική τεχνική για την ανίχνευση ανωμαλιών στα δεδομένα των συσκευών IoT. Αντί να επικεντρώνεται αποκλειστικά στην ανάλυση των απομονωμένων χαρακτηριστικών των δεδομένων, όπως κάνει η συμβατική στατιστική ανάλυση, η ανάλυση συμπεριφοράς εστιάζει στην παρακολούθηση της συμπεριφοράς των συσκευών και των δεδομένων τους στον χρόνο.

Αυτή η τεχνική επιδιώκει να ανιχνεύσει μη συνήθεις δραστηριότητες, απρόβλεπτες αλλαγές και ανωμαλίες στις συμπεριφορές των συσκευών. Στόχος της είναι να εντοπίσει αυτόματα τυχόν μοτίβα ή συμπεριφορές που αποκλίνουν από το αναμενόμενο και να τις ταυτοποιήσει ως πιθανές ανωμαλίες.

Για την επίτευξη αυτού του στόχου, η ανάλυση συμπεριφοράς χρησιμοποιεί μεθόδους μηχανικής μάθησης, όπως οι νευρωνικοί δίκτυα και οι αλγόριθμοι ομαδοποίησης, που επιτρέπουν την αυτόματη αναγνώριση παρατηρούμενων προτύπων. Επίσης, χρησιμοποιούνται τεχνικές όπως η ανάλυση ακολουθιών (sequence analysis) και οι χρονοσειρές (time series) για την ανίχνευση ανωμαλιών που προκύπτουν με την πάροδο του χρόνου.[\[26\]](#)

Η ανάλυση συμπεριφοράς είναι κρίσιμης σημασίας για την ασφάλεια και τη διαχείριση των δεδομένων IoT, καθώς μπορεί να εντοπίσει εγκαίρως ανωμαλίες ή καινούργιες απειλές που μπορεί να παρουσιάζουν οι συνδεδεμένες συσκευές. Επιπλέον, μπορεί να βοηθήσει στην πρόβλεψη προβλημάτων και στη λήψη μέτρων πρόληψης για την αποτροπή μελλοντικών ανωμαλιών.

2.3.8 Ανάλυση Δικτύου

Η ανάλυση δικτύου αποτελεί μια σημαντική προσέγγιση για την ανίχνευση ανωμαλιών στα δεδομένα του IoT. Καθώς οι συσκευές του IoT συνδέονται μεταξύ τους και ανταλλάσσουν δεδομένα, η παρατήρηση και η ανάλυση της κίνησης και της επικοινωνίας μπορούν να αποκαλύψουν σημαντικές πληροφορίες.

Η ανάλυση δικτύου παρακολουθεί τις μεταβολές της ροής των δεδομένων και εξετάζει τα μοτίβα κίνησης μεταξύ των διαφόρων συσκευών του IoT. Σε ένα κανονικό δίκτυο, οι συσκευές επικοινωνούν με συγκεκριμένα μοτίβα και συχνότητες. Η ανίχνευση ανωμαλιών βασίζεται στον εντοπισμό απρόβλεπτων αλλαγών στον τρόπο λειτουργίας του δικτύου. Μερικά παραδείγματα ανωμαλιών που μπορεί να ανιχνεύσει η ανάλυση δικτύου περιλαμβάνουν:

- Διακοπές συνδεσιμότητας
Αναγνωρίζει περιοχές όπου η συνδεσιμότητα μεταξύ συσκευών διακόπτεται ξαφνικά ή παρουσιάζει ασυνήθιστες διακυμάνσεις.[\[21\]](#)
- Μη εξουσιοδοτημένη πρόσβαση
Εντοπίζει τυχόν προσπάθειες μη εξουσιοδοτημένης πρόσβασης στο δίκτυο, όπως κακόβουλα προσπάθεια εισβολής.[\[34\]](#)
- Ασυνήθιστα μοτίβα επικοινωνίας
Αναγνωρίζει ανατυπώσεις στον τρόπο επικοινωνίας μεταξύ συσκευών που μπορεί να υποδεικνύουν παρείσακτες δραστηριότητες.[\[19\]](#)

Με την ανάλυση δικτύου, μπορούμε να ανιχνεύσουμε τις παραπάνω ανωμαλίες και να προλάβουμε δυνητικά προβλήματα ή απειλές στην ασφάλεια του δικτύου του IoT. Συνεπώς, η ανάλυση δικτύου αποτελεί μια σημαντική τεχνική για την επιτήρηση και την προστασία των δεδομένων και των συσκευών σε ένα περιβάλλον του IoT.

Οι παραπάνω τεχνικές ανάλυσης ανωμαλιών παρέχουν ισχυρά εργαλεία για την αναγνώριση και τον εντοπισμό ανωμαλιών στα συστήματα IoT. Μπορούν να χρησιμοποιηθούν μαζί ή ξεχωριστά για τον εντοπισμό και την αντιμετώπιση ανωμαλιών στα συστήματα IoT, ενισχύοντας την ασφάλεια, την αξιοπιστία και την απόδοση των συστημάτων. Η επιτυχημένη εφαρμογή αυτών των τεχνικών μπορεί να συμβάλει στην ενίσχυση της ασφάλειας, της αξιοπιστίας και της απόδοσης των συστημάτων IoT.

2.4 Εισαγωγή στα Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα αποτελούν τον βασικότερο πυλώνα της ανίχνευσης και ανάλυσης ανωμαλιών. Είναι μια ισχυρή κατηγορία αλγορίθμων μηχανικής μάθησης που εμπνέονται από την οργάνωση και τη λειτουργία του ανθρώπινου εγκεφάλου. Από τις πρώτες προσπάθειες που πραγματοποιήθηκαν πριν από δεκαετίες, τα νευρωνικά δίκτυα έχουν γνωρίσει ταχεία ανάπτυξη και αναγνώριση, έχοντας προσφέρει εντυπωσιακές επιδόσεις σε πολλούς τομείς, από την εικονική νοημοσύνη έως τη φυσική γλώσσα και την ρομποτική. Στην παρούσα ενότητα, θα παρουσιάσουμε μια λεπτομερή επισκόπηση των νευρωνικών δικτύων, από τη δομή τους έως τις εφαρμογές τους, καθώς και μια επισκόπηση των πιο σημαντικών μελετών και ερευνών που συνέβαλαν στην εξέλιξή τους.

2.4.1 Δομή των Νευρωνικών Δικτύων

Ένα νευρωνικό δίκτυο αποτελείται από διάφορα επίπεδα νευρώνων, τα οποία συνδέονται μεταξύ τους με βάση τις διαφορετικές συνδέσεις. Η δομή αυτή έχει υπερβολικά πολλά κοινά με τον τρόπο που λειτουργούν οι νευρώνες του ανθρώπινου εγκεφάλου, και αυτός είναι ο λόγος για τον οποίο τα νευρωνικά δίκτυα είναι εμπνευσμένα από τη βιολογία.[\[27\]\[35\]](#)

Κάθε επίπεδο του νευρωνικού δικτύου αποτελείται από έναν αριθμό νευρώνων. Κάθε νευρώνας σε ένα στρώμα συνδέεται με τους νευρώνες του προηγούμενου και του επόμενου στρώματος. Οι συνδέσεις αντιπροσωπεύονται από τα βάρη, τα οποία καθορίζουν τον βαθμό επιρροής του νευρώνα στον επόμενο.[\[36\]](#) Κάθε νευρώνας συνδέεται πλήρως με τους νευρώνες του επόμενου στρώματος, όμως υπάρχουν και διαφορετικές αρχιτεκτονικές, όπως οι αραιές συνδέσεις στα δίκτυα υπολογισμού δεξαμενής (reservoir computing).[\[37\]](#)

Το πρώτο στρώμα του νευρωνικού δικτύου αποτελεί την είσοδο του δικτύου, όπου οι αρχικές πληροφορίες εισέρχονται. Μετά την επεξεργασία των δεδομένων από τα ενδιάμεσα στρώματα, η έξοδος του δικτύου προκύπτει από το τελευταίο στρώμα νευρώνων. Η διαδικασία αυτή περιγράφεται συνοπτικά ως "προώθηση" (forward pass) και αποτελεί τον τρόπο με τον οποίο το δίκτυο εξάγει αποτελέσματα από τις εισόδους του.[\[38\]](#)

Η δομή αυτή, όπως και η λειτουργία του κάθε νευρώνα, επαναλαμβάνεται και προσαρμόζεται πολλές φορές κατά τη διάρκεια της εκπαίδευσης του δικτύου. Κατά τη διαδικασία αυτή, τα βάρη των συνδέσεων ρυθμίζονται κατάλληλα ώστε το δίκτυο να μπορεί να ανταποκριθεί αποτελεσματικά στις εισόδους και να προσεγγίσει την επιθυμητή συμπεριφορά.[\[39\]](#)

Η δομή των νευρωνικών δικτύων είναι κρίσιμη για την επίλυση πολύπλοκων προβλημάτων και την εξαγωγή χρήσιμων πληροφοριών από τα δεδομένα. Καθώς η έρευνα προχωράει, αναμένουμε να δούμε ακόμα πιο προηγμένες αρχιτεκτονικές που θα ενισχύουν την απόδοση των νευρωνικών δικτύων σε διάφορους τομείς.

2.4.2 Τύποι Νευρωνικών Δικτύων

2.4.2.1 Πολυεπίπεδο Πλήρως Συνδεδεμένο Δίκτυο (Fully Connected Neural Network):

Το πολυεπίπεδο πλήρως συνδεδεμένο δίκτυο είναι ένας από τους πιο βασικούς και απλούς τύπους νευρωνικών δικτύων. Σε αυτό το μοντέλο, κάθε νευρώνας σε ένα στρώμα συνδέεται με κάθε νευρώνα στο προηγούμενο και επόμενο στρώμα. Η λειτουργία του περιγράφεται από μια

σειρά παραμέτρων, γνωστών ως βάρδιες, οι οποίες ρυθμίζουν τον βαθμό επιρροής κάθε νευρώνα στους γειτονικούς.[36]

Κάθε στρώμα αποτελείται από έναν αριθμό νευρώνων, και κάθε νευρώνας συνδέεται με κάθε νευρώνα στο προηγούμενο και επόμενο στρώμα. Όταν τα δεδομένα εισέρχονται στο δίκτυο, οι τιμές προς τα εμπρός υπολογίζονται μέσω της διαδικασίας της προώθησης (forward pass).[38] Συγκεκριμένα, κάθε νευρώνας υπολογίζει την εξαγόμενη τιμή του ως συνάρτηση των εισερχόμενων δεδομένων και των βαρών που ρυθμίζουν τις συνδέσεις του με τους γειτονικούς νευρώνες.

Ένα βασικό μειονέκτημα του πολυεπίπεδου πλήρως συνδεδεμένου δικτύου είναι ο μεγάλος αριθμός παραμέτρων που απαιτείται για τη λειτουργία του, ειδικά όταν το δίκτυο γίνεται βαθύτερο με περισσότερα ενδιάμεσα στρώματα. Η αύξηση των παραμέτρων καθιστά τη διαδικασία εκπαίδευσης πιο χρονοβόρα και αυξάνει τον κίνδυνο υπερεκπαίδευσης, όπου το δίκτυο προσαρμόζεται υπερβολικά στα δεδομένα εκπαίδευσης και αποτυγχάνει να γενικεύσει σε νέα δεδομένα.

Παρ' όλα αυτά, το πολυεπίπεδο πλήρως συνδεδεμένο δίκτυο είναι αποτελεσματικό για απλά προβλήματα και είναι σε θέση να αντιμετωπίσει αρκετές εφαρμογές, ειδικά εκεί όπου οι συσχετίσεις μεταξύ των χαρακτηριστικών είναι απλές. Ωστόσο, για πιο πολύπλοκα προβλήματα, όπως η επεξεργασία εικόνων και η φυσική γλώσσα, η χρήση πιο εξελιγμένων αρχιτεκτονικών, όπως τα συνελκτικά νευρωνικά δίκτυα και τα αναδρομικά νευρωνικά δίκτυα, είναι πιο προτιμητέα για να επιτύχουμε ανώτερες επιδόσεις.

2.4.2.2 Συνελκτικό Νευρωνικό Δίκτυο (Convolutional Neural Network - CNN):

Τα Συνελκτικά Νευρωνικά Δίκτυα (CNNs) αντιπροσωπεύουν μια εξαιρετικά ισχυρή κατηγορία των νευρωνικών δικτύων που έχουν επιτύχει αναπάντεχες επιδόσεις σε εφαρμογές επεξεργασίας εικόνας και βίντεο.

Τα CNNs έχουν επανασχεδιάσει το τοπίο της τεχνητής όρασης, πραγματοποιώντας επιτυχημένες λύσεις σε πολλές προκλήσεις, όπως η αναγνώριση προσώπων, τον αυτόματο εντοπισμό αντικειμένων, την αυτόματη κατηγοριοποίηση εικόνων και πολλές άλλες επεξεργασίες εικόνας.[40]

Το κλειδί για την αποτελεσματικότητα των CNNs βρίσκεται στην αρχιτεκτονική τους, η οποία μιμείται τον τρόπο με τον οποίο λειτουργεί το ανθρώπινο οπτικό σύστημα. Είναι εμπνευσμένα από τον τρόπο με τον οποίο ο ανθρώπινος εγκέφαλος επεξεργάζεται τις εικόνες, ανιχνεύοντας και εστιάζοντας σε συγκεκριμένα χαρακτηριστικά για την αναγνώριση αντικειμένων.

Σε ένα CNN, οι εικόνες εισέρχονται στο δίκτυο ως είσοδος και περνούν από συνελκτικά φίλτρα (convolutional filters) ή πυρήνες. Αυτά τα φίλτρα αποτελούνται από μικρές περιοχές που κινούνται στην εικόνα για να εξάγουν τα χαρακτηριστικά της. Κατά τη διάρκεια της συνέλιξης, το φίλτρο προσαρμόζεται για να ανιχνεύσει συγκεκριμένα χαρακτηριστικά, όπως ακμές, γωνίες, χρώματα, περιγράμματα και άλλα.

Η συνέλιξη επιτρέπει στο CNN να μάθει να αναγνωρίζει συγκεκριμένα χαρακτηριστικά σε οποιαδήποτε θέση της εικόνας, καθιστώντας το ιδανικό για εικόνες με διαφορετικές θέσεις, μεγέθη και προοπτικές. Επίσης, η χρήση πολλαπλών φίλτρων σε διαφορετικά στρώματα επιτρέπει στο CNN να μάθει ιεραρχικές αναπαραστάσεις της εικόνας, με αυξανόμενη συνθετότητα των χαρακτηριστικών.[41]

Μετά την συνέλιξη, τα χαρακτηριστικά περνούν από τα επίπεδα συγκέντρωσης (pooling layers) όπου γίνεται μείωση των διαστάσεων των χαρακτηριστικών για να μειωθεί ο όγκος των

δεδομένων και να γίνει η παραλληλοποίηση της επεξεργασίας. Οι περισσότεροι αλγόριθμοι συγκέντρωσης επιλέγουν το μέγιστο ή το μέσον των τιμών ενός πλαισίου χαρακτηριστικών.

Μετά τα επίπεδα συγκέντρωσης, τα χαρακτηριστικά περνούν από πλήρως συνδεδεμένα επίπεδα (fully connected layers), τα οποία λειτουργούν όπως τα παραδοσιακά νευρωνικά δίκτυα και είναι υπεύθυνα για την τελική κατηγοριοποίηση και επεξεργασία των χαρακτηριστικών.[\[42\]](#) Το τελευταίο επίπεδο του CNN παράγει την έξοδο του δικτύου, όπου μπορεί να είναι μια πρόβλεψη, μια κατηγοριοποίηση ή ένας πίνακας πιθανοτήτων.

Είναι σημαντικό να σημειωθεί ότι τα CNNs μπορούν να εκπαιδευτούν με τη χρήση μεγάλων συλλογών δεδομένων και προηγμένων μεθόδων βελτιστοποίησης. Η διαδικασία εκπαίδευσης περιλαμβάνει την ρύθμιση των βαρών των φίλτρων και των παραμέτρων των επιπέδων, ώστε το CNN να μπορεί να αναγνωρίζει και να διαχωρίζει τις εικόνες της εκπαίδευσης σωστά.

Συνοψίζοντας, τα συνελκτικά νευρωνικά δίκτυα αποτελούν έναν από τους πιο ισχυρούς και αποτελεσματικούς τύπους νευρωνικών δικτύων για την επεξεργασία εικόνων και βίντεο. Η ιδέα των φίλτρων συνέλιξης, που είναι ικανά να ανιχνεύουν συγκεκριμένα χαρακτηριστικά στην εικόνα, κάνει τα CNNs ιδανικά για την αναγνώριση αντικειμένων, την ανίχνευση ανωμαλιών, την εξαγωγή χαρακτηριστικών και πολλές άλλες εργασίες επεξεργασίας εικόνας. Η συνεχής ανάπτυξη της τεχνολογίας CNNs υπόσχεται να έχει σημαντική επίδραση σε πολλούς τομείς, όπως η ρομποτική, η ιατρική εικονικής πραγματικότητας και η αυτόνομη οδήγηση.

2.4.2.3 Αναδρομικό Νευρωνικό Δίκτυο (Recurrent Neural Network - RNN):

Τα αναδρομικά νευρωνικά δίκτυα (RNNs) είναι μια ειδική κατηγορία νευρωνικών δικτύων που έχουν τη δυνατότητα να επεξεργαστούν ακολουθίες δεδομένων, όπως χρονοσειρές, μουσικά κομμάτια, φυσική γλώσσα, αλλά και άλλα είδη διατεταγμένων δεδομένων. Το κύριο χαρακτηριστικό που ξεχωρίζει τα RNNs από άλλους τύπους νευρωνικών δικτύων είναι η ικανότητά τους να διατηρούν εσωτερική κατάσταση ή μνήμη, η οποία μπορεί να ανανεώνεται και να επηρεάζεται καθ' όλη τη διάρκεια της ανάλυσης της ακολουθίας.[\[43\]](#)

Η αρχιτεκτονική των αναδρομικών νευρωνικών δικτύων επιτρέπει την επεξεργασία διαδοχικών στοιχείων μιας ακολουθίας, λαμβάνοντας υπόψη τις πληροφορίες που προήλθαν από τα προηγούμενα στοιχεία. Αυτό το χαρακτηριστικό καθιστά τα RNNs ιδανικά για καθήκοντα που απαιτούν κατανόηση της σειράς και της συσχέτισης μεταξύ των στοιχείων, όπως η μετάφραση φυσικής γλώσσας, η πρόβλεψη χρονοσειρών ή η ανάλυση κειμένων.[\[44\]](#)

Ένα RNN αποτελείται από μια αλυσίδα νευρώνων που συνδέονται μεταξύ τους. Κάθε νευρώνας λαμβάνει ως είσοδο την τρέχουσα τιμή της ακολουθίας και την κρυφή κατάσταση ή τη μνήμη από τον προηγούμενο χρόνο. Στη συνέχεια, αποφασίζει πώς θα ενημερώσει την κρυφή κατάστασή του για τον επόμενο χρόνο και πώς θα παράξει την έξοδο του. Η διαδικασία αυτή επαναλαμβάνεται για κάθε στοιχείο της ακολουθίας, δημιουργώντας ένα είδος αυτοαναφοράς που καθιστά το νευρωνικό δίκτυο ικανό να κρατά τη μνήμη της πληροφορίας που είναι χρήσιμη για την ανάλυση της συνολικής ακολουθίας.[\[45\]\[42\]](#)

Η δυνατότητα αναδρομής παρέχει στα αναδρομικά νευρωνικά δίκτυα την ικανότητα να αντιμετωπίζουν προβλήματα που απαιτούν μακροπρόθεσμες εξαρτήσεις και προβλέψεις.[\[43\]](#) Για παράδειγμα, στην πρόβλεψη χρονοσειρών, ένα RNN μπορεί να ανατρέξει στα παρελθόντα δεδομένα για να προβλέψει την επόμενη τιμή, λαμβάνοντας υπόψη τις προηγούμενες τιμές και μοτίβα. Αυτό το κύκλωμα αναδρομής μνήμης καθιστά τα RNNs πολύ ισχυρά για προβλήματα που απαιτούν διακριτή συμπεριφορά, όπως στη φυσική γλώσσα, όπου η σειρά των λέξεων είναι καθοριστική για την κατανόηση του νοήματος.

Παρόλα αυτά, οι παραδοσιακοί RNNs έχουν κάποια μειονεκτήματα. Ένα από τα κύρια προβλήματα είναι η προκαθορισμένη μνήμη που μπορούν να διατηρήσουν, καθώς η αναδρομή επεκτείνεται μεγάλο χρονικό διάστημα. Αυτό οδηγεί σε φαινόμενο που ονομάζεται "ξέχασμα των μακροπρόθεσμων εξαρτήσεων", όπου πληροφορίες από πολύ παλιά στοιχεία της ακολουθίας δεν μπορούν να επηρεάσουν σημαντικά τις τρέχουσες προβλέψεις. Επιπλέον, το πρόβλημα της εξαφάνισης ή εκρηκτικής διασποράς των κλίσεων μπορεί να προκαλέσει προβλήματα στην εκπαίδευση του δικτύου.

Για να αντιμετωπιστούν αυτά τα προβλήματα, έχουν αναπτυχθεί βελτιωμένες αρχιτεκτονικές RNN, όπως οι Long Short-Term Memory (LSTM) και Gated Recurrent Units (GRU). Αυτές οι αρχιτεκτονικές επιτρέπουν τον πιο αποτελεσματικό έλεγχο της ροής των πληροφοριών μέσα στο δίκτυο, επιτρέποντας την αποτελεσματική μάθηση μακροπρόθεσμων εξαρτήσεων και προβλέψεων.[\[44\]](#)

Συνοψίζοντας, τα αναδρομικά νευρωνικά δίκτυα είναι ισχυρά εργαλεία για την ανάλυση ακολουθιών δεδομένων, λόγω της ικανότητάς τους να διατηρούν μνήμη και να αντιμετωπίζουν μακροπρόθεσμες εξαρτήσεις. Με την εξέλιξη της έρευνας, οι βελτιωμένες αρχιτεκτονικές RNN έχουν βελτιώσει τις επιδόσεις τους σε πολλούς τομείς, καθιστώντας τα ακόμα πιο ικανά για απαιτητικά προβλήματα ανάλυσης ακολουθιών.

2.4.2.4 Ευρεία Πολυεπίπεδα Νευρωνικά Δίκτυα (Deep Neural Networks - DNN):

Τα Ευρεία Πολυεπίπεδα Νευρωνικά Δίκτυα (DNNs) αποτελούν μια εξέλιξη του πολυεπίπεδου πλήρως συνδεδεμένου δικτύου και αποτελούν τη βασική αρχιτεκτονική που βρίσκεται πίσω από την έννοια της "βαθιάς μάθησης" (deep learning). Το βασικό χαρακτηριστικό αυτών των δικτύων είναι ότι διαθέτουν πολλά ενδιάμεσα στρώματα (ή κρυφά στρώματα) ανάμεσα στο εισαγωγικό και το εξαγωγικό στρώμα.[\[27\]](#)[\[35\]](#)

Η προσθήκη αυτών των ενδιάμεσων στρωμάτων επιτρέπει στα DNN να δημιουργούν πιο σύνθετα και αφηρημένα μοντέλα που μπορούν να αντιπροσωπεύουν δομικά πιο πολύπλοκες συσχετίσεις μεταξύ των δεδομένων.[\[32\]](#) Αυτό τους επιτρέπει να μάθουν πιο ανεξάρτητα και χρήσιμα χαρακτηριστικά προγραμματισμού από τα δεδομένα και να εξάγουν βαθύτερη και πιο σύνθετη κατανόηση.

Η επιπλέον δυνατότητα των DNN να εκπαιδεύουν σύνθετα μοντέλα σημαίνει ότι μπορούν να προσαρμοστούν σε πολύ πιο απαιτητικά προβλήματα. Αντί να εξαρτώνται από προκαθορισμένες λειτουργίες και χαρακτηριστικά, τα DNN μπορούν να μάθουν από τα δεδομένα τις αναγκαίες αναπαραστάσεις για την επίλυση των προβλημάτων.[\[46\]](#) Τα DNN έχουν εξαιρετικές επιδόσεις σε πολύπλοκες εφαρμογές, όπως στην αναγνώριση αντικειμένων και προσώπων,[\[40\]](#) στην επεξεργασία ήχου[\[41\]](#) και την επεξεργασία φυσικής γλώσσας.[\[47\]](#)

Παρόλα αυτά, τα DNN χρειάζονται μεγάλο όγκο δεδομένων και υψηλές υπολογιστικές δυνατότητες για την εκπαίδευση, ενώ η πλήρης εκπαίδευσή τους μπορεί να απαιτήσει πολύ χρόνο.[\[48\]](#) Επίσης, η ανάπτυξη των δικτύων και η ρύθμιση των υπερπαραμέτρων μπορεί να είναι προκλητική.[\[33\]](#) Όμως, η συνεχής έρευνα και η προοδευτική βελτίωση των DNN αναμένεται να συμβάλει στην ανάπτυξη εκπληκτικών εφαρμογών σε περαιτέρω πεδία, επιτρέποντάς μας να επιλύουμε πιο πολύπλοκα προβλήματα και να κατανοήσουμε καλύτερα τον κόσμο γύρω μας.[\[49\]](#)

2.4.2.5 Δίκτυα Υπολογισμού Δεξαμενής (Reservoir Computing):

Τα δίκτυα υπολογισμού δεξαμενής (Reservoir Computing) ανήκουν στην κατηγορία των αναδρομικών νευρωνικών δικτύων και αποτελούν μια εξειδικευμένη αρχιτεκτονική με διαστήματα μνήμης.[42] Αυτή η προσέγγιση διαφέρει από την παραδοσιακή εκπαίδευση νευρωνικών δικτύων, όπου οι όροι της εκπαίδευσης αλλάζουν τις συνδέσεις μεταξύ των νευρώνων κατά την εκπαίδευση.[43] Στα δίκτυα υπολογισμού δεξαμενής, το σύνολο των νευρώνων, γνωστό ως "δεξαμενή", παραμένει σταθερό κατά τη διάρκεια της εκπαίδευσης και δεν υπόκειται σε αλλαγές. Αντ' αυτού, οι αλλαγές γίνονται μόνο στις συνδέσεις μεταξύ της δεξαμενής και της γραμμικής επιλογής εξόδου.[37]

Η αρχιτεκτονική των δικτύων υπολογισμού δεξαμενής αποτελείται από τρία βασικά στρώματα: το είσοδο, τη δεξαμενή και τη γραμμική επιλογή εξόδου. Το στρώμα εισόδου λαμβάνει τα δεδομένα εισόδου και τα μετατρέπει σε έναν διανυσματικό χώρο. Στη συνέχεια, τα δεδομένα εισέρχονται στην δεξαμενή, η οποία αποτελείται από έναν μεγάλο αριθμό νευρώνων. Κάθε νευρώνας της δεξαμενής είναι πλήρως συνδεδεμένος με τους γειτονικούς του. Αυτή η δομή δημιουργεί έναν πλούσιο χώρο καταστάσεων, που παρέχει την ικανότητα του δικτύου να αποθηκεύει και να αναπαριστά με πληροφορίες.[50]

Η σημαντικότερη ιδιότητα των δικτύων υπολογισμού δεξαμενής είναι η εκπληκτική τους ικανότητα να παραμείνουν αμετάβλητα κατά τη διάρκεια της εκπαίδευσης. Αυτό σημαίνει ότι η δεξαμενή μπορεί να ενεργεί ως ένας πλούσιος και σταθερός χώρος χαρακτηριστικών, αποτελώντας ένα είδος μνήμης.[50] Κατά τη διάρκεια της εκπαίδευσης, οι συνδέσεις ανάμεσα στη δεξαμενή και τη γραμμική επιλογή εξόδου ρυθμίζονται, συνήθως χρησιμοποιώντας απλούστατους αλγόριθμους προσαρμογής. Αυτή η απλή και αποτελεσματική διαδικασία εκπαίδευσης καθιστά τα δίκτυα υπολογισμού δεξαμενής πιο εύκολα στην υλοποίηση από πολύ βαθύτερα νευρωνικά δίκτυα.

Η ιδιότητα αυτή καθιστά τα δίκτυα υπολογισμού δεξαμενής ιδανικά για εφαρμογές που απαιτούν αναγνώριση χρονοσειρών και πρόβλεψη, καθώς η αντιμετώπιση δυναμικών δεδομένων αποτελεί μια πρόκληση για τα παραδοσιακά νευρωνικά δίκτυα. Επιπλέον, τα δίκτυα υπολογισμού δεξαμενής είναι ανθεκτικά σε διάφορα προβλήματα, όπως το πρόβλημα της εξαφάνισης του γραφήματος κατά την εκπαίδευση, που μπορεί να παρουσιαστεί σε πιο πολύπλοκες αρχιτεκτονικές νευρωνικών δικτύων.[48]

Συνοψίζοντας, τα δίκτυα υπολογισμού δεξαμενής αποτελούν μια εξειδικευμένη και ισχυρή κατηγορία νευρωνικών δικτύων που έχουν εφαρμογές σε πολλούς τομείς, ιδιαίτερα στην αναγνώριση και πρόβλεψη χρονοσειρών. Η απλότητα της εκπαίδευσης και η ικανότητά τους να παραμένουν ανθεκτικά τους καθιστούν ένα πολύτιμο εργαλείο για την επίλυση πολύπλοκων προβλημάτων στον τομέα της μηχανικής μάθησης και της τεχνητής νοημοσύνης.

2.4.3 Εφαρμογές των Νευρωνικών Δικτύων

Οι εφαρμογές των νευρωνικών δικτύων έχουν επεκταθεί σε πολλούς τομείς και έχουν επιφέρει σημαντικές επιδόσεις και καινοτομίες σε κάθε έναν από αυτούς. Ας αναλύσουμε λεπτομερώς τις εφαρμογές των νευρωνικών δικτύων σε κάθε τομέα:

2.4.3.1 Εικόνα και Βίντεο

- Αναγνώριση αντικειμένων

Τα νευρωνικά δίκτυα χρησιμοποιούνται ευρέως σε συστήματα αναγνώρισης αντικειμένων. Μπορούν να αναγνωρίσουν αντικείμενα και περιγραφές τους σε εικόνες και

βίντεο, καθιστώντας τα ιδανικά για εφαρμογές αυτόματης χαρακτηριστικής και ασφάλειας.[\[31\]](#)

- Ανίχνευση προσώπων

Οι αλγόριθμοι ανίχνευσης προσώπων βασιζόμενοι σε νευρωνικά δίκτυα έχουν αυξήσει σημαντικά την ακρίβεια της ανίχνευσης και την ταχύτητα επεξεργασίας. Αυτό έχει εφαρμογές στην ασφάλεια, την αυτοματοποίηση και την παρακολούθηση.[\[40\]](#)

- Αυτόματη όραση

Τα νευρωνικά δίκτυα επιτρέπουν την ανάπτυξη συστημάτων αυτόματης όρασης που μπορούν να επεξεργάζονται εικόνες και βίντεο, εξάγοντας πληροφορίες όπως το περιεχόμενο, η κίνηση και οι χρήστες.[\[41\]](#)

- Επεξεργασία εικόνας

Τα νευρωνικά δίκτυα εφαρμόζονται σε εφαρμογές επεξεργασίας εικόνας, όπως βελτίωση της ανάλυσης, απομάκρυνση θορύβου, εξαγωγή χαρακτηριστικών και δημιουργία φωτορεαλιστικών εικόνων.[\[51\]\[39\]](#)

2.4.3.2 Φυσική Γλώσσα

- Μηχανική μετάφραση

Τα νευρωνικά δίκτυα έχουν επιφέρει μεγάλες βελτιώσεις στην μηχανική μετάφραση. Επιτρέπουν τη μετάφραση μεγάλων κειμένων με περισσότερη ακρίβεια και φυσικότητα.[\[38\]\[45\]](#)

- Αυτόματη συνομιλία

Τα νευρωνικά δίκτυα χρησιμοποιούνται για την ανάπτυξη συστημάτων αυτόματης συνομιλίας και chatbots. Αυτά τα συστήματα μπορούν να ανταποκριθούν σε φυσική γλώσσα, παρέχοντας απαντήσεις και υποστηρίζοντας τους χρήστες.[\[50\]\[47\]](#)

- Ανάλυση συναισθήματος

Τα νευρωνικά δίκτυα μπορούν να αναγνωρίσουν και να αναλύσουν τα συναισθήματα που περιέχονται σε κείμενα, όπως θυμός, χαρά, λύπη κλπ. Αυτό έχει εφαρμογές σε κοινωνικά δίκτυα, παρακολούθηση αντιδράσεων πελατών κ.ά.[\[35\]\[51\]](#)

- Δημιουργία κειμένου

Τα νευρωνικά δίκτυα εφαρμόζονται σε εφαρμογές δημιουργίας κειμένου, όπως αυτόματη σύνθεση μουσικής, συνθετική γραφή κειμένων και παραγωγή περιεχομένου για ιστοσελίδες.[\[41\]\[42\]](#)

2.4.3.3 Ρομποτική

- Αυτόνομη οδήγηση

Τα αυτόνομα οχήματα εκμεταλλεύονται νευρωνικά δίκτυα για την αναγνώριση των περιβαλλοντικών συνθηκών και τον έλεγχο των κινήσεών τους, προσφέροντας ασφαλή και αποτελεσματική οδήγηση.[\[40\]\[49\]](#)

- Ρομπότ μάθησης

Τα νευρωνικά δίκτυα χρησιμοποιούνται σε ρομποτικά συστήματα μάθησης, τα οποία μπορούν να αποκτούν γνώση και εξελίσσονται με την εμπειρία τους. [\[50\]\[42\]](#)

- Ρομποτικά συστήματα εκτέλεσης

Τα νευρωνικά δίκτυα εφαρμόζονται σε ρομποτικά συστήματα εκτέλεσης, όπως στην αυτοματοποίηση βιομηχανικών διεργασιών και την ρομποτική χειρουργική. [\[48\]\[39\]](#)

Συμπερασματικά, τα νευρωνικά δίκτυα έχουν επανασχεδιάσει τον τρόπο που αντιλαμβανόμαστε και αντιμετωπίζουμε πολλά προβλήματα. Οι εφαρμογές τους στην εικόνα, τη φυσική γλώσσα και τη ρομποτική είναι μόνο μερικές από τις πολλές που διαμορφώνουν τη σύγχρονη τεχνολογία και την καθημερινή ζωή μας. Οι προοπτικές της εξέλιξής τους είναι ενθαρρυντικές και αναμένουμε ότι θα συνεχίσουν να προωθούν τα όρια της τεχνητής νοημοσύνης και της μηχανικής μάθησης.

Αποτελούν μια σημαντική πρόοδο στον τομέα της τεχνητής νοημοσύνης και της μηχανικής μάθησης. Με την εφαρμογή διαφόρων τύπων νευρωνικών δικτύων, μπορούμε να λύσουμε πολύπλοκα προβλήματα και να επιτύχουμε εξαιρετικές επιδόσεις σε πολλούς τομείς. Η εισαγωγή των ευρέων πολυεπίπεδων νευρωνικών δικτύων (DNN) έχει επιτρέψει την επίλυση ακόμα πιο περίπλοκων προβλημάτων και την εξαγωγή βαθύτερης κατανόησης από τα δεδομένα. Επιπλέον, τα δίκτυα υπολογισμού δεξαμενής (reservoir computing) παρέχουν μια ιδιαίτερα αποτελεσματική προσέγγιση για την επεξεργασία χρονοσειρών και την πρόβλεψη ανωμαλιών.

Καθώς η επιστήμη συνεχίζει να προχωράει, αναμένουμε να δούμε ακόμα περισσότερες καινοτομίες και εφαρμογές των νευρωνικών δικτύων στο μέλλον, επιτρέποντάς μας να αξιοποιήσουμε πλήρως τη δυνατότητά τους στο να μοντελοποιούν τα πολύπλοκα και αφαιρετικά στοιχεία της ανθρώπινης νοημοσύνης.

Τα νευρωνικά δίκτυα κατέχουν καθοριστικό ρόλο στην ανίχνευση ανωμαλιών, έναν τομέα υψίστης σημασίας σε διάφορους τομείς, όπως η ασφάλεια στον κυβερνοχώρο, ο βιομηχανικός έλεγχος ποιότητας και η υγειονομική περίθαλψη. Η σύνδεση μεταξύ των νευρωνικών δικτύων και της ανίχνευσης ανωμαλιών είναι σημαντική επειδή τα νευρωνικά δίκτυα μπορούν να μάθουν με αποτελεσματικότητα πολύπλοκα μοτίβα εντός των δεδομένων, καθιστώντας τα κατάλληλα για τον εντοπισμό ανωμαλιών ή αποκλίσεων από την αναμενόμενη συμπεριφορά. Εκπαιδεύοντας τα νευρωνικά δίκτυα σε μεγάλα σύνολα δεδομένων, μπορούν να συλλάβουν κανονικά μοτίβα και στη συνέχεια να επισημάνουν ασυνήθιστες ή ανώμαλες περιπτώσεις. Αυτή η ικανότητα είναι ανεκτίμητη για τη διασφάλιση κρίσιμων συστημάτων, τη διασφάλιση της ποιότητας των προϊόντων και την ανίχνευση μη φυσιολογικών ιατρικών καταστάσεων, μεταξύ άλλων εφαρμογών. Η συνεργασία μεταξύ των νευρωνικών δικτύων και της ανίχνευσης ανωμαλιών όχι μόνο ενισχύει την ικανότητά μας να αυτοματοποιούμε τον εντοπισμό παρατυπιών, αλλά δίνει επίσης τη δυνατότητα σε βιομηχανίες και ιδρύματα να αντιμετωπίζουν και να μετριάσουν προληπτικά πιθανά ζητήματα, βελτιώνοντας τελικά την ασφάλεια, την αποδοτικότητα και την αξιοπιστία.

Στη συνέχεια αυτής της εργασίας, θα αναλυθούν ειδικότερα τα Ευρεία Πολυεπίπεδα Νευρωνικά Δίκτυα (Deep Neural Networks - DNNs), Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks - RNNs) και Δίκτυα Υπολογισμού Δεξαμενής (Reservoir Computing), που είναι η βασικότερες κατηγορίες νευρωνικών δικτύων και κατα συνέπεια αυτές που χρησιμοποιούνται περισσότερο και στον τομέα της ανάλυσης και ανίχνευσης ανωμαλιών.

3

Deep Neural Networks (DNNs)

3.1 Γενική εικόνα

Τα Ευρεία Πολυεπίπεδα Νευρωνικά Δίκτυα (Deep Neural Networks ή DNNs) αποτελούν μια κατηγορία τεχνητών νευρωνικών δικτύων που έχουν σχεδιαστεί για να μοντελοποιούν και να κατανοούν πολύπλοκα πρότυπα και συσχετίσεις στα δεδομένα. Εμπνευσμένα από τη δομή και τη λειτουργία του ανθρώπινου εγκεφάλου, εστιάζουν ειδικά στον δικτυακό σύνδεσμο των νευρώνων.

Τα DNNs αποκαλούνται "βαθιά" επειδή αποτελούνται από πολλά επίπεδα αλληλεπίδρασης κόμβων, γνωστά και ως νευρώνες.[\[52\]](#) Αυτά τα επίπεδα είναι στοιβαγμένα ένα πάνω στο άλλο, σχηματίζοντας μια ιεραρχική δομή. Το πρώτο επίπεδο είναι το επίπεδο εισόδου, το οποίο λαμβάνει τα αρχικά δεδομένα. Το τελευταίο επίπεδο είναι το επίπεδο εξόδου, το οποίο παράγει το επιθυμητό αποτέλεσμα ή πρόβλεψη. Τα ενδιάμεσα επίπεδα ονομάζονται κρυφά επίπεδα, καθώς οι υπολογισμοί τους δεν παρατηρούνται άμεσα.

Κάθε νευρώνας σε ένα DNN λαμβάνει εισερχόμενα σήματα από τους νευρώνες του προηγούμενου επιπέδου, εκτελεί έναν υπολογισμό στα εισερχόμενα σήματα και παράγει ένα εξερχόμενο σήμα. Αυτός ο υπολογισμός περιλαμβάνει συνήθως τον πολλαπλασιασμό των εισερχόμενων σημάτων με βάρη, το άθροισμά τους, την εφαρμογή μιας συνάρτησης ενεργοποίησης και τη μετάδοση του αποτελέσματος στο επόμενο επίπεδο. Η συνάρτηση ενεργοποίησης εισάγει μη γραμμικότητα στο δίκτυο, επιτρέποντάς του να μάθει πολύπλοκα πρότυπα και να κάνει μη γραμμικές προβλέψεις.[\[53\]](#)

Η διαδικασία εκπαίδευσης ενός DNN περιλαμβάνει τη χρήση μεγάλου πλήθους ετικετοποιημένων δεδομένων, γνωστών ως σύνολο εκπαίδευσης. Κατά τη διάρκεια της εκπαίδευσης, το δίκτυο προσαρμόζει τα βάρη του μέσω μιας διαδικασίας που ονομάζεται backpropagation.[\[54\]](#) Η backpropagation υπολογίζει τις κλίσεις του σφάλματος του δικτύου ως προς τα βάρη, επιτρέποντας την εναλλακτική ενημέρωση των βαρών κατά τρόπο που ελαχιστοποιεί το σφάλμα πρόβλεψης. Αυτή η επαναληπτική διαδικασία βελτιστοποίησης εκτελείται συνήθως χρησιμοποιώντας την gradient descent ή τις παραλλαγές της.[\[53\]](#)

Τα DNNs έχουν κερδίσει σημαντική προσοχή και δημοτικότητα λόγω της ικανότητάς τους να μαθαίνουν αυτόματα περίπλοκα χαρακτηριστικά και ιεραρχικές αναπαραστάσεις από ακατέργαστα δεδομένα. Έχουν επιτύχει εξαιρετικά αποτελέσματα σε διάφορους τομείς, συμπεριλαμβανομένης της επεξεργασίας εικόνας, της επεξεργασίας φυσικής γλώσσας, της αναγνώρισης ομιλίας, των συστημάτων συστάσεων και πολλών άλλων.[\[55\]](#)

Ορισμένες αξιοσημείωτες αρχιτεκτονικές βαθιών νευρωνικών δικτύων περιλαμβάνουν τα συνελκτικά νευρωνικά δίκτυα (Convolutional Neural Networks ή CNNs) για την επεξεργασία εικόνων[\[51\]](#), τα αναδρομικά νευρωνικά δίκτυα (Recurrent Neural Networks ή RNNs) για σειριακά

δεδομένα[45] και τα μοντέλα μετασχηματιστών (transformer models) για εργασίες επεξεργασίας φυσικής γλώσσας[56].

3.2 Βασικές Αρχές των Deep Neural Networks

Τα βαθιά νευρωνικά δίκτυα (Deep Neural Networks ή DNNs) αποτελούν μια εξελιγμένη κατηγορία τεχνητών νευρωνικών δικτύων που βασίζονται στην αρχή της εκμάθησης επιπέδου προς επίπεδο (layer-by-layer learning).[55] Αυτό σημαίνει πως το δίκτυο αποτελείται από πολλά επίπεδα, ο καθένας από τα οποία εκτελεί συγκεκριμένες λειτουργίες επεξεργασίας των δεδομένων προκειμένου να εξάγει χρήσιμες πληροφορίες. Τα DNNs είναι εμπνευσμένα από τον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου, καθώς και από την δομή των νευρωνικών δικτύων που συνθέτουν τα βιολογικά νευρωνικά κυκλώματα.[43]

3.2.1 Βασικά Στοιχεία των DNNs

3.2.1.1 Αρχιτεκτονική

Η αρχιτεκτονική των Βαθιών Νευρωνικών Δικτύων (DNNs) αναφέρεται στη δομή και την οργάνωση των νευρωνικών δικτύων, που είναι ένα είδος τεχνητών νευρωνικών δικτύων που είναι εμπνευσμένα από τον τρόπο που λειτουργεί το ανθρώπινο εγkéφαλο. Τα DNNs είναι ιδιαίτερα αποτελεσματικά στην αντιμετώπιση προβλημάτων που σχετίζονται με την αναγνώριση προτύπων, την ταξινόμηση και την πρόβλεψη.[41]

Η βασική αρχιτεκτονική ενός DNN περιλαμβάνει τρία κύρια είδη επιπέδων: το επίπεδο εισόδου, τα κρυφά επίπεδα και το επίπεδο εξόδου.

- Επίπεδο Εισόδου (Input Layer)

Το επίπεδο εισόδου είναι το πρώτο επίπεδο του DNN και αποτελεί το σημείο εισαγωγής των δεδομένων στο νευρωνικό δίκτυο. Εννοείται ότι τα δεδομένα εισόδου παρέχονται σε μορφή που μπορεί να κατανοήσει το νευρωνικό δίκτυο, και αυτό μπορεί να ποικίλλει ανάλογα με το πρόβλημα που προσπαθεί να λύσει το DNN.[43]

- Κρυφά Επίπεδα (Hidden Layers)

Τα κρυφά επίπεδα αποτελούν τα ενδιάμεσα επίπεδα μεταξύ του επιπέδου εισόδου και του επιπέδου εξόδου. Κάθε κρυφό επίπεδο περιέχει πολλούς νευρώνες (ή μονάδες) που λειτουργούν συλλέγοντας και επεξεργάζοντας πληροφορίες από το επίπεδο εισόδου και δημιουργώντας χαρακτηριστικές αναπαραστάσεις των δεδομένων. Αυτές οι χαρακτηριστικές αναπαραστάσεις συνεισφέρουν στην ικανότητα του DNN να μάθει και να κατανοήσει τα πρότυπα και τα χαρακτηριστικά των δεδομένων.[42]

Οι πληροφορίες που διαδίδονται μέσω των κρυφών επιπέδων αναπτύσσουν περίπλοκες αναπαραστάσεις όσο το μοντέλο εκπαιδεύεται, επιτρέποντας στο δίκτυο να αντιμετωπίσει πολύπλοκα προβλήματα.

- Επίπεδο Εξόδου (Output Layer)

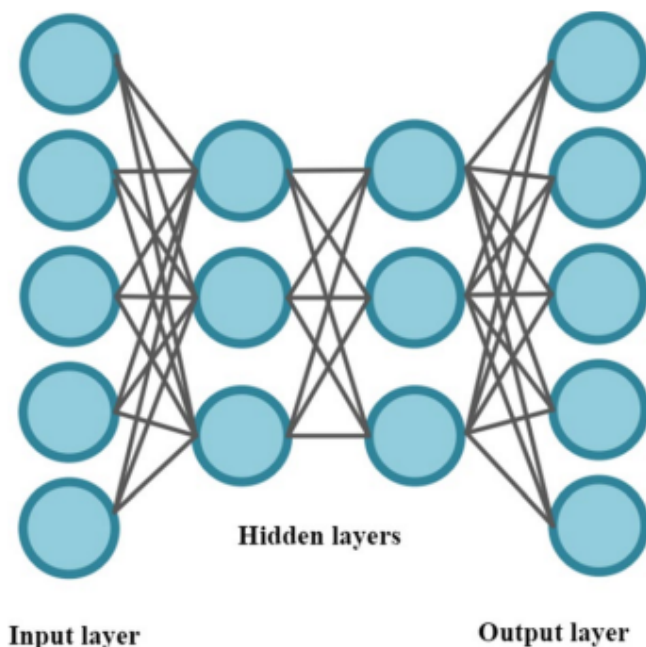
Το επίπεδο εξόδου αποτελεί το τελευταίο επίπεδο του DNN και παράγει τις τελικές προβλέψεις ή αποτελέσματα με βάση τις χαρακτηριστικές αναπαραστάσεις που εξάγονται από τα κρυφά επίπεδα.[42] Το αριθμό των νευρώνων στο επίπεδο εξόδου εξαρτάται από τον τύπο του προβλήματος που επιλύεται. Για παράδειγμα, αν το DNN χρησιμοποιείται για πρόβλεψη κατηγοριών, το επίπεδο εξόδου μπορεί να περιλαμβάνει έναν νευρώνα για κάθε κατηγορία.[57]

Η διαδικασία εκπαίδευσης του DNN περιλαμβάνει τη διόρθωση των παραμέτρων του μοντέλου μέσω αλγορίθμων βελτιστοποίησης, προκειμένου να μάθει να κάνει ακριβείς προβλέψεις για τα δεδομένα εισόδου.[58] Καθώς το μοντέλο εκπαιδεύεται, οι παράμετροι των κρυφών επιπέδων προσαρμόζονται για να αντιστοιχούν στην κατανομή των χαρακτηριστικών των δεδομένων και να βελτιώσουν την απόδοση του μοντέλου.[56]

Στο άρθρο “[Design possibilities and challenges of DNN models: a review on the perspective of end devices](#)” των H Hussain, PS Tamizharasan και CS Rahul αναλύονται πιο συγκεκριμένα κάποια από τα μοντέλα αρχιτεκτονικής DNN. Το μοντέλο Deep Learning αποτελείται από διάφορα DNNs. Μερικά από τα εξέχοντα μοντέλα είναι τα ακόλουθα:

→ Πλήρως συνδεδεμένο νευρωνικό δίκτυο (Fully connected neural network - FCNN)

Τα FCNN είναι μια επέκταση του παραδοσιακού τεχνητού νευρωνικού δικτύου (artificial neural network - ANN). Τα ANN έχουν μια ρηχή δομή και είναι λιγότερο βαθύτερα σε σύγκριση με άλλα DNN. Ωστόσο, τα FCNN έχουν συνήθως μια δομή βαθύτερης στρώσης για την εκτέλεση περίπλοκων εργασιών μάθησης. Όπως φαίνεται στο Σχήμα 1, έχουν τρεις τύπους επιπέδων: στρώμα εισόδου, πολλαπλά κρυφά στρώματα και στρώμα εξόδου.



Σχήμα 1. Fully Connected Neural Network (FCNN)

Εδώ οι νευρώνες στο ένα στρώμα συνδέονται με τους νευρώνες στο επόμενο στρώμα, αλλά κανείς από τους νευρώνες δεν συνδέονται μεταξύ τους στο ίδιο στρώμα. Η έξοδος κάθε στρώματος τροφοδοτείται στο επόμενο στρώμα με λειτουργίες ενεργοποίησης (activation functions). Μερικές από τις δημοφιλείς συναρτήσεις ενεργοποίησης είναι συνάρτηση δυαδικού βήματος (binary step function), γραμμική συνάρτηση (linear function), σιγμοειδής συνάρτηση (sigmoid function), συνάρτηση υπερβολικής εφαπτομένης (hyperbolic tangent function),

διορθωμένη γραμμική μονάδα (rectified linear unit - ReLU), διαρροή ReLU (leaky ReLU), παραμετροποιημένη ReLU (parameterized ReLU), εκθετική γραμμική μονάδα (exponential linear unit). Στο επίπεδο εξόδου, παράγεται το αποτέλεσμα της πρόβλεψης του μοντέλου. Ο υπολογισμός του στρώματος εξόδου του FCNN προχωρά σύμφωνα με τα ακόλουθα βήματα: Έστω $w_j(k)$ το βάρος για το perceptron j στο επίπεδο k για τον κόμβο i , το $O_i(k)$ είναι η έξοδος για τον κόμβο i στο επίπεδο k και το $b_i(k)$ η προκατάληψη για το perceptron i στο στρώμα k .

Αρχικοποίηση του επιπέδου εισόδου l_o : Ορίζονται οι τιμές των εξόδων $O_i(o)$ για τους κόμβους στο επίπεδο εισόδου l_o , οι σχετικές εισοδοί τους στο διάνυσμα $X = x_i, \dots, x_n$, εδώ $O_i(o) = x_i$.

Υπολογίζονται τα αθροίσματα γινομένων και οι έξοδοι κάθε κρυφού στρώματος με σειρά από l_1 έως l_m , υποθέτοντας m κρυφά επίπεδα και κόμβους r_k στο στρώμα k . Για k από 1 έως m , και i από 1 έως r_k ισχύει:

$$h_i^k = b_i^k + \sum_{j=1}^{r_{k-1}} w_{ji}^k o_j^{k-1}$$
$$o_i^k = g(h_i^k)$$

Υπολογίζεται η έξοδος y του πρώτου κόμβου του στρώματος εξόδου l_m :

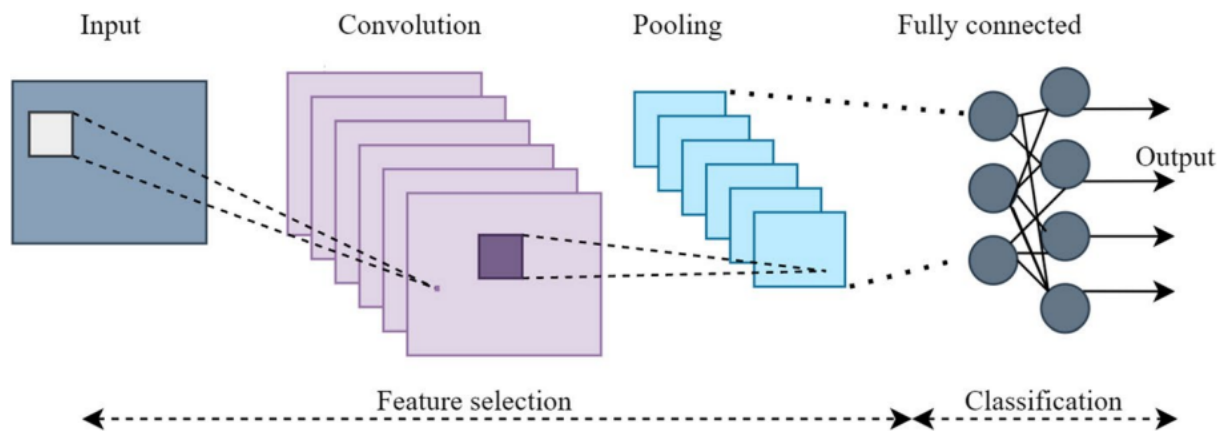
$$h_1^k = b_1^k + \sum_{j=1}^{r_{k-1}} w_{j1}^k o_j^{k-1}$$
$$o_1^k = g(h_1^k)$$

Το μοντέλο προσαρμόζεται σε ένα συγκεκριμένο σύνολο δεδομένων με τη βοήθεια αλγορίθμων εκπαίδευσης όπως η στοχαστική κλίση (stochastic gradient descent - SGD), ο αλγόριθμος Levenberg-Marquardt, η συζευγμένη κλίση (conjugate gradient) κ.λπ. Η εκμάθηση συνίσταται στη συνεχή ενημέρωση των τιμών των βαρών και της bias για την ελαχιστοποίηση του μέσου τετραγώνου σφάλματος (mean squared error - MSE). Αυτοί οι τύποι DNN έχουν υψηλή πολυπλοκότητα και χαμηλή απόδοση και σύγκλιση. Ως εκ τούτου, μπορούν να χρησιμοποιηθούν κυρίως για εξαγωγές χαρακτηριστικών και προσέγγιση.

→ Συνελικτικό νευρωνικό δίκτυο (Convolutional neural network - CNN)

Τα CNN έχουν σχεδιαστεί κυρίως για την επεξεργασία δεδομένων που έρχονται σε πολυδιάστατες συστοιχίες όπως εικόνες. Η βασική δομή ενός CNN, που δίνεται στο Σχήμα 2, αποτελείται από δύο κύρια μέρη, (α) Εξαγωγή χαρακτηριστικών, όπου οι τεχνικές συνέλιξης χρησιμοποιούνται για την εξαγωγή των χαρακτηριστικών των εικόνων εισόδου. (β) Ταξινόμηση,

στην οποία τα χαρακτηριστικά που μελετήθηκαν χρησιμοποιούνται για την πρόβλεψη της κλάσης της δεδομένης εικόνας.



Σχήμα 2. Η βασική δομή ενός CNN.

Διαφορετικοί τύποι επιπέδων που δίνονται παρακάτω στοιβάζονται για να σχηματίσουν ένα CNN, κάθε επίπεδο έχει τις δικές του λειτουργίες:

- **Συνελικτικό επίπεδο (Convolutional layer):**

Η λειτουργία συνέλιξης είναι ένας υπολογισμός μεταξύ της εικόνας εισόδου και του φίλτρου ενός δεδομένου μεγέθους ($M \times M$). Το φίλτρο ολισθαίνει πάνω από τις εικόνες για να υπολογίσει το γινόμενο κουκίδων (dot product) και να αποκτήσει χάρτες χαρακτηριστικών (feature maps). Σε αυτό το στάδιο, οι χάρτες χαρακτηριστικών δίνουν μια ιδέα για τις γωνίες και τις άκρες της εικόνας εισόδου. Στη συνέχεια τροφοδοτείται στο επόμενο στρώμα για να μάθει τα χαρακτηριστικά υψηλού επιπέδου της εικόνας.

- **Επίπεδο συγκέντρωσης (Pooling layer):**

Η λειτουργία συγκέντρωσης συμβάλλει στη μείωση του μεγέθους του σύνθετου χάρτη χαρακτηριστικών, μειώνοντας έτσι το υπολογιστικό κόστος. Υπάρχουν διάφοροι τύποι εργασιών συγκέντρωσης. Στο Max pooling, η μεγαλύτερη τιμή λαμβάνεται από τον χάρτη χαρακτηριστικών, ενώ στη μέση συγκέντρωση, η μέση τιμή των στοιχείων εντός μιας προκαθορισμένης ($N \times N$) επιλογής υπολογίζεται για μετάβαση στο επόμενο επίπεδο. Λειτουργεί ως γέφυρα μεταξύ της συνέλιξης και ενός πλήρως συνδεδεμένου στρώματος.

- **Πλήρως συνδεδεμένο επίπεδο (Fully connected layers - FC):**

Τα πλήρως συνδεδεμένα επίπεδα είναι συνήθως τα τελευταία επίπεδα μιας αρχιτεκτονικής CNN. Αποτελούνται από νευρώνες (neurons), βάρη (weights) και biases για τη σύνδεση διαφορετικών στρωμάτων μεταξύ τους. Η είσοδος στο στρώμα FC ισοπεδώνεται πάντα πριν τροφοδοτηθεί. Η ταξινόμηση της εικόνας πραγματοποιείται στο τέλος των στρώσεων FC.

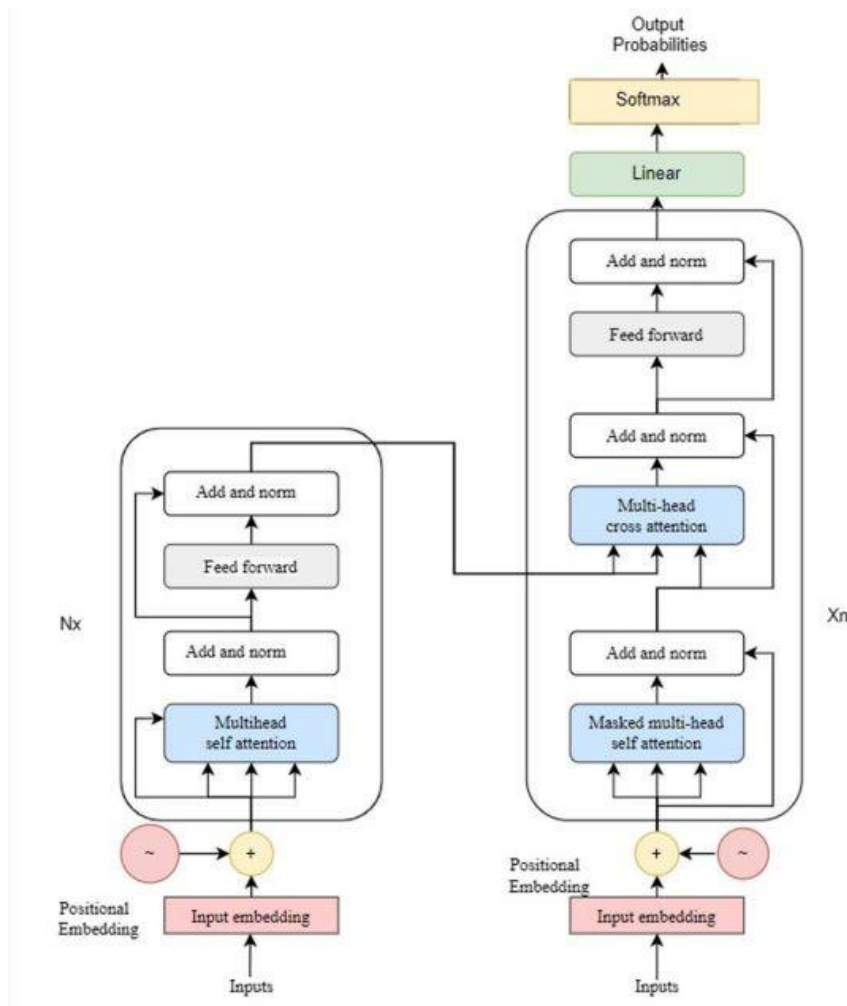
- **Ορισμένα από τα πρόσθετα στρώματα που προστέθηκαν περιλαμβάνουν στρώματα εγκατάλειψης (dropout layers) για τη μείωση over-fitting. Μερικοί νευρώνες**

απορρίπτονται από το νευρωνικό δίκτυο κατά τη διάρκεια της εκπαίδευσης, για να μειωθεί το μέγεθος του μοντέλου. Η αναλογία πτώσης (Drop ratio) μπορεί να δοθεί χειροκίνητα. Άλλα στρώματα περιλαμβάνουν στρώματα κανονικοποίησης παρτίδας (batch-normalization layers), στρώματα ενεργοποίησης (activation layers), στρώματα συνένωσης βάθους (depth concatenation layers) κ.λπ.

Γενικά, τα CNN μπορούν να εξάγουν χαρακτηριστικά αυτόματα με μειωμένη πολυπλοκότητα σε σύγκριση με τα FCNN.

→ Δίκτυα μετασχηματιστών (Transformer networks)

Οι μετασχηματιστές είναι αρχιτεκτονικές πολλαπλών επιπέδων που κατασκευάζονται με την τοποθέτηση μπλοκ μετασχηματιστών το ένα πάνω στο άλλο. Το Σχήμα 3 δείχνει τον κύριο μηχανισμό που είναι ενσωματωμένος σε μια αρχιτεκτονική μετασχηματιστή με μια διαδικασία αυτοπροσοχής πολλαπλών κεφαλών (multi-head self-attention process), ένα position-wise FFNN, μονάδες κανονικοποίησης στρώματος (layer normalization modules - LN) και υπολειπόμενες συνδέσεις (residual connections - RC).



Σχήμα 3. Λειτουργία του δικτύου Transformer

Η είσοδος στον μετασχηματιστή είναι ένας tensor σχήματος $B \times N$, όπου τα B και N είναι το μέγεθος παρτίδας (batch size) και το μήκος ακολουθίας (sequence length) αντίστοιχα. Αρχικά, οι είσοδοι τροφοδοτούνται σε ένα στρώμα ενσωμάτωσης για να σχηματίσουν έναν tensor διαστάσεων D σχήματος $B \times N \times D$. Στη συνέχεια συντίθεται προσθετικά έναντι κωδικοποίησης θέσης και τροφοδοτείται μέσω μιας μονάδας αυτοπροσοχής πολλαπλών κεφαλών (multi-headed self-attention module). Οι κωδικοποιήσεις θέσης είναι ημιτονοειδείς (sinusoidal) είσοδοι στις εκπαιδευσιμες ενσωματώσεις. Οι είσοδοι και οι εξοδοι της μονάδας αυτοπροσοχής πολλαπλών κεφαλών συνδέονται με RC και ένα LN. Η έξοδος της μονάδας αυτο-προσοχής πολλαπλών κεφαλών μεταβιβάζεται στη συνέχεια σε ένα FFNN δύο επιπέδων, το οποίο έχει τις εισόδους και τις εξόδους του συνδεδεμένες κατά έναν υπολειπόμενο τρόπο με κανονικοποίηση στρώματος.

Συνοπτικά, κάθε μπλοκ μετασχηματιστή μπορεί να αναπαρασταθεί ως:

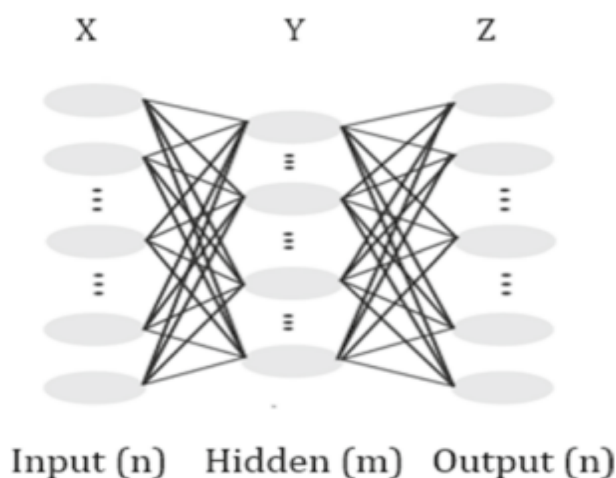
$$XA = \text{Layer Norm}(\text{Multihead Self Attention}(X)) + X$$

$$XB = \text{Layer Norm}(\text{Position FFNN}(XA)) + XA$$

Όπου X είναι η είσοδος που δίνεται στο μπλοκ μετασχηματιστή και XB είναι η έξοδος του μπλοκ μετασχηματιστή. Οι εφαρμογές του μετασχηματιστή περιλαμβάνουν (α) ταξινόμηση χρησιμοποιώντας μονάδα κωδικοποιητή (classification using encoder module), (β) μοντελοποίηση γλώσσας χρησιμοποιώντας μονάδα αποκωδικοποίησης (language modeling using decode module) και (γ) μηχανική μετάφραση με τη μονάδα κωδικοποιητή-αποκωδικοποιητή (machine translation with the encoder-decoder module).

→ Αυτόματος κωδικοποιητής (Auto-encoder - AE)

Ένας αυτόματος κωδικοποιητής είναι ένα τεχνητό νευρωνικό δίκτυο που παράγει έξοδο παρόμοια με τις εισόδους του. Έχει δύο μέρη, δηλαδή τον κωδικοποιητή (encoder) και τον αποκωδικοποιητή (decoder). Ο κωδικοποιητής εξάγει τα n χαρακτηριστικά εισόδου και τα μετατρέπει στα m μοναδικά κωδικοποιημένα χαρακτηριστικά του, όπου m και n είναι οι αριθμοί των νευρώνων στα εισερχόμενα και στα κρυφά επίπεδα αντίστοιχα. Ο αποκωδικοποιητής ανακατασκευάζει τις n εισόδους από τα χαρακτηριστικά m .



Σχήμα 4. Autoencoder Model

Από το Σχήμα 4, θεωρείται ότι X είναι το διάνυσμα εισόδου και $(x_1, x_2, \dots, x_n)$ οι δευτερεύουσες εισοδοι (sub inputs) ή χαρακτηριστικά (features), τότε η έξοδος που λαμβάνεται από το κρυφό

στρώμα $Y = (y_1, y_2, \dots, y_m)$ μετά τη φάση κωδικοποίησης είναι δίνεται στην ακόλουθη εξίσωση. Εδώ το w είναι το διάνυσμα βάρους ή πυρήνα (weight or kernel vector) και b είναι η τιμή πόλωσης (bias value).

$$f(x) = Y = \xi(wX + b)$$

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

Η ανακατασκευασμένη είσοδος μετά τη φάση αποκωδικοποίησης λαμβάνεται χρησιμοποιώντας την εξίσωση:

$$f(Y) = Z = \zeta(wY + b) \approx X$$

Ο αλγόριθμος κωδικοποίησης μειώνει το σφάλμα ανακατασκευής μεταξύ της εισόδου X και της εξόδου Z . Βοηθά επίσης στην επίτευξη ενός συνόλου παραμέτρων για τις φάσεις κωδικοποίησης και αποκωδικοποίησης. Το $J(X, Z)$ αντιπροσωπεύει το σφάλμα ανακατασκευής, το x_i αντιπροσωπεύει την τιμή εισόδου a . Το z_i αντιπροσωπεύει την έξοδο που προκύπτει.

$$J(X, Z) = \frac{1}{n} \sum_i^n \|x_i - z_i\|^2$$

Το ΑΕ μπορεί να σχηματίσει μια βαθιά αρχιτεκτονική στοιβάζοντας πολλαπλά επίπεδα ως κρυφό στρώμα από τον στοιβαγμένο αυτόματο κωδικοποιητή (stacked autoencoder - SAE). Περιλαμβάνει βαθιές παραλλαγές νευρωνικών δικτύων που αποτελούν υψηλότερο επίπεδο πολυπλοκότητας και απόδοσης. Ο σχεδιασμός ενός μοντέλου βαθιάς εκμάθησης με στοίβαξη αυτόματου κωδικοποιητή έδειξε εξαιρετικά αποτελέσματα σε διαφορετικές εφαρμογές. Τα βαθιά νευρωνικά δίκτυα Stacked Autoencoder (SAE) έχουν αποδείξει τις καλύτερες υπολογιστικές δυνατότητες όταν εκπαιδεύονται χρησιμοποιώντας την καλύτερη λειτουργία εκπαίδευσης.

3.2.1.2 Νευρώνες

Κάθε επίπεδο περιλαμβάνει πολλούς νευρώνες (κόμβους). Οι νευρώνες είναι οι βασικές μονάδες επεξεργασίας πληροφορίας στο νευρικό σύστημα των ζώων, συμπεριλαμβανομένων και των ανθρώπων. Είναι εξειδικευμένα κύτταρα που έχουν την ικανότητα να επικοινωνούν μεταξύ τους και να μεταφέρουν τις πληροφορίες από ένα σημείο του νευρικού συστήματος σε ένα άλλο.[\[59\]](#)

Κάθε επίπεδο στο νευρικό σύστημα, όπως για παράδειγμα στον εγκέφαλο ή το νωτιαίο μυελό, περιλαμβάνει πολλούς νευρώνες, οι οποίοι οργανώνονται σε πυκνούς δικτυώσεις.[\[60\]](#) Κάθε νευρώνας έχει μια σώματος, που περιέχει τον κυτταρικό πυρήνα, και προεκτάσεις που του επιτρέπουν να επικοινωνεί με άλλα κύτταρα.

Οι νευρώνες λαμβάνουν εισερχόμενα σήματα από άλλους νευρώνες μέσω των δενδριτών, οι οποίοι αποτελούν τις διεγερτικές εισόδους του νευρώνα.[\[57\]](#) Όταν αυτά τα σήματα φτάσουν σε ένα νευρώνα, θα εξετάσει το συνολικό εισερχόμενο ερέθισμα και, εάν το σήμα υπερβεί ένα κατώφλι, θα δημιουργήσει και θα αποστείλει ένα εξερχόμενο σήμα.[\[56\]](#)

Οι εξερχόμενες επικοινωνίες του νευρώνα γίνονται μέσω των αξόνων. Οι άξονες είναι προεκτάσεις του νευρώνα που μεταφέρουν το σήμα από το σώμα του νευρώνα και τις δενδριτικές του διακλαδώσεις σε άλλους νευρώνες ή σε κύτταρα-στόχους, όπως οι μύες ή οι γλωχίνες.[\[61\]](#) Τα

σήματα μεταφέρονται στα αξόνια με τη βοήθεια ηλεκτροχημικών διεργασιών, καθιστώντας δυνατή την επικοινωνία και την ανταλλαγή πληροφοριών ανάμεσα στους νευρώνες.[\[62\]](#)

Αυτή η συλλογιστική διαδικασία επιτρέπει την επεξεργασία των πληροφοριών σε διάφορα επίπεδα του νευρικού συστήματος, διευκολύνοντας τις λειτουργίες όπως ο συντονισμός κινήσεων, η αντίληψη, η μάθηση και η μνήμη, μεταξύ άλλων. Αυτός ο δικτυακός χαρακτήρας του νευρικού συστήματος και η ικανότητα των νευρώνων να επικοινωνούν μεταξύ τους καθιστούν το νευρικό σύστημα ένα εξαιρετικά πολύπλοκο και ισχυρό σύστημα πληροφορικής που κυριαρχεί στη λειτουργία του ανθρώπινου σώματος.

3.2.1.3 Συνάρτηση Ενεργοποίησης

Οι νευρώνες εφαρμόζουν μια συνάρτηση ενεργοποίησης στον συνδυασμό των εισερχομένων σημάτων τους προκειμένου να εισάγουν μη γραμμικότητα (nonlinearity) στο δίκτυο. Οι πιο γνωστές συναρτήσεις ενεργοποίησης είναι η σιγμοειδής (sigmoid), η υπερβολική εφαπτομένη (hyperbolic tangent - tanh) και η ενεργοποίηση ReLU (Rectified Linear Unit).[\[55\]](#)

Οι συναρτήσεις ενεργοποίησης είναι κρίσιμα στοιχεία των νευρωνικών δικτύων και χρησιμοποιούνται για να εισάγουν μη γραμμικότητα στο δίκτυο. Χωρίς μη γραμμικότητα, το νευρωνικό δίκτυο θα μπορούσε απλώς να μάθει γραμμικούς συνδυασμούς των εισόδων του, οδηγώντας σε περιορισμένη εκφραστική ικανότητα και αποδοτικότητα. Ας εξετάσουμε λεπτομερέστερα τις τρεις πιο γνωστές συναρτήσεις ενεργοποίησης:

- Σιγμοειδής (Sigmoid)

Η σιγμοειδής συνάρτηση είναι μια καμπύλη συνάρτηση που παίρνει μια εύρος τιμών και μετατρέπει τις εισόδους σε μια πιθανότητα που βρίσκεται μεταξύ 0 και 1. Η μορφή της συνάρτησης μοιάζει με "S" και δίνει τιμές που αντιστοιχούν σε μια καμπύλη πιθανοκρατικής κατανομής. Παλαιότερα, η σιγμοειδής συνάρτηση χρησιμοποιούνταν συχνά, αλλά πλέον αποφεύγεται λόγω των προβλημάτων που προκαλεί στο φαινόμενο vanishing/exploding gradients κατά την εκπαίδευση βαθιών νευρωνικών δικτύων.[\[41\]](#)

Μαθηματικά, η σιγμοειδής συνάρτηση ορίζεται ως:

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

- Υπερβολική Εφαπτομένη (Hyperbolic tangent - Tanh)

Η συνάρτηση υπερβολικής εφαπτομένης (tanh) είναι μια άλλη καμπύλη συνάρτηση που παίρνει τιμές μεταξύ -1 και 1. Είναι παραλλαγή της σιγμοειδούς συνάρτησης και επιτρέπει την κεντράριση των δεδομένων γύρω από το μηδέν. Αυτό μπορεί να βοηθήσει στην αποφυγή του προβλήματος vanishing/exploding gradients, αλλά παραμένει πιθανό να προκαλέσει αυτό το φαινόμενο σε βαθιά νευρωνικά δίκτυα.[\[41\]](#)

Μαθηματικά, η συνάρτηση υπερβολικής εφαπτομένης ορίζεται ως:

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

- Ενεργοποίηση ReLU (Rectified Linear Unit)

Η συνάρτηση ενεργοποίησης ReLU είναι απλή και αποτελεσματική. Εάν η είσοδος είναι θετική, η ReLU επιστρέφει απλώς την είσοδο, ενώ εάν η είσοδος είναι αρνητική, επιστρέφει μηδέν.

Μαθηματικά, αυτό ορίζεται ως:

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

Η ReLU αποφεύγει το πρόβλημα vanishing/exploding gradients και όταν συνδυάζεται με κατάλληλη διαδικασία εκπαίδευσης, έχει αποδειχθεί ότι ενισχύει την απόδοση των νευρωνικών δικτύων.[\[43\]](#)

Πρόσφατα, έχουν αναπτυχθεί πολλές άλλες συναρτήσεις ενεργοποίησης, καθώς η έρευνα στον τομέα της μηχανικής μάθησης συνεχίζεται. Οι παραπάνω συναρτήσεις ενεργοποίησης είναι μερικές από τις πιο βασικές και πολύ χρησιμοποιούμενες, αλλά σε κάθε περίπτωση, η επιλογή της κατάλληλης συνάρτησης εξαρτάται από το πρόβλημα που προσπαθείτε να λύσετε και την αρχιτεκτονική του δικτύου σας.

3.2.1.4 Βάρη

Κάθε σύνδεση μεταξύ δύο νευρώνων διαθέτει ένα σύνολο βαρών (weights). Στο πλαίσιο της νευρωνικής δικτύωσης, ο όρος "βάρη" αναφέρεται στις παραμέτρους που χρησιμοποιούνται για να περιγράψουν την δύναμη και τη σημαντικότητα της σύνδεσης μεταξύ δύο νευρώνων σε ένα νευρωνικό δίκτυο.[\[41\]\[40\]](#) Σε ένα τυπικό νευρωνικό δίκτυο, οι νευρώνες συνδέονται μεταξύ τους μέσω συνάψεων, και αυτές οι συνάψεις διαθέτουν βάρη που καθορίζουν τον τρόπο με τον οποίο τα σήματα διασχίζουν το δίκτυο. Καθορίζουν επι της ουσίας, την σημαντικότητα και τη συμβολή των εισερχομένων σημάτων σε κάθε νευρώνα.[\[43\]](#)

Κάθε νευρώνας λαμβάνει εισερχόμενα σήματα από άλλους νευρώνες μέσω αξόνων σύνδεσης. Κάθε σύνδεση περιλαμβάνει ένα βάρος που καθορίζει πόσο επηρεάζεται ο νευρώνας από τον συγκεκριμένο νευρώνα.[\[61\]](#) Αν το βάρος είναι υψηλό, τότε η επιρροή του εισερχομένου σήματος είναι ισχυρή, ενώ αν είναι χαμηλό, η επιρροή είναι ασθενής.[\[60\]](#)

Οι τιμές των βαρών προσαρμόζονται κατά τη διαδικασία εκπαίδευσης του νευρωνικού δικτύου, ώστε να μάθει να αναγνωρίζει πρότυπα και να επιλύει συγκεκριμένα προβλήματα. Κατά την εκπαίδευση, το δίκτυο προσπαθεί να προσαρμόσει τις τιμές των βαρών έτσι ώστε οι έξοδοι του να συμφωνούν με τις επιθυμητές εξόδους. Με άλλα λόγια, προσπαθεί να μάθει τις συνδέσεις μεταξύ των νευρώνων έτσι ώστε να παρέχει τις σωστές απαντήσεις στις εισόδους που δέχεται.[\[45\]](#)

Ένας τρόπος για να φανταστεί κανείς τη λειτουργία των βαρών είναι μέσω ενός αναλογικού παραδείγματος:

Σκεφτόμαστε την σύνδεση μεταξύ δύο νευρώνων ως μία σωλήνωση με νερό που ρέει μέσα της. Τα βάρη αντιστοιχούν στο πώς είναι ανοιχτή η βαλβίδα της σωλήνωσης. Όταν η βαλβίδα είναι πλήρως ανοιχτή, το νερό ρέει ελεύθερα, και η σύνδεση είναι πολύ ισχυρή. Αντίστοιχα, όταν η βαλβίδα είναι μερικώς κλειστή, το νερό ρέει πιο αργά και η σύνδεση είναι ασθενέστερη. Στο νευρωνικό δίκτυο, οι τιμές των βαρών προσαρμόζονται κατά την εκπαίδευση για να καταλήξουν στις βέλτιστες τιμές, ώστε να πετύχει την καλύτερη δυνατή επίδοση στο πρόβλημα που προσπαθεί να επιλύσει.

Τα βάρη αποτελούν σημαντικό μέρος της νευρωνικής δικτύωσης και συμβάλλουν στην ικανότητα του δικτύου να μάθει, να γενικεύει και να παράγει ενδιαφέρουσες εξόδους σε νέα δεδομένα που δεν είχε δει κατά τη διάρκεια της εκπαίδευσής του.[\[54\]\[59\]](#)

Συνοψίζοντας, τα βαθιά νευρωνικά δίκτυα έχουν αναδειχθεί σε έναν από τους πιο ισχυρούς και πολλά υποσχόμενους τύπους τεχνητών νευρωνικών δικτύων. Η δυνατότητά τους να εξάγουν περίπλοκα χαρακτηριστικά και να μάθουν από μεγάλα σύνολα δεδομένων τους καθιστά κατάλληλα για ποικίλες εφαρμογές. Παρόλα αυτά, η επιτυχημένη χρήση των DNNs απαιτεί πολυάριθμους υπολογιστικούς πόρους και προσεκτική εκπαίδευση προκειμένου να αποφευχθεί η υπερεκπαίδευση. Παρ' όλα αυτά, η συνεχής έρευνα και ανάπτυξη σε αυτό το πεδίο αναμένεται να διαμορφώσει μια λαμπρή και πρωτοποριακή προοπτική για το μέλλον των DNNs.

3.3 Εκπαίδευση των DNNs

Η εκπαίδευση των Deep Neural Networks (DNNs) αποτελεί κρίσιμο στάδιο για την απόκτηση της ικανότητάς τους να πραγματοποιούν ακριβείς προβλέψεις και να εκτελούν πολύπλοκες εργασίες. Η διαδικασία της εκπαίδευσης αναφέρεται στον τρόπο με τον οποίο το νευρωνικό δίκτυο προσαρμόζεται ώστε να επιτυγχάνει αυτό που ονομάζουμε "μάθηση" από τα δεδομένα εκπαίδευσης. Για να γίνει αυτό, το DNN χρησιμοποιεί ένα μεγάλο σύνολο δεδομένων εκπαίδευσης, που περιλαμβάνει εισόδους και τις αντίστοιχες επιθυμητές εξόδους.

Κατά την εκπαίδευση των DNNs, το δίκτυο προσαρμόζει τα βάρη του, που είναι καθοριστικά για τη λειτουργία του, μέσω επαναληπτικών βελτιστοποιήσεων. Οι βελτιστοποιήσεις αυτές βασίζονται στον υπολογισμό του σφάλματος πρόβλεψης (prediction error) που εμφανίζεται μεταξύ των προβλέψεων του δικτύου και των πραγματικών τιμών των δεδομένων εκπαίδευσης.[\[41\]](#)

Είναι σημαντικό να σημειωθεί πως η επιτυχία της εκπαίδευσης εξαρτάται από πολλούς παράγοντες, όπως η σωστή επιλογή της αρχιτεκτονικής του δικτύου, η κατάλληλη ρύθμιση των υπερπαραμέτρων, η ποσότητα και ποιότητα των δεδομένων εκπαίδευσης και πολλοί άλλοι παράγοντες.[\[55\]](#) Η εκπαίδευση DNNs είναι συχνά χρονοβόρα και απαιτεί μεγάλη υπολογιστική ισχύ, αλλά έχει αποτελέσει καθοριστικό παράγοντα στην πρόοδο της τεχνητής νοημοσύνης και της μηχανικής μάθησης.[\[48\]](#)

Ο στόχος της εκπαίδευσης είναι να επιτύχουμε τη μείωση του σφάλματος πρόβλεψης ή του λάθους σε όσο το δυνατόν χαμηλότερο επίπεδο, έτσι ώστε το δίκτυο να μπορεί να γενικεύει και να κάνει ακριβείς προβλέψεις για νέα, μη-διαθέσιμα δεδομένα.[\[55\]](#)

3.3.1 Διαδικασία Εκπαίδευσης

Η διαδικασία εκπαίδευσης των DNNs αναδεικνύει μια ενδιαφέρουσα σύγκλιση μεταξύ της τεχνολογίας και της βιολογίας. Όπως και το ανθρώπινο νευρικό σύστημα μαθαίνει από εμπειρίες και προσαρμόζει τη συμπεριφορά του με βάση τις αλληλεπιδράσεις με το περιβάλλον, έτσι και τα DNNs μαθαίνουν από τα δεδομένα και προσαρμόζουν τις εσωτερικές τους αναπαραστάσεις και παραμέτρους.[\[48\]](#) Αναδεικνύεται ως ένας σύνθετος και εξαιρετικά κρίσιμος αλγόριθμος που επιτρέπει στα νευρωνικά δίκτυα να αποκτήσουν την ικανότητα να αντιμετωπίζουν πολυδιάστατα προβλήματα με εξαιρετική ακρίβεια.[\[41\]](#) Το βασικό χαρακτηριστικό αυτής της διαδικασίας είναι η συνεχής προσαρμογή των βαρών του δικτύου προκειμένου να επιτυγχάνεται καλύτερη προσέγγιση των επιθυμητών αποτελεσμάτων.

Η βασική ιδέα είναι ότι το DNN προσπαθεί να μάθει τις βέλτιστες παραμέτρους των βαρών του, ώστε να μετατρέψει τις εισόδους σε εξόδους με όσο το δυνατόν μεγαλύτερη συσχέτιση με τα επιθυμητά αποτελέσματα. Για να επιτευχθεί αυτό, το δίκτυο προσαρμόζει τα βάρη του με βάση το σφάλμα μεταξύ των προβλέψεων του και των πραγματικών επιθυμητών αποτελεσμάτων.[\[55\]](#) Στον

πυρήνα της διαδικασίας εκπαίδευσης, η αντίστροφη διάδοση (backpropagation) αποτελεί ένα βασικό εργαλείο που διευκολύνει το δίκτυο να προσαρμόσει τη δομή του ώστε να αντικατοπτρίζει τα πρότυπα των δεδομένων.[43] Η gradient descent, που βασίζεται στις παραγώγους του σφάλματος, λειτουργεί σαν τον "οδηγό" που καθοδηγεί την προσαρμογή των βαρών προς την κατεύθυνση της ελαχιστοποίησης του σφάλματος.[51] Στην πορεία αυτής της προσαρμογής, το DNN συγκλίνει προς πιο ακριβείς προβλέψεις για τα δεδομένα εκπαίδευσης.

Η διαδικασία αυτή βασίζεται στην αντίστροφη διάδοση, η οποία αποτελεί τον μηχανισμό κλειδί για την ενημέρωση των βαρών. Κατά την αντίστροφη διάδοση, το σφάλμα πρόβλεψης που προκύπτει από τη σύγκριση των προβλέψεων με τα πραγματικά αποτελέσματα διαδίδεται προς τα πίσω από το δίκτυο. Σε κάθε στρώμα, υπολογίζονται οι παράγωγοι του σφάλματος ως προς τα βάρη του. Οι παράγωγοι αυτές δείχνουν τον τρόπο με τον οποίο θα πρέπει να προσαρμοστούν τα βάρη για να μειωθεί το σφάλμα. Αυτός ο μηχανισμός προσαρμογής των βαρών λέγεται gradient descent.[51]

Κατά την εκπαίδευση, το δίκτυο εκτελεί πολλές επαναλήψεις (epochs) με μικρότερα σύνολα δεδομένων (mini-batches) για την ενημέρωση των βαρών. Ο αριθμός των επαναλήψεων αποτελεί ένα κρίσιμο κομμάτι της διαδικασίας εκπαίδευσης. Κάθε επανάληψη επιτρέπει στο δίκτυο να βελτιώσει την ακρίβεια των προβλέψεών του, καθώς προσαρμόζει τις παραμέτρους του. Οι πολλαπλές επαναλήψεις επιτρέπουν στο DNN να αναπτύξει βαθύτερη κατανόηση των δεδομένων, ενώ η αναδιατύπωση των δεδομένων σε μικρότερα σύνολα συμβάλλει στην εξομάλυνση της προσαρμογής των βαρών και τη βελτίωση της απόδοσης.[52] Αυτή η επαναληπτική διαδικασία είναι κρίσιμη για την ανάπτυξη του μοντέλου, καθώς εξασφαλίζει όχι μόνο την καλή εφαρμογή στα δεδομένα εκπαίδευσης, αλλά και την ικανότητα γενίκευσης σε νέα, προηγουμένως αγνώστika δεδομένα.

Πέρα από την τεχνική πολυπλοκότητα, η εκπαίδευση των DNNs απαιτεί επίσης δημιουργικότητα και επιστημονική διαίσθηση. Η επιλογή της κατάλληλης αρχιτεκτονικής του δικτύου, η ρύθμιση των υπερπαραμέτρων, και η εξερεύνηση της βέλτιστης στρατηγικής εκπαίδευσης αποτελούν μέρος της διαδικασίας που απαιτεί την εμπειρία και την εξειδίκευση των ερευνητών.[41][55] Παρότι η διαδικασία εκπαίδευσης των DNNs είναι απαιτητική, η τελική επίτευξη ενός αξιόπιστου και αποδοτικού μοντέλου ανταμείβει την προσπάθεια. Τα DNNs έχουν αναβαθμίσει σημαντικά την ικανότητά μας να αντιμετωπίζουμε πολυπλοκότητες και να εξαγάγουμε σημαντικές πληροφορίες από τα δεδομένα, καθώς και να ανοίγουμε νέες προοπτικές σε πεδία όπως η ιατρική, η ρομποτική και η τεχνητή νοημοσύνη.[48][49]

Στο άρθρο “[Brief Introduction of Back Propagation \(BP\) Neural Network Algorithm and Its Improvement](#)” των Jing Li, Ji-hang Cheng, Jing-yuan Shi και Fei Huang, αναλύεται περαιτέρω η διαδικασία της προαναφερόμενης αντίστροφης διάδοσης (backpropagation - BP).

Στην πραγματικότητα, ο αλγόριθμος BP είναι μια μέθοδος παρακολούθησης της μάθησης. Χρησιμοποιεί τις μεθόδους του square error και της gradient descent για να πραγματοποιήσει την τροποποίηση του βάρους σύνδεσης του δικτύου. Η τροποποίηση του βάρους σύνδεσης του δικτύου στοχεύει στην επίτευξη του ελάχιστου αθροίσματος των square error. Σε αυτόν τον αλγόριθμο, δίνεται μια μικρή τιμή στην τιμή σύνδεσης του δικτύου πρώτα και, στη συνέχεια, επιλέγεται ένα δείγμα εκπαίδευσης για τον υπολογισμό της διαβάθμισης σφάλματος σε σχέση με αυτό το δείγμα.

Η διαδικασία μάθησης BP μπορεί να περιγραφεί ως εξής:

- Forward propagation of operating signal:

Το σήμα εισόδου διαδίδεται από το στρώμα εισόδου (input layer), μέσω του στρώματος απόκρυψης (hide layer), στο στρώμα εξόδου (output layer). Κατά την Forward propagation of operating signal, η τιμή βάρους και η τιμή μετατόπισης του δικτύου διατηρούνται σταθερές και η κατάσταση κάθε στρώματος νευρώνα θα ασκήσει επίδραση μόνο σε αυτήν του επόμενου στρώματος νευρώνα. Σε περίπτωση που η αναμενόμενη έξοδος δεν μπορεί να επιτευχθεί στο επίπεδο εξόδου, μπορεί να μεταφερθεί στην Back propagation of error signal.

•Back propagation of error signal:

Η διαφορά μεταξύ της πραγματικής εξόδου και της αναμενόμενης εξόδου του δικτύου ορίζεται ως το σήμα σφάλματος (error signal). Στην Back propagation of error signal, το σήμα σφάλματος διαδίδεται από το άκρο εξόδου στο στρώμα εισόδου με τρόπο στρώμα προς στρώμα (layer-by-layer). Κατά την Back propagation of error signal, η τιμή βάρους του δικτύου ρυθμίζεται από την ανάδραση σφάλματος (error feedback). Η συνεχής τροποποίηση της τιμής βάρους και της τιμής μετατόπισης εφαρμόζεται για να κάνει την πραγματική έξοδο του δικτύου πιο κοντά στην αναμενόμενη.

Η ιδεολογία που καθοδηγεί τους κανόνες εκμάθησης του δικτύου BP είναι η εξής, η τροποποίηση της τιμής βάρους (weight value) και της τιμής threshold του δικτύου γίνεται κατά μήκος της negative gradient κατεύθυνσης (negative gradient direction) που αντικατοπτρίζει την ταχύτερη πτώση της συνάρτησης.

$$x_{k+1} = x_k - \eta_k g_k$$

Στον τύπο που αναφέρθηκε παραπάνω, το x_k αντιπροσωπεύει τον πίνακα της τρέχουσας τιμής βάρους και της τιμής threshold. Το g_k αντιπροσωπεύει τη διαβάθμιση της τρέχουσας συνάρτησης. το η_k αντιπροσωπεύει το ρυθμό μάθησης. Εδώ, το δίκτυο BP τριών επιπέδων λαμβάνεται ως παράδειγμα για να περιγράψει λεπτομερώς ο αλγόριθμος BP.

Όσο για το δίκτυο BP τριών επιπέδων, ας υποθέσουμε ότι ο κόμβος εισόδου του είναι x_i , ο κόμβος του στρώματος απόκρυψης είναι y_j και ο κόμβος του στρώματος εξόδου είναι z_l . Η τιμή βάρους του δικτύου μεταξύ του κόμβου εισόδου και του κόμβου του στρώματος απόκρυψης είναι w_{ij} και η τιμή βάρους του δικτύου μεταξύ των κόμβων του στρώματος απόκρυψης και του στρώματος εξόδου είναι v_{lj} . Όταν η αναμενόμενη τιμή του κόμβου εξόδου είναι t_l , η ενεργή συνάρτηση είναι η $f(\cdot)$. Ο υπολογιστικός τύπος του μοντέλου εκφράζεται ως εξής:

•Forward propagation: output of computer network

Output of the node of hide layer:

$$y_j = f(\sum_i w_{ij} x_i - \theta_j) = f(net_j) \text{ συμπεριλαμβανομένου } net_j = \sum_i w_{ij} x_i - \theta_j$$

Computational output of the output node:

$$z_l = f(\sum_j v_{lj} y_j - \theta_l) = f(net_l) \text{ συμπεριλαμβανομένου } net_l = \sum_j v_{lj} y_j - \theta_l$$

Error of the output node:

$$E = \frac{1}{2} \sum_l (t_l - z_l)^2 = \frac{1}{2} \sum_l (t_l - f(\sum_i v_{lj} y_j - \theta_l))^2 = \frac{1}{2} \sum_l (t_l - f(\sum_i v_{lj} f(\sum_i w_{ij} x_i - \theta_j) - \theta_l))^2$$

•Back propagation:

Η μέθοδος gradient descent υιοθετείται για τη ρύθμιση της τιμής βάρους όλων των στρωμάτων και ο αλγόριθμος εκμάθησης της τιμής βάρους εκφράζεται ως εξής:

Παραγωγή του κόμβου εξόδου μέσω της συνάρτησης σφάλματος:

$$\frac{\partial E}{\partial v_{lj}} = \sum_{k=l}^n \frac{\partial E}{\partial z_k} \cdot \frac{\partial z_k}{\partial v_{lj}} = \frac{\partial E}{\partial z_l} \cdot \frac{\partial z_l}{\partial v_{lj}}$$

Το E είναι μια συνάρτηση που περιέχει πολλά z_k , αλλά μόνο ένα z_l σχετίζεται με το v_{lj} και όλα τα z_k είναι ανεξάρτητα μεταξύ τους, σε αυτόν τον τύπο,

$$\frac{\partial E}{\partial z_l} = \frac{1}{2} \sum_k \left[-2(t_k - z_k) \cdot \frac{\partial z_k}{\partial z_l} \right] = -(t_l - z_l) \quad \text{και}$$

$$\frac{\partial z_l}{\partial v_{lj}} = \frac{\partial z_l}{\partial net_l} \cdot \frac{\partial net_l}{\partial v_{lj}} = f'(net_l) \cdot y_j$$

$$\text{Έτσι, } \frac{\partial E}{\partial v_{lj}} = -(t_l - z_l) \cdot f'(net_l) \cdot y_j$$

Ας υποθέσουμε ότι το σφάλμα του κόμβου εισόδου είναι:

$$\delta_l = (t_l - z_l) \cdot f'(net_l)$$

$$\text{Έτσι, } \frac{\partial E}{\partial v_{lj}} = -\delta_l \cdot y_j$$

Απόκλιση του κόμβου του στρώματος απόκρυψης από τη συνάρτηση σφάλματος:

$$\frac{\partial E}{\partial w_{ij}} = \sum_i \sum_j \frac{\partial E}{\partial z_l} \cdot \frac{\partial z_l}{\partial y_j} \cdot \frac{\partial y_j}{\partial w_{ij}}$$

Το E είναι μια συνάρτηση που περιέχει πολλά z_l . στοχεύει σε συγκεκριμένο w_{ji} , που αντιστοιχεί σε ένα y_j , και σχετίζεται με όλα τα z_l , σε αυτόν τον τύπο,

$$\frac{\partial E}{\partial z_l} = \frac{1}{2} \sum_k \left[-2(t_k - z_k) \cdot \frac{\partial z_k}{\partial z_l} \right] = -(t_l - z_l)$$

Έτσι,

$$\frac{\partial E}{\partial w_{ij}} = - \sum_l (t_l - z_l) \cdot f'(net_l) \cdot v_{lj} \cdot f'(net_j) \cdot x_i = - \sum_l \delta_l \cdot v_{lj} \cdot f'(net_j) \cdot x_i$$

Ας υποθέσουμε ότι το σφάλμα του κόμβου του στρώματος απόκρυψης είναι:

$$\delta'_j = f'(net_j) \cdot \sum_l \delta_l \cdot v_{lj}$$

$$\text{Συνεπώς, } \frac{\partial E}{\partial w_{ij}} = - \delta'_j \cdot x_i$$

Δεδομένου ότι η τροποποίηση του βάρους Δv_{lj} και Δw_{ij} είναι ανάλογη με τις συναρτήσεις σφάλματος και κατεβαίνει κατά μήκος της κλίσης, ο τύπος που δείχνει την τροποποίηση του βάρους του στρώματος δέρματος και του στρώματος εξόδου εκφράζεται ως εξής:

$$\Delta v_{lj} = - \eta \cdot \frac{\partial E}{\partial v_{lj}} = \eta \cdot \delta_l \cdot y_j$$

Σε αυτόν τον τύπο, το η αντιπροσωπεύει το ρυθμό μάθησης. Ο τύπος που δείχνει την τροποποίηση μεταξύ του επιπέδου εισόδου και του επιπέδου απόκρυψης εκφράζεται ως εξής:

$$\Delta w_{ij} = - \eta' \cdot \frac{\partial E}{\partial w_{ij}} = \eta' \cdot \delta'_j \cdot x_i \text{ και } \delta'_j = f'(net_j) \cdot \sum_l \delta_l \cdot v_{lj}$$

Σε αυτόν τον τύπο, το η' αντιπροσωπεύει το ρυθμό εκμάθησης. Το $\sum_l \delta_l \cdot v_{lj}$ στο σφάλμα κόμβου του στρώματος απόκρυψης, δ'_j εκφράζει ότι το σφάλμα δ_l του κόμβου εξόδου z_l διαδίδεται ξανά μέσω της τιμής βάρους v_{lj} στον κόμβο y_j για να γίνει το σφάλμα του κόμβου του στρώματος απόκρυψης.

Τέλος, Η τιμή threshold θ είναι επίσης μια τιμή παραλλαγής και πρέπει επίσης να τροποποιηθεί ενώ τροποποιείται η τιμή βάρους. η θεωρία που εφαρμόζεται είναι η ίδια με αυτή που χρησιμοποιείται για την τροποποίηση της τιμής βάρους.

• Παραγωγή του ορίου του κόμβου εξόδου με συνάρτηση σφάλματος:

$$\frac{\partial E}{\partial \theta_l} = \frac{\partial E}{\partial z_l} \cdot \frac{\partial z_l}{\partial \theta_l}$$

$$\text{όπου ισχύει } \frac{\partial z_l}{\partial \theta_l} = \frac{\partial z_l}{\partial net_l} \cdot \frac{\partial net_l}{\partial \theta_l} = f'(net_l) \cdot (-1) = -f'(net_l)$$

Με αυτόν τον τρόπο, ο τύπος που εκφράζει την τροποποίηση της τιμής threshold είναι:

$$\Delta \theta_l = \eta \cdot \frac{\partial E}{\partial \theta_l} = \eta \cdot \delta_l$$

Πιο συγκεκριμένα, $\theta_l(k + 1) = \theta_l(k) + \Delta\theta_l = \theta_l(k) + \eta \cdot \delta_l$

• Παραγωγή του ορίου του κόμβου του στρώματος απόκρυψης από τη συνάρτηση σφάλματος:

$$\frac{\partial E}{\partial \theta_j} = \sum_l \frac{\partial E}{\partial z_l} \cdot \frac{\partial z_l}{\partial y_j} \cdot \frac{\partial y_j}{\partial \theta_j}$$

όπου ισχύει:

$$\frac{\partial z_j}{\partial \theta_j} = \frac{\partial z_j}{\partial net_j} \cdot \frac{\partial net_j}{\partial \theta_j} = f'(net_j) \cdot (-1) = -f'(net_j)$$

Έτσι:

$$\frac{\partial E}{\partial \theta_j} = \sum_l (t_l - z_l) \cdot f'(net_l) \cdot v_{lj} \cdot f'(net_j) = - \sum_l \delta_l \cdot v_{lj} \cdot f'(net_j) = \delta'_j$$

Ο τύπος που εκφράζει την τροποποίηση της τιμής threshold είναι:

$$\Delta\theta_j = \eta' \cdot \frac{\partial E}{\partial \theta_j} = \eta' \cdot \delta'_j$$

Συγκεκριμένα: $\theta_j(k + 1) = \theta_j(k) + \Delta\theta_j = \theta_j(k) + \eta' \cdot \delta'_j$.

Συνολικά, η διαδικασία εκπαίδευσης των DNNs αντιπροσωπεύει έναν σημαντικό σταυροδρόμι όπου η μαθηματική θεωρία, ο υπολογιστικός χειρισμός και η δημιουργική σκέψη συνεργάζονται για να δημιουργήσουν ευφυή συστήματα που μπορούν να αντιμετωπίσουν προκλήσεις και να προσεγγίσουν λύσεις σε ποικίλα προβλήματα.

3.3.2 Προκλήσεις στην Εκπαίδευση των DNNs

Πέρα από την αναγνώριση των Deep Neural Networks (DNNs) ως ισχυρών εργαλείων στην ανάλυση, επεξεργασία και πρόβλεψη πολύπλοκων δεδομένων, προκύπτουν πολλές προκλήσεις που διαμορφώνουν τη διαδικασία εκπαίδευσής τους. Αυτές οι προκλήσεις, είναι σημαντικές για τη δημιουργία αξιόπιστων και αποτελεσματικών DNN μοντέλων και αντικατοπτρίζουν την ευρύτερη πολυπλοκότητα της εκπαίδευσης τέτοιων δικτύων.

Ένα από τα βασικά ζητήματα είναι η υπερεκπαίδευση, μια κατάσταση κατά την οποία το DNN μαθαίνει να "μνημονεύει" τα δεδομένα εκπαίδευσης αντί να γενικεύει.[\[55\]](#) Αυτό οδηγεί σε εξαιρετικά καλές επιδόσεις πάνω στα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση, αλλά και σε δυσκολίες στην αναγνώριση νέων δεδομένων. Τεχνικές κανονικοποίησης, όπως η dropout, εφαρμόζονται για να αντιμετωπίσουν το πρόβλημα αυτό, μειώνοντας την πιθανότητα υπερεκπαίδευσης και προωθώντας την παραγωγή γενικών χαρακτηριστικών.

Εκτός αυτού, η εκπαίδευση των βαθιών νευρωνικών δικτύων απαιτεί μεγάλες ποσότητες δεδομένων και υψηλή υπολογιστική ισχύ. Βαθιά δίκτυα με πολλά επίπεδα επιτείνουν αυτές τις

απαιτήσεις, με τον υπολογιστικό φόρτο και τον αριθμό επαναλήψεων να αυξάνονται. Η τεχνική των μίνι-παρτίδων (mini-batch training) βοηθά στην αποδοτικότερη εκπαίδευση, επιτρέποντας τη χρήση πιο αποτελεσματικών αλγορίθμων βελτιστοποίησης και μειώνοντας την απαιτούμενη μνήμη.[52]

Για την αντιμετώπιση αυτών των προκλήσεων, η εξέλιξη της τεχνολογίας προσφέρει νέα εργαλεία. Η εκμετάλλευση υλικού επιτάχυνσης, όπως γραφικές κάρτες, βελτιστοποιεί την εκπαίδευση πολύπλοκων μοντέλων.[49] Παράλληλα, η συνεχής έρευνα στην μηχανική μάθηση αναπτύσσει νέες μεθοδολογίες και τεχνικές προσέγγισης για την αντιμετώπιση των προκλήσεων.

Στο σύνολο του, η διαδικασία εκπαίδευσης DNNs αποδεικνύεται περίπλοκη και απαιτητική, αλλά αναμφίβολα απαραίτητη. Η κατανόηση και η αντιμετώπιση των προκλήσεων αυτών είναι ουσιώδους σημασίας για τη συνεχή βελτίωση των DNNs και την επίλυση πολύπλοκων προβλημάτων.

3.4 Εφαρμογές των DNNs

Τα Deep Neural Networks (DNNs) έχουν επιτύχει εντυπωσιακά αποτελέσματα σε πολλούς τομείς. Ο τρόπος με τον οποίο αυτά τα δίκτυα μπορούν να μοντελοποιήσουν πολύπλοκα πρότυπα, να αντλούν υψηλού επιπέδου χαρακτηριστικά και να προσαρμόζονται σε νέα δεδομένα έχει μετασχηματίσει την τεχνολογική και επιστημονική κοινότητα.

Αποτελούν μία από τις πιο σημαντικές εξελίξεις στον τομέα της τεχνητής νοημοσύνης και της μηχανικής μάθησης. Βασίζονται στο μοντέλο του ανθρώπινου εγκεφάλου και έχουν τη δυνατότητα να εξάγουν υψηλού επιπέδου αναπαραστάσεις από πολύπλοκα δεδομένα.[43] Τα DNNs έχουν επηρεάσει σημαντικά την τεχνολογία, την επιστήμη και την κοινωνία, επιτρέποντας την ανάπτυξη καινοτόμων εφαρμογών σε διάφορους τομείς.

Ανάμεσα στις εφαρμογές τους περιλαμβάνονται η επεξεργασία εικόνας, όπως η αναγνώριση αντικειμένων και προσώπων[40], η επεξεργασία φυσικής γλώσσας, όπως η μετάφραση κειμένου και η ανάλυση συναισθήματος[55], καθώς και η ρομποτική και τα αυτόνομα οχήματα[49]. Επίσης, τα DNNs χρησιμοποιούνται ευρέως σε συστήματα συστάσεων για την παροχή εξατομικευμένων συστάσεων στους χρήστες.[50]

Σε αυτή την ενότητα, θα εξετάσουμε αναλυτικά κάποιες από τις κυριότερες εφαρμογές των DNNs που έχουν διαμορφώσει την σύγχρονη τεχνολογία και κοινωνία.

3.4.1 Εικόνα και Βίντεο

Μία από τις πιο εντυπωσιακές εφαρμογές των DNNs βρίσκεται στην επεξεργασία εικόνας και βίντεο. Με τη χρήση συνελκτικών νευρωνικών δικτύων, είναι δυνατή η αναγνώριση αντικειμένων, προσώπων, σκηνών και περιβάλλοντων σε εικόνες και βίντεο. Αυτό βρίσκει εφαρμογή σε αυτόνομα οχήματα, παρακολούθηση ασφαλείας, ανάλυση κινήσεων και οπτική αναγνώριση. Η επανάσταση που προκάλεσαν τα DNNs στον τομέα της επεξεργασίας εικόνας και βίντεο είναι ανυπολόγιστη. Η ικανότητά τους να εξάγουν υψηλού επιπέδου χαρακτηριστικά από τα οπτικά δεδομένα έχει ανοίξει νέες προοπτικές σε πολλούς τομείς.[51][40]

Στην ανίχνευση και ταξινόμηση αντικειμένων σε εικόνες, τα DNNs έχουν επιτύχει απίστευτα αποτελέσματα. Αυτό δημιουργεί τη δυνατότητα για αυτόματη επεξεργασία και ανάλυση των οπτικών δεδομένων, με εφαρμογές που εκτείνονται από την αυτόνομη οδήγηση μέχρι την ασφαλή αναγνώριση πεζών σε πραγματικό χρόνο. Είναι εντυπωσιακό να παρατηρεί κανείς πώς η

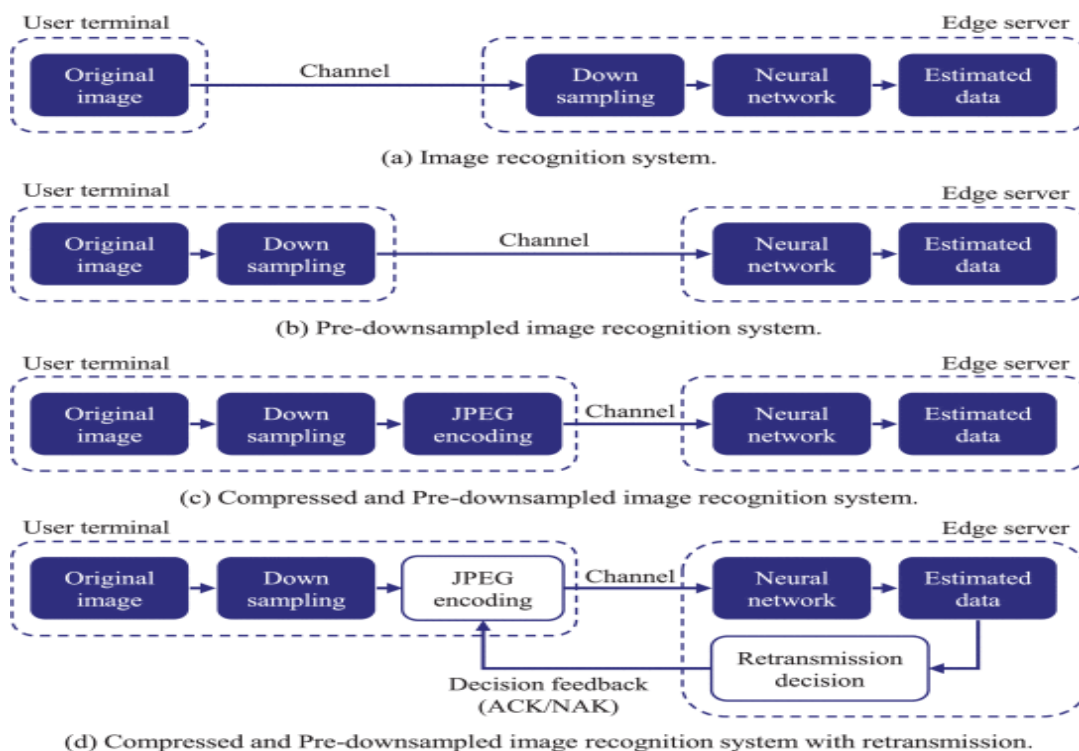
τεχνολογία αυτή μπορεί να εντοπίσει και να κατηγοριοποιήσει αντικείμενα με ανθρώπινη ακρίβεια, αναπαριστώντας ένα βασικό στοιχείο της ανθρώπινης νοημοσύνης. [51][40]

Εκτός από την ταξινόμηση, η ανάλυση εικόνων με τη χρήση DNNs έχει κάνει σημαντική πρόοδο στη διάγνωση ιατρικών προβλημάτων. Η ικανότητα αυτή να αναγνωρίζει σημάνσεις προειδοποίησης σε ιατρικές εικόνες, όπως ακτινογραφίες, απεικονίσεις μαγνητικής τομογραφίας και μικροσκοπικές εικόνες, έχει δυνητικά επαναστατική επίδραση στον τομέα της ιατρικής. Οι ιατροί μπορούν να χρησιμοποιήσουν αυτές τις τεχνικές για την πρόωρη ανίχνευση ασθενειών και για την προσδιορισμό ακόμα και των πιο περίπλοκων κλινικών σεναρίων.

Η εξέλιξη αυτή στην επεξεργασία εικόνων και βίντεο με τη χρήση DNNs διευρύνει το φάσμα των δυνατοτήτων σε περιβάλλοντα που απαιτούν υψηλή ακρίβεια και αυξημένη αξιοπιστία. Είναι πραγματικά συναρπαστικό να παρακολουθεί κανείς τον τρόπο με τον οποίο η τεχνολογία αυτή συμβάλλει στη βελτίωση της ανθρώπινης ζωής, επιτρέποντας την ανάπτυξη εξειδικευμένων εφαρμογών που ανταποκρίνονται σε κρίσιμες ανάγκες της κοινωνίας.

Το άρθρο “[Edge Computing-Assisted DNN Image Recognition System With Progressive Image Retransmission](#)” των Mutsuki Nakahara, Mai Nishimura, Yoshitaka Ushiku, κ.α., δίνει ένα παράδειγμα της εφαρμογής των DNN στον αναφερόμενο τομέα, προτείνοντας μία edge-assisted μέθοδο αναγνώρισης εικόνας με συμπίεση εικόνας και προοδευτική αναμετάδοση.

Παρακάτω περιγράφεται η έννοια του προτεινόμενου συστήματος. Το Σχήμα 1 δείχνει το υπονήφιο μοντέλο για το edge computing σύστημα για τη μείωση της κυκλοφορίας. Το Σχήμα 1 (α) είναι η διαμόρφωση της κανονικής αναγνώρισης εικόνας με edge computing. Το τερματικό χρήστη έχει την αρχική εικόνα και τη στέλνει στον edge server. Πριν από την αναγνώριση της εικόνας με το DNN, ο edge server χρησιμοποιεί μια μέθοδο downsampling για να αντιστοιχίσει το μέγεθος της εικόνας με το μέγεθος εισόδου του DNN. Για παράδειγμα, το μέγεθος εισόδου στο ResNet, ένα δημοφιλές μοντέλο του DNN, είναι 224×224 pixel. Παρόλο που το Σχήμα 1 δείχνει μόνο τη μέθοδο downsampling, η εικόνα γίνεται up-sampled όταν το μέγεθος της εικόνας είναι μικρότερο από το μέγεθος εισόδου.

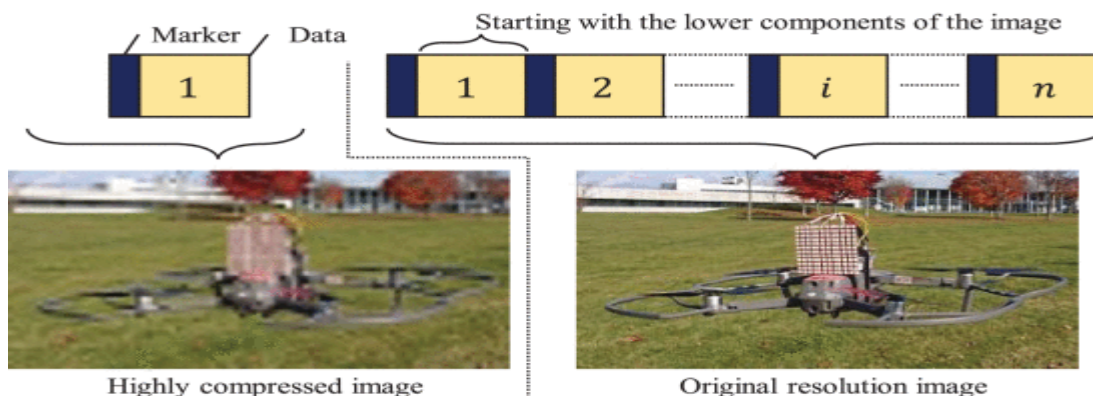


Σχήμα 1. Υποψήφιο μοντέλο για edge computing system για τη μείωση της κυκλοφορίας.

Στο Σχήμα 1 (β), το τερματικό χρήστη πραγματοποιεί εκ των προτέρων το downsampling. Αναμένεται μείωση του μεγέθους της εικόνας. Επιπλέον, το σύστημα δεν έχει καμία επίδραση στην ακρίβεια αναγνώρισης εικόνας. Εν τω μεταξύ, το μέγεθος των δεδομένων είναι ίσο με τον συνολικό αριθμό των pixel πολλαπλασιαζόμενο επί 24 bit. Ως αποτέλεσμα, αναμένεται μεγαλύτερη μείωση της κυκλοφορίας όταν το τερματικό χρήστη εκτελεί κωδικοποίηση JPEG πριν από τη μετάδοση, όπως φαίνεται στο Σχήμα 1 (γ). Ωστόσο, όταν χρησιμοποιείται συμπίεση με απώλειες, η ακρίβεια αναγνώρισης μειώνεται.

Το Σχήμα 1 (δ) δείχνει το προτεινόμενο σχέδιο αναμετάδοσης. Στο προτεινόμενο σχήμα, οι χρήστες κάνουν downsampling και συμπιέζουν εικόνες πριν τις στείλουν σε έναν edge server. Ο edge server χρησιμοποιεί το DNN για να αναγνωρίσει τις εικόνες. Στέλνει το μήνυμα αιτήματος αναμετάδοσης Image-NAK όταν η ακρίβεια της εκτίμησης πέσει κάτω από το προκαθορισμένο όριο. Τα Image-NAK-received user terminals επαναφέρουν τον ρυθμό συμπίεσης σε χαμηλότερη τιμή και στέλνουν ξανά τις εικόνες στον edge server. Όταν ο edge server επιτυγχάνει επαρκή ακρίβεια, μεταδίδει ένα μήνυμα επιβεβαίωσης γνωστό ως Image-ACK και τερματίζει τη διαδικασία προώθησης. Επιπλέον, ο edge server τερματίζει τη διαδικασία όταν ο αριθμός των αναμεταδόσεων εικόνας φτάσει σε ένα συγκεκριμένο όριο. Αξίζει να σημειωθεί ότι τα μηνύματα ACK και NAK σύνδεσης TCP επικοινωνούνται στο δίκτυο. Το ACK και το NAK είναι διαφορετικά μηνύματα του Image-ACK και του Image-NAK.

Στη συνέχεια, εισάγονται δύο τύποι μορφής συμπίεσης εικόνας. Η πρώτη είναι μια βασική κωδικοποίηση JPEG τυποποιημένη από το ISO/IEC JTC 1/SC 29. Η άλλη είναι μια προοδευτική μορφή JPEG. Το προοδευτικό JPEG αποθηκεύει τα δυαδικά δεδομένα με τη σειρά, ξεκινώντας από τα στοιχεία χαμηλότερης ανάλυσης της εικόνας. Με άλλα λόγια, η εικόνα μπορεί να ανοίξει ακόμα και όταν τα δυαδικά δεδομένα κόβονται από την αρχή στη μέση. Όσο πιο σύντομα κόβονται τα δυαδικά δεδομένα, τόσο πιο τραχιά είναι η εικόνα. Εν τω μεταξύ, η τυπική μορφή JPEG δεν μπορεί να ανοίξει την εικόνα όταν τα δυαδικά δεδομένα κόβονται στη μέση. Το προοδευτικό JPEG ενσωματώνει δείκτες στα δυαδικά δεδομένα. Ο ρυθμός συμπίεσης υπολογίζεται στη θέση του δείκτη. Για παράδειγμα, το Σχήμα 2 δείχνει την εξαιρετικά συμπιεσμένη εικόνα με μια πρώτη θέση δείκτη από την αρχή των δυαδικών δεδομένων και την αρχική εικόνα με όλα τα δυαδικά δεδομένα. Στο Σχήμα 2, υποθέτουμε ότι ο μέγιστος αριθμός δεικτών είναι n . Τα στοιχεία υψηλότερης συχνότητας απουσιάζουν από την εξαιρετικά συμπιεσμένη εικόνα. Ο ρυθμός συμπίεσης του προτεινόμενου σχήματος ελέγχεται με αναφορά στους δείκτες. Σε αυτό το άρθρο, ο αριθμός του δείκτη αναφέρεται ως το βήμα συμπίεσης.



Σχήμα 2. Παράδειγμα προοδευτικής μορφής JPEG.

Έπειτα παρουσιάζεται η Λειτουργία των User Terminals.

•Baseline JPEG Case

Εκ των προτέρων, τα User terminals εκτελούν downsampling και μετατρέπουν τη μορφή εικόνας σε βασικό JPEG. Το τερματικό χρήστη ρυθμίζει την ποιότητα στο εύρος 0–100% του JPEG για τη συμπίεση της εικόνας. Το επίπεδο ποιότητας είναι προκαθορισμένο. Όταν το user terminal λαμβάνει το Image-NAK, επανασυμπίεζει την εικόνα χρησιμοποιώντας τον επόμενο καθορισμένο ρυθμό συμπίεσης. Αυτή η φάση επαναλαμβάνεται μέχρι να ληφθεί Image-ACK ή να επιτευχθεί το όριο αναμετάδοσης.

•Progressive JPEG Case

Ο Αλγόριθμος 1 δείχνει τη λειτουργία των user terminals στο προτεινόμενο σχήμα. Η εικόνα σε προοδευτική μορφή JPEG μετατρέπεται σε δυαδικά δεδομένα D_{orig} . Στη συνέχεια, ο αλγόριθμος διαβάζει το τρέχον βήμα συμπίεσης σ_c , όπου c είναι ο αριθμός του βήματος συμπίεσης. Το αρχικό βήμα συμπίεσης θεωρείται ως σ_i . Η λήψη του Image-NAK υποδηλώνει ότι η διαδικασία προώθησης έχει ήδη πραγματοποιηθεί αρκετές φορές. Έτσι, ο αλγόριθμος εξάγει τα δυαδικά δεδομένα D_p σε θέσεις από το προηγούμενο βήμα συμπίεσης σ_c στο τρέχον καθορισμένο βήμα συμπίεσης σ_f . Όταν το Image-NAK δεν έχει ληφθεί ποτέ και αυτή είναι η πρώτη φορά που προωθείται η εικόνα, ο χρήστης προωθεί τα δυαδικά δεδομένα από την αρχή των δεδομένων στο αρχικό βήμα συμπίεσης σ_i .

Αλγόριθμος 1. Τερματικά χρήστη σε προοδευτική αναμετάδοση:

Input: Binary data D_{orig} with or without NAK N_f

Output: Requested partial binary data D_p

Read: Current compression step σ_c

if Receive NAK N_f ? **then**

Extract: Binary data D_p between comp. steps σ_c and σ_f in D_{orig}

Update: $\sigma_c \leftarrow \sigma_f$

else

return Requested partial binary data D_p

Ενώ το user terminal συνεχίζει να στέλνει τη συμπιεσμένη εικόνα στον edge server μέχρι να ολοκληρωθεί η αναγνώριση εικόνας, τα επικαλυπτόμενα δεδομένα (overlapped data) δεν προωθούνται λόγω της προοδευτικής μετάδοσης. Από αυτή την άποψη, το προτεινόμενο σχέδιο συμβάλλει στη μείωση της κυκλοφορίας.

3.4.2 Φυσική Γλώσσα

Στον τομέα της επεξεργασίας φυσικής γλώσσας (Natural Language Processing - NLP), τα DNNs έχουν επίσης επαναστατήσει στην προσέγγιση προβλημάτων όπως η μηχανική μετάφραση, η αυτόματη δημιουργία κειμένου και η ανάλυση συναισθηματικού τόνου. Αυτό έχει εφαρμογή σε πλατφόρμες κοινωνικών μέσων, ανάλυση κειμένου και διαχείριση περιεχομένου. Η επεξεργασία φυσικής γλώσσας αποτελεί έναν από τους πιο ενδιαφέροντες τομείς της τεχνητής νοημοσύνης, και τα DNNs έχουν επανασχεδιάσει εντελώς τον τρόπο με τον οποίο προσεγγίζουμε και επεξεργαζόμαστε τον φυσικό λόγο.[\[55\]\[47\]](#)

Η μηχανική μετάφραση αποτελεί ένα από τα πιο εντυπωσιακά παραδείγματα εφαρμογής των DNNs στην επεξεργασία φυσικής γλώσσας.[\[56\]](#) Αντιμετωπίζοντας την πρόκληση της μετάφρασης μεταξύ διαφορετικών γλωσσών, τα DNNs έχουν καταφέρει να δημιουργήσουν μοντέλα που μπορούν να μεταφράσουν κείμενα με εντυπωσιακή ακρίβεια. Αυτό συμβαίνει χάρη στην ικανότητα τους να αντλούν υψηλού επιπέδου χαρακτηριστικά από τον φυσικό λόγο και να επιλέγουν τις κατάλληλες λέξεις και δομές για τη μετάφραση.[\[52\]](#)

Επιπλέον, τα DNNs έχουν καταφέρει να αντιμετωπίσουν προκλήσεις στην αναγνώριση εξελισσόμενου λόγου.[\[60\]](#) Δηλαδή, μπορούν να καταλάβουν τι σκοπό έχει ένα κείμενο, ποια είναι η πρόθεση του συγγραφέα. Αυτό είναι ιδιαίτερα χρήσιμο για την ανάλυση κειμένων στα κοινωνικά δίκτυα, την αναζήτηση πληροφοριών και την αυτόματη ταξινόμηση άρθρων ή κειμένων.[\[50\]\[42\]](#)

Η σύνθεση κειμένου είναι μία ακόμα πρωτοφανής εφαρμογή των DNNs.[\[57\]](#) Τα μοντέλα αυτά μπορούν να δημιουργήσουν νέα κείμενα, ποιητικά κείμενα, άρθρα και ακόμα και σενάρια, προσαρμόζοντας το στυλ και το περιεχόμενο σύμφωνα με τις απαιτούμενες ανάγκες.[\[54\]](#)

Τέλος, η δημιουργία αυτόματων απαντήσεων σε ερωτήσεις αποτελεί έναν ακόμα τρόπο εφαρμογής των DNNs. Αυτά τα μοντέλα μπορούν να αναλύσουν το περιεχόμενο μιας ερώτησης και να παράγουν ακριβείς και συνοπτικές απαντήσεις, βασισμένες στις πληροφορίες που έχουν αντλήσει από το διαδίκτυο. Συνολικά, οι εφαρμογές των DNNs στην επεξεργασία φυσικής γλώσσας έχουν ανοίξει νέες προοπτικές για την αυτοματοποίηση και βελτίωση της επικοινωνίας μεταξύ ανθρώπων και μηχανών.

3.4.3 Υγεία

Οι εφαρμογές των DNNs στον τομέα της υγείας είναι εξαιρετικά ενθαρρυντικές. Με τη χρήση βαθιών νευρωνικών δικτύων, είναι δυνατή η ανίχνευση και διάγνωση ασθενειών από εικόνες ιατρικών εξετάσεων, όπως αξονικές και μαγνητικές τομογραφίες.[\[41\]](#) Επίσης, χρησιμοποιούνται για την πρόβλεψη ιατρικών εκβάσεων και την αξιολόγηση αποτελεσμάτων θεραπειών.[\[48\]](#) Η εξέλιξη αυτή έχει επιφέρει δραματικές αλλαγές στη διάγνωση, την πρόληψη και τη θεραπεία των ασθενειών, ανοίγοντας νέους ορίζοντες για τη βελτίωση της υγείας του ανθρώπου.

Η ανάλυση εικόνων ιατρικών εξετάσεων αποτελεί ένα από τα πιο σημαντικά πεδία εφαρμογής των DNNs στην υγεία. Με τη βοήθεια της βαθιάς μηχανικής μάθησης, αυτά τα δίκτυα μπορούν να αναγνωρίσουν σημαντικά χαρακτηριστικά στις ιατρικές εικόνες που είναι δύσκολο για τον ανθρώπινο εγκέφαλο να ανιχνεύσει. Αυτό καθιστά δυνατή την έγκαιρη ανίχνευση παθήσεων, όπως καρκίνος, λοιμώξεις και νευρολογικές διαταραχές.[\[51\]](#) Τα DNNs μπορούν να προσδιορίσουν ανωμαλίες και να παρέχουν πολύτιμες πληροφορίες στους ιατρούς, κάνοντας τη διαγνωστική διαδικασία ακριβέστερη και πιο αξιόπιστη.

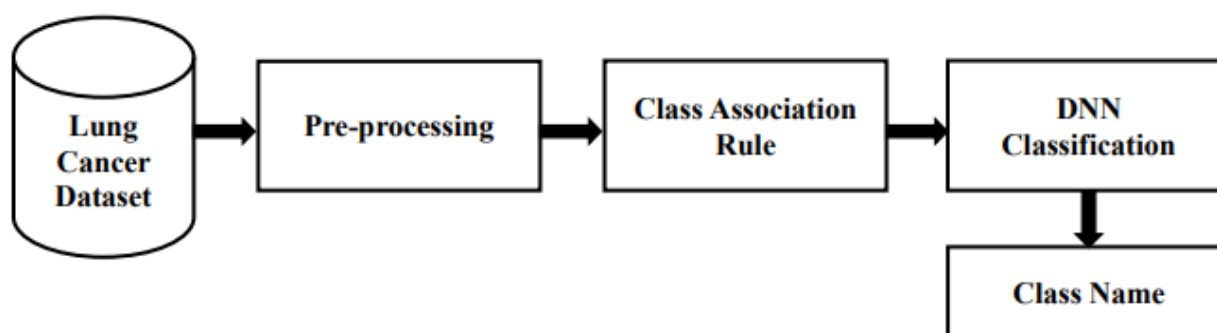
Επιπλέον, τα DNNs έχουν τη δυνατότητα να αναλύουν ιατρικά δεδομένα για την πρόβλεψη εξελίξεων των ασθενειών.[\[55\]](#) Χρησιμοποιώντας την ισχύ της μηχανικής μάθησης, αυτά τα δίκτυα μπορούν να αντλήσουν πολύτιμα πρότυπα από μεγάλα σύνολα δεδομένων και να

προβλέψουν την πιθανότητα εξέλιξης μιας ασθένειας. Αυτό βοηθά τους ιατρούς να λαμβάνουν πιο ενημερωμένες αποφάσεις για τη θεραπεία και τη διαχείριση των ασθενειών.

Επιπλέον, τα DNNs συνεισφέρουν στη βελτιστοποίηση των θεραπειών με βάση τα ιατρικά δεδομένα. Μπορούν να αναλύουν την απόκριση των ασθενών σε διάφορες θεραπευτικές προσεγγίσεις και να προτείνουν την καλύτερη δυνατή λύση. Αυτό έχει σημαντική σημασία στην παραμετροποίηση των θεραπειών, διασφαλίζοντας όχι μόνο την αποτελεσματικότητά τους, αλλά και τη μείωση των πιθανών παρενεργειών.

Το άρθρο “[Implementation of Lung Cancer Diagnosis based on DNN in Healthcare System](#)” του Suhair Mohammed Zeki Abd Alsammed δίνει ένα παράδειγμα εφαρμογής των DNN στον τομέα της υγείας, το οποίο προτείνει ένα σύστημα που βασίζεται στο DNN με στόχο την ακριβή διάγνωση του καρκίνου του πνεύμονα, βοηθώντας τους γιατρούς να αποτρέψουν σφάλματα ενώ ταυτοποιούν και ταξινομούν τους όγκους από διαφορετικά σύνολα δεδομένων.

Στο προτεινόμενο σύστημα διάγνωσης, η είσοδος είναι ένα σύνολο δεδομένων και η έξοδος είναι η ταξινόμηση του καρκίνου του πνεύμονα. Το γενικό block diagram του προτεινόμενου συστήματος απεικονίζεται στο Σχήμα 1.



Σχήμα 1. General block diagram of the proposed DNN based system.

Αυτό το προτεινόμενο σύστημα περιλαμβάνει διάφορες φάσεις, απόκτηση δεδομένων για τον καρκίνο του πνεύμονα (lung cancer dataset acquisition)- αντιπροσωπεύει πόσα δεδομένα παρέχονται, προεπεξεργασία (pre-processing)- βελτιώνει σημαντικά την ποιότητα των μοτίβων και μειώνει τον χρόνο που απαιτείται για την πραγματική εξόρυξη δεδομένων και εφαρμογή class association rule και ταξινόμηση DNN (DNN classification).

•Lung Cancer Dataset Pre-processing:

Η διαδικασία της προεπεξεργασίας δεδομένων λειτουργεί στον χειρισμό των δεδομένων εισόδου πριν από τις κύριες διεργασίες. Γενικά, ορισμένοι τύποι δεδομένων είναι ανεπεξέργαστα (raw), ασυνεπή (inconsistent), ελλιπή (incomplete) και θορυβώδη (noisy) και δεν είναι δυνατό να χρησιμοποιηθούν άμεσα στην εξόρυξη δεδομένων υγειονομικής περίθαλψης, ή ακόμα και τα αποτελέσματα μεταγενέστερης ταξινόμησης τους είναι απαράδεκτα. Επομένως, για τη βελτίωση της απόδοσης της εξόρυξης δεδομένων υγειονομικής περίθαλψης, μια φάση προεπεξεργασίας παρουσιάζεται ως σημαντική φάση σε ολόκληρο το σύστημα. Αυτή η φάση περιλαμβάνει δύο διαδικασίες, τον χειρισμό των δεδομένων που λείπουν (handling missing data) και τον μετασχηματισμό δεδομένων με χρήση κανονικοποίησης (transforming data using normalization). Η πρώτη διαδικασία ολοκληρώνεται με την αντικατάσταση κάθε τιμής που λείπει με μηδενική

τιμή, ενώ η δεύτερη διεργασία υλοποιείται για τον προσδιορισμό της τιμής των χαρακτηριστικών που είναι αναποτελεσματική στη διαδικασία ταξινόμησης και αξίζει να είναι μηδέν. Η διαδικασία κανονικοποίησης περιλαμβάνει διάφορα βήματα.

Βήμα 1:

Εφαρμογή της κανονικοποίησης zero-mean (zero-mean normalization) στην οποία οι τιμές για ένα χαρακτηριστικό (p) μετασχηματίζονται ανάλογα με τον μέσο όρο (mean - m) και την τυπική απόκλιση (standard deviation - d) του p . Κάθε τιμή x_i του p μετατρέπεται σε y_i μέσω της ακόλουθης εξίσωσης.

$$y_i = \frac{x_i - m(p)}{d(p)}$$

Βήμα 2:

Εφαρμόζεται άλλη μία κανονικοποίηση, όπως δίνεται στην ακόλουθη εξίσωση.

$$y'_i = maximum \cdot \frac{(y_i + 1)}{2}$$

Βήμα 3:

Υπολογίζεται η συχνότητα κάθε τιμής που δημιουργήθηκε από το προηγούμενο βήμα.

Βήμα 4:

Υπολογίζεται η μέση τιμή των συχνοτήτων και συγκρίνεται η τιμή κάθε συχνότητας με τη μέση τιμή και, στη συνέχεια, λαμβάνονται μόνο οι τιμές μικρότερες από τη μέση τιμή.

Βήμα 5:

Εντοπίζονται οι ελάχιστες και μέγιστες τιμές και υπολογίζεται το εύρος ως εξής.

$$Range = maximum - minimum$$

Βήμα 6:

Εφαρμόζεται η συνάρτηση πυκνότητας πιθανότητας (Probability Density Function) για να βρεθεί η τιμή του εύρους κάθε ομάδας και να οριστούν οι τιμές της πρώτης ομάδας ως μηδέν.

•The Rules Generation:

Σε αυτή τη φάση, χρησιμοποιείται ένας αλγόριθμος που προέρχεται από τον αλγόριθμο Apriori που βασίζεται στην εξόρυξη κανόνων συσχέτισης (association rule mining) για τη δημιουργία ισχυρών κανόνων. Ο σκοπός της χρήσης ενός τέτοιου αλγορίθμου είναι να βρεθεί μια σχέση μεταξύ των χαρακτηριστικών και του χαρακτηριστικού κλάσης. Οι κατάλληλοι κανόνες είναι αυτοί με υψηλές confidence values. Αυτό συμβαίνει επειδή αυτοί οι κανόνες είναι συχνά μοτίβα επαγωγή από το σύνολο δεδομένων και υποτίθεται ότι περιέχουν χαρακτηριστικά που είναι σημαντικά για το χαρακτηριστικό της κλάσης. Τα απαιτούμενα βήματα για τη δημιουργία των κανόνων είναι τα εξής.

Βήμα 1:

Δόμηση κανόνων συσχέτισης για χαρακτηριστικά συχνών μοτίβων που λαμβάνονται από την προηγούμενη φάση. Συνεπώς, ένας class association rule έχει τη μορφή:

$$C \rightarrow c_i$$

όπου c_i υποδηλώνει το χαρακτηριστικό κλάσης και C υποδηλώνει $\{c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_n\}$, λαμβάνοντας υπόψη ότι η βάση δεδομένων $B = \{b_1, b_2, \dots, b_i\}$ περιλαμβάνει ένα σύνολο δεδομένων που κατέχουν n χαρακτηριστικό και k ετικέτα κλάσης, $b = \{c_1, c_2, \dots, c_n, l_k\}$, όπου $k=1, 2, \dots, m$;

Βήμα 2:

Υπολογίζεται το support και το confidence για όλους τους παραγόμενους κανόνες, όπου η τιμή support αντιπροσωπεύει τη συχνότητα όλων των δεδομένων που περιλαμβάνουν το σύνολο στοιχείων C και την ετικέτα κλάσης l και η τιμή confidence αντιπροσωπεύει τη συχνότητα των δεδομένων που περιλαμβάνουν την ετικέτα κλάσης l όταν τα δεδομένα κατέχουν το σύνολο στοιχείων C .

Βήμα 3:

Διαγράφονται οι κανόνες που έχουν χαρακτηριστικά στο consequence part τους που διαφωνούν με ένα χαρακτηριστικό κλάσης. Αυτό το βήμα είναι για την επιλογή των συχνών μοτίβων από κανόνες συσχέτισης που περιέχουν τόσο προγνωστικά χαρακτηριστικά όσο και χαρακτηριστικά κλάσης για την ικανοποίηση ελάχιστου support και ελάχιστων confidence thresholds.

•DNN Classification:

Σε αυτή τη φάση, η έξοδος της διαδικασίας class association rule χρησιμοποιείται ως είσοδος στον ταξινομητή DNN. Αυτό γίνεται με τη μετατροπή των παραγόμενων κανόνων σε ψηφιακά διανύσματα, όπου το μηδέν υποδηλώνει το χαρακτηριστικό που δεν είναι διαθέσιμο στον κανόνα, ενώ το ένα υποδεικνύει το χαρακτηριστικό που είναι διαθέσιμο στον κανόνα και αυτά τα διανύσματα αποθηκεύονται στη βάση δεδομένων.

Το σύστημα που βασίζεται στο DNN περιλαμβάνει δύο κύρια στάδια, την εκπαίδευση (training) και τη δοκιμή (testing). Το σύνολο δεδομένων χωρίζεται σε δύο μέρη, το πρώτο χρησιμοποιείται για την εκπαίδευση δεδομένων καρκίνου του πνεύμονα και το δεύτερο χρησιμοποιείται ως δοκιμαστικές περιπτώσεις για τον έλεγχο της απόδοσης του συστήματος ταξινόμησης. Σε αυτό το σύστημα που βασίζεται στο DNN, τα αποτελέσματα της διαδικασίας κανόνα συσχέτισης κλάσεων αποθηκεύονται στη βάση δεδομένων αφού μετατραπούν σε αριθμητικά διανύσματα. Αυτά τα διανύσματα χρησιμοποιούνται ως είσοδος στο μοντέλο DNN. Το προτεινόμενο μοντέλο DNN περιλαμβάνει οκτώ επίπεδα, επίπεδα εισόδου (input) και εξόδου (output) και έξι κρυφά επίπεδα (hidden layers).

Η διαδικασία εκπαίδευσης DNN βασίζεται στον αλγόριθμο backpropagation. Αυτή η διαδικασία αντιπροσωπεύει ένα σημαντικό μέρος του προτεινόμενου συστήματος. Η εκπαίδευση ξεκινά με τη διάδοση προς τα εμπρός (forward-propagation) που είναι μέρος της backpropagation όπου η τιμή κάθε νευρώνα στο πρώτο κρυφό στρώμα έχει ως αποτέλεσμα ένα άθροισμα όλων των εισόδων του πολλαπλασιαζόμενο με το αντίστοιχο βάρος, τότε το αποτέλεσμα αθροίζεται με μία bias value. Η forward-propagation τερματίζεται με τον υπολογισμό όλων των νευρώνων στο στρώμα εξόδου. Η backpropagation ξεκινά όταν η πραγματική έξοδος δεν ικανοποιεί την έξοδο

στόχο, όπου τα σφάλματα υπολογίζονται στο επίπεδο εξόδου για να προσαρμόσουν τα βάρη κάθε επιπέδου σύμφωνα με την κάθοδο βαθμίδας σφάλματος και να διαδοθούν πίσω αυτά τα σφάλματα στα κρυφά στρώματα και στο στρώμα εισόδου. Αυτή η διαδικασία τερματίζεται όταν τα σφάλματα εξόδου μειωθούν σε αποδεκτά σφάλματα. Η έξοδος αυτής της φάσης είναι το βάρος που χρησιμοποιείται στη φάση της δοκιμής.

Προκειμένου να αξιολογηθεί η ποιότητα της προτεινόμενης εκπαίδευσης του συστήματος, όπου το υποσύνολο δεδομένων χρησιμοποιείται για δοκιμή, αυτό το στάδιο είναι σημαντικό, καθώς ταξινομεί κάθε περίπτωση χωρίς ετικέτα. Έτσι, ο ταξινομητής (classifier) χρησιμοποιεί την ίδια διαδικασία που χρησιμοποιήθηκε στο στάδιο εκπαίδευσης (μόνο forward-propagation). Ο πίνακας των παρατηρήσεων αντιπροσωπεύει το σύνολο των περιπτώσεων όπου κάθε στήλη συσχετίζει ένα χαρακτηριστικό. Επομένως, το instance θα πρέπει να έχει τον ίδιο αριθμό χαρακτηριστικών με τα δεδομένα εκπαίδευσης. Τα κύρια βήματα του ταξινομητή DNN στο στάδιο εκπαίδευσης είναι τα εξής:

Βήμα 1:

Επιλέγονται τα βάρη αρχικοποίησης (ρυθμίζονται σε μικρές τυχαίες τιμές, όχι μηδενικά).

Βήμα 2:

Εφαρμόζεται η forward-propagation, στην οποία πρώτα κάθε μονάδα εισόδου ($x_i, i = 1, \dots, n$) λαμβάνει το σήμα εισόδου x_i και το μεταδίδει σε όλες τις μονάδες στο πρώτο κρυφό στρώμα (hidden units).

Βήμα 2.1:

Υπολογίστε την τιμή κάθε κρυφής μονάδας ($Z_j, j = 1, \dots, n$) χρησιμοποιώντας την ακόλουθη εξίσωση.

$$v_j(n) = w_{oj} + \sum_{i=0}^m w_{ij}(n)y_i(n)$$

όπου το m υποδηλώνει τον πλήρη αριθμό των εισόδων που εφαρμόζονται στον νευρώνα j (excepting bias), το $w_{ij}(n)$ δείχνει το βάρος που συνδέει τον νευρώνα i με τον νευρώνα j , το w_{oj} υποδεικνύει την τιμή της πόλωσης και το $y_i(n)$ δείχνει το σήμα εισόδου του νευρώνα i . Στη συνέχεια, εφαρμόζει μια ενεργή λειτουργία ($\tanh$) για την εύρεση του σήματος εξόδου και στέλνει αυτό το σήμα σε όλες τις μονάδες στο επόμενο κρυφό στρώμα.

Βήμα 2.2:

Επαναλαμβάνεται το βήμα 2.1 σε όλα τα κρυφά επίπεδα, το σήμα εξόδου από το τελευταίο κρυφό στρώμα στέλνεται σε όλες τις μονάδες στο επόμενο επίπεδο (output units).

Βήμα 2.3:

Υπολογίζεται η τιμή κάθε μονάδας εξόδου ($y_k, k=1, \dots, n$) και εφαρμόζεται η συνάρτηση softmax για να υπολογίσετε το σήμα εξόδου χρησιμοποιώντας την ακόλουθη εξίσωση.

$$y_i(n) = f_j(v_j(n))$$

Κάθε μονάδα εξόδου (y_k , $k=1, \dots, n$) λαμβάνει μια επιθυμητή κλάση που αντιστοιχεί στην κλάση εκπαίδευσης εισόδου και συγκρίνει με την έξοδο στόχο $t_j(n)$ για να βρει το σφάλμα της $e_j(n)$ για τον j^{th} νευρώνα εξόδου όπως στην ακόλουθη εξίσωση

$$e_j(n) = t_j(n) - y_j(n)$$

Βήμα 3:

Εφαρμογή Back-propagation που ξεκινά με το επίπεδο εξόδου περνώντας τα σήματα σφάλματος προς τα αριστερά μέσω του δικτύου, από το ένα στρώμα στο άλλο.

Βήμα 3.1:

Υπολογίζεται η διαβάθμιση $\delta_j(n)$ για τον νευρώνα j είναι κόμβος εξόδου και για τον νευρώνα j είναι ένας κρυφός κόμβος, όπως δίνεται στις ακόλουθες εξισώσεις.

$$\delta_j(n) = e_j(n)f'_j(v_j(n)) \text{ και } \delta_j(n) = f'_j(v_j(n)) \sum_k \delta_k(n)w_{kj}(n)$$

Βήμα 3.2:

Υπολογίζεται ο κανόνας $\Delta w_{ij}(n)$ που διορθώνει το βάρος που συνδέει τον νευρώνα i με τον νευρώνα j και υπολογίστε τον term of bias correction που χρησιμοποιήθηκε για την ενημέρωση του b_{kj} , όπως δίνεται στις ακόλουθες εξισώσεις.

$$\Delta w_{ji} = \alpha \delta_j(n)y_i(n) \text{ και } \Delta w_{oj} = \alpha \delta_j(n)$$

Βήμα 3.3:

Ενημερώνονται τα βάρη του bias του δικτύου στο επίπεδο L , όπως δίνεται στις ακόλουθες εξισώσεις.

$$w_{jk}^{(L)}(n+1) = w_{jk}^{(L)}(n) + \alpha \delta_k^{(L)} z_j + \eta [w_{jk}^{(L)}(n) - w_{jk}^{(L)}(n-1)]$$
$$v_{ij}^{(L)}(n+1) = v_{ij}^{(L)}(n) + \alpha \delta_j^{(L)} x_i + \eta [v_{ij}^{(L)}(n) - v_{ij}^{(L)}(n-1)]$$

Όπου το α δείχνει το ρυθμό μάθησης και το η την ορμή.

Βήμα 3.4:

Αυξάνεται η επανάληψη και επαναλαμβάνεται η διαδικασία μέχρι να γίνει σταθερό το βάρος.

Κατά τη διάρκεια του σταδίου της δοκιμής, τα βήματα forward propagation επαναλαμβάνονται στα υπόλοιπα δεδομένα από το σύνολο δεδομένων χρησιμοποιώντας τα αποτελέσματα βαρών από τη φάση εκπαίδευσης.

Συνολικά, οι εφαρμογές των DNNs στον τομέα της υγείας έχουν αλλάξει τον τρόπο που αντιλαμβανόμαστε και προσεγγίζουμε την ιατρική περίθαλψη. Με την εξελιγμένη αυτή τεχνολογία, είμαστε σε θέση να εντοπίζουμε ασθένειες σε πρώιμα στάδια, να παρέχουμε ακριβείς προβλέψεις και να προσαρμόζουμε ατομικές θεραπείες. Αυτό ενισχύει την ποιότητα ζωής των ανθρώπων και συμβάλλει στην προαγωγή της υγείας και της ευημερίας.

3.4.4 Αυτόνομα Οχήματα

Τα DNNs αποτελούν τον πυρήνα της τεχνολογίας των αυτόνομων οχημάτων.[51] Με τη χρήση συνελκτικών νευρωνικών δικτύων και αισθητήρων, είναι δυνατή η αναγνώριση αντικειμένων και η πρόβλεψη της συμπεριφοράς των άλλων οδηγών. Αυτό είναι καίριο για την ασφάλεια και την αποτελεσματικότητα της αυτόνομης οδήγησης.

Η τεχνολογία των DNNs έχει επιφέρει μεταστροφή στον τομέα της αυτονομίας των οχημάτων.[49] Η αυτόνομη οδήγηση είναι ένας φιλόδοξος στόχος που απαιτεί την ακριβή ανίχνευση του περιβάλλοντος, την πρόβλεψη της κίνησης άλλων οχημάτων και πεζών, καθώς και τη λήψη ασφαλών αποφάσεων σε πραγματικό χρόνο. Σε αυτόν τον τομέα, τα DNNs αποτελούν τον πυρήνα της τεχνολογίας που επιτρέπει στα αυτόνομα οχήματα να λειτουργούν με ασφάλεια και αποτελεσματικότητα.

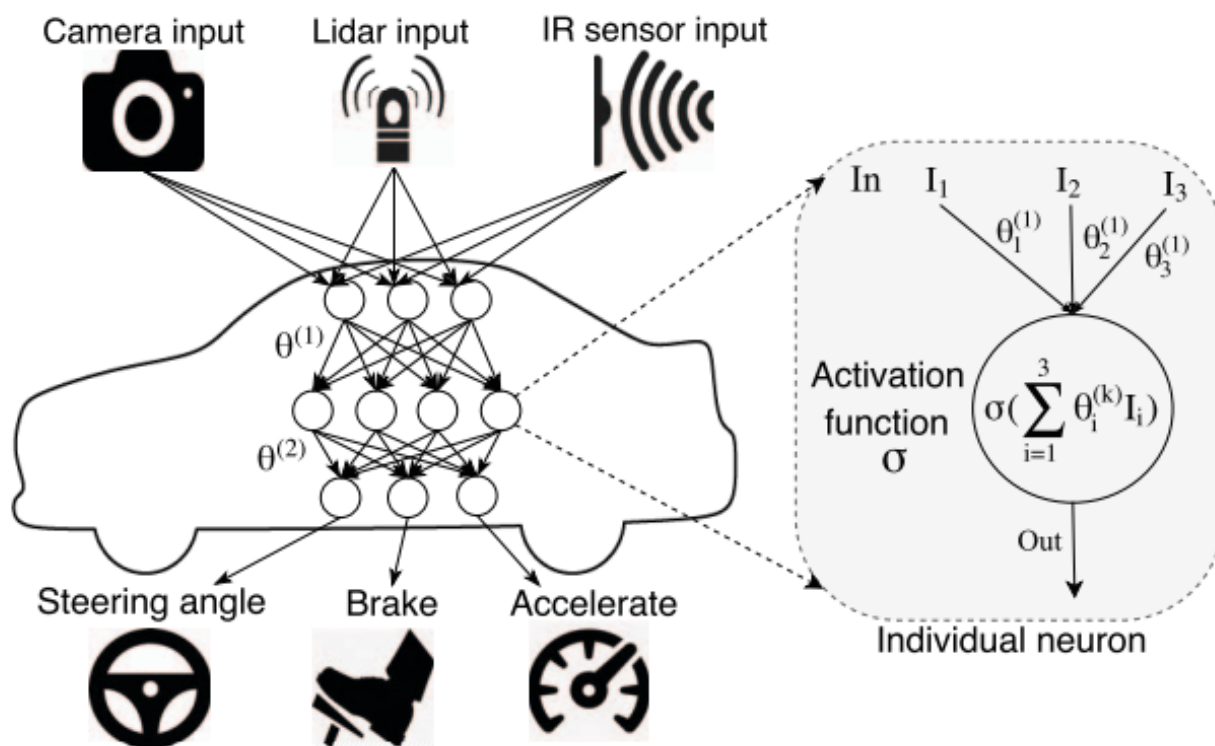
Ένα από τα βασικά καθήκοντα των DNNs στα αυτόνομα οχήματα είναι η ανίχνευση και αναγνώριση αντικειμένων στο περιβάλλον.[40] Χάρη στην ικανότητά τους να αντλούν υψηλού επιπέδου χαρακτηριστικά από τις εικόνες και τα δεδομένα αισθητήρων, τα DNNs μπορούν να ανιχνεύσουν οχήματα, πεζούς, ποδηλάτες και άλλα αντικείμενα στον οδικό χώρο με εντυπωσιακή ακρίβεια.[52]

Εκτός από την ανίχνευση, τα DNNs είναι σε θέση να προβλέπουν την κίνηση των αντικειμένων αυτών.[42] Με την ανάλυση του ιστορικού κίνησης και την πρόβλεψη των πιθανών πορειών τους, τα αυτόνομα οχήματα μπορούν να λάβουν αποφάσεις για τον τρόπο που θα αλληλεπιδράσουν με το περιβάλλον.[59] Επιπλέον, τα DNNs αναλαμβάνουν τον υπολογισμό ασφαλών διαδρομών για τα αυτόνομα οχήματα.[53] Αυτό περιλαμβάνει τον υπολογισμό των βέλτιστων πορειών για την αποφυγή αντικειμένων, τη διατήρηση ασφαλούς απόστασης από άλλα οχήματα και την προσαρμογή της ταχύτητας στις συνθήκες του οδικού δικτύου.

Το άρθρο “[Securing DNN for smart vehicles: an overview of adversarial attacks, defenses, and frameworks](#)” των S Almutairi και A Barnawi δίνει ένα παράδειγμα της εφαρμογής των DNN στον αναφερόμενο τομέα.

Τα αυτόνομα οχήματα χρησιμοποιούν τεχνικές τεχνητής νοημοσύνης, όπως ευφυείς πράκτορες (intelligent agents). Ένας πράκτορας αντιλαμβάνεται το περιβάλλον του μέσω αισθητήρων, αποφασίζει για τις κατάλληλες ενέργειες και εφαρμόζει αυτές τις ενέργειες στο περιβάλλον μέσω ενεργοποιητών (actuators). Η τεχνολογία αυτοματοποιημένης οδήγησης αποτελείται από τρία βασικά λειτουργικά επίπεδα: ένα επίπεδο αίσθησης (sensing layer), ένα επίπεδο αντίληψης (perception layer) και ένα επίπεδο ελέγχου ή απόφασης (control or decision layer). Το στρώμα ανίχνευσης έχει πολλούς τύπους αισθητήρων, συμπεριλαμβανομένων αισθητήρων lidar, κάμερας και ραντάρ. Αυτοί οι αισθητήρες βρίσκονται στο μπροστινό και πίσω μέρος του οχήματος. Οι αισθητήρες Lidar χρησιμοποιούνται για πολλούς σκοπούς, όπως η ανίχνευση αντικειμένων, κατά την οποία ο αισθητήρας ανιχνεύει τα κύματα φωτός που αντανακλώνται από τα αντικείμενα στο περιβάλλον. Οι αισθητήρες κάμερας χρησιμοποιούνται για τη λήψη βίντεο του περιβάλλοντος περιβάλλοντος και εικόνων, όπως οδικές πινακίδες. Οι αισθητήρες ραντάρ χρησιμοποιούνται για την προσομοίωση της όρασης, την παρακολούθηση καιρικών συνθηκών και την αποφυγή οδικών αντικειμένων. Οι αισθητήρες είναι υπεύθυνοι για τη συλλογή δεδομένων από το περιβάλλον. Το επίπεδο αντίληψης περιέχει DNN για την ανάλυση και την ερμηνεία των δεδομένων που συλλέγονται από τους αισθητήρες και την εξαγωγή λογικών πληροφοριών, όπως φαίνεται στο Σχήμα 1. Αυτό το επίπεδο λαμβάνει τα σήματα εισόδου και τα επεξεργάζεται για να κατανοήσει τις πληροφορίες που παρέχουν στο σύστημα. Το επίπεδο απόφασης είναι υπεύθυνο για τη λήψη αποφάσεων, όπως η δρομολόγηση, και ελέγχει την

οδήγηση, την αυτοστάθμευση, τον προσδιορισμό της γωνίας διεύθυνσης και την ανίχνευση λωρίδας.



Σχήμα 1. Επισκόπηση του ρόλου του DNN σε ένα μοντέλο αυτόνομης οδήγησης.

Η SAE International έχει εισαγάγει έξι επίπεδα για την ταξινόμηση αυτοματοποιημένων συστημάτων σε οχήματα. Αυτά τα επίπεδα κυμαίνονται από το επίπεδο 0 (χωρίς αυτοματισμό) έως το επίπεδο 5 (πλήρης αυτοματισμός). Ο ακόλουθος πίνακας παρουσιάζει μια επισκόπηση αυτών των επιπέδων και των κριτηρίων που τα διακρίνουν. Για περισσότερες πληροφορίες σχετικά με τα αυτοματοποιημένα οχήματα.

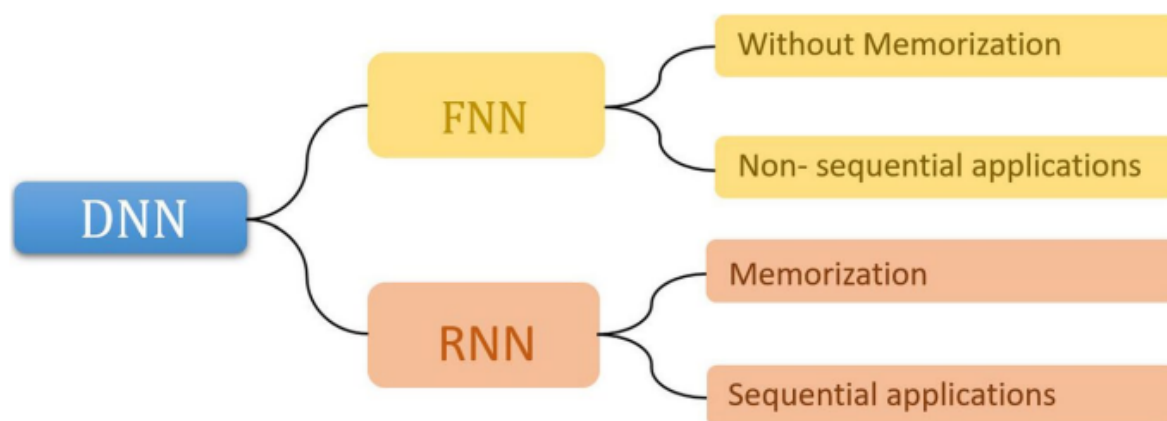
Level	Name	Controller	Require human monitoring	Driving features	Example of features
0	No automation	Human	Yes	Support	Automatic emergency braking, blind spot warning, lane departure warning
1	Driver assistance	Human and system	Yes	Support	Lane centering or adaptive cruise control
2	Partial automation	System	Yes	Support	Simultaneous lane centering and adaptive cruise control
3	Conditional automation	System	No	Automated	Traffic jam chauffeur
4	High automation	System	No	Automated	Features can drive the car in limited conditions
5	Full automation	System	No	Automated	Features can drive the car everywhere in all conditions

Τα έξι επίπεδα συστημάτων αυτοματισμού οχημάτων.

Η βαθιά μάθηση (Deep Learning) είναι ένα υποσύνολο της μηχανικής μάθησης στο οποίο η αφαίρεση της υποκείμενης γνώσης μαθαίνεται από το σύνολο δεδομένων. Η κύρια διαφορά μεταξύ των αλγορίθμων μηχανών και των DNN είναι ότι τα DNN δεν απαιτούν γνώση τομέα ή μηχανική χαρακτηριστικών επειδή είναι διεργασίες εκμάθησης από άκρο σε άκρο. Αυτή η διαδικασία μάθησης, η οποία αναφέρεται ως μάθηση αναπαράστασης, την καθιστά πιο μεταβιβάσιμη σε άλλα μοντέλα. Οι αλγόριθμοι βαθιάς μάθησης χρησιμοποιούν πολλαπλά επίπεδα για την εκμάθηση χαρακτηριστικών δεδομένων. Επιπλέον, οι αλγόριθμοι απαιτούν τεράστια σύνολα δεδομένων για εκμάθηση και επεξεργασία, επειδή η αποτελεσματικότητα των DNN εξαρτάται από το μέγεθος των δεδομένων και τους τεράστιους πόρους του υπολογιστή. Τα DNN σε αυτόνομα οχήματα μπορούν να ταξινομηθούν σε δύο βασικούς τύπους:

- Feed-forward neural networks (FNNs)
- Recurrent neural networks (RNNs)

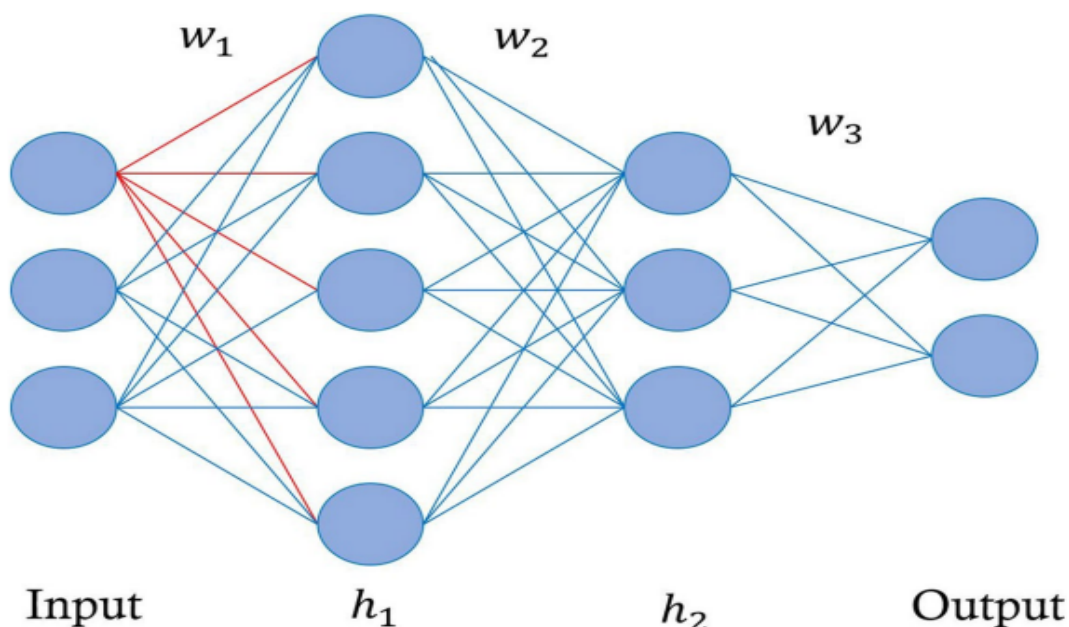
Η κύρια διαφορά μεταξύ των FNN και των RNN, όπως φαίνεται στο Σχήμα 2, είναι ότι τα FNN δεν διατηρούν τις τιμές των νευρώνων του τελευταίου στρώματος, πράγμα που σημαίνει ότι οι τιμές των νευρώνων διαδίδονται σε ένα κατεύθυνση. Αυτή η feed-forward λειτουργία καθιστά τα FNN πιο κατάλληλα για μη διαδοχικές εφαρμογές, όπως εικόνες. Τα RNN, εν τω μεταξύ, απομνημονεύουν την έξοδο του τελευταίου στρώματος νευρώνων, γεγονός που τους καθιστά κατάλληλους για διαδοχικές εφαρμογές, όπως ο ήχος.



Σχήμα 2. Η διαφορά μεταξύ RNN και FNN.

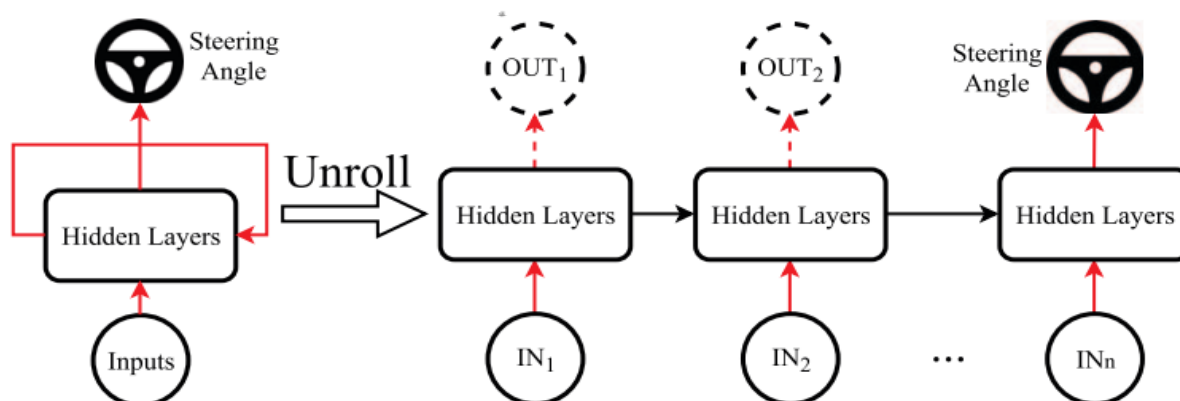
Τα συνελκτικά νευρωνικά δίκτυα (Convolutional neural networks - CNN) είναι παραδείγματα FNN. Η αρχιτεκτονική του CNN έχει δύο μέρη: πλήρως συνδεδεμένα νευρωνικά στρώματα (fully connected neural layers) και συνελκτικά στρώματα (convolutional layers). Τα πλήρως συνδεδεμένα νευρωνικά δίκτυα είναι συνδεδεμένα δίκτυα επιπέδων. Η επίσημη αρχιτεκτονική αποτελείται από ένα επίπεδο εισόδου, ένα ή περισσότερα κρυφά επίπεδα και ένα επίπεδο εξόδου. Το επίπεδο εισόδου είναι υπεύθυνο για τη μετάδοση των δεδομένων εισόδου στα κρυφά επίπεδα. Το κρυφό στρώμα ή τα στρώματα είναι υπεύθυνα για την εξαγωγή των χαρακτηριστικών και για την ανάλυση πληροφοριών. Το επίπεδο εξόδου είναι υπεύθυνο για την πρόβλεψη της κλάσης εισόδου. Για παράδειγμα, όπως φαίνεται στο Σχήμα 3, κάθε νευρώνας (μπλε κύκλος) σε οποιοδήποτε στρώμα L_i συνδέεται με κάθε νευρώνα στο επόμενο στρώμα L_{i+1} (κόκκινες άκρες). Αυτή η δομή εφαρμόζεται σε κάθε νευρώνα σε κάθε στρώμα, εκτός από το τελευταίο στρώμα (έξοδος). Τα κρυφά στρώματα συμβολίζονται στο Σχήμα 3 ως h1 και h2. Οι κύκλοι αντιπροσωπεύουν τους νευρώνες και οι άκρες αντιπροσωπεύουν το εσωτερικό γινόμενο

μεταξύ των αντίστοιχων βαρών και των νευρώνων του προηγούμενου στρώματος. Εάν το αποτέλεσμα αυτού του πολλαπλασιασμού είναι υψηλό, οι πληροφορίες ή το χαρακτηριστικό που σχετίζονται με αυτόν τον νευρώνα είναι σημαντικές και ο νευρώνας θα πρέπει να ενεργοποιηθεί. Τα στρώματα συνέλιξης έχουν νευρώνες που συνδέονται μόνο με ορισμένους νευρώνες στο επόμενο στρώμα και τα ίδια βάρη μοιράζονται μεταξύ διαφορετικών συνδέσεων για διαφορετικούς νευρώνες.



Σχήμα 3. Η απλή αρχιτεκτονική των πλήρως συνδεδεμένων στρωμάτων.

Η αρχιτεκτονική των RNN φαίνεται στο Σχήμα 4. Εκτός από τη διαδικασία λειτουργίας στα CNN, τα RNN επιτρέπουν να λαμβάνεται υπόψη η απομνημονευμένη έξοδος του τελευταίου νευρώνα στον υπολογισμό της πρόβλεψης του τρέχοντος νευρώνα. Δηλαδή, η έξοδος του νευρώνα τροφοδοτείται στο προηγούμενο στρώμα και στο επόμενο στρώμα. Αυτή η λειτουργία καθορίζει την έξοδο πρόβλεψης για την προηγούμενη είσοδο. Αυτή η προσέγγιση είναι χρήσιμη για πολλές εφαρμογές, όπως οι ακολουθίες καρέ βίντεο, όπου το τρέχον καρέ προβλέπεται από το προηγούμενο καρέ.



Σχήμα 4. Η βασική αρχιτεκτονική των RNN.

Ο παράγοντας στην ενισχυτική μάθηση (Reinforcement learning - RL) μαθαίνει πώς να βελτιώνει τη συμπεριφορά του σε ένα συγκεκριμένο περιβάλλον αλληλεπιδρώντας με αυτό το περιβάλλον. Σε αντίθεση με την εποπτευόμενη μάθηση, η σχέση και η αντιστοίχιση από την είσοδο στην έξοδο δεν αναφέρεται ρητά στον πράκτορα. Αντί να χρησιμοποιείται δοκιμή και σφάλμα, χρησιμοποιείται μια συνάρτηση ανταμοιβής για την αξιολόγηση των ενεργειών και την ενημέρωση της απόδοσης. Υπάρχουν αρκετές προσεγγίσεις DRL για τη λήψη αποφάσεων από αυτόνομα οχήματα σε αντίθετες συνθήκες.

Συνολικά, η τεχνολογία των DNNs έχει επιτρέψει την εξέλιξη της αυτόνομης οδήγησης από μια φανταστική ιδέα σε μια πραγματικότητα που είναι ολοένα και πιο κοντά. Η ικανότητα των DNNs να αντιλαμβάνονται το περιβάλλον, να προβλέπουν και να λαμβάνουν ασφαλείς αποφάσεις είναι ζωτικής σημασίας για την επίτευξη μιας ασφαλούς και αποτελεσματικής αυτοκινητικής εμπειρίας. Οι εφαρμογές των DNNs στα αυτόνομα οχήματα διαμορφώνουν το μέλλον της μεταφορικής βιομηχανίας, δημιουργώντας ένα περιβάλλον μείωσης των ατυχημάτων, βελτιωμένης κυκλοφορίας και πιο αποδοτικής μεταφοράς.

3.4.5 Κοινωνικά Δίκτυα και Διαδίκτυο

Τα DNNs επηρεάζουν επίσης τον τομέα των κοινωνικών δικτύων και του διαδικτύου. Χρησιμοποιούνται για την προσαρμογή περιεχομένου στα ενδιαφέροντα των χρηστών, την αναγνώριση ανεπιθύμητων μηνυμάτων και την ανάλυση συμπεριφοράς χρηστών.[\[41\]\[40\]](#)

Στον σύγχρονο κόσμο, τα κοινωνικά δίκτυα και το διαδίκτυο έχουν επηρεάσει σε μεγάλο βαθμό τον τρόπο με τον οποίο επικοινωνούμε, αλληλεπιδρούμε και αποκτούμε πληροφορίες. Η χρήση DNNs στον τομέα αυτόν έχει ανοίξει νέους τρόπους βελτίωσης της εμπειρίας των χρηστών και της αλληλεπίδρασής τους με τις πληροφορίες που τους παρέχονται.[\[48\]\[55\]](#)

Ένα από τα κύρια πλεονεκτήματα της χρήσης των DNNs στα κοινωνικά δίκτυα είναι η δυνατότητα παρακολούθησης και ανάλυσης της συμπεριφοράς των χρηστών. Μέσω της συλλογής και ανάλυσης μεγάλου όγκου δεδομένων, τα DNNs μπορούν να αναγνωρίζουν μοτίβα στην δραστηριότητα των χρηστών, προβλέποντας τις προτιμήσεις τους, τις ανάγκες τους και τη συμπεριφορά τους στην πλατφόρμα. Αυτό το είδος ανάλυσης δίνει τη δυνατότητα στις εταιρείες κοινωνικών δικτύων να προσαρμόζουν το περιεχόμενο και τις υπηρεσίες τους με βάση τις ατομικές προτιμήσεις και ανάγκες κάθε χρήστη. Επιτρέποντας έτσι τη δημιουργία εξατομικευμένων εμπειριών για τους χρήστες και βελτιώνει την αποδοτικότητα των πλατφορμών.[\[60\]\[42\]](#)

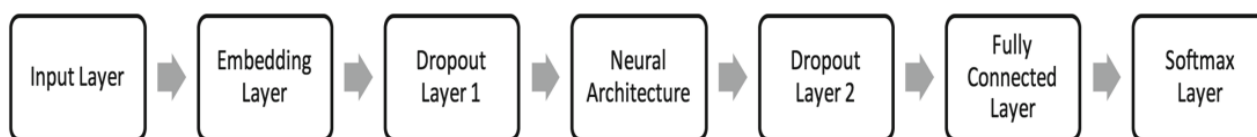
Επιπλέον, η αλληλεπίδραση των χρηστών με την πληροφορία βελτιώνεται μέσω των DNNs. Τα δίκτυα αυτά μπορούν να αναγνωρίζουν συναισθηματικές αντιδράσεις, να προβλέπουν συμπεριφορές και να αντλούν σχετικές πληροφορίες από τα δεδομένα των χρηστών. Αυτό μπορεί να χρησιμοποιηθεί για να προσαρμοστεί η αλληλεπίδραση, όπως για παράδειγμα να προβλέπονται αυτόματα απαντήσεις σε σχόλια και μηνύματα, να ανιχνεύονται καταναλωτικές τάσεις και να δημιουργούνται περιεχόμενα που ανταποκρίνονται στις ανάγκες των χρηστών.

Στο ακόλουθο άρθρο “[Deep Learning for Detecting Cyberbullying Across Multiple Social Media Platforms](#)” των Sweta Agrawal και Amit Awekar παρουσιάζεται ένα παράδειγμα εφαρμογής των DNN στον τομέα των κοινωνικών δικτύων. Ο διαδικτυακός εκφοβισμός αναγνωρίζεται ως φαινόμενο τουλάχιστον από το 2003. Η χρήση των μέσων κοινωνικής δικτύωσης εκτοξεύτηκε με την κυκλοφορία πολλαπλών πλατφορμών όπως η Wikipedia (2001), το

MySpace (2003), το Orkut (2004), το Facebook (2004) και το Twitter (2005). Μέχρι το 2006, οι ερευνητές είχαν επισημάνει ότι ο διαδικτυακός εκφοβισμός ήταν εξίσου σοβαρό φαινόμενο με τον εκφοβισμό εκτός σύνδεσης. Ωστόσο, η αυτόματη ανίχνευση του διαδικτυακού εκφοβισμού αντιμετωπίστηκε μόνο από το 2009.

Ως ερευνητικό θέμα, ο εντοπισμός του διαδικτυακού εκφοβισμού είναι ένα πρόβλημα ταξινόμησης κειμένου. Τα περισσότερα από τα υπάρχοντα έργα ταιριάζουν στο ακόλουθο πρότυπο: λαμβάνεται σύνολο δεδομένων εκπαίδευσης από μεμονωμένο SMP, δημιουργείται μια ποικιλία χαρακτηριστικών με στόχο συγκεκριμένο στυλ διαδικτυακού εκφοβισμού, εφαρμόζονται μερικές παραδοσιακές μέθοδοι μηχανικής μάθησης και αξιολογείται η επιτυχία με βάση μέτρα όπως η βαθμολογία F1 και ακρίβεια. Αυτά τα έργα βασίζονται σε μεγάλο βαθμό σε χειροποίητα χαρακτηριστικά, όπως η χρήση βρισιάς. Αυτές οι μέθοδοι τείνουν να έχουν χαμηλή ακρίβεια για την ανίχνευση του διαδικτυακού εκφοβισμού, καθώς τα χειροποίητα χαρακτηριστικά δεν είναι ανθεκτικά έναντι των παραλλαγών του στυλ εκφοβισμού σε SMPs και θέματα εκφοβισμού. Μόλις πρόσφατα, το deep learning έχει εφαρμοστεί για τον εντοπισμό του διαδικτυακού εκφοβισμού.

Πειραματιστήκαμε με τέσσερα μοντέλα που βασίζονται στο DNN για τον εντοπισμό διαδικτυακού εκφοβισμού: Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Bidirectional LSTM (BLSTM) και BLSTM with attention. Αυτά τα μοντέλα παρατίθενται στην αυξανόμενη πολυπλοκότητα της νευρωνικής τους αρχιτεκτονικής και του όγκου των πληροφοριών που χρησιμοποιούνται από αυτά τα μοντέλα. Στο Σχήμα 1 φαίνεται η γενική αρχιτεκτονική που χρησιμοποιήθηκε σε τέσσερα μοντέλα.



Σχήμα 1. Model architecture

Διάφορα μοντέλα διαφέρουν μόνο στο επίπεδο Νευρωνικής Αρχιτεκτονικής ενώ έχουν τα ίδια υπόλοιπα επίπεδα. Τα CNN παρέχουν κορυφαία αποτελέσματα για την εξαγωγή χαρακτηριστικών συμφραζομένων για εργασίες ταξινόμησης σε εικόνες, βίντεο, ήχους και κείμενο. Πρόσφατα, τα CNN χρησιμοποιήθηκαν για την ταξινόμηση συναισθημάτων. Τα Long Short Term Memory networks είναι ένα ειδικό είδος RNN, ικανό να μάθει μακροπρόθεσμες εξαρτήσεις. Η ικανότητά τους να χρησιμοποιούν την εσωτερική τους μνήμη για την επεξεργασία αυθαίρετων ακολουθιών εισόδων έχει βρεθεί ότι είναι αποτελεσματική για την ταξινόμηση κειμένου. Τα Bidirectional LSTM αυξάνουν περαιτέρω τον όγκο των πληροφοριών εισόδου που είναι διαθέσιμες στο δίκτυο κωδικοποιώντας πληροφορίες τόσο προς τα εμπρός όσο και προς τα πίσω. Χρησιμοποιώντας δύο κατευθύνσεις, μπορούν να χρησιμοποιηθούν πληροφορίες εισόδου τόσο από το παρελθόν όσο και από το μέλλον του τρέχοντος χρονικού πλαισίου. Οι μηχανισμοί προσοχής επιτρέπουν μια πιο άμεση εξάρτηση μεταξύ της κατάστασης του μοντέλου σε διαφορετικά χρονικά σημεία. Είναι σημαντικό ότι ο μηχανισμός προσοχής επιτρέπει στο μοντέλο να μάθει τι πρέπει να προσέχει με βάση την πρόταση εισαγωγής και τι έχει δημιουργήσει μέχρι τώρα.

Το στρώμα ενσωμάτωσης (embedding layer) επεξεργάζεται μια ακολουθία λέξεων σταθερού μεγέθους. Κάθε λέξη αναπαρίσταται ως διάνυσμα πραγματικής αξίας, γνωστό και ως ενσωματώσεις λέξεων. Πειραματιστήκαμε με τρεις μεθόδους για την αρχικοποίηση ενσωματώσεων λέξεων: τυχαία, GloVe και SSWE. Κατά τη διάρκεια της εκπαίδευσης, το μοντέλο βελτιώνει τις αρχικές ενσωματώσεις λέξεων για να μάθει ενσωματώσεις λέξεων για

συγκεκριμένες εργασίες. Παρατηρήθηκε ότι αυτές οι ενσωματώσεις λέξεων για συγκεκριμένες εργασίες αποτυπώνουν το στυλ του διαδικτυακού εκφοβισμού που αφορά το συγκεκριμένο SMP και το συγκεκριμένο θέμα. Η χρήση διανυσμάτων GloVe πάνω από τυχαία αρχικοποίηση διανυσμάτων έχει αναφερθεί ότι βελτιώνει την απόδοση για ορισμένες εργασίες NLP. Οι περισσότερες από τις μεθόδους ενσωμάτωσης λέξεων, όπως το GloVe, λαμβάνουν υπόψη μόνο το συντακτικό πλαίσιο της λέξης, ενώ αγνοούν το συναίσθημα που μεταφέρεται από το κείμενο. Η μέθοδος SSWE ξεπερνά αυτό το πρόβλημα ενσωματώνοντας το συναίσθημα κειμένου ως μία από τις παραμέτρους για τη δημιουργία ενσωμάτωσης λέξεων. Πειραματιστήκαμε με διάφορα μεγέθη διαστάσεων για ενσωματώσεις λέξεων. Τα πειραματικά αποτελέσματα που αναφέρονται εδώ έχουν μέγεθος διάστασης 50. Δεν υπήρξε σημαντική διαφοροποίηση στα αποτελέσματα με μέγεθος διάστασης που κυμαίνεται από 30 έως 200.

Για να αποφευχθεί η υπερβολική προσαρμογή, χρησιμοποιήσαμε δύο στρώματα εγκατάλειψης, ένα πριν από το στρώμα νευρωνικής αρχιτεκτονικής και ένα μετά, με ποσοστά εγκατάλειψης 0,25 και 0,5 αντίστοιχα. Το πλήρως συνδεδεμένο στρώμα είναι ένα πυκνό στρώμα εξόδου με τον αριθμό των νευρώνων ίσο με τον αριθμό των κλάσεων, ακολουθούμενο από το στρώμα softmax που παρέχει ενεργοποίηση softmax. Όλα τα μοντέλα μας εκπαιδεύονται χρησιμοποιώντας backpropagation. Ο optimizer που χρησιμοποιείται για την προπόνηση είναι ο Adam και η συνάρτηση απώλειας είναι η categorical cross-entropy. Εκτός από την εκμάθηση των βαρών του δικτύου, αυτές οι μέθοδοι μαθαίνουν επίσης ενσωματώσεις λέξεων για συγκεκριμένες εργασίες προσαρμοσμένες στις ετικέτες εκφοβισμού.

Δείχθηκε ότι τα μοντέλα DNN μπορούν να χρησιμοποιηθούν για τον εντοπισμό διαδικτυακού εκφοβισμού σε διάφορα θέματα σε πολλαπλά SMP. Αυτά τα μοντέλα σε συνδυασμό με transfer learning ξεπέρασαν τα τελευταία τεχνολογία αποτελέσματα και για τα τρία σύνολα δεδομένων. Αυτά τα μοντέλα μπορούν να βελτιωθούν περαιτέρω με επιπλέον δεδομένα, όπως πληροφορίες σχετικά με το προφίλ και το κοινωνικό γράφημα των χρηστών. Τα περισσότερα από τα τρέχοντα σύνολα δεδομένων δεν παρέχουν καμία πληροφορία σχετικά με τη σοβαρότητα του εκφοβισμού. Εάν διατίθενται τέτοιες λεπτομερείς πληροφορίες, τότε τα μοντέλα ανίχνευσης διαδικτυακού εκφοβισμού μπορούν να βελτιωθούν περαιτέρω για να αναλάβουν μια ποικιλία ενεργειών ανάλογα με την αντιληπτή σοβαρότητα των αναρτήσεων.

Η χρήση των Βαθιών Νευρωνικών Δικτύων στα κοινωνικά δίκτυα και το διαδίκτυο έχει επιφέρει μια νέα διάσταση στην αλληλεπίδραση των ανθρώπων με την τεχνολογία. Μέσω της αναγνώρισης των μοτίβων, της προσαρμογής του περιεχομένου και της βελτιωμένης αλληλεπίδρασης, τα DNNs δημιουργούν εξατομικευμένες και εμπλουτισμένες εμπειρίες για τους χρήστες, ανοίγοντας τον δρόμο για την εξέλιξη της ψηφιακής κοινωνίας. Με την δυνατότητα εξατομίκευσης της εμπειρίας των χρηστών και την παρακολούθηση της συμπεριφοράς τους, δημιουργείται ένα πιο εξατομικευμένο, ασφαλές και θετικό περιβάλλον στον κυβερνοχώρο.

Συνοψίζοντας, οι εφαρμογές των Deep Neural Networks είναι ποικίλες και πολυδιάστατες, καλύπτουν ένα ευρύ φάσμα τομέων, από την υγεία και την τεχνολογία έως τον πολιτισμό και την κοινωνία. Η τεχνολογία αυτή έχει επανασχεδιάσει τον τρόπο με τον οποίο αντιμετωπίζουμε προκλήσεις και δημιουργούμε λύσεις σε ποικίλους τομείς. Η ικανότητά τους να αναγνωρίζουν μοτίβα και να αντιλαμβάνονται το περιβάλλον τους έχει ανοίξει τον δρόμο για καινοτόμες εφαρμογές που αλλάζουν τον τρόπο ζωής και εργασίας μας. Καθώς οι τεχνολογίες των DNNs εξελίσσονται περαιτέρω, αναμένεται ότι θα ανακαλύπτουμε ακόμα περισσότερες νέες καινοτόμες εφαρμογές που θα διαμορφώνουν τον τρόπο ζωής, την καθημερινότητά μας και τον τρόπο με τον οποίο αλληλεπιδρούμε με τον κόσμο.

4

Recurrent Neural Networks (RNNs)

4.1 Εισαγωγικά

Τα Αναδρομικά Νευρωνικά Δίκτυα (RNNs) εκπροσωπούν μια σημαντική εξέλιξη στον κόσμο των τεχνητών νευρωνικών δικτύων. Η πρωτοτυπία τους βρίσκεται στην ικανότητά τους να διαχειρίζονται και να εξάγουν σημαντικές πληροφορίες από σειριακά δεδομένα, όπως χρονοσειρές, κείμενα, ομιλία και άλλες ακολουθίες. Σε αντίθεση με τις παραδοσιακές αρχιτεκτονικές νευρωνικών δικτύων, τα RNNs εισάγουν έναν δυναμικό και εσωτερικό μηχανισμό που επιτρέπει την αποθήκευση και τη χρήση πληροφοριών από προηγούμενα χρονικά βήματα.[\[41\]\[55\]](#)

Στον πυρήνα της λειτουργίας των RNNs βρίσκονται δύο κύριες ιδιότητες: η αναδρομή και η μνήμη. Ο μηχανισμός της αναδρομής επιτρέπει στο δίκτυο να διατηρεί και να χρησιμοποιεί πληροφορίες από προηγούμενες επαναλήψεις, κάτι που είναι ιδιαίτερα χρήσιμο σε ακολουθιακά δεδομένα. Η μνήμη ενός RNN λαμβάνει τη μορφή της κρυφής κατάστασης (hidden state), η οποία συνεχίζει να εξελίσσεται καθώς το δίκτυο προχωρά στην ανάλυση της ακολουθίας.[\[61\]](#)

Ένα από τα πλεονεκτήματα των RNNs είναι η ικανότητα να αναγνωρίζουν μοτίβα και να δημιουργούν συνδέσεις ανάμεσα στα διάφορα στοιχεία μιας ακολουθίας.[\[59\]](#) Αυτό τα καθιστά ιδανικά για εφαρμογές που εξαρτώνται από την χρονική διάσταση, καθώς μπορούν να ανακαλύπτουν τις διαδοχικές συσχετίσεις και να προβλέπουν τις μελλοντικές εξελίξεις με βάση τα παρελθοντικά δεδομένα.[\[62\]](#)

Είναι όμως σημαντικό να σημειωθεί ότι αυτοί καθαυτοί οι αναδρομικοί νευρώνες που χρησιμοποιούνται στα απλά RNNs μπορούν να αντιμετωπίσουν δυσκολίες σε περιπτώσεις μεγάλης χρονικής απόστασης μεταξύ σχετικών γεγονότων.[\[58\]](#) Αυτό οφείλεται στο φαινόμενο της "εξαφάνισης της αναδρομής", καθώς η επίδραση των προηγούμενων βημάτων μειώνεται όσο προχωράμε πίσω στο χρόνο. Αυτό το πρόβλημα αναδεικνύει την ανάγκη για πιο προηγμένες αρχιτεκτονικές, όπως τα Μακροχρόνια Αναδρομικά Νευρωνικά Δίκτυα (Long Short-Term Memory - LSTM)[\[53\]](#) και τα Δίκτυα Μνήμης Μετασχηματιστών (Transformer)[\[55\]](#), τα οποία αντιμετωπίζουν τις προκλήσεις της αναλυτικής διατήρησης πληροφοριών από μεγάλες χρονικές αποστάσεις.

Συνολικά, τα Αναδρομικά Νευρωνικά Δίκτυα είναι ένα ισχυρό εργαλείο για την ανάλυση και τη μοντελοποίηση σειριακών δεδομένων, ανοίγοντας νέες προοπτικές σε πεδία όπως η πρόβλεψη, η κατανόηση και η αναγνώριση μοτίβων σε πολυπλοκότερες ακολουθίες. Ανοίγουν πόρτες σε ανεξερεύνητους τομείς, όπου η ανάλυση σειριακών δεδομένων είναι κρίσιμη. Η ικανότητά τους να αποτυπώνουν τη χρονική εξέλιξη των γεγονότων είναι καίρια για την κατανόηση και πρόβλεψη ποικίλων σεναρίων. Αυτό το εργαλείο έχει εφαρμογές από την ανακάλυψη νέων τάσεων στις χρονοσειρές μέχρι τη βελτιστοποίηση των αυτόματων μεταφραστικών συστημάτων[\[56\]](#) και τη

δημιουργία μουσικής βάσει συνεχόμενων ήχων ποταμού.[54] Με τον συνεχή εξελικτικό χαρακτήρα της τεχνητής νοημοσύνης, η εξέλιξη αυτών των αρχιτεκτονικών είναι αναπόφευκτη και ανοίγει νέες προοπτικές για την αναγνώριση, την ανάλυση και την αξιοποίηση σειριακών δεδομένων.

4.2 Θεμελιώδης Αρχή των RNNs

Τα Αναδρομικά Νευρωνικά Δίκτυα αποτελούν μια σημαντική κατηγορία νευρωνικών δικτύων που έχει επανασχεδιαστεί για την επεξεργασία ακολουθιών δεδομένων, όπως κείμενα, ήχος και χρονοσειρές. Η θεμελιώδης αρχή που διέπει τα RNNs είναι η ικανότητά τους να διατηρούν την εσωτερική τους κατάσταση ενόσω εξελίσσεται καθώς επεξεργάζονται σταδιακά τα στοιχεία μιας ακολουθίας. Αυτή η ιδιότητα τους, τα καθιστά ιδιαίτερα αποτελεσματικά στην αντιμετώπιση προβλημάτων που σχετίζονται με τον χρόνο και τη σειρά των γεγονότων.

4.2.1 Αναδρομική Λειτουργία

Η αναδρομική λειτουργία των RNNs προκύπτει από την ιδέα της επαναχρησιμοποίησης της εσωτερικής κατάστασής τους κατά την επεξεργασία κάθε στοιχείου μιας εκάστοτε ακολουθίας. Κάθε επίπεδο ενός RNN διαθέτει έναν εσωτερικό χώρο κατάστασης, που μπορεί να αποθηκεύσει πληροφορίες από προηγούμενα στοιχεία της ακολουθίας.[63] Κατά την επεξεργασία κάθε νέου στοιχείου, το RNN συνδυάζει την εισερχόμενη πληροφορία με την υπάρχουσα κατάσταση για να δημιουργήσει μια νέα κατάσταση, η οποία στη συνέχεια θα χρησιμοποιηθεί κατά την επεξεργασία του επόμενου στοιχείου.

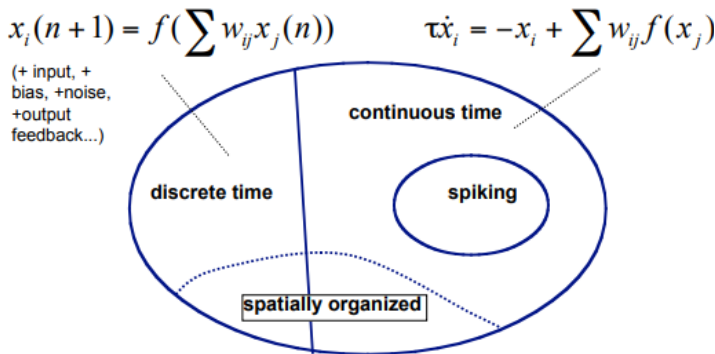
Η αναδρομική λειτουργία αποτελεί τον πυρήνα της συμπεριφοράς των Αναδρομικών Νευρωνικών Δικτύων και αντικατοπτρίζει τον τρόπο με τον οποίο διαχειρίζονται και αποθηκεύουν πληροφορία σε ακολουθιακά δεδομένα.[64] Βασίζεται στην αρχή της επαναχρησιμοποίησης της εσωτερικής κατάστασης. Ας εξετάσουμε αναλυτικά τα στοιχεία αυτής της αναδρομικής λειτουργίας.

Κάθε επίπεδο του RNN διαθέτει έναν εσωτερικό χώρο κατάστασης, γνωστός και ως "κρυφό διάνυσμα", που λειτουργεί ως μνήμη για τις πληροφορίες που έχουν εισαχθεί από προηγούμενα στοιχεία της ακολουθίας. Κατά τη διάρκεια επεξεργασίας κάθε νέου στοιχείου της ακολουθίας, το RNN λαμβάνει την εισερχόμενη πληροφορία από αυτό το στοιχείο και συνδυάζει την πληροφορία αυτή με την υπάρχουσα κατάσταση του εσωτερικού χώρου. Αυτή η συνδυαστική διαδικασία δημιουργεί μια νέα κατάσταση, η οποία κατοχυρώνει τις σημαντικές πτυχές της τρέχουσας πληροφορίας και εξελίσσεται με βάση τις επεξεργασμένες τιμές των προηγούμενων στοιχείων.[65]

Αυτή η δυνατότητα να διατηρείται μια εσωτερική κατάσταση που μεταβάλλεται με την επεξεργασία των δεδομένων είναι κρίσιμη για την αντιμετώπιση ακολουθιών με μεταβαλλόμενες δομές και συνθήκες. Ενώ τα απλά RNNs μπορεί να αντιμετωπίζουν δυσκολίες όπως την εξαφάνιση του κύριου σφάλματος κατά την εκπαίδευση, προηγμένες αρχιτεκτονικές όπως τα LSTM και Gated Recurrent Unit (GRU) έχουν δημιουργηθεί για να διορθώσουν αυτά τα προβλήματα και να ενισχύσουν την ικανότητα των RNNs να αντιμετωπίζουν πιο πολύπλοκες ακολουθίες.[63][64]

Βασιζόμενοι στο άρθρο “[A tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach](#)” του Herbert Jaeger, αναλύεται παρακάτω η βασική δομή και λειτουργία των RNNs.

Τα στοιχειώδη δομικά στοιχεία ενός RNN είναι οι νευρώνες (χρησιμοποιείται ο όρος μονάδες) που συνδέονται με συναπτικούς συνδέσμους (συνδέσεις) των οποίων η συναπτική ισχύς κωδικοποιείται από ένα βάρος. Συνήθως διακρίνει κανείς μονάδες εισόδου, εσωτερικές (ή κρυφές) μονάδες και μονάδες εξόδου. Σε μια δεδομένη χρονική στιγμή, μια μονάδα έχει μια ενεργοποίηση. Οι ενεργοποιήσεις των μονάδων εισόδου συμβολίζονται με $u(n)$, των εσωτερικών μονάδων με $x(n)$ και των μονάδων εξόδου με $y(n)$.



Εικόνα 1: Μια τυπολογία μοντέλων RNN (ημιτελής).

Υπάρχουν πολλοί τύποι επίσημων μοντέλων RNN (Σχήμα 1). Τα μοντέλα διακριτού χρόνου είναι μαθηματικά ως χάρτες που επαναλαμβάνονται σε διακριτά χρονικά βήματα $n = 1, 2, 3, \dots$. Τα μοντέλα συνεχούς χρόνου ορίζονται μέσω διαφορικών εξισώσεων των οποίων οι λύσεις ορίζονται σε συνεχή χρόνο t . Ειδικά για τους σκοπούς της βιολογικής μοντελοποίησης, τα συνεχή δυναμικά μοντέλα μπορεί να είναι αρκετά περίπλοκα και να περιγράφουν σήματα ενεργοποίησης σε επίπεδο μεμονωμένων δυναμικών δράσης (spikes). Συχνά το μοντέλο ενσωματώνει ένα προσδιορισμό μιας χωρικής τοπολογίας, συχνότερα μιας 2D επιφάνειας όπου οι μονάδες είναι τοπικά συνδεδεμένες σε δομές που μοιάζουν με τον αμφιβληστροειδή.

Στο άρθρο αυτό εξετάζεται μόνο ένα συγκεκριμένο είδος μοντέλου διακριτού χρόνου χωρίς χωρική οργάνωση. Το μοντέλο αποτελείται από K μονάδες εισόδου με μια ενεργοποίηση (στήλη) διάνυσμα: $u(n) = (u_1(n), \dots, u_K(n))^t$ (1),

N εσωτερικών μονάδων με διάνυσμα ενεργοποίησης: $x(n) = (x_1(n), \dots, x_N(n))^t$ (2),

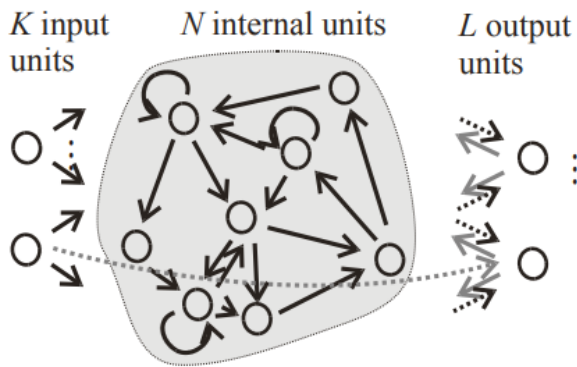
και των μονάδων εξόδου L με διάνυσμα ενεργοποίησης: $y(n) = (y_1(n), \dots, y_L(n))^t$ (3),

όπου το t υποδηλώνει τη μετατόπιση. Τα βάρη των συνδέσεων εισόδου/εσωτερικής/εξόδου είναι συλλέγονται σε πίνακες βαρών $N \times K / N \times N / L \times (K+N)$:

$$W^{in} = W_{ij}^{in}, W = W_{ij}, W^{out} = W_{ij}^{out} \quad (4).$$

Οι μονάδες εξόδου μπορούν προαιρετικά να προβάλλονται σε εσωτερικές μονάδες με συνδέσεις των οποίων βάρη συγκεντρώνονται σε έναν πίνακα βαρών οπισθοπροβολής $N \times L$: $W^{back} = W_{ij}^{back}$.

Σχήμα 2. Η βασική αρχιτεκτονική δικτύου που χρησιμοποιείται σε αυτό το σεμινάριο.



Σκιασμένα βέλη υποδεικνύουν προαιρετικές συνδέσεις. Τα διακεκομμένα βέλη επισημαίνουν συνδέσεις που εκπαιδεύονται σε την προσέγγιση του "δικτύου κατάστασης ηχούς" (σε άλλες προσεγγίσεις, όλες οι συνδέσεις μπορούν να εκπαιδευτούν).

Μια μηδενική τιμή βάρους μπορεί να ερμηνευθεί ως "καμία σύνδεση". Σημειώστε ότι οι μονάδες εξόδου μπορεί να έχουν συνδέσεις όχι μόνο από εσωτερικές μονάδες αλλά και (συνήα) από μονάδες εισόδου και (σπάνια) από μονάδες εξόδου. Η ενεργοποίηση των εσωτερικών μονάδων ενημερώνεται σύμφωνα με την:

$$x(n+1) = f(W^{in}u(n+1) + Wx(n) + W^{back}y(n)) \quad (6),$$

όπου $u(n+1)$ είναι η εξωτερικά δοσμένη είσοδος και f δηλώνει τη συνισταμένη-σοφή εφαρμογή της συνάρτησης μεταφοράς της μεμονωμένης μονάδας, f (επίσης γνωστή ως συνάρτηση ενεργοποίησης, συνάρτηση εξόδου μονάδας ή συνάρτηση συμπίεσης). Θα χρησιμοποιήσουμε κυρίως τη σιγμοειδή συνάρτηση $f = \tanh$, αλλά μερικές φορές εξετάζουμε επίσης γραμμικά δίκτυα με $f = 1$. Η έξοδος είναι υπολογίζεται σύμφωνα με:

$$y(n+1) = f^{out}(W^{out}(u(n+1), x(n+1), y(n))) \quad (7),$$

όπου $(u(n+1), x(n+1), y(n))$ δηλώνει το συνυφασμένο διάνυσμα που αποτελείται από την είσοδο, το εσωτερικό, και τα διανύσματα ενεργοποίησης εξόδου. Θα χρησιμοποιηθούν συναρτήσεις μεταφοράς εξόδου $f^{out} = \tanh$ ή $f^{out} = 1$, στην τελευταία περίπτωση έχουμε γραμμικές μονάδες εξόδου.

Συνοψίζοντας, η αναδρομική λειτουργία αποτελεί το θεμέλιο των RNNs και τους παρέχει την ικανότητα να αντιμετωπίζουν ακολουθιακά δεδομένα, διατηρώντας τη μνήμη των προηγούμενων στοιχείων και ενσωματώνοντας τις πληροφορίες σε νέες καταστάσεις.

4.2.2 Χρήση στη Μοντελοποίηση Χρονοσειρών

Ένα από τα βασικά πεδία εφαρμογής των RNNs είναι η μοντελοποίηση χρονοσειρών. Με την ικανότητά τους να λαμβάνουν υπόψη τη σειρά των γεγονότων, τα RNNs μπορούν να ανιχνεύσουν πολύπλοκα πρότυπα και εξελίξεις σε χρονοσειρές δεδομένων, όπως οικονομικές προβλέψεις, αναγνώριση φωνής και πρόβλεψη καιρού. [63][64]

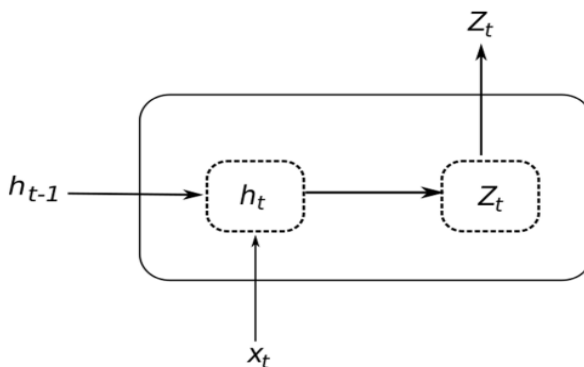
Διαθέτουν την εντυπωσιακή ικανότητα να ανιχνεύουν μη γραμμικές σχέσεις και αλληλεπιδράσεις μεταξύ των δεδομένων, αξιοποιώντας τα προηγούμενα στοιχεία της ακολουθίας για την εκτίμηση των μελλοντικών συμπεριφορών. Αυτό αποτελεί σημαντικό πλεονέκτημα, ειδικά όταν διαχειριζόμαστε δεδομένα που παρουσιάζουν απρόβλεπτες μεταβολές και πολύπλοκες δομές. [66][67] Επιπλέον, η ικανότητα των RNNs να διατηρούν πληροφορίες για προηγούμενα χρονικά βήματα επιτρέπει την ανίχνευση μακροπρόθεσμων εξελίξεων και προτύπων, κάτι που συχνά αποτελεί δύσκολο εγχείρημα για παραδοσιακές στατιστικές μεθόδους. [68]

Εξαιρετικά παραδείγματα εφαρμογών των RNNs στη μοντελοποίηση χρονοσειρών περιλαμβάνουν την πρόβλεψη οικονομικών εξελίξεων, την αναγνώριση και μετάφραση φωνητικών εντολών, την πρόβλεψη καιρού, καθώς και την αναγνώριση προτύπων σε ιατρικά δεδομένα.[69][70][71] Οι τεχνολογικές εξελίξεις στη βελτίωση των RNNs, με την εισαγωγή μονάδων όπως οι μονάδες LSTM και GRU, έχουν ενισχύσει την αποτελεσματικότητά τους σε αυτό το πεδίο, καθιστώντας τη χρήση τους πιο ευέλικτη και αποτελεσματική.[68][72]

Μελετώντας το άρθρο “[Recurrent Neural Networks for Time Series Forecasting: Current status and future directions](#)” εμβαθύνουμε στη μεθοδολογία μιας εκτεταμένης εμπειρικής μελέτης και ενός πλαισίου λογισμικού ανοικτού κώδικα των υφιστάμενων αρχιτεκτονικών RNN για πρόβλεψη (forecasting). Υλοποιήθηκε ένας αριθμός αρχιτεκτονικών RNN μαζί με διαφορετικές μονάδες RNN, χρησιμοποιώντας τις τρεις ακόλουθες για να αποτελέσουν τα στρώματα των RNN στο πειραματικό μέρος:

- Αναδρομική μονάδα Elman (Elman recurrent unit)
- Αναδρομική μονάδα με πύλη (Gated recurrent unit)
- Μακρά βραχυπρόθεσμη μνήμη με συνδέσεις ματιού (Long short-term memory with peephole connections)

Η βασική επαναλαμβανόμενη μονάδα εισήχθη από τον Elman (1990). Η δομή της βασικής κυψέλης ERNN παρουσιάζεται στο Σχήμα 1.



Σχήμα 1. Elman recurrent unit.

Στις ακόλουθες εξισώσεις:

$$h_t = \sigma(W_i \cdot h_{t-1} + V_i \cdot x_t + b_i) \quad (1)$$

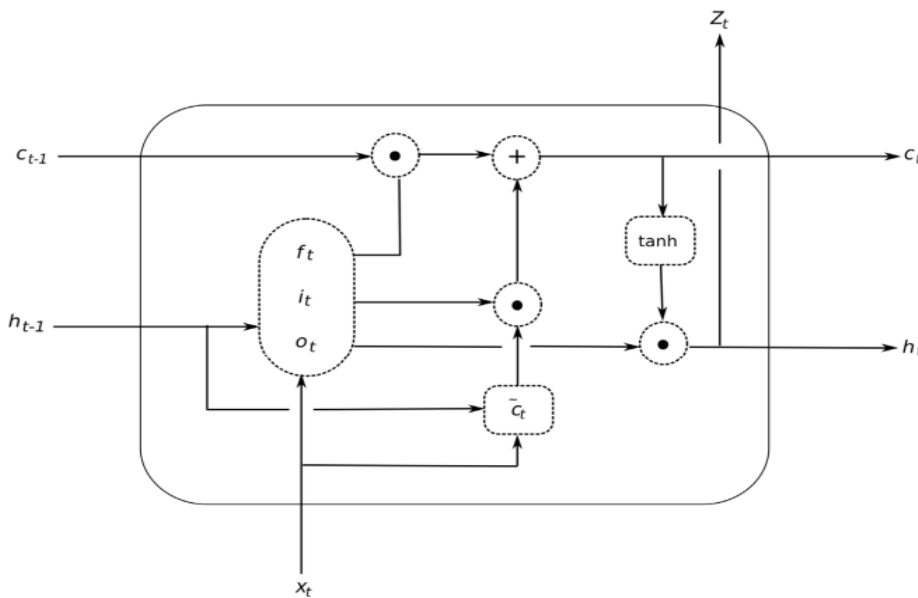
$$z_t = \tanh(W_o \cdot h_t + b_o) \quad (2)$$

όπου $h_t \in \mathbb{R}^d$ συμβολίζει την κρυφή κατάσταση του κυττάρου RNN (όπου d είναι η διάσταση του κελιού). Αυτή είναι η μόνη μορφή μνήμης στο κύτταρο ERNN. $x_t \in \mathbb{R}^m$ (όπου m είναι το μέγεθος της εισόδου) και $z_t \in \mathbb{R}^d$ συμβολίζουν αντίστοιχα την είσοδο και την έξοδο του κελιού στο χρονικό βήμα t . $W_i \in \mathbb{R}^{d \times d}$ και $V_i \in \mathbb{R}^{d \times m}$ υποδηλώνουν τους πίνακες βαρών, ενώ $b_i \in \mathbb{R}^d$ συμβολίζει το διάνυσμα μεροληψίας για την κρυφή κατάσταση. Ομοίως, $W_o \in \mathbb{R}^{d \times d}$ και $b_o \in \mathbb{R}^d$ δηλώνουν αντίστοιχα τον πίνακα βαρών και το διάνυσμα προκατάληψης της εξόδου του κελιού.

Η τρέχουσα κρυφή κατάσταση εξαρτάται από την κρυφή κατάσταση του προηγούμενου χρονικού βήματος καθώς και από την τρέχουσα είσοδο. Αυτό υποστηρίζεται με τους βρόχους ανατροφοδότησης στο κύτταρο RNN που συνδέουν την τρέχουσα κατάστασή του με την επόμενη κατάσταση. Αυτές οι συνδέσεις είναι εξαιρετικά σημαντικές προκειμένου να λαμβάνονται υπόψη οι πληροφορίες του παρελθόντος κατά την ενημέρωση της τρέχουσας κατάστασης του κελιού. Στα πειράματα, χρησιμοποιήσαμε τη sigmoid function (υποδεικνύεται με σ) ως ενεργοποίηση της κρυφής κατάστασης και την hyperbolic tangent function (υποδεικνύεται με $\tanh$) ως ενεργοποίηση της εξόδου.

Το κύτταρο ERNN υποφέρει από τα γνωστά προβλήματα της vanishing gradient και της exploding gradient σε πολύ μεγάλες ακολουθίες. Αυτό σημαίνει ότι τα απλά κύτταρα RNN δεν είναι ικανά να μεταφέρουν μακροχρόνιες εξαρτήσεις στο μέλλον. Όταν οι ακολουθίες είναι αρκετά μεγάλες, οι κλίσεις που διαδίδονται προς τα πίσω τείνουν να μειώνονται (εξαφανίζονται) και, κατά συνέπεια, τα βάρη δεν ενημερώνονται επαρκώς. Από την άλλη πλευρά, όταν οι κλίσεις είναι τεράστιες, μπορούν να “εκραγούν” σε μεγάλες ακολουθίες φέροντας ως αποτέλεσμα ασταθείς πίνακες βαρών. Και τα δύο αυτά ζητήματα απορρέουν από το γεγονός ότι οι κλίσεις είναι δύσκολες και εμποδίζουν την ικανότητα των κυττάρων RNN να καταγράφουν τις μακροπρόθεσμες εξαρτήσεις.

Με την πάροδο των ετών, έχουν εισαχθεί διάφορες παραλλαγές αυτής της βασικής επαναλαμβανόμενης μονάδας για να αντιμετωπιστούν οι αδυναμίες της. Η κυψέλη LSTM που εισήχθη από τους Hochreiter και Schmidhuber (1997) είναι ίσως η πιο δημοφιλής κυψέλη για εργασίες επεξεργασίας φυσικής γλώσσας, επειδή μπορεί να συλλάβει τις μακροπρόθεσμες εξαρτήσεις στην ακολουθία, ενώ παράλληλα μετριάξει τα προβλήματα vanishing gradient. Η δομή του κυττάρου LSTM απεικονίζεται στο Σχήμα 2.



Σχήμα 2. Βασική εικόνα long short-term memory Unit (LSTM).

$$i_t = \sigma(W_i \cdot h_{t-1} + V_i \cdot x_t + b_i)$$

$$o_t = \sigma(W_o \cdot h_{t-1} + V_o \cdot x_t + b_o)$$

$$f_t = \sigma(W_f \cdot h_{t-1} + V_f \cdot x_t + b_f)$$

$$\bar{c}_t = \tanh(W_c \cdot h_{t-1} + V_c \cdot x_t + b_c)$$

$$C_t = i_t \odot \bar{C}_t + f_t \odot C_{t-1}$$

$$h_t = o_t \odot \tanh(C_t)$$

$$z_t = h_t$$

Σε σύγκριση με το βασικό στοιχείο RNN, το στοιχείο LSTM έχει δύο συνιστώσες στην κατάστασή του: την κρυφή κατάσταση (hidden state) και την εσωτερική κατάσταση του στοιχείου (internal cell state).

Η κρυφή κατάσταση αντιστοιχεί στη συνιστώσα της βραχυπρόθεσμης μνήμης, και η κατάσταση του στοιχείου αντιστοιχεί στη μακροπρόθεσμη μνήμη. Με την ικανότητα σταθερού σφάλματος καρουσέλ (constant error carousel - CEC) που υποστηρίζεται από την εσωτερική κατάσταση των κελιών, η LSTM αποφεύγει τα προβλήματα vanishing και exploding gradient. Επιπλέον, εισάγεται ένας μηχανισμός πύλης που περιλαμβάνει τις πύλες εισόδου (input), forget και εξόδου (output). Στις παραπάνω εξισώσεις, $h_t \in \mathbb{R}^d$ είναι ένα διάνυσμα που δηλώνει την κρυφή κατάσταση του στοιχείου, όπου d είναι η διάσταση του. Ομοίως, $C_t \in \mathbb{R}^d$ είναι η κατάσταση του στοιχείου και $\bar{C}_t \in \mathbb{R}^d$ είναι η υπονήφια κατάσταση του στο χρονικό βήμα t , η οποία συλλαμβάνει σημαντικές πληροφορίες που θα παραμείνουν μέχρι το μέλλον. $x_t \in \mathbb{R}^m$ και $z_t \in \mathbb{R}^d$ είναι τα ίδια με το βασικό στοιχείο RNN. $W_i, W_o, W_f, W_c \in \mathbb{R}^{d \times d}$ δηλώνουν τους πίνακες βαρών της πύλης εισόδου, της πύλης εξόδου, της πύλης forget και της κατάστασης του στοιχείου, αντίστοιχα. Ομοίως, οι $V_i, V_o, V_f, V_c \in \mathbb{R}^{d \times m}$ και $b_i, b_o, b_f, b_c \in \mathbb{R}^d$ υποδηλώνουν τους πίνακες βαρών που αντιστοιχούν στην τρέχουσα είσοδο και στα διανύσματα πόλωσης, αντίστοιχα. $i_t, o_t, f_t \in \mathbb{R}^d$ είναι τα διανύσματα της πύλης input, της πύλης output και της forget gate.

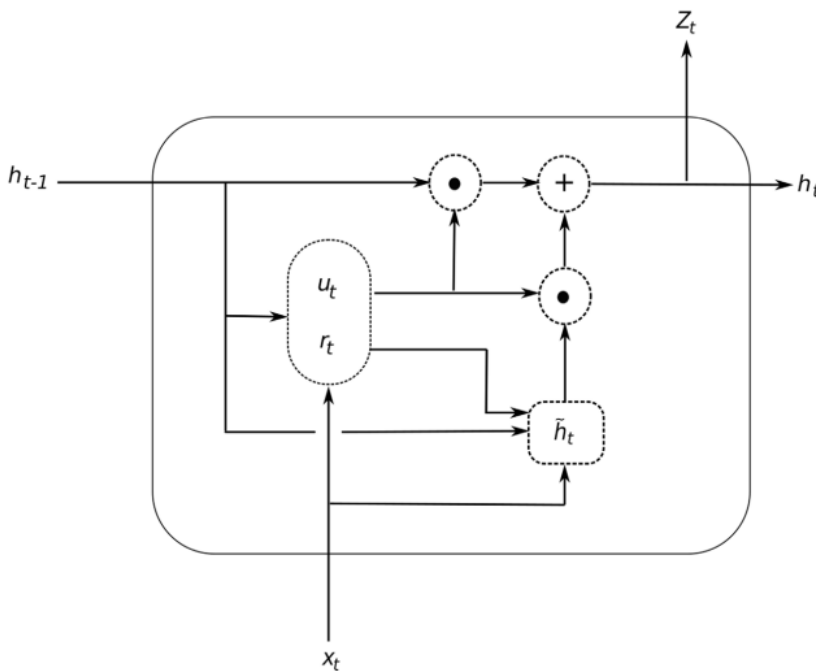
Η συνάρτηση ενεργοποίησης σ των πυλών δηλώνει τη σιγμοειδή συνάρτηση, η οποία δίνει τιμές στο εύρος $[0, 1]$. Στην εξίσωση $C_t = i_t \odot \bar{C}_t + f_t \odot C_{t-1}$, οι πύλες input και οι πύλες forget καθορίζουν από κοινού πόση από την πληροφορία του παρελθόντος θα διατηρηθεί στην τρέχουσα κατάσταση του στοιχείου και πόση από το τρέχον πλαίσιο θα διαδοθεί στα μελλοντικά χρονικά βήματα. $\odot$ συμβολίζει τον στοιχειομετρικό πολλαπλασιασμό, ο οποίος είναι γνωστός ως γινόμενο Hadamard. Η τιμή 0 στην forget gate f_t δηλώνει ότι τίποτα δεν πρέπει να μεταφερθεί από την προηγούμενη κατάσταση στοιχείου. Με άλλα λόγια, η προηγούμενη κατάσταση στοιχείου πρέπει να ξεχαστεί εντελώς στην τρέχουσα κατάσταση στοιχείου. Ακολουθώντας αυτό το επιχειρήμα, η τιμή 1 υποδηλώνει ότι η προηγούμενη κατάσταση στοιχείου πρέπει να διατηρηθεί πλήρως. Η ίδια έννοια ισχύει και για τις άλλες δύο πύλες, i_t και o_t .

Μια τιμή μεταξύ των δύο ακραίων τιμών 0 και 1 τόσο για τις πύλες input όσο και για τις forget gates μπορεί να ελέγξει προσεκτικά την τιμή της τρέχουσας κατάστασης του στοιχείου χρησιμοποιώντας μόνο τις πληροφορίες που είναι σημαντικές τόσο από την προηγούμενη κατάσταση του στοιχείου όσο και από την τρέχουσα υπονήφια κατάσταση του. Για την κατάσταση του υπονήφιου στοιχείου (candidate cell state), η συνάρτηση ενεργοποίησης είναι μια συνάρτηση υπερβολικής εφαπτομένης (hyperbolic tangent function) που εξάγει τιμές στο εύρος $[-1, 1]$. Μια σημαντική διαφορά μεταξύ του στοιχείου LSTM και ενός απλού στοιχείου RNN είναι ότι η έξοδος z_t του πρώτου είναι ίση με την κρυφή κατάσταση h_t .

Σε αυτή την ανασκόπηση, χρησιμοποιείται μια παραλλαγή του στοιχείου LSTM που είναι γνωστό ως στοιχείο LSTM με peephole συνδέσεις. Σε αυτή τη μονάδα LSTM, οι καταστάσεις ενημερώνονται ως εξής:

$$\begin{aligned} i_t &= \sigma(W_i \cdot h_{t-1} + V_i \cdot x_t + P_i \cdot C_{t-1} + b_i) \\ o_t &= \sigma(W_o \cdot h_{t-1} + V_o \cdot x_t + P_o \cdot C_t + b_o) \\ f_t &= \sigma(W_f \cdot h_{t-1} + V_f \cdot x_t + P_f \cdot C_{t-1} + b_f) \\ \bar{C}_t &= \tanh(W_c \cdot h_{t-1} + V_c \cdot x_t + b_c) \\ C_t &= i_t \odot \bar{C}_t + f_t \odot C_{t-1} \\ h_t &= o_t \odot \tanh(C_t) \\ z_t &= h_t \end{aligned}$$

Η διαφορά είναι ότι οι peephole συνδέσεις επιτρέπουν στην πύλη forget και στην πύλη εισόδου του στοιχείου να εξετάσουν την προηγούμενη κατάσταση του στοιχείου C_{t-1} πριν την ενημερώσουν. Στις προαναφερόμενες ανανεωμένες εξισώσεις, $P_i, P_o, P_f, P_c \in \mathbb{R}^{d \times d}$ αντιπροσωπεύουν τους πίνακες βαρών των πυλών εισόδου, εξόδου και forget, αντίστοιχα. Όσον αφορά την πύλη εξόδου, μπορεί τώρα να επιθεωρήσει την τρέχουσα κατάσταση του στοιχείου για τη δημιουργία της εξόδου. Η υλοποίηση που χρησιμοποιείται στην παρούσα έρευνα είναι η προεπιλεγμένη υλοποίηση των συνδέσεων peephole στο πλαίσιο Tensorflow.



Σχήμα 3. Gated Recurrent Unit (GRU).

Η μονάδα GRU είναι μια άλλη παραλλαγή, που εισήχθη από τους Cho et al. (2014), η οποία είναι συγκριτικά απλούστερη από τη μονάδα LSTM, καθώς και ταχύτερη στους υπολογισμούς. Αυτό οφείλεται στο γεγονός ότι η μονάδα LSTM διαθέτει τρεις πύλες εντός του εσωτερικού μηχανισμού πύλης, ενώ η GRU διαθέτει μόνο δύο (δηλαδή την πύλη ενημέρωσης (update gate) και την πύλη επαναφοράς (reset gate)). Η πύλη ενημέρωσης σε αυτή τη μονάδα παίζει το συνδυαστικό ρόλο της πύλης forget και της πύλης εισόδου. Επιπλέον, παρόμοια με το στοιχείο RNN, το στοιχείο GRU έχει μόνο μία συνιστώσα της κατάστασης, δηλαδή την κρυφή κατάσταση. Το Σχήμα 3 και οι ακόλουθες εξισώσεις παρουσιάζουν τη λειτουργικότητα της κυψέλης GRU:

$$\begin{aligned}u_t &= \sigma(W_u \cdot h_{t-1} + V_u \cdot x_t + b_u) \\r_t &= \sigma(W_r \cdot h_{t-1} + V_r \cdot x_t + b_r) \\ \bar{h}_t &= \tanh(W_h \cdot r_t \cdot h_{t-1} + V_h \cdot x_t + b_h) \\h_t &= u_t \odot \bar{h}_t + (1 - u_t) \odot h_{t-1} \\z_t &= h_t\end{aligned}$$

όπου $u_t, r_t \in \mathbb{R}^d$ υποδηλώνουν τις πύλες ενημέρωσης και επαναφοράς, αντίστοιχα. $\bar{h}_t \in \mathbb{R}^d$ υποδηλώνει την υποψήφια κρυφή κατάσταση και $h_t \in \mathbb{R}^d$ υποδηλώνει την τρέχουσα κρυφή κατάσταση στο χρονικό βήμα t . Τα βάρη και οι προκαταλήψεις ακολουθούν τον ίδιο συμβολισμό που αναφέρθηκε παραπάνω. Η πύλη επαναφοράς αποφασίζει πόσο η προηγούμενη κρυφή κατάσταση συμβάλλει στην υποψήφια κατάσταση του τρέχοντος βήματος. Δεδομένου ότι η πύλη ενημέρωσης λειτουργεί μόνη της χωρίς πύλη forget, $(1 - u_t)$ χρησιμοποιείται ως εναλλακτική λύση. Η GRU είναι δημοφιλής, λόγω της απλότητας της, έχοντας λιγότερες παραμέτρους από το κύτταρο LSTM και λόγω της αποτελεσματικότητας της στην εκπαίδευση.

4.2.3 Προκλήσεις και Εξελίξεις

Παρά τις εξαιρετικές δυνατότητες που προσφέρουν, τα απλά Αναδρομικά Νευρωνικά Δίκτυα αντιμετωπίζουν προκλήσεις τόσο κατά τη διαδικασία της εκπαίδευσης όσο και κατά τη φάση της λειτουργίας τους. Ένα από τα πρωτεύοντα προβλήματα που αναδύονται είναι το γνωστό vanishing gradient problem.[63] Κατά την εκπαίδευση των RNNs, ιδίως όταν αυτά είναι μεγάλα σε μέγεθος, οι αναδρομικοί υπολογισμοί μπορεί να προκαλέσουν πολύ μικρές τιμές στις παραγώγους των συναρτήσεων ενεργοποίησης, με αποτέλεσμα η διαδικασία εκπαίδευσης να προχωρά με αργούς ρυθμούς ή ακόμη και να καταλήγει σε αδυναμία εκμάθησης. Αυτό έχει αναστατώσει την ικανότητα των απλών RNNs να αντιμετωπίσουν ακολουθίες δεδομένων με μεγάλο χρονικό ορίζοντα.

Για την αντιμετώπιση αυτών των προκλήσεων, έχουν προταθεί προηγμένες αρχιτεκτονικές όπως τα Μακροσκοπικά Μνημονικά Δίκτυα (LSTMs) και τα Δίκτυα GRU (Gated Recurrent Units).[68] Τα LSTMs αντιπροσωπεύουν μια πιο πολύπλοκη προσέγγιση στη διαχείριση της εσωτερικής κατάστασης και των κλίσεων των RNNs. Από την άλλη, τα Δίκτυα GRU παρέχουν μια πιο απλοποιημένη δομή, με λιγότερες πύλες από τα LSTMs, όπως αναφέρεται και παραπάνω.

Αυτές οι εξελίξεις στον τομέα των RNNs έχουν επιτρέψει τη μοντελοποίηση ακόμα μεγαλύτερων και πιο πολύπλοκων ακολουθιών δεδομένων. Ενώ τα LSTMs και τα Δίκτυα GRU

είναι απλώς μερικά παραδείγματα από τις πολλές διαθέσιμες αρχιτεκτονικές, αυτή η προσέγγιση έχει ενισχύσει την απόδοση των RNNs σε ποικίλους τομείς, όπως η επεξεργασία φυσικής γλώσσας, η αναγνώριση φωνής και η πρόβλεψη χρονοσειρών.[\[67\]](#)

Συμπερασματικά, η θεμελιώδης αρχή των Αναδρομικών Νευρωνικών Δικτύων, που στηρίζεται στην έννοια της αναδρομής και της εσωτερικής κατάστασης, έχει ανοίξει νέους δρόμους στον χώρο της μηχανικής μάθησης. Η δυνατότητα αυτών των δικτύων να αποθηκεύουν και να αναλύουν πληροφορίες από προηγούμενα στοιχεία μιας ακολουθίας τους παραχωρεί μια εξαιρετική ικανότητα να αντιμετωπίζουν προβλήματα που σχετίζονται με τη χρονική σειρά των δεδομένων. Με την εξέλιξη της τεχνολογίας, παρατηρούμε συνεχή βελτίωση στην απόδοση των RNNs. Αυτές οι εξελίξεις διευρύνουν τις δυνατότητες των RNNs στο να αναγνωρίζουν πολύπλοκα πρότυπα και να εκτελούν ακόμα πιο απαιτητικές εργασίες πρόβλεψης και μοντελοποίησης. Προβλέπεται ότι τα RNNs και τα συναφή μοντέλα θα συνεχίσουν να διαδραματίζουν ζωτικό ρόλο σε πολλούς τομείς, όπως η φυσική γλώσσα επεξεργασία, η αναγνώριση φωνής, η ανάλυση κειμένου, η πρόβλεψη χρονοσειρών και η ιατρική διάγνωση. Με την αξιοποίηση της δύναμης της αναδρομικής λειτουργίας και της εσωτερικής κατάστασης, αναμένουμε ότι η πορεία των RNNs θα συμβάλει στην περαιτέρω καινοτομία και πρόοδο στον τομέα της μηχανικής μάθησης και της τεχνητής νοημοσύνης.

4.3 Μοντέλα-Αρχιτεκτονικές RNN

Τα Αναδρομικά Νευρωνικά Δίκτυα αποτελούν μια σημαντική κατηγορία νευρωνικών δικτύων που έχει αποδειχθεί ότι είναι εξαιρετικά αποτελεσματική στην αντιμετώπιση ακολουθιακών δεδομένων, όπως κείμενο, ήχος και χρονοσειρές. Ενώ οι παραδοσιακοί προσεγγιστικοί αλγόριθμοι βασίζονται στην υπόθεση της ανεξαρτησίας των δειγμάτων, τα RNNs αντιμετωπίζουν τη δομή του χρόνου και τις εξαρτήσεις μεταξύ των δειγμάτων. Ωστόσο, παρουσιάζουν προκλήσεις όπως τη δυσκολία στην εκμάθηση μακροπρόθεσμων εξαρτήσεων και το πρόβλημα της εξαφάνισης ή διασποράς της πληροφορίας κατά τη διάρκεια της αναδρομής.[\[67\]\[68\]](#)

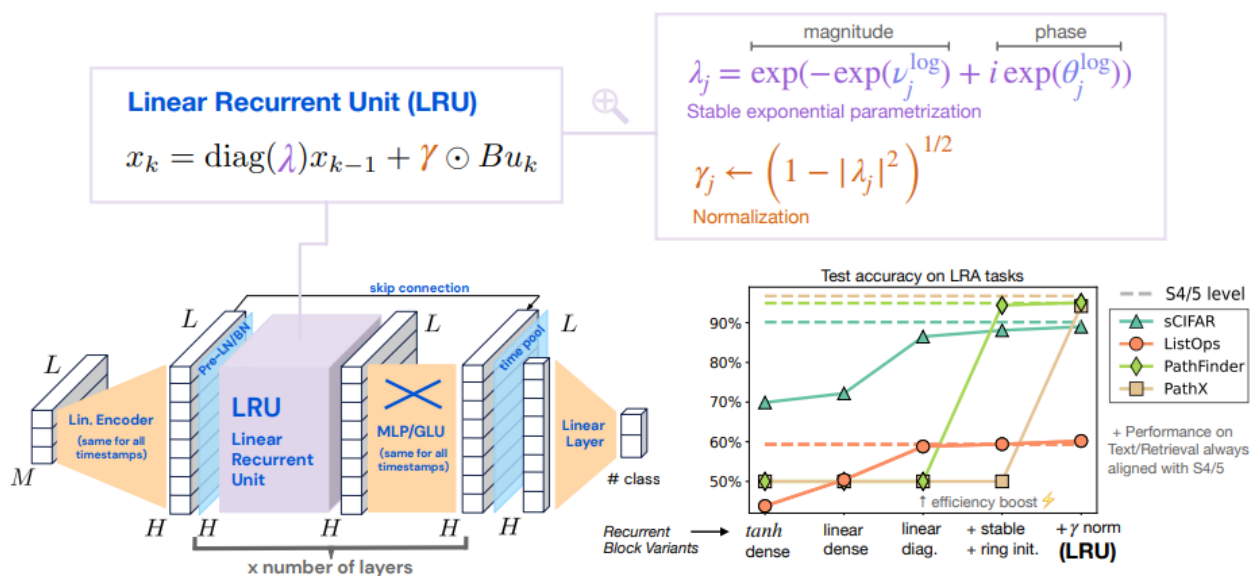
Κατά τη διάρκεια της εξέλιξής τους, έχουν αναπτυχθεί διάφορες αρχιτεκτονικές RNN που στοχεύουν στην αντιμετώπιση των προκλήσεων αυτών. Στην παρούσα εργασία, θα εξετάσουμε με περισσότερη λεπτομέρεια τρεις σημαντικές αρχιτεκτονικές RNN: τις Γραμμικές Αναδρομικές Μονάδες (Linear Recurrent Units - LRUs)[\[63\]](#), τις Μονάδες με Μοναδική Μνήμη (Unique Memory Units - UMUs)[\[70\]](#) και τις Αναδρομικές Μονάδες Μακράς Προθεσμίας (Long Short-Term Memory - LSTM).[\[68\]\[69\]](#)

4.3.1 Γραμμικές Αναδρομικές Μονάδες (LRUs)

Οι Γραμμικές Αναδρομικές Μονάδες (Linear Recurrent Units - LRUs) ανήκουν στην κατηγορία των αναδρομικών νευρωνικών δικτύων (RNNs) και αποτελούν μία από τις πρώτες αρχιτεκτονικές που εισήχθησαν για την επεξεργασία ακολουθιακών δεδομένων. Η βασική τους ιδέα είναι να διατηρούν μια αναδρομή στο χρόνο, κρατώντας παράλληλα μια απλή δομή που επιτρέπει την ανανέωση των πληροφοριών. Είναι κατάλληλες για απλά προβλήματα όπου οι συσχετίσεις στο παρελθόν είναι σημαντικές, αλλά χρειάζονται περαιτέρω βελτίωση για πιο πολύπλοκες εργασίες.

Οι LRUs λειτουργούν με την ανανέωση των εσωτερικών καταστάσεών τους καθώς και την μεταφορά της πληροφορίας προς τα εμπρός. Κάθε LRUs δέχεται είσοδο από την προηγούμενη χρονική στιγμή και τη συνδυάζει με την τρέχουσα είσοδο. Αυτή η αναδρομική διαδικασία τους επιτρέπει να αντιλαμβάνονται το παρελθόν και τη συνεχόμενη εξέλιξη των δεδομένων. Παρ' όλα αυτά, έχουν μια σημαντική αδυναμία στην αντιμετώπιση μακροπρόθεσμων εξαρτήσεων σε ακολουθίες, γνωστή και ως πρόβλημα vanishing gradient.

Ακολουθεί μία εικόνα αντλημένη από το άρθρο “[Resurrecting Recurrent Neural Networks for Long Sequences](#)” των Antonio Orvieto, Samuel L Smith, Albert Gu2, κ.α., η εικόνα περιγράφει την αρχιτεκτονική Deep Linear Recurrent Units (LRU) που εισάγεται στην προαναφερόμενη εργασία. Το μοντέλο είναι μια στοίβα από μπλοκ LRU, με μη γραμμικές προβολές ενδιάμεσα, και χρησιμοποιεί επίσης συνδέσεις παράλειψης και μεθόδους κανονικοποίησης όπως η κανονικοποίηση παρτίδας/στρώματος (batch/layer normalization). Στην δεξιά πλευρά της εικόνας δίνεται μία σύνοψη αποτελεσμάτων για τα κύρια βήματα που περιγράφονται στην εισαγωγή της εργασίας για το σχεδιασμό LRUs ξεκινώντας από tanh RNNs.



Ο κύριος λόγος που οι LRUs δυσκολεύονται να αντιμετωπίσουν μακροπρόθεσμες εξαρτήσεις είναι η τοποθέτηση γραμμικών συναρτήσεων εντός της αναδρομικής δομής. Οι γραμμικές συναρτήσεις είναι περιορισμένες στην ικανότητα τους να μάθουν πολύπλοκες αναλογίες και εξαρτήσεις στα δεδομένα. Αυτό το περιορισμένο εύρος αναπαράστασης δεδομένων συνεπάγεται ότι οι LRUs δυσκολεύονται να διατηρήσουν σημαντικές πληροφορίες για μεγάλα χρονικά διαστήματα.

Συνεπώς, παρότι οι LRUs μπορούν να είναι χρήσιμες για απλά προβλήματα όπου η συσχέτιση στο παρελθόν είναι κρίσιμη, αντιμετωπίζουν δυσκολίες όταν αντιμετωπίζουν πιο πολύπλοκες ακολουθιακές εργασίες που απαιτούν την ανίχνευση μακροπρόθεσμων εξαρτήσεων. Η περαιτέρω βελτίωση των LRUs ή η χρήση πιο προηγμένων αρχιτεκτονικών μπορεί να αντιμετωπίσει αυτήν την περιοριστική πτυχή της απόδοσής τους.

Με βάση το άρθρο “[Presenting a Novel Page Replacement Algorithm Based on LRU](#)” των Ali Khosrozadeh, Sanaz Pashmforoush, Abolfazl Akbari, Maryam Bagheri, Neda Beikmahdavi, εμβαθύνουμε λίγο περισσότερο στην αρχιτεκτονική των LRUs.

Η διαχείριση της μνήμης σε ένα υπολογιστικό σύστημα απαιτεί τον έλεγχο δύο ή περισσότερων επιπέδων μνήμης. Μεταξύ αυτών των επιπέδων, επιλέγεται ως πρώτο επίπεδο αυτό με τον μικρότερο χρόνο πρόσβασης, το οποίο ονομάζεται επίσης πρωτογενής μνήμη. Λόγω του σύντομου χρόνου πρόσβασης, η μνήμη αυτή είναι ακριβή ενώ έχει μικρότερο αποθηκευτικό χώρο. Αυτός ο περιορισμένος χώρος, με τη σειρά του, οδηγεί σε γρήγορη υπερχειλίση και έτσι, εάν απαιτείται περισσότερη μνήμη, ορισμένα τμήματα αυτής της μνήμης πρέπει να είναι κενά.

Μια σημαντική μέθοδος για τη διαχείριση του χώρου μνήμης και την κατανομή του περιορισμένου χώρου μεταξύ των εφαρμογών που θα ήταν η τμηματοποίηση της μνήμης (memory segmentation). Όμως, εξαιτίας του φαινομένου του εξωτερικού κατακερματισμού (external fragmentation) που απέδωσε σε προφανή σπατάλη του χώρου μνήμης, η μέθοδος αυτή αντικαταστάθηκε από τη μέθοδο σελιδοποίησης (paging method). Επίσης, η λανθάνουσα μνήμη σπατάλη μνήμης ήταν ο λόγος που η μέθοδος σελιδοποίησης ήταν αναποτελεσματική. Έτσι, μια μέθοδος που βασίζεται σε συνδυασμό αυτών των δύο μοντέλων εισήχθη για να επιλύσει τα προβλήματά τους. Στη βάση της σελιδοποίησης, ο διαθέσιμος χώρος μνήμης χωρίζεται σε μπλοκ γνωστά ως πλαίσια σελίδων. Κάθε εφαρμογή που ζητάει τη μνήμη χωρίζεται σε κάποιες σελίδες και θα της δοθεί ένας αριθμός πλαισίων σελίδων για να περιέχει μερικές από αυτές τις σελίδες. Ο αριθμός αυτός καθορίζεται από το λειτουργικό σύστημα και μπορεί να είναι διαφορετικός για εφαρμογές. Μέχρι η μνήμη να περιέχει τουλάχιστον ένα κενό πλαίσιο σελίδας, ο καθορισμός των πλαισίων για τις εφαρμογές και η φόρτωση των σελίδων σε αυτές είναι σχετικά εύκολη.

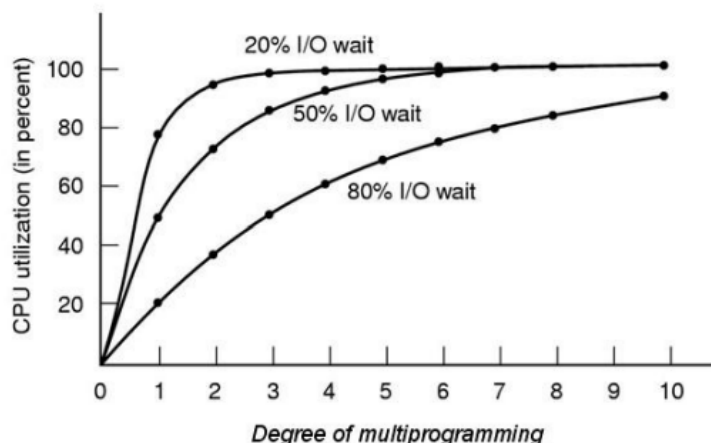
Όμως, το κύριο πρόβλημα στη διαχείριση των πλαισίων θα εξελιχθεί όταν όλα τα διαθέσιμα πλαίσια έχουν οριστεί σε εφαρμογές και δεν υπάρχουν κενά πλαίσια. Η λύση γι' αυτό, θα είναι η διαγραφή μιας ή περισσότερων σελίδων από τη μνήμη και η απελευθέρωση κάποιων πλαισίων που θα οριστούν για νέες σελίδες. Είναι σημαντικό να καθοριστεί το ποιες σελίδες πρέπει να διαγραφούν. Η κατάργηση μιας σελίδας που μπορεί να απαιτηθεί στο εγγύτερο μέλλον μπορεί να επηρεάσει την απόδοση του συστήματος, επειδή επιβάλλει επιβάρυνση χρόνου επαναφόρτωσης.

Αυτές οι προκλήσεις έχουν παρακινήσει τους ερευνητές να εισαγάγουν αλγόριθμους αντικατάστασης σελίδων, οι οποίοι επιλέγουν μια σελίδα που θα αντικατασταθεί με μια νέα, βάσει καθορισμένων κριτηρίων, πολιτικών και σκοπών. Σε διαφορετικές συνθήκες, διαφορετικοί αλγόριθμοι έχουν διαφορετικές επιδόσεις. Μεταξύ των προτεινόμενων μεθόδων, μπορούν να αναφερθούν οι αλγόριθμοι FIFO, NRU, Second Chance, Clock, LRU και Optimal. Επίσης, μεταξύ αυτών, ο Least Frequently Used (LFU) είναι γνωστός ως ο πιο αποτελεσματικός αλγόριθμος. Αυτός ο αλγόριθμος λαμβάνει υπόψη τον αριθμό των φορών στις οποίες έχει αναφερθεί κάθε μόνιμη σελίδα στη μνήμη και επιλέγει μια σελίδα με ελάχιστη τιμή μετρητή που θα μετακινηθεί εκτός μνήμης. Δεδομένου ότι αυτός ο αλγόριθμος ήταν η πιο αποδοτική μέθοδος μεταξύ των πρακτικών αλγορίθμων, οι περισσότερες έρευνες έχουν επικεντρωθεί σε αυτόν για τη βελτίωση των επιδόσεων.

Όπως αναφέρθηκε προηγουμένως, κατά τη διάρκεια της εκτέλεσης ενός προγράμματος, συνήθως απαιτείται η φόρτωση ορισμένων δεδομένων στην κύρια μνήμη. Στη διαχείριση της μνήμης, με βάση το μηχανισμό σελιδοποίησης, τα απαιτούμενα δεδομένα εισάγονται στη μνήμη ως περιεχόμενο σελίδων. Όταν μια απαιτούμενη σελίδα δεν βρίσκεται στη μνήμη, εμφανίζεται ένα συμβάν σφάλματος σελίδας. Σε αυτές τις περιπτώσεις, η απαιτούμενη σελίδα πρέπει να φορτωθεί στη μνήμη. Εάν όμως όλα τα πλαίσια σελίδων είναι γεμάτα, το κύριο πρόβλημα θα είναι η εύρεση ενός κατάλληλου πλαισίου που θα περιέχει την απαιτούμενη σελίδα. Πραγματικά, το λειτουργικό σύστημα πρέπει να επιλέξει ένα πλαίσιο σελίδας στην πρωτογενή μνήμη και να μετακινήσει το περιεχόμενό του από αυτό, έτσι ώστε να παρέχεται ένα κενό πλαίσιο σελίδας για τη νέα σελίδα, που εισέρχεται στη μνήμη.

Παρόλο που, όταν παρουσιάζεται σφάλμα σελίδας, ένα πλαίσιο μπορεί να επιλεγεί τυχαία, προκειμένου να αυξηθεί η απόδοση του συστήματος, θα είναι καλύτερο να επιλέξετε μια σελίδα που δεν επηρεάζει τις πολιτικές εκτέλεσης συστήματος και προγράμματος, επειδή, εάν μια σελίδα

με υψηλή επιλέγεται η αναλογία χρήσης και απομακρύνεται από την κύρια μνήμη, ενδέχεται να χρειαστεί να φορτωθεί ξανά σύντομα. Όταν τα περιεχόμενα μιας σελίδας φορτώνονται στην πρωτογενή μνήμη, στην πραγματικότητα διαβάζονται από μια δευτερογενή μνήμη (input operation). Επίσης, όταν μια σελίδα μετακινείται από την πρωτογενή μνήμη, πρέπει να εγγραφεί σε μια δευτερογενή μνήμη (output operation). Έτσι, ένα σφάλμα σελίδας μπορεί να επιβάλει λειτουργίες εισόδου/εξόδου στο σύστημα. Όπως φαίνεται στο Σχήμα 1, η μείωση της εισόδου/εξόδου σε μοντέλα πολλαπλού προγραμματισμού αυξάνει την απόδοση της CPU. Η κατάλληλη επιλογή των σελίδων μπορεί να μειώσει τα σφάλματα σελίδων και έτσι, μπορεί να μειώσει το I/O, το οποίο με τη σειρά του, αυξάνει την απόδοση του συστήματος.



Σχήμα 1. Αύξηση της απόδοσης της CPU με μείωση I/O.

Η ιδέα πίσω από αυτόν τον αλγόριθμο (LRU) βασίζεται στο γεγονός ότι οι σελίδες που χρησιμοποιήθηκαν ευρέως κατά τις τελευταίες οδηγίες, πιθανόν να χρησιμοποιηθούν ξανά ευρέως κατά τις επόμενες. Και, αντίθετα, οι σελίδες που δεν έχουν χρησιμοποιηθεί για μεγάλο χρονικό διάστημα, πιθανόν να παραμείνουν αχρησιμοποίητες στη συνέχεια. Έτσι, επιλέγει μια σελίδα που η τελευταία της χρήση είναι πριν από όλες τις άλλες μόνιμες σελίδες. Η ιδέα αυτή βασίζεται στη θεωρία της εντοπιότητας (Locality Theory). Ο αλγόριθμος LRU έχει χρησιμοποιηθεί ως κοινός αλγόριθμος αντικατάστασης σελίδας και έχει θεωρηθεί ως επιτυχημένος αλγόριθμος. Το Σχήμα 2 δείχνει αυτή τη λειτουργικότητα του αλγορίθμου λαμβάνοντας υπόψη μια τυχαία ακολουθία κλήσης σελίδων.

reference string

7 0 1 2 0 3 0 4 2 3 0 3 2 1 2 0 1 7 0 1

7	7	7	2	2	4	4	4	0	1	1	1
	0	0	0	0	0	0	3	3	3	0	0
		1	1	3	3	2	2	2	2	2	7

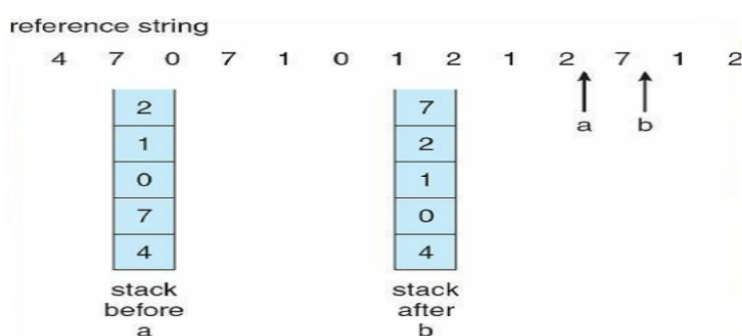
page frames

Σχήμα 2. Εκτέλεση αλγορίθμου LRU λαμβάνοντας υπόψη 3 πλαίσια σελίδας

Αν και αυτός ο αλγόριθμος έχει μια κατάλληλη απόδοση, ως μειονέκτημα, μπορεί να θεωρηθεί η δύσκολη και δαπανηρή υλοποίησή του. Η πλήρης εφαρμογή του LRU απαιτεί τη διατήρηση μιας συνδεδεμένης λίστας όλων των σελίδων στη μνήμη, στην οποία η σελίδα που χρησιμοποιήθηκε πιο πρόσφατα τοποθετείται στο μπροστινό μέρος και η λιγότερο πρόσφατα χρησιμοποιημένη σελίδα στο πίσω μέρος. Αυτός ο κατάλογος πρέπει να ενημερώνεται σε κάθε αναφορά μνήμης. Αυτό περιλαμβάνει την εύρεση μιας σελίδας στη λίστα, τη διαγραφή της και στη συνέχεια τη μετακίνησή της στο μπροστινό μέρος. Πρόκειται όμως για μια πολύ χρονοβόρα λειτουργία, ακόμη και σε hardware (αν υποθέσουμε ότι θα μπορούσε να κατασκευαστεί τέτοιο hardware). Ως άλλος τρόπος εφαρμογής του LRU, μπορούν να εξεταστούν δύο μέθοδοι:

→ Χρήση στοίβας (stack)

Σε αυτή την περίπτωση, κάθε σελίδα που καλείται εισάγεται στην κορυφή της στοίβας. Έτσι, οι σελίδες που έχουν τη μεγαλύτερη χρήση, πρόσφατα, τοποθετούνται πάντα στην κορυφή της στοίβας. Το Σχήμα 3 δείχνει πώς λειτουργεί αυτή η στοίβα.



Σχήμα 3. Hardware implementation για LRU από τη στοίβα

→ Χρήση μετρητή (counter)

Σε αυτή τη μέθοδο, χρησιμοποιείται ένας μετρητής 64-bit (C). Μετά την εκτέλεση κάθε εντολής, αυτός ο μετρητής αυξάνεται αυτόματα. Η καταχώρηση του πίνακα σελίδων για κάθε μόνιμη σελίδα περιλαμβάνει ένα πεδίο που μπορεί να αποθηκεύσει την τιμή του μετρητή, και όταν γίνεται αναφορά σε μια σελίδα, η τρέχουσα τιμή του C αποθηκεύεται στην αντίστοιχη καταχώρηση του πίνακα σελίδων. Όταν προκύψει σφάλμα σελίδας, εξετάζονται όλες οι τιμές του μετρητή που είναι αποθηκευμένες στις καταχωρήσεις του πίνακα σελίδων και επιλέγεται η χαμηλότερη. Αυτή η ελάχιστη τιμή αντιστοιχεί στη σελίδα που έχει χρησιμοποιηθεί λιγότερο πρόσφατα. Έτσι, αυτή η σελίδα θα μετακινηθεί από τη μνήμη και η θέση της θα οριστεί για μια νέα σελίδα.

Η μελέτη έδειξε ότι ο LRU είχε την καλύτερες επιδόσεις. Στην παρούσα εργασία, παρουσιάστηκε ένας αλγόριθμος αντικατάστασης σελίδων με βάση τον LRU. Λαμβάνοντας υπόψη τα αποτελέσματα των πειραμάτων δείχνει ότι η προσθήκη αυτής της παραμέτρου μπορεί να βελτιώσει την απόδοση της διαδικασίας αντικατάστασης σελίδων.

4.3.2 Μονάδες με Μοναδική Μνήμη (UMUs)

Οι μονάδες με Μοναδική Μνήμη (Unique Memory Units - UMUs) αναδεικνύονται ως μια πρωτοποριακή προσέγγιση στο πλαίσιο των αναδρομικών νευρωνικών δικτύων.[63] Η βασική τους αρχή βασίζεται στην ιδέα της διατήρησης πληροφορίας από προηγούμενες χρονικές στιγμές

με τρόπο που να αντιμετωπίζει την πρόκληση vanishing gradient[73]. Αυτό σημαίνει ότι οι UMUs διατηρούν ένα είδος "μοναδικής μνήμης" που επιτρέπει την αναγνώριση παραγόντων που έχουν επίδραση σε μακροπρόθεσμη βάση.[64] Επιτρέποντας την αποθήκευση πληροφορίας για περισσότερες χρονικές στιγμές, οι UMUs είναι σε θέση να αναγνωρίζουν πρότυπα και σχέσεις που εκτείνονται σε μεγάλα χρονικά διαστήματα.[65]

Αυτό το χαρακτηριστικό είναι εξαιρετικά σημαντικό για προβλήματα όπου οι εξαρτήσεις και οι συνδέσεις μεταξύ των δεδομένων είναι πολύπλοκες και απαιτούν περαιτέρω ανάλυση.[74] Ωστόσο, αν και οι UMUs αντιμετωπίζουν το πρόβλημα vanishing gradient με τρόπο αποτελεσματικό, μπορεί να υπάρχουν περιπτώσεις όπου απαιτείται πολύπλοκη ανάλυση των δεδομένων.[75] Αυτό συμβαίνει όταν τα δεδομένα είναι πολύ ανεπαίσθητα ή οι συσχετίσεις μεταξύ τους είναι εξαιρετικά περίπλοκες.

Συνεπώς, παρά την επιτυχία των UMUs στην αντιμετώπιση προβλημάτων, είναι σημαντικό να αναγνωριστεί ότι απαιτείται περισσότερη σύνθετη ανάλυση.[76] Παρ' όλα αυτά, οι UMUs αποτελούν μια σημαντική εξέλιξη στον τομέα των αναδρομικών νευρωνικών δικτύων και ανοίγουν νέες προοπτικές για την ανάλυση και την εκμάθηση ακολουθιακών δεδομένων.

4.3.3 Αναδρομικές Μονάδες Μακράς Προθεσμίας (LSTM)

Οι Αναδρομικές Μονάδες Μακράς Προθεσμίας (Long Short-Term Memory - LSTM) αποτελούν μια προηγμένη αρχιτεκτονική RNN που έχει επιλύσει επιτυχώς το σημαντικό πρόβλημα vanishing gradient, όπως έχει αναλυθεί παραπάνω.

Βασίζόμενοι στο άρθρο “[A Novel Stochastic LSTM Model Inspired by Quantum Machine Learning](#)” των Joseph Lindsay και Ramtin Zand, εμβαθύνουμε περεταίρω στην ιδέα των LSTM. Ως παραλλαγή του RNN, το μοντέλο LSTM είναι καταλληλότερο για την ανάλυση ακολουθιών δεδομένων χρονοσειρών για την εκμάθηση χρονικών εξαρτήσεων. Οι κύριες διαφορές που κάνουν τα δίκτυα LSTM να διαφέρουν από τα τυπικά RNN είναι το σύνολο των εξισώσεων που χρησιμοποιούνται εντός του στοιχείου του δικτύου και η προσθήκη μιας άλλης παραμέτρου, που αναφέρεται ως κατάσταση του στοιχείου, η οποία περνάει στα επόμενα χρονικά βήματα της τρέχουσας ακολουθίας του δικτύου, όπως η κρυφή του κατάσταση.

Όπου v_t είναι η προηγούμενη κρυφή κατάσταση συνδεδεμένη με τα τρέχοντα δεδομένα s.t. $v_t = [h_t - 1, x_t]$, W_f και b_f είναι τα βάρη και οι προκαταλήψεις για τη λειτουργία forget, W_i και b_i είναι τα βάρη και οι προκαταλήψεις για τη λειτουργία input, $W_{\bar{c}}$ και $b_{\bar{c}}$ είναι τα βάρη και οι προκαταλήψεις για τη λειτουργία update candidate, και W_o και b_o είναι τα βάρη και οι προκαταλήψεις για τη λειτουργία output, ένα κλασικό στοιχείο LSTM μπορεί να εκφραστεί μαθηματικά από τις ακόλουθες εξισώσεις:

$$f_t = \sigma(W_f \cdot v_t + b_f)$$

$$i_t = \sigma(W_i \cdot v_t + b_i)$$

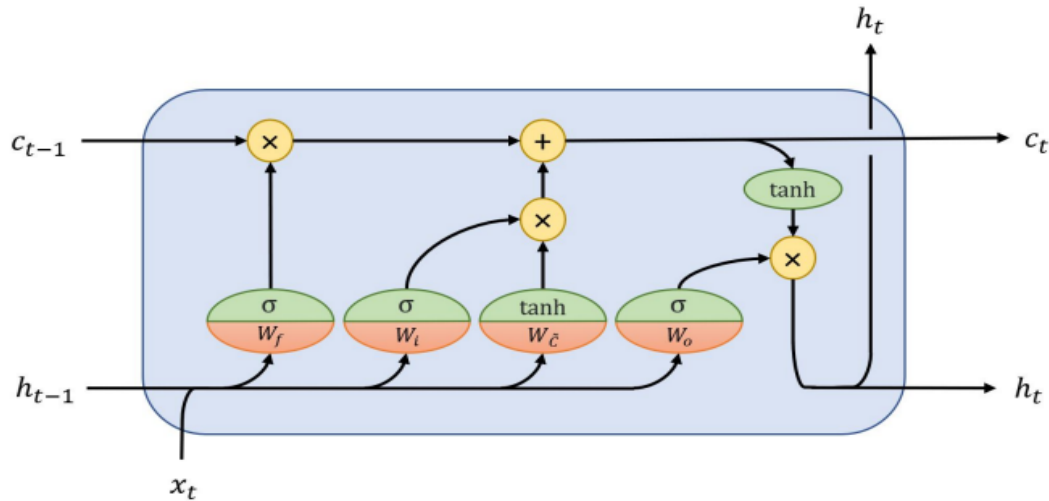
$$\bar{c}_t = \tanh(W_{\bar{c}} \cdot v_t + b_{\bar{c}})$$

$$c_t = (f_t \cdot c_{t-1}) + (i_t \cdot \bar{c}_t)$$

$$o_t = \sigma(W_o \cdot v_t + b_o)$$

$$h_t = o_t \cdot \tanh(c_t)$$

Αυτή η κατασκευή ενός στοιχείου LSTM όπως περιγράφεται παραπάνω απεικονίζεται στο Σχήμα 1.



Σχήμα 1. Γραφική αναπαράσταση στοιχείου LSTM.

Αξίζει να αναφερθεί ότι κάθε μία από τις πράξεις πολλαπλασιασμού και συσσώρευσης (multiply-and-accumulate - MAC) πραγματοποιείται με ξεχωριστά σύνολα βαρών και προκαταλήψεων, τα οποία υποδεικνύονται από τους δείκτες τους. Σε υψηλό επίπεδο, η λειτουργία ενός στοιχείου LSTM μπορεί να περιγραφεί από τις τρεις πύλες από τις οποίες αποτελείται: την πύλη forget, την πύλη input and update και την πύλη output.

Σύμφωνα με την πρώτη εξίσωση και το πρώτο μέρος της τέταρτης εξίσωσης, η πύλη forget έρχεται πρώτη και λειτουργεί για να ενημερώσει τη συνάφεια των δεδομένων από την προηγούμενη κατάσταση του στοιχείου ανάλογα με τα δεδομένα του τρέχοντος χρονικού βήματος.

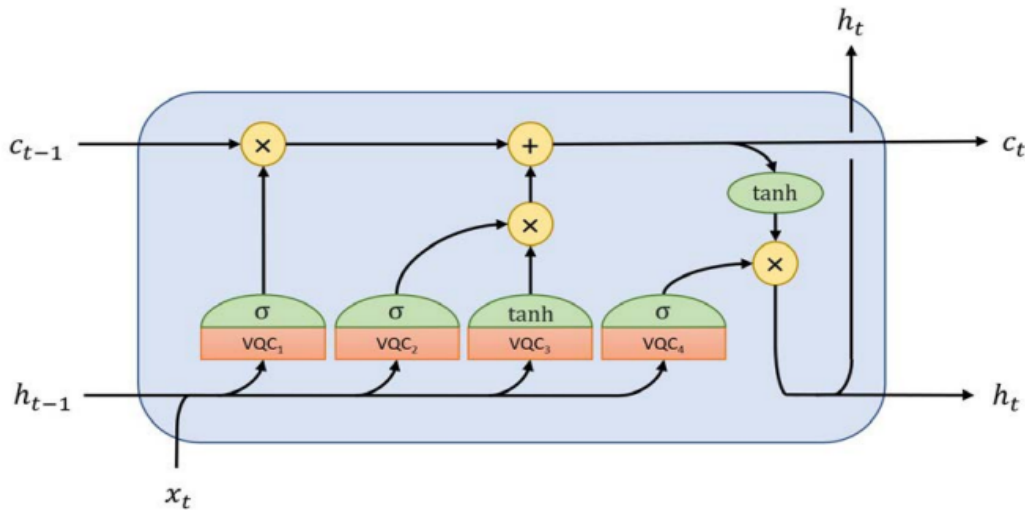
Το επόμενο τμήμα του στοιχείου αποτελούμενο από τη δεύτερη, τρίτη και το υπόλοιπο της τέταρτης εξίσωσης, η πύλη input and update είναι το επόμενο τμήμα του στοιχείου LSTM και λειτουργεί για να αποφασίσει με ποιες νέες πληροφορίες θα πρέπει να ενημερωθεί η τρέχουσα κατάσταση του. Αφού καθοριστεί ποιες τιμές δεδομένων της κατάστασης του πρέπει να ενημερωθούν μέσω της σιγμοειδούς ενεργοποίησης και αφού δημιουργηθεί ένας αρχικός υποψήφιος με τον οποίο θα ενημερωθεί η κατάσταση του στοιχείου μέσω της ενεργοποίησης της υπερβολικής εφαπτομένης, τα αποτελέσματα πολλαπλασιάζονται μεταξύ τους στοιχειωδώς για να παραχθούν κλιμακωτές υποψήφιες τιμές που θα προσδώσουν την ενημέρωση στην κατάσταση του στοιχείου που προκύπτει από την πύλη forget.

Αποτελούμενο από τις δύο τελευταίες εξισώσεις, την πέμπτη και έκτη, το τελικό τμήμα του στοιχείου LSTM, η πύλη output, θέτει την τρέχουσα κρυφή κατάσταση του δικτύου πριν τη μεταβιβάσει μαζί με την κατάσταση του στοιχείου στο επόμενο χρονικό βήμα και ενδεχομένως επιστρέφοντας αυτή την κρυφή κατάσταση ως έξοδο.

Συνεχίζοντας με την κβαντική εκδοχή της αρχιτεκτονικής LSTM (Quantum LSTM - QLSTM), όπως έδειξαν οι Chen et al , μια τέτοια εκδοχή μπορεί να υλοποιηθεί με πολύ παρόμοιο

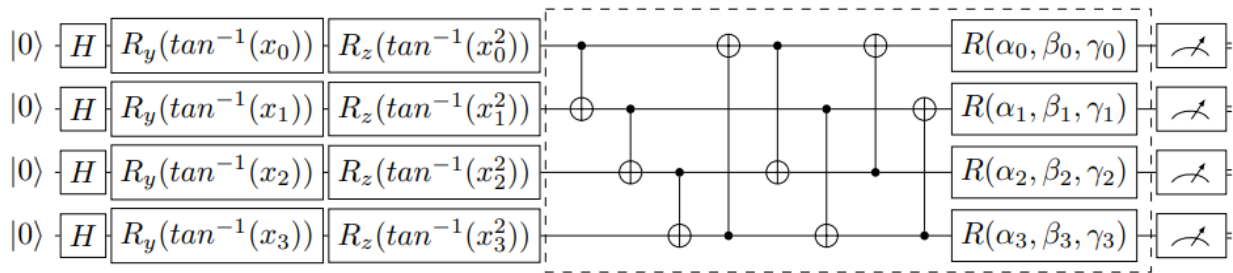
τρόπο με την κλασική κατασκευή της. Η κύρια διαφορά είναι ότι οι λειτουργίες MAC εντός του στοιχείου LSTM έχουν αντικατασταθεί από VQCs. Οι συγγραφείς αυτής της εργασίας εφάρμοσαν επίσης τα VQCs στην εξερχόμενη κρυφή κατάσταση και στην έξοδο του στοιχείου LSTM-ωστόσο, η έρευνα που έγινε εδώ έδειξε ότι αυτό δεν είναι απαραίτητο για να επιτύχει υψηλή απόδοση αυτό το QLSTM. Το στοιχείο QLSTM μπορεί στη συνέχεια να απεικονιστεί από το Σχήμα 2 και οι εξισώσεις του μοντέλου γίνονται ως εξής:

$$\begin{aligned} f_t &= \sigma(VQC_1(v_t; \theta_f)) \\ i_t &= \sigma(VQC_2(v_t; \theta_i)) \\ \bar{c}_t &= \tanh(VQC_3(v_t; \theta_{\bar{c}})) \\ c_t &= (f_t \cdot c_{t-1}) + (i_t \cdot \bar{c}_t) \\ o_t &= \sigma(VQC_4(v_t; \theta_o)) \\ h_t &= o_t \cdot \tanh(c_t) \end{aligned}$$



Σχήμα 2. Γραφική αναπαράσταση ενός στοιχείου QLSTM.

Ο VQC που χρησιμοποιήθηκε για την παρούσα κατασκευή έχει την αρχιτεκτονική που φαίνεται στο Σχήμα 3. Αυτή μπορεί να περιγραφεί ως υπάρχουσα σε τρία τμήματα: ένα στρώμα κωδικοποίησης δεδομένων (data encoding layer), ένα μεταβλητό στρώμα (variational layer) και ένα στρώμα μέτρησης (measurement layer). Το κύκλωμα ξεκινά με την τοποθέτηση των αρχικών qubits της βασικής κατάστασης σε μια αμερόληπτη κατάσταση μέσω μιας τράπεζας πυλών Hadamard. Στη συνέχεια, τα δεδομένα εισόδου κωδικοποιούνται ως γωνίες περιστροφής των qubits μέσω του μετασχηματισμού των τιμών τους με τη χρήση της συνάρτησης $\arctan$ και την εφαρμογή των πυλών R_y και R_z .



Σχήμα 3. Η αρχιτεκτονική QLSTM VQC που παρουσιάστηκε- η μεταβλητή ansatz στο διακεκομμένο πλαίσιο μπορεί να εφαρμοστεί όσες φορές χρειάζεται.

Στη συνέχεια, το μεταβλητό στρώμα, το οποίο υποδεικνύεται με το διακεκομμένο πλαίσιο στο σχήμα του κυκλώματος, παράγει multi-qubit διεμπλοκή μέσω δύο δακτυλίων πυλών CNOT. Στη συνέχεια, οι ρυθμιζόμενες παράμετροι α , β και γ κάθε qubit κωδικοποιούνται με τη χρήση περιστροφικών πυλών ενός qubit, δηλαδή $R(\alpha, \beta, \gamma)$. Αυτό το μεταβλητό στρώμα μπορεί να εφαρμοστεί πολλές φορές, καθορίζοντας μια υπερπαράμετρο βάθους στο κύκλωμα- για την παρούσα εργασία χρησιμοποιήθηκε μόνο ένα τέτοιο στρώμα. Το τελικό τμήμα, όπως πρέπει να κάνει κάθε κβαντικό κύκλωμα για να επιστρέψει αποτελέσματα, μετρά στη συνέχεια κάθε qubit. Αυτό γίνεται στην υπολογιστική βάση ως τιμές προσδοκίας Pauli-Z.

Τέλος αναλύεται το στοχαστικό LSTM, πιο συγκεκριμένα η στοχαστικότητα μέσω κλασικής κβαντοποίησης. Ο κβαντισμός βάρους είναι μια τεχνική που έχει χρησιμοποιηθεί κυρίως για την ανάπτυξη μοντέλων ML σε συσκευές χαμηλής ακρίβειας που έχουν περιορισμένους υπολογιστικούς πόρους και χωρητικότητα μνήμης, όπως τα ενσωματωμένα συστήματα και οι κινητές συσκευές. Τα δεδομένα που χρησιμοποιούνται από αυτά τα συστήματα περιορίζονται συνήθως σε μικρότερο αριθμό συστατικών bits από τα δεδομένα υψηλής ακρίβειας πραγματικών τιμών που επιτρέπει το τυπικό υπολογιστικό υλικό. Επιπλέον, η αριθμητική κινητής υποδιαστολής μπορεί να μην είναι καν δυνατή σε ορισμένες από αυτές τις συσκευές. Η κβάντιση λειτουργεί επομένως για τη λογική μείωση του εύρους των bit των δεδομένων, σε αντίθεση με την απλή αποκοπή των πλεοναζόντων bit. Αυτές οι τεχνικές μετατρέπουν γενικά τις αναπαραστάσεις κινητής υποδιαστολής είτε σε αριθμούς σταθερής υποδιαστολής είτε σε διακριτούς ακέραιους αριθμούς- για την παρούσα εργασία, θα εξετάσουμε μόνο την περίπτωση όπου οι τιμές που προκύπτουν είναι ακέραιοι αριθμοί. Οι εφαρμογές αυτής της μεθόδου μπορούν σε γενικές γραμμές να κατηγοριοποιηθούν σε περιπτώσεις όπου ένα προ-εκπαιδευμένο μοντέλο κβαντίζεται και σε περιπτώσεις όπου ένα μοντέλο κβαντίζεται κατά τη διάρκεια της εκπαίδευσής του.

Υπάρχουν μερικές μέθοδοι κβαντισμού που έχουν αποδειχθεί αποτελεσματικές, αν και οι αφελείς προσεγγίσεις και η χρήση μικρού αριθμού bit που προκύπτουν τείνουν να παράγουν φτωχά αποτελέσματα. Αυτό οφείλεται σε μεγάλο βαθμό στο ότι οι μικρές ενημερώσεις κατά τη βελτιστοποίηση στρογγυλοποιούνται, προκαλώντας στασιμότητα στην εκπαίδευση- αυτό είναι παρόμοιο με το ζήτημα της vanishing gradient στην εκπαίδευση μοντέλων πλήρους ακρίβειας και μπορεί να έχει ουσιαστικά τον ίδιο αντίκτυπο στη βελτιστοποίηση. Η μέθοδος που χρησιμοποιείται εδώ είναι η στοχαστική στρογγυλοποίηση κβάντισης, η οποία είναι τόσο μια κοινή επιλογή για την κβάντιση βάρους όσο και κατάλληλη για την παρούσα εργασία. Ενώ οι ντετερμινιστικές στρατηγικές είναι πιο επιρρεπείς στο προαναφερθέν ζήτημα βελτιστοποίησης, η στοχαστική στρογγυλοποίηση επηρεάζεται λιγότερο από αυτό, δεδομένης της πιθανοτικής φύσης της μεθόδου.

Η μαθηματική διατύπωση αυτής της μεθόδου παρουσιάζεται στην παρακάτω εξίσωση, όπου $Q_s(\bullet)$ είναι η στοχαστική συνάρτηση στρογγυλοποίησης, w είναι το βάρος που πρέπει να κβαντιστεί και p είναι ένας ομοιόμορφα παραγόμενος τυχαίος αριθμός. Αυτές οι συναρτήσεις

συνήθως ακολουθούνται από μια συνάρτηση αποκοπής (clipping) ή σύσφιξης (clamping) για να διασφαλιστεί ότι το αποτέλεσμα βρίσκεται εντός του εύρους (προσημασμένου ή μη προσημασμένου) που καθορίζεται από το εύρος bit (bitwidth).

$$Q_s(w) = \begin{cases} [w]+1, & \text{for } pw+[w] \\ [w], & \text{otherwise} \end{cases}$$

Συμπερασματικά, οι αρχιτεκτονικές RNN αναπτύχθηκαν για να αντιμετωπίσουν τις προκλήσεις της επεξεργασίας ακολουθιακών δεδομένων. Κάθε αρχιτεκτονική έχει τα πλεονεκτήματά της και τα μειονεκτήματά της, ανάλογα με τα χαρακτηριστικά του προβλήματος. Οι LSTM μονάδες, με τον προηγμένο τρόπο που χειρίζονται την πληροφορία και τις συνδέσεις μεταξύ των δειγμάτων, αναδεικνύονται ως μια εξαιρετική επιλογή για πολλές ακολουθιακές εργασίες, ανοίγοντας τον δρόμο για περαιτέρω εξελίξεις στον τομέα της επεξεργασίας ακολουθιακών δεδομένων.

4.4 Εφαρμογές των RNNs

Τα Επαναλαμβανόμενα Νευρωνικά Δίκτυα αναπαριστούν μια εξαιρετικά σημαντική και ενδιαφέρουσα κατηγορία των νευρωνικών δικτύων, ανοίγοντας την πόρτα προς ένα νέο επίπεδο ανάλυσης και κατανόησης σειριακών δεδομένων.[\[63\]](#) Ο κύριος και εξαιρετικά ενδιαφέρον παράγοντας των RNNs είναι η δυνατότητά τους να χειρίζονται και να επεξεργάζονται δεδομένα με χρονική δομή. Αυτή η ιδιότητα ανοίγει την πόρτα για την αντιμετώπιση ακολουθιών δεδομένων που εκφράζουν χρονικές εξαρτήσεις και πολύπλοκες δομές, δημιουργώντας ένα ευρύ φάσμα δυνατοτήτων.

Είναι σημαντικό να κατανοήσουμε πως η ικανότητα αυτή επιτρέπει στα RNNs να αντιμετωπίζουν ακολουθίες δεδομένων που παρουσιάζουν ποικίλες χρονικές εξαρτήσεις, όπως χρονοσειρές, κείμενα, αλλά και ακόμα πιο πολύπλοκα σενάρια όπως η ομιλία ή η μελωδία.[\[64\]](#) Αυτή η δυνατότητα ανοίγει νέους δρόμους σε μια ποικιλία εφαρμογών σε διάφορους τομείς, από την τεχνητή νοημοσύνη και τη βιολογία, μέχρι την οικονομία και την πρόβλεψη τάσεων.[\[65\]](#)

Σε αυτήν την ενότητα, θα εστιάσουμε στον τρόπο με τον οποίο τα RNNs εφαρμόζονται σε πρακτικά προβλήματα και συνεισφέρουν στην αντιμετώπιση προκλήσεων με έμφαση στη χρονική διάσταση. Θα εξεταστούν ενδεικτικά μερικές από τις βασικές και πιο ενδιαφέρουσες εφαρμογές των RNNs, αναδεικνύοντας την ικανότητα των RNNs να ανακαλύπτουν πρότυπα, να εντοπίζουν τις χρονικές δομές και να προσφέρουν λύσεις σε πολύπλοκότερα προβλήματα που συνδέονται με τον χρόνο.

4.4.1 Επεξεργασία Φυσικής Γλώσσας

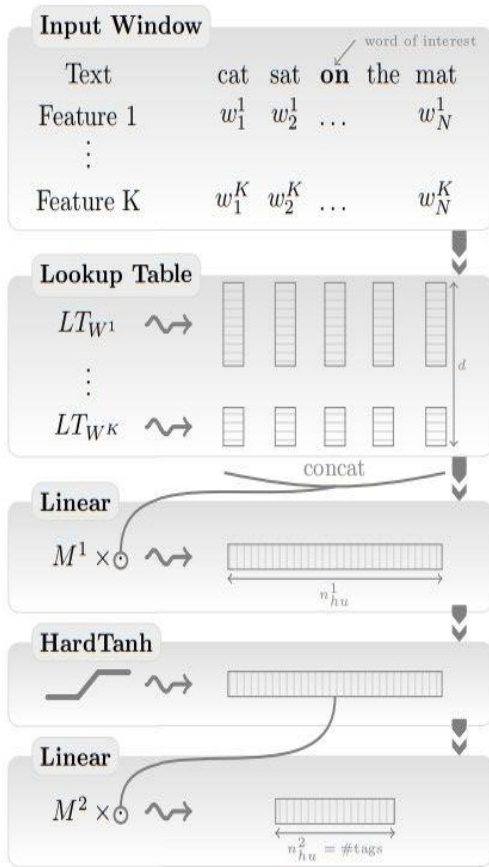
Ένας από τους πλέον προβεβλημένους τομείς εφαρμογής των RNNs είναι η Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP).[\[63\]](#) Αναπτύχθηκε σαν πεδίο επιστημονικής έρευνας και τεχνολογικής ανάπτυξης με σκοπό την κατανόηση και την

αλληλεπίδραση ανθρώπου-υπολογιστή μέσω της φυσικής γλώσσας. Οι RNNs διαθέτουν την ικανότητα να αντιλαμβάνονται και να αναλύουν την χρονική δομή των γλωσσικών ακολουθιών, προσφέροντας λύσεις σε διάφορα προβλήματα που αφορούν τη γλώσσα. Εδώ, τα RNNs έχουν αποδειχθεί εξαιρετικά χρήσιμα για την ανάλυση κειμένων και προτάσεων. Η ικανότητα τους να αναγνωρίζουν και να αποσαφηνίζουν τις δομές των φράσεων, των παραγράφων και των κειμένων, κάνει δυνατή την αυτόματη μετάφραση, την αναγνώριση ονοματικών οντοτήτων, την ανάλυση συναισθηματικού τόνου, αλλά και την αυτόματη παραγωγή κειμένου.^[73]

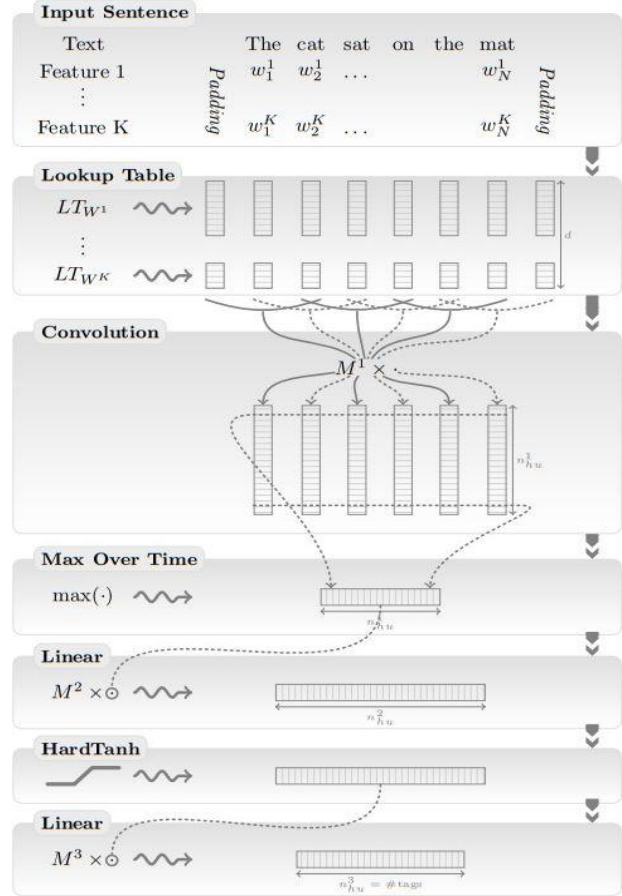
Σύμφωνα με το άρθρο “[Natural Language Processing \(Almost\) from Scratch](#)” των R Collobert, J Weston, L Bottou, κ.α. , όλες οι εργασίες NLP μπορούν να θεωρηθούν ως εργασίες που αποδίδουν ετικέτες σε λέξεις. Η παραδοσιακή προσέγγιση NLP είναι: εξάγετε από την πρόταση ένα πλούσιο σύνολο από χαρακτηριστικά σχεδιασμένα με το χέρι, τα οποία στη συνέχεια τροφοδοτούνται σε έναν τυπικό αλγόριθμο ταξινόμησης, για παράδειγμα, μια μηχανή διανυσμάτων υποστήριξης (Support Vector Machine - SVM), συχνά με γραμμικό πυρήνα. Η επιλογή των χαρακτηριστικών είναι μια εντελώς εμπειρική διαδικασία, βασισμένη κυρίως στη γλωσσική διαίσθηση και στη συνέχεια στη δοκιμή και το σφάλμα, και η επιλογή χαρακτηριστικών εξαρτάται από την εργασία, υποδηλώνοντας πρόσθετη έρευνα για κάθε νέα εργασία NLP. Οι σύνθετες εργασίες όπως το Semantic Role Labeling (SRL) απαιτούν στη συνέχεια έναν μεγάλο αριθμό πιθανώς πολύπλοκων χαρακτηριστικών (π.χ. εξάγονται από ένα δέντρο ανάλυσης) που μπορεί να επηρεάσουν το υπολογιστικό κόστος που μπορεί να είναι σημαντικό για εφαρμογές μεγάλης κλίμακας ή εφαρμογές που απαιτούν απόκριση σε πραγματικό χρόνο.

Αντίθετα, υποστηρίζεται μια ριζικά διαφορετική προσέγγιση: ως είσοδος γίνεται προσπάθεια να προεπεξεργαστούν τα χαρακτηριστικά όσο το δυνατόν λιγότερο και στη συνέχεια να χρησιμοποιηθεί μια αρχιτεκτονική πολυεπίπεδων νευρωνικών δικτύων (multilayer neural network - NN), εκπαιδευμένη με τρόπο end-to-end. Η αρχιτεκτονική παίρνει την πρόταση εισόδου και μαθαίνει πολλά επίπεδα εξαγωγής χαρακτηριστικών που επεξεργάζονται τις εισόδους. Τα χαρακτηριστικά που υπολογίζονται από τα βαθιά στρώματα του δικτύου εκπαιδεύονται αυτόματα με backpropagation ώστε να σχετίζονται με την εργασία. Σε αυτό το άρθρο περιγράφεται μια γενική αρχιτεκτονική πολλαπλών επιπέδων κατάλληλη για όλες τις εργασίες NLP, η οποία μπορεί να γενικευτεί.

Η παρουσιαζόμενη αρχιτεκτονική συνοψίζεται στο Σχήμα 1 και στο Σχήμα 2. Το πρώτο στρώμα εξάγει χαρακτηριστικά για κάθε λέξη. Το δεύτερο επίπεδο εξάγει χαρακτηριστικά από ένα παράθυρο λέξεων ή από ολόκληρη την πρόταση, αντιμετωπίζοντάς το ως μια ακολουθία με τοπική και σφαιρική δομή (δηλαδή, δεν αντιμετωπίζεται σαν ένα σάκο λέξεων). Τα ακόλουθα επίπεδα είναι τυπικά επίπεδα NN.



Σχήμα 1 (Αριστερά). Window approach network.



Σχήμα 2 (Δεξιά). Sentence approach network.

Θεωρούμε ένα νευρωνικό δίκτυο $f_{\theta}(\bullet)$, με παραμέτρους θ . Οποιοδήποτε νευρωνικό δίκτυο τροφοδοσίας με στρώματα L , μπορεί να θεωρηθεί ως σύνθεση συναρτήσεων $f_{\theta}^l(\bullet)$, που αντιστοιχεί σε κάθε στρώμα l :

$$f_{\theta}(\bullet) = f_{\theta}^L(f_{\theta}^{L-1}(\dots f_{\theta}^1(\bullet) \dots)).$$

Στη συνέχεια, περιγράφεται κάθε επίπεδο που χρησιμοποιείται στα προαναφερόμενα δίκτυα των σχημάτων 1 και 2. Με δεδομένο έναν πίνακα A , που συμβολίζουμε $[A]_{ij}$ τον συντελεστή στη σειρά i και τη στήλη j στον πίνακα. Δηλώνεται επίσης $\langle A \rangle_i^{d_{win}}$ το διάνυσμα που λαμβάνεται με τη σύνδεση των διανυσμάτων στήλης d_{win} γύρω από το διάνυσμα της i στήλης του πίνακα $\langle A \rangle \in R^{d_1 \times d_2}$:

$$\left[\langle A \rangle_i^{d_{win}} \right]^T = \left([A]_{1,i-d_{win}/2} \dots [A]_{d_1,i-d_{win}/2}, \dots, [A]_{1,i+d_{win}/2} \dots [A]_{d_1,i+d_{win}/2} \right)$$

Ως ειδική περίπτωση, το $\langle A \rangle_i^1$ αντιπροσωπεύει την i στήλη του πίνακα A . Για ένα διάνυσμα v , συμβολίζουμε $[v]_i$ τον βαθμωτό στον δείκτη i στο διάνυσμα. Τέλος, μια ακολουθία του στοιχείου $\{x_1, x_2, \dots, x_T\}$ γράφεται x_1^T . Το i ο στοιχείο της ακολουθίας είναι το $[x]_i$.

Ένα από τα βασικά σημεία της αρχιτεκτονικής αυτής είναι η ικανότητα του να αποδίδει καλά με τη χρήση (σχεδόν) ακατέργαστων λέξεων. Η ικανότητα της μελετούμενης μεθόδου να μαθαίνει καλές αναπαραστάσεις λέξεων είναι επομένως κρίσιμη για την προσέγγισή αυτή. Για λόγους αποτελεσματικότητας, οι λέξεις τροφοδοτούνται στην αρχιτεκτονική ως δείκτες που λαμβάνονται από ένα πεπερασμένο λεξικό D . Ωστόσο, το πρώτο επίπεδο του δικτύου αντιστοιχίζει κάθε έναν από αυτούς τους δείκτες λέξεων σε ένα διάνυσμα χαρακτηριστικών, με μια λειτουργία πίνακα αναζήτησης. Δεδομένης μιας εργασίας ενδιαφέροντος, μια σχετική αναπαράσταση κάθε λέξης δίνεται στη συνέχεια από το αντίστοιχο διάνυσμα χαρακτηριστικών του πίνακα αναζήτησης, το οποίο εκπαιδεύεται με backpropagation, ξεκινώντας από μια τυχαία αρχικοποίηση. Η αρχιτεκτονική μας επιτρέπει να εκμεταλλευόμαστε καλύτερα εκπαιδευμένες αναπαραστάσεις λέξεων, αρχικοποιώντας απλώς τον πίνακα αναζήτησης λέξεων με αυτές τις αναπαραστάσεις (αντί για τυχαία).

Πιο τυπικά, για κάθε λέξη $w \in D$, δίνεται μια εσωτερική αναπαράσταση διανυσμάτων χαρακτηριστικών d_{wrd} -διάστασης από το επίπεδο πίνακα αναζήτησης $LT_W(\bullet)$:

$$LT_W(w) = \langle W \rangle_w^1,$$

όπου $W \in R^{d_{wrd} \times |D|}$ είναι ένας πίνακας παραμέτρων προς εκμάθηση, $\langle W \rangle_w^1 \in R^{d_{wrd}}$ είναι η w η στήλη του W και d_{wrd} είναι το μέγεθος του διανύσματος λέξης (μια υπερπαραμέτρος που επιλέγει ο χρήστης). Δεδομένης μιας πρότασης ή οποιασδήποτε ακολουθίας T λέξεων $[w]_1^T$ σε D , το επίπεδο πίνακα αναζήτησης εφαρμόζει την ίδια λειτουργία για κάθε λέξη της ακολουθίας, παράγοντας τον ακόλουθο πίνακα εξόδου. Αυτός ο πίνακας μπορεί στη συνέχεια να τροφοδοτηθεί σε περαιτέρω στρώματα νευρωνικών δικτύων.

$$LT_W([w]_1^T) = \left(\langle W \rangle_{[w]_1}^1 \quad \langle W \rangle_{[w]_2}^1 \quad \dots \quad \langle W \rangle_{[w]_T}^1 \right)$$

4.4.1.1 Αυτόματη Μετάφραση

Ένα από τα εντυπωσιακά παραδείγματα εφαρμογής των RNNs στην NLP είναι η αυτόματη μετάφραση. Η πρόκληση της μετάφρασης από μία γλώσσα σε μία άλλη είναι σύνθετη, καθώς απαιτεί την κατανόηση της σημασίας των λέξεων και την αποτύπωσή τους σε νέες φράσεις με σεβασμό στη γραμματική και τη συντακτική δομή της γλώσσας προορισμού.[\[63\]](#) Τα RNNs, ειδικότερα οι Long Short-Term Memory (LSTM) και οι Gated Recurrent Units (GRUs), έχουν επιτρέψει τη δημιουργία μοντέλων που μπορούν να αντιμετωπίσουν αυτήν την πολυπλοκότητα. Τα μοντέλα αυτά μπορούν να "μάθουν" τις συντακτικές δομές και τα σημασιολογικά περιεχόμενα

των γλωσσικών ακολουθιών σε διάφορες γλώσσες, επιτρέποντας την αποτελεσματική μετάφραση.[64]

Η αυτόματη μετάφραση από μία γλώσσα σε μία άλλη αποτελεί ένα πολύπλοκο πρόβλημα της Επεξεργασίας Φυσικής Γλώσσας, που έχει απασχολήσει την επιστημονική κοινότητα και τη βιομηχανία για πολλές δεκαετίες.[66] Η πρόκληση αυτή εμπεριέχει την αντιστοίχιση με νόημα των λέξεων, φράσεων και προτάσεων από τη γλώσσα πηγής στη γλώσσα προορισμού, λαμβάνοντας υπόψη τη συντακτική και σημασιολογική δομή. Τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNNs) έχουν αποδειχθεί εξαιρετικά χρήσιμα για την αντιμετώπιση αυτής της πρόκλησης, προσφέροντας επιτυχημένες λύσεις στον τομέα της αυτόματης μετάφρασης.

Συνοψίζοντας, τα Επαναλαμβανόμενα Νευρωνικά Δίκτυα αποτελούν ισχυρό εργαλείο για την αυτόματη μετάφραση, επιτρέποντας τη δημιουργία προηγμένων μοντέλων που είναι ικανά να αντιμετωπίσουν την πολυπλοκότητα της γλωσσικής μετάφρασης. Η δυνατότητά τους να εξάγουν νόημα από τις γλωσσικές δομές και να διατηρούν τις συνδέσεις μεταξύ των λέξεων αποτελεί τη βάση για την επίλυση ενός από τα πιο προκλητικά προβλήματα της NLP, συμβάλλοντας στην επικοινωνία και την κατανόηση μεταξύ διαφορετικών γλωσσών και πολιτισμών.

Με βάση το άρθρο “[Natural Language Processing and Its Applications in Machine Translation: A Diachronic Review](#)” των Kai Jiang και Xi Lu γίνεται μία εκτενέστερη αναφορά στις εφαρμογές της Επεξεργασίας Φυσικής Γλώσσας στη Μηχανική Μετάφραση.

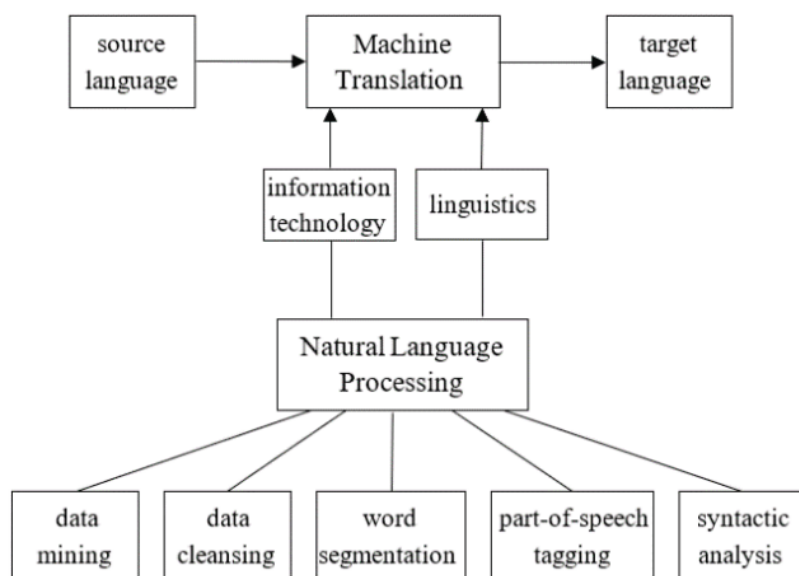
Τα τελευταία χρόνια, η τεχνολογία επεξεργασίας φυσικής γλώσσας έχει σημειώσει ταχεία πρόοδο και μια σειρά από συστήματα αυτόματης μετάφρασης έχουν επινοηθεί και χρησιμοποιηθεί ευρέως. Το 2006, η Google άρχισε να αναπτύσσει το σύστημα αυτόματης μετάφρασης και τελικά έφερε το Google Translate στην αγορά. Το 2011, Η Baidu κυκλοφόρησε το σύστημα αυτόματης μετάφρασης που μπορεί να υποστηρίξει 27 γλώσσες. Αυτά τα συστήματα αυτόματης μετάφρασης χρησιμοποιούνται εκτενώς στην καθημερινή μας εργασία και ζωή και η ποιότητα της μετάφρασης έχει βελτιωθεί σημαντικά.

Η μηχανική μετάφραση είναι ένα σύστημα που χρησιμοποιεί τον υπολογιστή για τη μετάφραση, δηλαδή ένα πρόγραμμα υπολογιστή που έχει σχεδιαστεί για να μεταφράζει κείμενο από μια γλώσσα (γλώσσα πηγής - source language) σε μια άλλη γλώσσα (γλώσσα στόχου - target language) χωρίς τη βοήθεια του ανθρώπου. Η μηχανική μετάφραση αποτελεί μια από τις σημαντικότερες εφαρμογές της επεξεργασίας της φυσικής γλώσσας - ένα πεδίο που συνδυάζει τα στοιχεία της τεχνολογίας της πληροφορίας με τη γλωσσολογία. Περιλαμβάνει πολλά ζητήματα κλασικής επεξεργασίας φυσικής γλώσσας κατά τη διάρκεια της λειτουργίας, όπως η εξόρυξη (data mining) και ο καθαρισμός (data cleansing) δεδομένων, η κατάτμηση λέξεων, η επισήμανση μέρους του λόγου και συντακτική ανάλυση. Επιπλέον, η μηχανική μετάφραση σχετίζεται επίσης στενά με την εφαρμογή αλγορίθμων μηχανικής μάθησης.

Η μηχανική μετάφραση χωρίζεται ευρέως σε μηχανική μετάφραση βασισμένη σε κανόνες (rule-based machine translation) και μηχανική μετάφραση που βασίζεται σε σώματα (corpus-based machine translation). Σύμφωνα με τις διαφορετικές μεθόδους μοντελοποίησης, η corpus-based μηχανική μετάφραση χωρίζεται περαιτέρω σε μηχανική μετάφραση βάσει παραδειγμάτων (example-based machine translation), στατιστική μηχανική μετάφραση (statistical machine translation) και νευρωνική μηχανική μετάφραση (neural machine translation). Στη μηχανική μετάφραση, το corpus αναφέρεται ως δεδομένα (data), πράγμα που σημαίνει ότι η μηχανική μετάφραση απαιτεί μεγάλη ποσότητα corpus για την εκπαίδευση του μοντέλου. Διαφορετικοί τύποι corpus χρησιμοποιούνται για την εκπαίδευση διαφορετικών μοντέλων. Για παράδειγμα, το σώμα της γλώσσας-στόχου χρησιμοποιείται για την εκπαίδευση μοντέλων δημιουργίας γλώσσας (για τη βελτίωση της ευχέρειας προτάσεων) και το παράλληλο σώμα

(parallel corpus) χρησιμοποιείται για την εκπαίδευση μεταφραστικών μοντέλων (για την απόκτηση μεταφραστικής γνώσης).

Όπως αναφέρεται παρακάτω, ο αλγόριθμος μηχανικής μάθησης χρησιμοποιείται τόσο στη μοντελοποίηση πιθανοτήτων όσο και στην εκπαίδευση μοντέλων. Το model building στη μηχανική μετάφραση χρειάζεται επίσης παραμέτρους για περαιτέρω βελτιστοποίηση των αλγορίθμων. Μπορεί να φανεί ότι η μηχανική μετάφραση είναι ένας σημαντικός ερευνητικός κλάδος της επεξεργασίας φυσικής γλώσσας. Μέσα από την ανάλυση, είναι προφανές ότι η τεχνολογία επεξεργασίας φυσικής γλώσσας παίζει καθοριστικό ρόλο στη μηχανική μετάφραση.



Σχήμα. Η εφαρμογή του NLP στην αυτόματη μετάφραση.

4.4.1.2 Αναγνώριση Ονοματικών Οντοτήτων

Η αναγνώριση ονοματικών οντοτήτων αποτελεί ένα σημαντικό κομμάτι της Επεξεργασίας Φυσικής Γλώσσας (NLP) και έχει εξαιρετική σημασία για πολλές εφαρμογές, όπως την αναζήτηση πληροφοριών, την ανάλυση κειμένου και την εξαγωγή γνώσης από κείμενα. Η διαδικασία της αναγνώρισης ονοματικών οντοτήτων αφορά τον εντοπισμό και την κατηγοριοποίηση συγκεκριμένων τμημάτων του κειμένου που αναφέρονται σε οντότητες όπως πρόσωπα, τόποι, οργανισμοί, ημερομηνίες και άλλα.

Τα Επαναλαμβανόμενα Νευρωνικά Δίκτυα αποτελούν ισχυρή εργαλειοθήκη για την αντιμετώπιση αυτού του προβλήματος. Η βασική ιδέα πίσω από τη χρήση των RNNs σε αυτήν την εφαρμογή είναι η ικανότητά τους να αντιλαμβάνονται την συνεχή συνεκτικότητα των λέξεων σε ένα κείμενο. Κατά τη διάρκεια της εκπαίδευσης, τα RNNs μαθαίνουν τα πολύπλοκα περιβάλλοντα και τα συμφραζόμενα όπου εμφανίζονται οι ονοματικές οντότητες.

Ένας τρόπος υλοποίησης της αναγνώρισης ονοματικών οντοτήτων με τη χρήση RNNs είναι μέσω της εφαρμογής δικτύων με επαναλαμβανόμενες μονάδες, όπως τα LSTM και τα GRU.[\[68\]](#)

Είναι σημαντικό να σημειωθεί ότι, παρά την αποτελεσματικότητα των RNNs στην αναγνώριση ονοματικών οντοτήτων, υπάρχουν και προκλήσεις. Η αντιμετώπιση πολυπλοκότητας γλωσσικών ακολουθιών, η αντιμετώπιση των ποικίλων μορφών ονοματικών οντοτήτων και η αποφυγή υπερεκπαίδευσης είναι μερικές από τις προκλήσεις που μπορεί να εμφανίζονται.

Συμπερασματικά, η αναγνώριση ονοματικών οντοτήτων αποτελεί μία ακόμη ένδειξη της ευρείας και ισχυρής εφαρμογής των Επαναλαμβανόμενων Νευρωνικών Δικτύων στον τομέα της Επεξεργασίας Φυσικής Γλώσσας. Η ικανότητά τους να λαμβάνουν υπόψη το περιβάλλον και τη συνεκτικότητα των λέξεων καθιστά τα RNNs ιδανικά για την επίλυση αυτού του σύνθετου προβλήματος στον τομέα της γλωσσικής ανάλυσης και κατανόησης.

Σύμφωνα με το paper “[A Survey on Named Entity Recognition](#)” των Y Wen, C Fan, G Chen, X Chen και M Chen, η Επεξεργασία Φυσικής Γλώσσας (NLP), γνωστή και ως Εξαγωγή Πληροφοριών (Information Extraction - IE), είναι μια σημαντική ερευνητική κατεύθυνση στον τομέα της επιστήμης των υπολογιστών και της τεχνητής νοημοσύνης. Μελετά κυρίως διάφορες θεωρίες και μεθόδους που μπορούν να πραγματοποιήσουν αποτελεσματική επικοινωνία μεταξύ ανθρώπου και υπολογιστή σε φυσική γλώσσα. Η NLP περιλαμβάνει πολλαπλά επιμέρους καθήκοντα, όπως η κατάτμηση λέξεων (word segmentation), η αναγνώριση ονοματικών οντοτήτων (named entity recognition), η περίληψη κειμένου (text summarization), η μηχανική μετάφραση (machine translation), η ανάλυση συναισθήματος (sentiment analysis), η αναγνώριση ομιλίας (speech recognition) και άλλα.

Η αναγνώριση ονοματικών οντοτήτων (NER) είναι ένα από τα βασικά καθήκοντα. Οι κατονομαζόμενες οντότητες προτάθηκαν αρχικά στο 6ο συνέδριο MUC (Message Understanding Conferences) το 1995, αναφερόμενες κυρίως σε λέξεις ή φράσεις με συγκεκριμένα ονόματα στο κείμενο. Γενικά αποτελείται από τρεις κύριες κατηγορίες (κλάσεις οντοτήτων (Entity classes), κλάσεις χρόνου (Time classes) και κλάσεις αριθμών (Number classes)) και επτά υποκλάσεις (όνομα προσώπου (Person name), όνομα τόπου (Place name), όνομα ιδρύματος (Institution name), ώρα (Time), ημερομηνία (Date), νόμισμα (Currency) και ποσοστό (Percentage)). Η NER έχει σχεδιαστεί για τον εντοπισμό αυτών των κύριων ουσιαστικών στο κείμενο και την ταξινόμησή τους στις κατάλληλες κατηγορίες. Αποτελεί τη βάση για πολλές προηγμένες εργασίες στην επεξεργασία φυσικής γλώσσας, όπως η συντακτική ανάλυση, η περίληψη κειμένου, η ανάκτηση πληροφοριών, η αυτόματη ερώτηση και απάντηση κ.ο.κ.

Οι μέθοδοι που βασίζονται σε στατιστικές μεθόδους δημιουργούν μοντέλα στατιστικής μάθησης, όπως τα μοντέλα μέγιστης εντροπίας (Maximum Entropy Models - MEM), το κρυφό μοντέλο Markov (Hidden Markov Model - HMM), το μοντέλο τυχαίου πεδίου υπό συνθήκη (Conditional Random Field Model - CRF), κ.λπ. στο χειροκίνητα επισημειωμένο σώμα κειμένων και σε έναν αριθμό επιλεγμένων χαρακτηριστικών, και τα μοντέλα αυτά χρησιμοποιούνται για την εξαγωγή νέων ονομαστικών οντοτήτων.

Το HMM είναι ένα στατιστικό μοντέλο που χρησιμοποιείται ευρέως και είναι κλασικό. Στο HMM, $\{x_1, x_2, \dots, x_{n+1}\}$ είναι η κρυφή κατάσταση της διαδικασίας μετάβασης Markov και $\{y_1, y_2, \dots, y_{n+1}\}$ είναι η έξοδος, δηλαδή η κατάσταση παρατήρησης. Το μοντέλο του HMM εκφράζεται ως εξής:

$$\lambda = (A, B, \pi)$$

όπου A είναι ο πίνακας πιθανοτήτων μετάβασης στην κατάσταση και B είναι ο πίνακας πιθανοτήτων παρατήρησης.

Λόγω της υψηλής απόδοσης του αλγορίθμου Viterbi, το HMM είναι σχετικά απλό και γρήγορο για εκπαίδευση. Ωστόσο, λόγω της παραδοχής ανεξαρτησίας της εξόδου, το μοντέλο έχει σχετικά χαμηλή απόδοση.

Προχωρώντας, το CRF είναι επίσης ένα κλασικό μοντέλο μάθησης με επίβλεψη για την επισήμανση ακολουθιών. Το γραμμικό αλυσιδωτό τυχαίο πεδίο υπό συνθήκη ήταν το κυρίαρχο

μοντέλο της NER πριν από την επικράτηση των μοντέλων που βασίζονται σε βαθιά νευρωνικά δίκτυα.

Έστω $X=(X_1, X_2, \dots, X_n)$, $Y=(Y_1, Y_2, \dots, Y_n)$, οι οποίες είναι όλες ακολουθίες τυχαίων μεταβλητών που αντιπροσωπεύονται από γραμμικές αλυσίδες. Δεδομένης μιας ακολουθίας X τυχαίων μεταβλητών, η υπό συνθήκη κατανομή πιθανότητας $P(Y|X)$ της τυχαίας μεταβλητής Y αποτελεί το CRF. Η έκφραση του μοντέλου έχει ως εξής:

$$P(Y|X) = \frac{1}{Z(x)} \exp \left(\sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_l s_l(y_i, x, i) \right)$$

$$Z(x) = \sum_Y \exp \left(\sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_l s_l(y_i, x, i) \right)$$

όπου t_k και s_l είναι χαρακτηριστικές συναρτήσεις. Τα λ_k και μ_l είναι αντίστοιχα βάρη. Το $Z(x)$ είναι ένας παράγοντας κανονικοποίησης.

Η απόδοση του αλγορίθμου CRF είναι καλύτερη από αυτή του HMM. Δεν κανονικοποιεί κάθε κόμβο, αλλά όλα τα χαρακτηριστικά κανονικοποιούνται συνολικά, επομένως μπορεί να ληφθεί η συνολική βέλτιστη τιμή. Αλλά η εκπαίδευση του μοντέλου CRF είναι πιο περίπλοκη, με αργή σύγκλιση και μεγάλο χρόνο εκπαίδευσης.

Αν και οι μέθοδοι που βασίζονται σε στατιστικά είναι αποτελεσματικές, χρειάζονται καλά επιλεγμένα χαρακτηριστικά και πρέπει να περάσουν από πιο σύνθετη εκπαίδευση. Επιπλέον, αυτές οι μέθοδοι βασίζονται σε μεγάλο βαθμό στο σώμα, αλλά το επισημασμένο σώμα (labelled corpus) που μπορεί να χρησιμοποιηθεί για την κατασκευή και την αξιολόγηση του ονομαζόμενου συστήματος αναγνώρισης οντοτήτων είναι σχετικά μικρό.

4.4.1.3 Ανάλυση Συναισθηματικού Τόνου

Τα RNNs έχουν επίσης χρησιμοποιηθεί για την ανάλυση του συναισθηματικού τόνου του κειμένου. Σε αυτήν την εφαρμογή, τα δίκτυα μπορούν να εκπαιδευτούν να αναγνωρίζουν τα συναισθηματικά χαρακτηριστικά των λέξεων και των προτάσεων, βοηθώντας στην αυτόματη κατανόηση της συναισθηματικής φόρτισης του κειμένου. [\[65\]\[75\]](#)

Η ανάλυση του συναισθηματικού τόνου του κειμένου αποτελεί μια σημαντική πτυχή της φυσικής γλώσσας επεξεργασίας (NLP) και έχει προκαλέσει μεγάλο ενδιαφέρον λόγω των δυνητικών εφαρμογών της σε διάφορους τομείς. [\[71\]](#) Η ύπαρξη της συναισθηματικής πλευράς στο κείμενο μπορεί να παρέχει σημαντική πληροφορία για την αντίληψη, την αντίδραση και τη συμπεριφορά του αναγνώστη.

Τα RNNs, και ειδικότερα οι παραλλαγές τους όπως τα Long Short-Term Memory (LSTM) και τα Gated Recurrent Units (GRUs), έχουν την ικανότητα να αντιλαμβάνονται τις χρονικές σχέσεις μεταξύ των λέξεων στο κείμενο. [\[68\]\[74\]](#) Αυτό είναι ιδιαίτερα σημαντικό στην ανάλυση του συναισθηματικού περιεχομένου, καθώς ο συναισθηματικός τόνος μπορεί να αλλάζει μέσα στο κείμενο και να επηρεάζεται από το πλαίσιο.

Η ανάλυση του συναισθηματικού τόνου με χρήση RNNs έχει εφαρμογές σε πολλούς τομείς. Στο χώρο των κοινωνικών μέσων, μπορεί να βοηθήσει στον αντίλογο και την ανάλυση της

συναισθηματικής ανταπόκρισης στις δημοσιεύσεις.[70] Επίσης, μπορεί να χρησιμοποιηθεί στη βιομηχανία για την αξιολόγηση της συναισθηματικής αντίδρασης των καταναλωτών σε προϊόντα και υπηρεσίες.[77] Συνολικά, η ανάλυση του συναισθηματικού τόνου με χρήση RNNs ανοίγει νέες δυνατότητες για την κατανόηση και αξιοποίηση της ανθρώπινης συναισθηματικής έκφρασης μέσω του γραπτού λόγου.[76]

Συμπερασματικά, τα Επαναλαμβανόμενα Νευρωνικά Δίκτυα αποτελούν κρίσιμο εργαλείο για την επίλυση ποικίλων προβλημάτων στον τομέα της φυσικής γλώσσας επεξεργασίας. Η ικανότητά τους να αντιμετωπίζουν τη χρονική δομή των γλωσσικών ακολουθιών έχει οδηγήσει σε αποτελεσματικές λύσεις για προβλήματα όπως η αυτόματη μετάφραση, η αναγνώριση ονοματικών οντοτήτων και η ανάλυση συναισθηματικού τόνου. Η συνεχής εξέλιξη των RNNs αναμένεται να συμβάλει σημαντικά στη βελτίωση των εφαρμογών τους στον τομέα της NLP.

4.4.2 Πρόβλεψη Χρονοσειρών

Οι χρονοσειρές αντιπροσωπεύουν μια κατηγορία δεδομένων που εμφανίζουν χρονικές εξαρτήσεις και τάσεις. Τα RNNs αποτελούν ισχυρά εργαλεία για την ανάλυση και την πρόβλεψη χρονοσειρών, καθώς μπορούν να αναγνωρίσουν πρότυπα και να αντιληφθούν τις δομές που αλλάζουν με τον χρόνο. Αυτό τα καθιστά αποτελεσματικά για προβλήματα όπως η πρόβλεψη των χρηματιστηριακών αγορών[68], η παρακολούθηση καιρικών φαινομένων[77], αλλά και η ανίχνευση ανωμαλιών σε χρονοσειριακά δεδομένα[71].

Η πρόβλεψη των χρηματιστηριακών αγορών είναι μία σημαντική εφαρμογή των RNNs. Η ικανότητά τους να αντιλαμβάνονται τις χρονικές συναρτήσεις και να αναγνωρίζουν προηγούμενες τάσεις μπορεί να συνδράμει στην πρόβλεψη των τιμών και των κινήσεων των χρηματιστηριακών αγορών.[68] Η ικανότητα να αντιλαμβάνονται τις ανεπιτήδευτες αλλαγές και τις περίπλοκες αλληλεπιδράσεις των χρηματιστηριακών δεδομένων, τα καθιστά χρήσιμα εργαλεία για τους επενδυτές και τους επαγγελματίες της οικονομίας.

Επίσης, τα RNNs είναι ιδιαίτερα κατάλληλα για την παρακολούθηση μετεωρολογικών μοτίβων και την πρόβλεψη καιρικών συνθηκών. Ο χρονικός χαρακτήρας των μετεωρολογικών δεδομένων απαιτεί τη χρήση μοντέλων που μπορούν να αντιληφθούν τις διακυμάνσεις και τις τάσεις στον καιρό.[77] Τα RNNs, με τη δυνατότητά τους να αναγνωρίζουν προηγούμενα μοτίβα και να προβλέπουν τις μελλοντικές καταστάσεις με βάση αυτά, μπορούν να βελτιώσουν την ακρίβεια των προβλέψεων και να συνδράμουν στην επιστημονική και εμπορική χρήση των μετεωρολογικών δεδομένων.

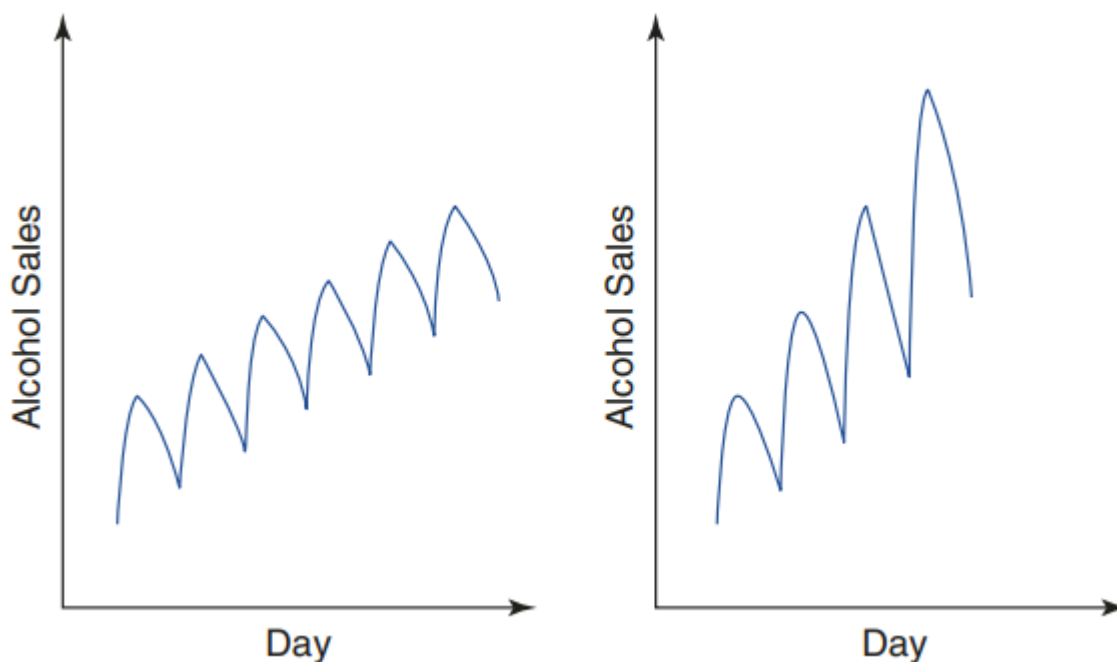
Πέραν τούτου, η ανίχνευση ανωμαλιών σε δεδομένα αποτελεί άλλη σημαντική εφαρμογή των RNNs. Η χρονική τους ευαισθησία, τους επιτρέπει να αναγνωρίζουν ανεπιθύμητες αλλαγές, έκτακτα γεγονότα και παρατηρήσεις που αποκλίνουν από τα συνηθισμένα μοτίβα.[67] Αυτό μπορεί να είναι ιδιαίτερα χρήσιμο για εφαρμογές όπως η ασφάλεια δικτύων[76], η ανίχνευση απάτης[71], και η παρακολούθηση βιομηχανικών διεργασιών[77].

Στο άρθρο “[Overview of Algorithms for Natural Language Processing and Time Series Analyses](#)” των James Feghali, Adrian E. Jimenez, Andrew T. Schilling και Tej D. Azad, αναφέρεται ότι η επεξεργασία φυσικής γλώσσας (NLP) και οι αναλύσεις χρονοσειρών (TSA) έχουν αναδειχθεί ως εξαιρετικά χρήσιμες υπολογιστικές τεχνικές στη σύγχρονη επιστήμη δεδομένων και τη μηχανική μάθηση, με ολοένα και πιο σχετικές εφαρμογές στην ιατρική.

Μπορούν να χρησιμοποιηθούν αρκετοί αλγόριθμοι μηχανικής μάθησης για την επίτευξη εργασιών NLP και την εκτέλεση αποτελεσματικών TSA. Στη συγκεκριμένη εργασία, παρέχεται μια λεπτομερή επισκόπηση των σημαντικών αλγορίθμων βαθιάς μάθησης, υποθέτοντας ότι δεν υπάρχουν προηγούμενες γνώσεις σε προηγμένη γραμμική άλγεβρα ή λογισμό. Περιγράφονται επίσης σχετικές τεχνικές στην προεπεξεργασία δεδομένων κατά την προετοιμασία για μοντελοποίηση. Εστιάζουμε σε εννοιολογικές βάσεις και βασικές υποκείμενες μαθηματικές πράξεις αντί για γλώσσα κωδικοποίησης για να διασφαλιστεί η προσβασιμότητα.

Με μια γενική κατανόηση του τρόπου με τον οποίο μια ποικιλία νευρωνικών δικτύων επεξεργάζονται δεδομένα προκειμένου να βελτιστοποιήσουν τις παραμέτρους του μοντέλου στο πλαίσιο του NLP, η εφαρμογή αυτών των ίδιων αλγορίθμων στο TSA μπορεί εύκολα να εκτιμηθεί. Τα νευρωνικά δίκτυα έχουν επιτύχει ευνοϊκά αποτελέσματα σε πολλές εργασίες TSA, ειδικά στην πρόβλεψη. Παρακάτω, παρουσιάζονται ορισμένες αρχιτεκτονικές νευρωνικών δικτύων που χρησιμοποιούνται στην πρόβλεψη.

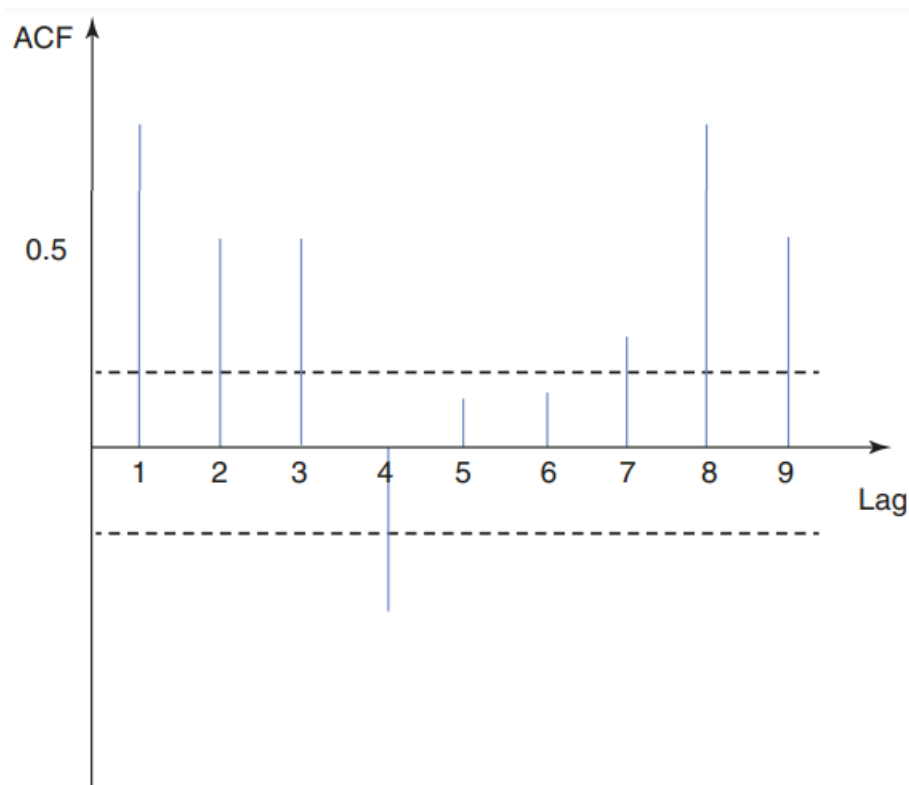
Πριν από την εφαρμογή αλγορίθμων, τα δεδομένα χρονοσειρών διερευνώνται συχνά για τον εντοπισμό γενικών τάσεων, οι οποίες είναι μια μακροπρόθεσμη αύξηση ή μείωση των τιμών σε μια σειρά (π.χ. μείωση στις πωλήσεις τσιγάρων με την πάροδο του χρόνου) και η εποχικότητα, η οποία είναι μια σταθερή επανάληψη βραχυπρόθεσμος κύκλος της σειράς (π.χ. σταθερή αύξηση στις πωλήσεις αλκοόλ κάθε Σαββατοκύριακο). Η εποχικότητα μπορεί να είναι αθροιστική (additive) (σταθερό μέγεθος διακύμανσης) ή πολλαπλασιαστική (multiplicative) (σταθερό ποσοστό διακύμανσης) όπως φαίνεται στην ακόλουθη εικόνα.



Χρονοσειρές με αυξητική τάση και εποχιακή διακύμανση. Προσθετική εποχικότητα (αριστερά) έναντι πολλαπλασιαστικής εποχικότητας (δεξιά)

Επιπλέον, η σειρά μπορεί να αναλυθεί για την παρουσία αυτοσυσχέτισης (autocorrelation), η οποία είναι η σημαντική συσχέτιση μεταξύ των τιμών τη φορά (t) και των προηγούμενων τιμών με μια ορισμένη υστέρηση ($t - 1$, $t - 2$, $t - 3$, κ.λπ.). Για τη μέτρηση της αυτοσυσχέτισης, μπορούν να δημιουργηθούν μεταβλητές καθυστέρησης της μέτρησης ενδιαφέροντος (measurement of interest) και η συσχέτιση μεταξύ της αρχικής μεταβλητής (original variable) και της μεταβλητής υστέρησης (lag variable) μπορεί να αξιολογηθεί με ένα τεστ συσχέτισης Pearson

για παράδειγμα. Κάποιος μπορεί επίσης να παράγει μια γραφική παράσταση αυτοσυσχέτισης (συντελεστής συσχέτισης έναντι τιμής καθυστέρησης), γνωστή και ως συνάρτηση αυτοσυσχέτισης, η οποία μπορεί επίσης να βοηθήσει στον εντοπισμό της εποχικότητας (Εικ. 26.27).



Γράφημα
αυτοσυσχέτισης.

Οι ράβδοι που
ξεπερνούν το όριο
αντιπροσωπεύουν
στατιστικά σημαντικές
αυτοσυσχετίσεις.
Παρατηρείται αρνητική
αυτοσυσχέτιση για την
τιμή υστέρησης 4 (t
έναντι $t - 4$) και η
εποχικότητα είναι
εμφανής κάθε επτά
χρονικά βήματα (π.χ. 1
εβδομάδα εάν οι
μετρήσεις είναι
καθημερινές)

Οι αυτοσυσχετίσεις που συναντώνται συνήθως είναι μια θετική και μια αρνητική συσχέτιση μεταξύ των τιμών τη στιγμή (t) και των τιμών από το προηγούμενο χρονικό βήμα ($t - 1$), γνωστές ως «stickiness» και «swings», αντίστοιχα. Σε αυτή τη φάση, είναι επίσης απαραίτητο να εντοπιστούν ασυνήθιστες περίοδοι ή ακραίες τιμές για να αποφευχθεί η εκπαίδευση του δικτύου σε αυτά τα μη συστηματικά στοιχεία της χρονοσειράς. Στο TSA, είναι επιτακτική ανάγκη να σημειωθεί ότι ο διαχωρισμός των δεδομένων σε εκπαίδευση και επικύρωση δεν μπορεί να γίνει τυχαία επειδή οι τιμές δεν είναι ανεξάρτητες. Απαιτούνται διαδοχικές παρατηρήσεις για την εκπαίδευση και τη δοκιμή ενός μοντέλου, επομένως τυπικά, το πρώτο $n\%$ των διαδοχικών παρατηρήσεων επιλέγεται για εκπαίδευση, αφήνοντας το επόμενο $100-n\%$ για επικύρωση.

Μια σειρά από αλγόριθμους μηχανικής μάθησης έχουν χρησιμοποιηθεί για την εκτέλεση πολλών διαφορετικών εργασιών στο NLP και το TSA. Πριν από την εφαρμογή αυτών των αλγορίθμων, απαιτείται κάποιος βαθμός προεπεξεργασίας δεδομένων. Οι προσεγγίσεις βαθιάς μάθησης που χρησιμοποιούν πολυστρωματικά perceptrons, RNN και CNN αντιπροσωπεύουν τεχνικές που χρησιμοποιούνται συνήθως. Σε εφαρμογές εποπτευόμενης εκμάθησης, όλα αυτά τα μοντέλα αντιστοιχίζουν τις εισόδους σε μια προβλεπόμενη έξοδο και στη συνέχεια μοντελοποιούν την απόκλιση μεταξύ των προβλεπόμενων τιμών και της πραγματικής παραγωγής σύμφωνα με μια συνάρτηση απώλειας. Στη συνέχεια, οι παράμετροι της συνάρτησης χαρτογράφησης βελτιστοποιούνται μέσω της διαδικασίας gradient descent και backpropagation προκειμένου να ελαχιστοποιηθεί αυτή η απώλεια. Αυτή είναι η κύρια προϋπόθεση πίσω από πολλούς αλγόριθμους εποπτευόμενης μάθησης. Καθώς αυξάνεται η εμπειρία με αυτούς τους αλγόριθμους, αναμένονται αυξημένες εφαρμογές στους τομείς της ιατρικής και της νευροεπιστήμης.

Συνολικά, οι εφαρμογές των RNNs στην ανάλυση και πρόβλεψη χρονοσειρών επιφέρουν εξαιρετικά θετικά αποτελέσματα σε πολλούς τομείς, επιτρέποντας την αξιοποίηση της χρονικής πληροφορίας για τη λήψη ακριβών και προβλεπτικών αποφάσεων.

4.4.3 Επεξεργασία Ήχου και Μουσικής

Η επεξεργασία ήχου και μουσικής αποτελεί έναν άλλον τομέα όπου τα RNNs έχουν εντυπωσιακές εφαρμογές. Αυτά τα δίκτυα μπορούν να αναγνωρίσουν ακουστικά πρότυπα, να δημιουργήσουν μουσική με βάση μια σειρά νοτών ή ακόμη και να βοηθήσουν στην ανάλυση της ακουστικής πληροφορίας.[\[63\]\[64\]](#)

Οι εφαρμογές των RNNs στην επεξεργασία ήχου και μουσικής ανοίγουν νέες προοπτικές για την δημιουργία καινοτόμων μουσικών εμπειριών, την ανάλυση των χαρακτηριστικών του ήχου και την ανακάλυψη νέων μουσικών τάσεων μέσω της τεχνολογίας.[\[65\]](#) Ακολούθως παρατίθενται επιγραμματικά κάποιες από αυτές τις εφαρμογές του συγκεκριμένου τομέα.

- Αναγνώριση Προσφωνήματος

Η αναγνώριση φωνής είναι μια από τις πρωταρχικές εφαρμογές των RNNs στην επεξεργασία ήχου. Τα RNNs μπορούν να εκπαιδευτούν να αναγνωρίζουν λέξεις, φράσεις και συναισθηματικούς τόνους στη φωνητική είσοδο, επιτρέποντας τη δημιουργία αποτελεσματικών συστημάτων αυτόματης αναγνώρισης φωνής.

- Παραγωγή Μουσικής με Βάση Μοτίβα

Ένα από τα σημαντικότερα πεδία εφαρμογής των RNNs στη μουσική είναι η παραγωγή μουσικής. Με τη χρήση RNNs, μπορούμε να εκπαιδευσουμε μοντέλα να αναγνωρίζουν μουσικά μοτίβα και δομές σε μουσικά κομμάτια και στη συνέχεια να δημιουργούν νέες μουσικές συνθέσεις βασισμένες σε αυτά τα μοτίβα.

- Δημιουργία Πρωτότυπων Μελωδιών

Με την χρήση RNNs, είναι δυνατό να δημιουργηθούν πρωτότυπα μουσικά κομμάτια και μελωδίες. Τα μοντέλα μπορούν να εκπαιδευτούν σε μεγάλο όγκο δεδομένων μουσικής και να παράγουν νέα μουσικά φράγματα που προκύπτουν από τις μοτίβα και τις δομές που έχουν μάθει.

- Επεξεργασία Ήχου και Φωνητική Αναγνώριση

Τα RNNs χρησιμοποιούνται για την ανάλυση των πολύπλοκων δομών του ήχου. Ενδεικτικά, μπορούν να αναγνωρίσουν τις διάφορες συχνότητες και τον τρόπο με τον οποίο ο ήχος μεταβάλλεται στο χρόνο. Αυτό βοηθά στην αναγνώριση και αντιμετώπιση προβλημάτων όπως η αναγνώριση φωνητικών εντολών, η κατανόηση της προφορικής γλώσσας και η απομόνωση φωνητικών μερών από άλλα ηχητικά στοιχεία.

- Αναγνώριση Μουσικών Ειδών και Χαρακτηριστικών

Η επεξεργασία ήχου μπορεί να εφαρμοστεί και στη μουσική. Τα RNNs μπορούν να αναγνωρίσουν μουσικά είδη, συνθετικά στοιχεία, ρυθμούς και άλλα χαρακτηριστικά από μουσικά κομμάτια. Αυτή η αναγνώριση μπορεί να χρησιμοποιηθεί για την κατασκευή συστημάτων αυτόματης ταξινόμησης μουσικής, τη δημιουργία playlists με βάση τις προτιμήσεις του χρήστη και την ανάλυση μουσικών τάσεων.

- Παραγωγή Συνθετικών Ήχων και Μουσικής

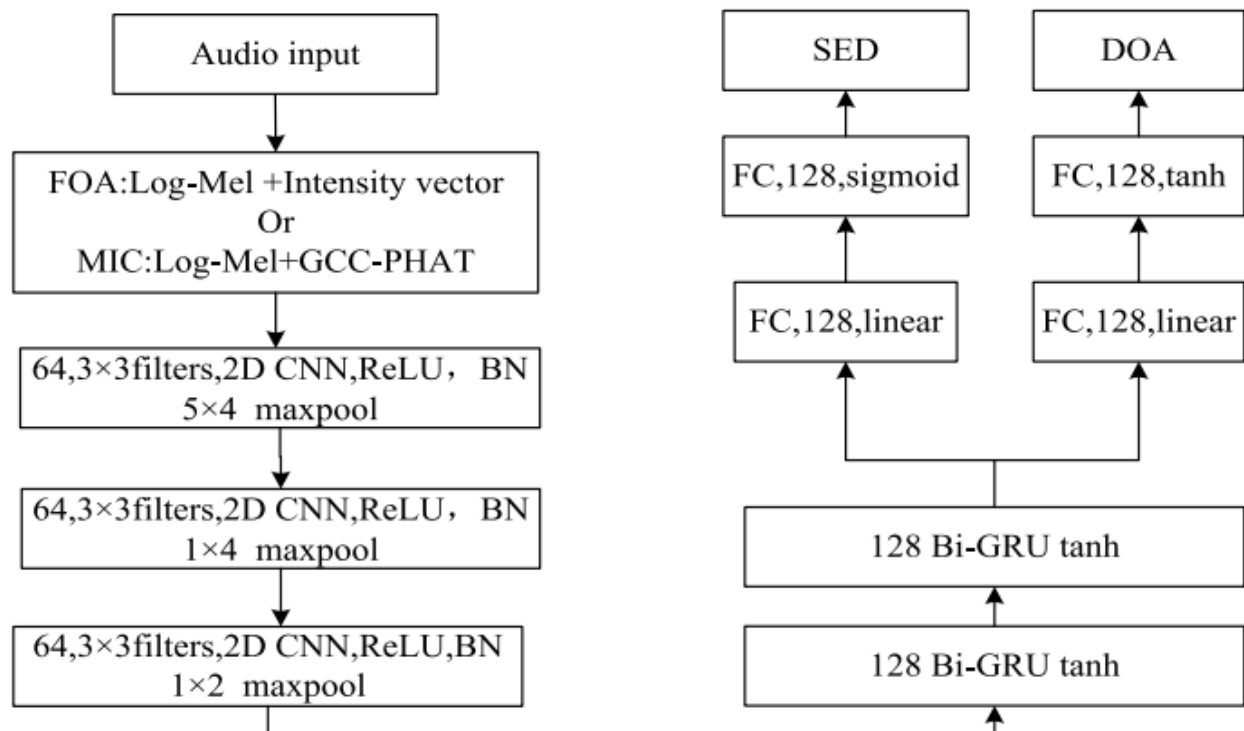
Ένα ακόμη πολύ ενδιαφέρον πεδίο, είναι η παραγωγή συνθετικών ήχων και μουσικής. Χρησιμοποιώντας RNNs, μπορούμε να δημιουργήσουμε νέους, πρωτότυπους ήχους και μελωδίες. Τα μοντέλα μπορούν να εκπαιδευτούν σε δείγματα ήχων και μουσικής και να δημιουργήσουν νέους ήχους που παρουσιάζουν τα χαρακτηριστικά του εκπαιδευτικού υλικού.

Η εφαρμογή των RNNs στην επεξεργασία ήχου και μουσικής δείχνει τον τρόπο με τον οποίο η τεχνολογία μπορεί να αλλάξει τον τρόπο που δημιουργούμε, αντιλαμβανόμαστε και απολαμβάνουμε τον ήχο και τη μουσική. Αυτές οι εφαρμογές ανοίγουν νέες προοπτικές για τη δημιουργία νέων ήχων, την ανάλυση μουσικής και την εξερεύνηση δημιουργικών δυνατοτήτων σε αυτούς τους τομείς.^[77]

Πιο συγκεκριμένα ένα ενδεικτικό παράδειγμα αναφέρεται στο άρθρο “[Attention mechanism combined with residual recurrent neural network for sound event detection and localization](#)” των C Lan, L Zhang, Y Zhang, L Fu, κ.α..

Ξεκινώντας το Convolutional neural network (CNN) αποδίδει καλά στην εφαρμογή ταξινόμησης και τμηματοποίησης εικόνων. Στην περίπτωση των δυνατοτήτων ήχου, είναι επίσης μια δισδιάστατη εικόνα που τροφοδοτείται στον υπολογιστή για επεξεργασία. Επομένως, διαφορετικές εικόνες χαρακτηριστικών γεγονότων ήχου επεξεργάζονται με τη μέθοδο επεξεργασίας εικόνας και οι επεξεργασμένες λειτουργίες εικόνας αποστέλλονται στο CNN για εκπαίδευση για την έξοδο των ετικετών κατηγορίας των συμβάντων. Εκτός από το CNN, το RNN έχει ισχυρές επιδόσεις στην επεξεργασία δεδομένων ακολουθίας. Η Gated Recurrent Unit (GRU) είναι μια παραλλαγή δικτύου στο RNN που μπορεί να αποτυπώσει αποτελεσματικά τα συμφοραζόμενα χαρακτηριστικά των χρονοσειρών και να αποδώσει καλύτερα σε μεγαλύτερες σειρές. Η δομή της Bidirectional Gating Recurrent Unit (Bi-GRU) κάνει όλο το GRU πιο ισχυρό, κάνοντας την τρέχουσα κρυφή κατάσταση όχι μόνο να σχετίζεται με την κρυφή κατάσταση της προηγούμενης στιγμής αλλά και με την κρυφή κατάσταση της μελλοντικής στιγμής.

Εξαγωγή πληροφοριών γενικού πλαισίου (global context information) και εξαγωγή μακροπρόθεσμων και βραχυπρόθεσμων χαρακτηριστικών εξάρτησης (long-term and short-term dependency characteristics). Στο μοντέλο SEDL, τα χαρακτηριστικά εισόδου είναι φασματικά χαρακτηριστικά που εξαρτώνται από χρονοσειρές. Χρησιμοποιώντας το δίκτυο που βασίζεται στη χρονική αλληλεξάρτηση, μπορούν να εκπαιδευτούν οι χαρακτηριστικές πληροφορίες κάθε στιγμής και η αλληλουχία της. Σε σύγκριση με το απλό δίκτυο CNN, το δίκτυο σε συνδυασμό με το CNN και το RNN έχει ισχυρότερη ικανότητα επεξεργασίας πληροφοριών χρόνου. Το σχηματικό διάγραμμα της δομής του δικτύου CRNN φαίνεται στο παρακάτω σχήμα.



Σχήμα 1. Structure diagram of CRNN

Η δομή του δικτύου CRNN στο Σχήμα 1 περιλαμβάνει τέσσερα κύρια μέρη: Το πρώτο μέρος είναι η εξαγωγή χαρακτηριστικών (feature extraction), η οποία λαμβάνει κυρίως δύο ομάδες διαφορετικών χαρακτηριστικών ταξινόμησης από δύο διαφορετικά σύνολα δεδομένων, συγκεκριμένα, το φάσμα log-mel και το διανυσματικό φάσμα έντασης ήχου (sound intensity vector) από το σύνολο δεδομένων FOA και το log-mel και γενικευμένο φάσμα διασυσχέτισης (generalized cross-correlation) από το σύνολο δεδομένων MIC.

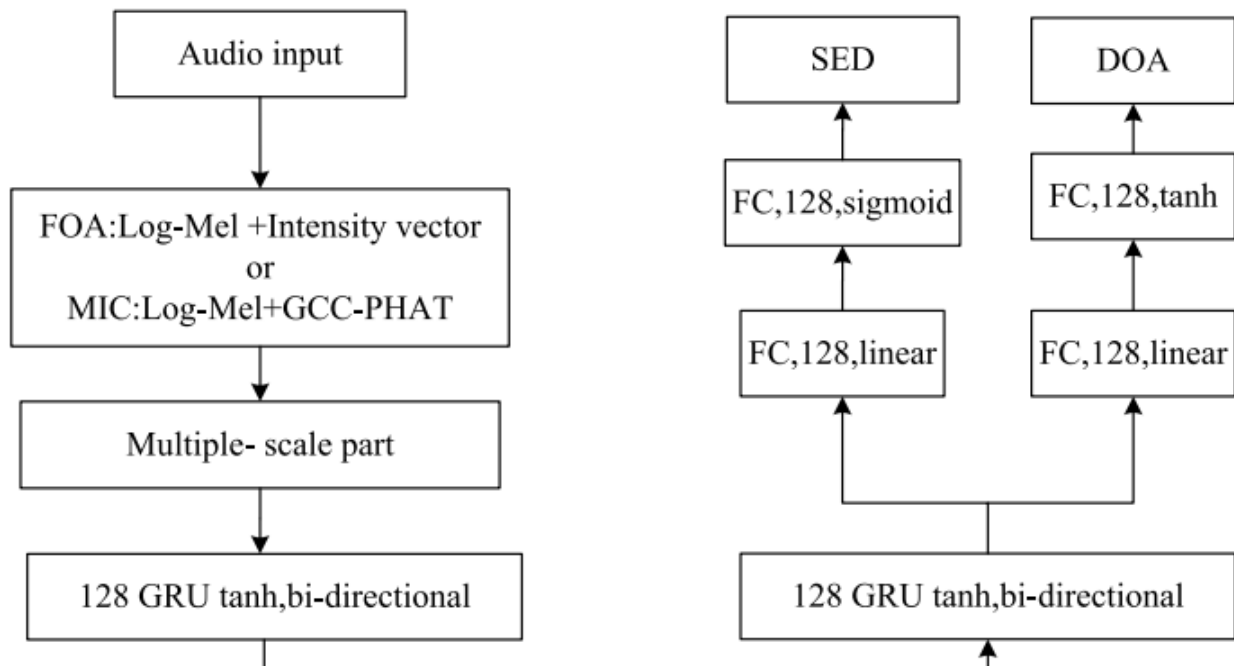
Το δεύτερο μέρος είναι ένα δισδιάστατο στρώμα συνέλιξης 3 επιπέδων και ο πυρήνας συνέλιξης κάθε επιπέδου συνέλιξης είναι 3×3 . Το συνελικτικό επίπεδο μαθαίνει τα αμετάβλητα χαρακτηριστικά μετατόπισης από τα χαρακτηριστικά εισόδου και μετά τη μη γραμμική λειτουργία της συνάρτησης ενεργοποίησης ReLU, τα χαρακτηριστικά επεξεργάζονται με κανονικοποίηση παρτίδας και μέγιστη συγκέντρωση.

Το τρίτο μέρος είναι το στρώμα GRU δύο επιπέδων, το οποίο μαθαίνει τη δομή χρόνου-συχνότητας από το χαρακτηριστικό του οποίου η έξοδος είναι 128 στο ανώτερο επίπεδο και λαμβάνει βαθιά είσοδο πληροφοριών στο επόμενο επίπεδο δικτύου μέσω της μη γραμμικής λειτουργίας της συνάρτησης ενεργοποίησης Tanh.

Το τέταρτο μέρος είναι η εκμάθηση της εκτίμησης SED και DOA μέσω δύο ανεξάρτητων διακλαδώσεων δύο πλήρως συνδεδεμένων επιπέδων. Το επίπεδο εξόδου SED έχει μια γραμμική συνάρτηση ενεργοποίησης (linear activation function) και μια συνάρτηση σιγμοειδούς ενεργοποίησης (sigmoid activation function) και η έξοδος αντιστοιχεί στη χρονική δραστηριότητα των ειδών C ηχητικών συμβάντων υπό χρονική συχνότητα. Ομοίως, το επίπεδο εξόδου DOA έχει επίσης μια συνάρτηση γραμμικής ενεργοποίησης και λειτουργία ενεργοποίησης Tanh, η οποία εξάγει τροχιές DOA που αντιστοιχούν σε είδη C ακουστικών γεγονότων ταυτόχρονα.

Δεδομένου ότι οι πυρήνες συνέλιξης σε διαφορετικές κλίμακες μπορούν να μάθουν deep features σε διαφορετικές κλίμακες και λόγω της ανεπαρκούς εξαγωγής χαρακτηριστικών της

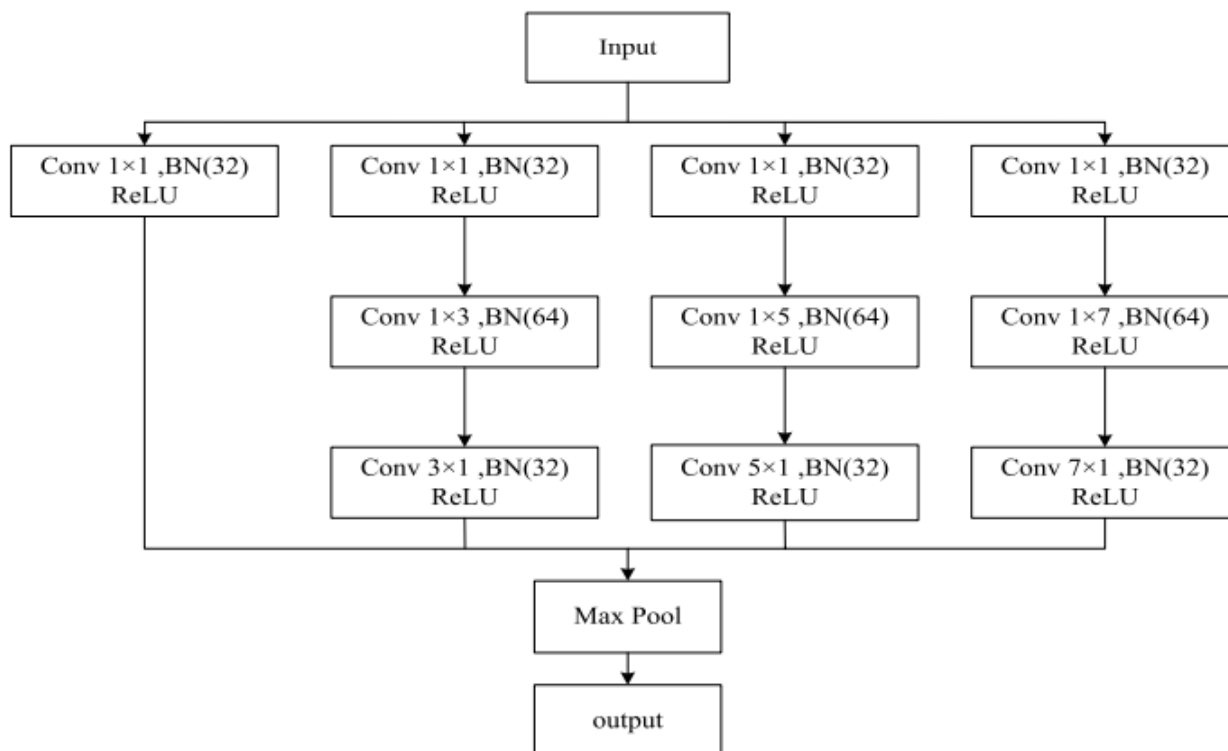
συνέλιξης μιας κλίμακας στο CRNN, αυτή η εργασία αντικατέστησε το στρώμα συνέλιξης του CRNN με μια μονάδα συνέλιξης πολλαπλής κλίμακας και πρότεινε το M- Μοντέλο CRNN, η δομή του οποίου φαίνεται στο Σχήμα 2.



Σχήμα 2. M-CRNN model

Στο Σχήμα 2, τέσσερις ομάδες πυρήνων παράλληλης συνέλιξης διαφορετικών κλιμάκων χρησιμοποιούνται για την εξαγωγή χαρακτηριστικών από τον τομέα χρόνου (time domain) και συχνότητας (frequency domain) αντίστοιχα, και τα συγχωνευμένα χαρακτηριστικά εισάγονται στο δίκτυο GRU. Τα SED και DOA εξάγονται μετά τη συνάρτηση ενεργοποίησης σιγμοειδούς (sigmoid activation function) και τη συνάρτηση ενεργοποίησης tanh.

Λόγω των περιορισμών σχεδίασης μονής κλίμακας (single-scale design), η ικανότητα εξαγωγής χαρακτηριστικών του μοντέλου σε κάποιο βαθμό, ακόμη και αν ο ίδιος τύπος συμβάντος μπορεί να έχει διαφορετική διάρκεια και τα στοιχεία του χρόνου και της συχνότητας του ήχου κατανέμονται διαφορετικά. Προκειμένου να ικανοποιηθούν οι απαιτήσεις συμπλήρωσης και ενίσχυσης χαρακτηριστικών του ίδιου ακουστικού συμβάντος σε διαφορετικές κλίμακες, η δομή και η παράμετρος δικτύου Ρυθμίσεις του μοντέλου συνέλιξης πολλαπλής κλίμακας που προτείνεται στο Σχήμα 2 φαίνονται στο Σχήμα 3.



Σχήμα 3. Multi-scale fusion module

Στο Σχήμα 3, η μονάδα σύντηξης πολλαπλής κλίμακας υιοθετεί τέσσερα σετ διακλαδώσεων παράλληλης συνέλιξης για να εξάγει χαρακτηριστικά από γραφήματα χρόνου-συχνότητας. Η συνέλιξη σε διαφορετικές κλίμακες είναι συμπληρωματική μεταξύ τους για τη βελτίωση της μαθησιακής ικανότητας του μοντέλου για διαφορετικά χαρακτηριστικά. Η συνέλιξη $n \times n$ αντικαθίσταται από $1 \times n$, η συνέλιξη και η συνέλιξη $n \times 1$, και το n αντιπροσωπεύει τον αριθμό των πυρήνων συνέλιξης, εξοικονομώντας το κόστος του χρόνου υπολογισμού.

Η πρώτη ομάδα παράλληλων διακλαδώσεων CNN έχει μόνο ένα επίπεδο συνέλιξης και το μέγεθος του πυρήνα συνέλιξης είναι 1×1 . Μετά τη συνέλιξη, τα χαρακτηριστικά αποστέλλονται στο επίπεδο BN για ομαδική εκπαίδευση (batch training). Η παράμετρος BN ορίζεται στο 32 και, στη συνέχεια, εξάγεται στο Max Pool για μείωση διαστάσεων μέσω της μη γραμμικής λειτουργίας της λειτουργίας ενεργοποίησης ReLU.

Η δεύτερη ομάδα παράλληλων διακλαδώσεων CNN έχει τρία επίπεδα συνέλιξης και το μέγεθος του πυρήνα συνέλιξης είναι 1×1 , 1×3 και 3×1 , αντίστοιχα. Το BN πραγματοποιείται μετά από κάθε λειτουργία συνέλιξης και οι ρυθμίσεις παραμέτρων είναι 32, 64 και 32, αντίστοιχα. Η λειτουργία ReLU χρησιμοποιείται για τη λειτουργία ενεργοποίησης.

Η τρίτη ομάδα παράλληλων διακλαδώσεων CNN έχει επίσης τρία επίπεδα συνέλιξης και το μέγεθος του πυρήνα συνέλιξης είναι 1×1 , 1×5 και 5×1 , αντίστοιχα. Τα χαρακτηριστικά εξακολουθούν να αποστέλλονται στο επίπεδο BN μετά από κάθε λειτουργία συνέλιξης και η ρύθμιση παραμέτρων και η λειτουργία ενεργοποίησης είναι ίδιες με αυτές της δεύτερης ομάδας.

Η τέταρτη ομάδα παράλληλων διακλαδώσεων CNN έχει επίσης τρία επίπεδα συνέλιξης και το μέγεθος του πυρήνα συνέλιξης είναι 1×1 , 1×7 και 7×1 , αντίστοιχα. Το BN εκτελείται μετά από κάθε λειτουργία συνέλιξης και η ρύθμιση παραμέτρων και η λειτουργία ενεργοποίησης είναι ίδια με την προηγούμενη ομάδα. Τα χαρακτηριστικά του τομέα συχνότητας λαμβάνονται με συνέλιξη τομέα συχνότητας και τα χαρακτηριστικά τομέα χρόνου με συνέλιξη τομέα χρόνου. Το 3×1 , το

$1 \times n(n=3,5,7)$ χρησιμοποιείται για τη λήψη χαρακτηριστικών του τομέα συχνότητας και το $n(n=3,5,7) \times 1$ για τη λήψη χαρακτηριστικών τομέα χρόνου.

Με στόχο την ανεπαρκή εξαγωγή χαρακτηριστικών του μοντέλου CRNN σε μια ενιαία κλίμακα, καθιερώθηκε μια μέθοδος ανίχνευσης και εντοπισμού ηχητικών συμβάντων βασισμένη στο M-CRNN με βάση τη δομή M-CNN της σύντηξης συνέλιξης πολλαπλής κλίμακας σε συνδυασμό με RNN. Σε σύγκριση με άλλα μοντέλα SEDL, το μοντέλο έδειξε σημαντική βελτίωση στην επίδραση ανίχνευσης και εντοπισμού, ειδικά στην απόδοση εντοπισμού.

Συμπερασματικά, τα Επαναλαμβανόμενα Νευρωνικά Δίκτυα αντιπροσωπεύουν μία ισχυρή τεχνολογία με αναμφίβολες εφαρμογές σε ποικίλους τομείς. Η δυνατότητά τους να λαμβάνουν υπόψη τη χρονική δομή των δεδομένων τους καθιστά ανεκτίμητα στην αντιμετώπιση προβλημάτων που εμπλέκουν την χρονική πτυχή. Από την Επεξεργασία Φυσικής Γλώσσας μέχρι την πρόβλεψη χρονοσειρών και την μηχανική μετάφραση, τα RNNs συνεισφέρουν στην επίλυση ποικίλων προκλήσεων. Με τη συνεχή ανάπτυξη της τεχνολογίας, αναμένεται ότι θα ανακαλύπτονται ολοένα και πιο νέες και πιο καινοτόμες εφαρμογές τους που θα ενισχύουν την δυνατότητα αντιμετώπισης προβλημάτων με χρονική διάσταση.

5

Reservoir Computing

5.1 Εισαγωγή

Στον ευρύτερο χώρο της τεχνητής νοημοσύνης και της μηχανικής μάθησης, ο υπολογισμός δεξαμενής (Reservoir Computing) αναδύεται ως μια εξαιρετικά πρωτοποριακή προσέγγιση, αντιμετωπίζοντας αποτελεσματικά τις προκλήσεις που αντιμετωπίζουν τα αναδρομικά νευρωνικά δίκτυα (RNNs) και ανοίγοντας νέες προοπτικές για την επεξεργασία σημάτων και την αντιμετώπιση πολύπλοκων προβλημάτων.[\[63\]](#)

Ο υπολογισμός δεξαμενής αναπτύχθηκε ως μια ανατρεπτική προσέγγιση στον τομέα της μηχανικής μάθησης και της επεξεργασίας σημάτων, βασιζόμενος στην ιδέα των αναδρομικών νευρωνικών δικτύων.[\[64\]](#) Το βασικό σκέλος αυτής της προσέγγισης είναι η χρήση μιας "δεξαμενής" ή "reservoir", μιας αυθόρμητα δυναμικής και ετερογενούς συστοιχίας νευρώνων, που είναι τυχαία συνδεδεμένοι μεταξύ τους.[\[69\]](#) Αυτή η δεξαμενή λειτουργεί ως μια είδους "μνήμη" ή "ενδιάμεση πλατφόρμα" που μπορεί να αποθηκεύσει και να επεξεργαστεί πληροφορίες καθ' όλη τη διάρκεια της εκτέλεσης του αλγορίθμου.

Αυτό το πλαίσιο αναδείχθηκε ως μια εναλλακτική προσέγγιση στις παραδοσιακές αρχιτεκτονικές των RNNs που αντιμετωπίζουν προβλήματα όπως vanishing gradient.[\[73\]](#) Αυτά τα προβλήματα εμφανίζονται κατά την εκπαίδευση των αναδρομικών δικτύων, όπου η πορεία των δεδομένων και των μαθησιακών σημάτων μπορεί να οδηγήσει σε εξαιρετικά μεγάλα ή μικρά κέρδη κλίσης. Ο υπολογισμός δεξαμενής αποφεύγει αυτά τα προβλήματα χρησιμοποιώντας την ιδιότητα της "ελαστικότητας" της δεξαμενής για να αποθηκεύσει και να επεξεργαστεί τις πληροφορίες, επιτρέποντας παράλληλα στον αλγόριθμο εκπαίδευσης να επικεντρωθεί στον προσανατολισμό των συνδέσμων προς τα επιθυμητά αποτελέσματα.

Έτσι, ο υπολογισμός δεξαμενής ανοίγει νέες προοπτικές για την ανάπτυξη προηγμένων μοντέλων που μπορούν να επεξεργαστούν πολύπλοκα σήματα και να αντιμετωπίσουν προκλήσεις που παρουσιάζονται σε πραγματικά περιβάλλοντα, όπως η αναγνώριση προτύπων, η πρόβλεψη και η αναγνώριση σημάτων με θόρυβο ή ανασφάλεια.

5.2 Αρχές του Υπολογισμού Δεξαμενής

Ο υπολογισμός δεξαμενής είναι μια μέθοδος μηχανικής μάθησης που αναπτύχθηκε με στόχο την ανάπτυξη αποτελεσματικών μοντέλων πρόβλεψης και αναγνώρισης προτύπων. Η βασική ιδέα είναι να χρησιμοποιήσουμε ένα δίκτυο νευρώνων, το οποίο ονομάζεται "δεξαμενή", για την ανάλυση και επεξεργασία εισόδου και τη δημιουργία προβλέψεων.

Στον υπολογισμό δεξαμενής, το αναδρομικό νευρωνικό δίκτυο διαιρείται σε δύο κύρια μέρη: το επίπεδο εισόδου και τη δεξαμενή (επίσης γνωστή ως κρυφό επίπεδο). Το επίπεδο εισόδου λαμβάνει τα δεδομένα εισόδου, και η δεξαμενή είναι ένα τυχαία αρχικοποιημένο δίκτυο που λειτουργεί ως δυναμικό σύστημα μνήμης. Οι συνδέσεις της δεξαμενής είναι σταθερές και τυχαία αντιστοιχισμένες, και δεν αλλάζουν κατά τη διάρκεια της διαδικασίας μάθησης.

5.2.1 Αρχές Λειτουργίας

Οι αρχές λειτουργίας της δεξαμενής (reservoir) αποτελούν ένα εξαιρετικά ενδιαφέρον κομμάτι της επιστήμης των νευρωνικών δικτύων, καθώς προσφέρουν μια νέα προσέγγιση για την επεξεργασία πληροφοριών με αναδρομικές δομές.

Η δεξαμενή αποτελείται από έναν μεγάλο αριθμό νευρώνων, οι οποίοι συνδέονται τυχαία μεταξύ τους μέσω συνάψεων μεταβλητών βαρών. Στην ουσία, αυτό δημιουργεί ένα εσωτερικό νευρωνικό δίκτυο με συνδέσεις που αντανakλούν τη δομή της εξωτερικής εισόδου. Κατά τη διάρκεια της φάσης εκπαίδευσης, η δεξαμενή "εκπαιδεύεται" να αποθηκεύει τις παραμέτρους ενός δυναμικού σήματος εισόδου, προσαρμόζοντας τα βάρη των συνάψεων του εσωτερικού νευρωνικού δικτύου.[\[64\]](#)

Έπειτα, το εσωτερικό δίκτυο λειτουργεί ως μια κινητή δεξαμενή γνώσης, καταφέροντας να αποτυπώσει τη δομή των εισερχόμενων δεδομένων. Μετά την εκπαίδευση, το δίκτυο αυτό μπορεί να χρησιμοποιηθεί για την πρόβλεψη ή αναγνώριση προτύπων με βάση τα εσωτερικά του δυναμικά.[\[67\]](#) Ουσιαστικά, τα δεδομένα εισόδου επεξεργάζονται παράλληλα από πολλούς νευρώνες του εσωτερικού δικτύου, δημιουργώντας μια εκτεταμένη αναπαράσταση της εισόδου με πολύπλοκες αλληλεπιδράσεις.

Η σημαντικότητα της δεξαμενής έγκειται στο γεγονός ότι μπορεί να προσεγγίσει πολύπλοκες συναρτήσεις και να αναδείξει πρότυπα που δεν είναι εύκολο να ανακαλυφθούν με άλλες μεθόδους.[\[70\]](#) Αυτή η προσέγγιση διαχωρίζει τον επεξεργαστή από τον "εκπαιδευτή", δηλαδή τον υπολογιστικό χώρο όπου διεξάγεται η εκπαίδευση των βαρών. Εν τέλει, η δεξαμενή αναδεικνύει ένα παράδοξο παράδειγμα ενός νευρωνικού δικτύου που μπορεί να εξάγει χρήσιμες πληροφορίες από έναν εξαιρετικά ανεπιθύμητο αρχικό σχεδιασμό.

Είναι πραγματικά αξιοσημείωτο πώς οι αρχές λειτουργίας της δεξαμενής συνιστούν ένα σημαντικό βήμα προς την κατανόηση της δυναμικής επεξεργασίας πληροφοριών στα νευρωνικά δίκτυα.[\[69\]](#) Αυτή η πρωτοποριακή προσέγγιση δημιουργεί έναν εσωτερικό χώρο επεξεργασίας, ο οποίος μπορεί να προσεγγίσει την πολυπλοκότητα των συσχετίσεων στα δεδομένα.[\[59\]](#)

Το μεγάλο πλεονέκτημα της δεξαμενής είναι η δυνατότητά της να ανακαλύπτει πρότυπα και συσχετίσεις που μπορεί να μην είναι εμφανείς με τις συμβατικές μεθόδους επεξεργασίας.[\[78\]](#) Καθώς το εσωτερικό δίκτυο λειτουργεί παράλληλα και αλληλεπιδρά μεταξύ των νευρώνων του, μπορεί να αποκτήσει μια πιο αφηρημένη και πλούσια αναπαράσταση των πληροφοριών.[\[71\]](#)

Στην ουσία, η δεξαμενή μας υπενθυμίζει πόσο σημαντικό είναι να είμαστε ανοιχτοί σε νέες προσεγγίσεις και να αναζητούμε άλλους τρόπους για την ανάλυση και την κατανόηση των δεδομένων.[\[79\]](#) Μέσα από αυτήν την πρωτοποριακή ιδέα, διαφαίνεται η ικανότητα των νευρωνικών δικτύων να αντιμετωπίζουν πολύπλοκα προβλήματα και να αναδεικνύουν μοτίβα που θα μπορούσαν να περάσουν απαρατήρητα με άλλες προσεγγίσεις.[\[75\]](#)

5.2.2 Επιλεγμένες Παραλλαγές του Υπολογισμού Δεξαμενής

Οι παραλλαγές του υπολογισμού δεξαμενής, όπως οι Echo State Networks (ESN) και οι Liquid State Machines (LSM), παρέχουν προσεγγίσεις που εμπνέονται από τη φύση και τη λειτουργία των νευρωνικών δικτύων στον εγκέφαλο. Αυτές οι παραλλαγές αποδεικνύουν την ποικιλία τρόπων με τους οποίους ένα νευρωνικό δίκτυο μπορεί να ανταποκριθεί σε πολυπλοκότητα και δομές χρονικής πληροφορίας.

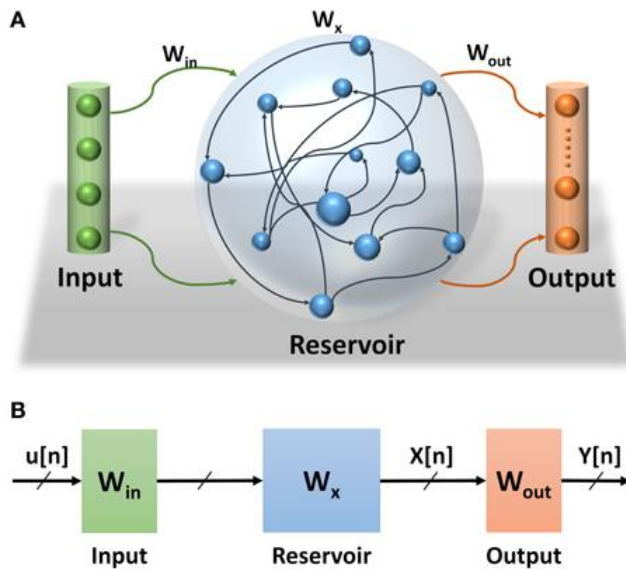
Στο πλαίσιο των Echo State Networks (ESN), η τυχαία σύνδεση των νευρώνων στη δεξαμενή αντιπροσωπεύει την ποικιλία των συνδέσεων που υπάρχουν μεταξύ νευρώνων στον ανθρώπινο εγκέφαλο. Στο μοντέλο αυτό, οι νευρώνες στη δεξαμενή είναι τυχαία συνδεδεμένοι μεταξύ τους, με τα βάρη των συνάψεων προκαθορισμένα. Οι νευρώνες αυτοί αποτελούν το "αντικείμενο" της δεξαμενής και παραμένουν σταθεροί κατά τη διάρκεια της εκπαίδευσης. Αυτή η δομή επιτρέπει τη δημιουργία ενός χώρου καταστάσεων που μπορεί να αντιμετωπίσει πολυπλοκότερες χρονικές δομές και σχέσεις.[73] Η εκπαίδευση εστιάζει στον εκμάθηση ενός γραμμικού τμήματος που συνδέει την είσοδο με τον αντικειμενικό χώρο της δεξαμενής. Επικεντρώνεται στον προσδιορισμό του γραμμικού τμήματος που συνδέει την είσοδο με τον αντικειμενικό χώρο της δεξαμενής, επιτρέποντας την ανάδειξη μοτίβων και προβλέψεων.[68][70]

Αναλυτικότερα, σύμφωνα με το άρθρο “[Design and Analysis of a Neuromemristive Reservoir Computing Architecture for Biosignal Processing](#)” των D Kudithipudi, Q Saleh και C Merkel, Το Echo State Network (ESN) είναι μια κατηγορία μοντέλου υπολογισμού δεξαμενών που παρουσιάστηκε από τον Jaeger (2001). Τα ESN θεωρούνται ως μερικώς εκπαιδευμένα τεχνητά νευρωνικά δίκτυα (ANN) με τοπολογία επαναλαμβανόμενου δικτύου. Χρησιμοποιούνται για χωροχρονικά προβλήματα επεξεργασίας σήματος. Το μοντέλο ESN είναι εμπνευσμένο από την αναδυόμενη δυναμική του τρόπου με τον οποίο ο εγκέφαλος χειρίζεται τα χρονικά ερεθίσματα.

Αποτελείται από ένα στρώμα εισόδου, ένα στρώμα δεξαμενής και ένα στρώμα εξόδου (βλ. Σχήμα 1Α). Το στρώμα δεξαμενής, είναι η καρδιά του δικτύου, με πλούσιες επαναλαμβανόμενες συνδέσεις. Αυτές οι συνδέσεις δημιουργούνται τυχαία και κάθε σύνδεση έχει ένα τυχαίο βάρος που συνδέεται με αυτήν. Μόλις δημιουργηθούν, αυτά τα τυχαία βάρη δεν αλλάζουν ποτέ κατά τη διάρκεια των φάσεων εκπαίδευσης ή δοκιμής του δικτύου. Το στρώμα εξόδου του ESN συνδυάζει γραμμικά το επιθυμητό σήμα εξόδου από την πλούσια ποικιλία διεγερμένων σημάτων του στρώματος δεξαμενής.

Η κεντρική ιδέα είναι ότι μόνο τα βάρη του στρώματος εξόδου πρέπει να εκπαιδευτούν, χρησιμοποιώντας απλούς αλγόριθμους γραμμικής παλινδρόμησης. Το ESN παρέχει ένα μαθηματικό πλαίσιο υψηλών επιδόσεων για την επίλυση πολλών προβλημάτων. Συγκεκριμένα, μπορούν να εφαρμοστούν σε αναδρομικά τεχνητά νευρωνικά δίκτυα χωρίς εσωτερικό θόρυβο. Τα ESN έχουν απλουστευμένους αλγόριθμους εκπαίδευσης σε σύγκριση με άλλα αναδρομικά ANN και είναι πιο αποδοτικά από τις μεθόδους που βασίζονται σε πυρήνες (π.χ. Support Vector Machines) λόγω της ικανότητάς τους να ενσωματώνουν χρονικά ερεθίσματα (Lukoševičius and Jaeger, 2009). Λόγω των αναδρομικών συνδέσεων, η έξοδος της δεξαμενής εξαρτάται από την τρέχουσα κατάσταση εισόδου και όλες τις προηγούμενες καταστάσεις εισόδου εντός της μνήμης του συστήματος.

Η τοπολογία του αναδρομικού δικτύου της δεξαμενής επιτρέπει την εξαγωγή χαρακτηριστικών χωροχρονικών σημάτων. Αυτή η ιδιότητα έχει χρησιμοποιηθεί σε διάφορους τομείς εφαρμογών, όπως η αναγνώριση κίνησης (Ishu et al., 2004), η ανάλυση φυσικής γλώσσας (Tong et al., 2007) και η αναγνώριση ομιλίας (Skowronski and Harris, 2007).



ΕΙΚΟΝΑ 1.

(Α) Το δίκτυο Echo State αποτελείται από τρία επίπεδα: στρώμα εισόδου, στρώμα δεξαμενής και στρώμα εξόδου.
(Β) Αφηρημένη δομή του Δικτύου Echo State. Πώς διαδίδονται τα σήματα μέσω του ESN και τα αποτελέσματα διαφορετικών συνόλων βάρους στο δίκτυο.

Τρία κύρια σύνολα βαρών σχετίζονται με τα ESN (Εικόνα 1B). Τα βάρη στα στρώματα εισόδου και δεξαμενής δημιουργούνται τυχαία. Αυτά τα στρώματα χρησιμοποιούνται για την εξαγωγή χρονικών χαρακτηριστικών του σήματος εισόδου. Μπορούν να θεωρηθούν ως ένα εσωτερικό στάδιο προεπεξεργασίας που προετοιμάζει το σήμα για το πραγματικό στρώμα επεξεργασίας όπου μαθαίνεται η ταξινόμηση στο στρώμα ανάγνωσης. Το Σχήμα 1B δείχνει επίσης τη διάδοση των σημάτων μέσω των ESN. Το σήμα εισόδου στο ESN $u(n)$ υφίσταται προεπεξεργασία στα στρώματα εισόδου και δεξαμενής για την εξαγωγή του σήματος χρονικού χαρακτηριστικού $x(n)$, το οποίο τροφοδοτείται στο στρώμα ανάγνωσης για την ολοκλήρωση της διαδικασίας ταξινόμησης. Λαμβάνοντας υπόψη ότι τα στρώματα εισόδου και δεξαμενής δεν αποτελούν πραγματικά μέρη αυτής της διαδικασίας, τα βάρη τους δεν εκπαιδεύονται, γεγονός που καθιστά την εκπαίδευση των ESN πολύ πιο εύκολη από άλλους τύπους RNN.

Στόχος του αλγορίθμου εκπαίδευσης είναι ο υπολογισμός των βαρών στο στρώμα εξόδου με βάση τη δυναμική απόκριση (καταστάσεις) του στρώματος δεξαμενής (Jaeger, 2002). Οι καταστάσεις του στρώματος δεξαμενής υπολογίζονται με βάση τα διανύσματα εισόδου και τα βάρη του στρώματος εισόδου και της δεξαμενής, όπως φαίνεται στην εξίσωση (1).

$$x[n+1] = f^{res}(W_{in} u[n+1] + W_x x[n]) \quad (1)$$

όπου $u[n]$ είναι η είσοδος του ESN, W_{in} είναι ο πίνακας βαρών μεταξύ του στρώματος εισόδου και του στρώματος δεξαμενής, W_x είναι ο πίνακας βαρών μεταξύ των νευρώνων εντός του στρώματος δεξαμενής και f^{res} είναι η συνάρτηση ενεργοποίησης του στρώματος δεξαμενής.

Οι καταστάσεις του στρώματος δεξαμενής για όλα τα διανύσματα εισόδου χρησιμοποιούνται ως είσοδος σε μια εκπαίδευση με επίβλεψη για τον υπολογισμό των βαρών εξόδου W_{out} . Υπάρχουν διάφορες μέθοδοι γραμμικής παλινδρόμησης για τον υπολογισμό των βαρών στο στρώμα εξόδου. Στην παρούσα εργασία χρησιμοποιείται η κανονική εξίσωση για την υλοποίηση της εκπαίδευσης με επίβλεψη του ESN (Σχήμα 1B).

$$W_{out} = (YX')(XX')^{-1} \quad (2)$$

όπου X είναι ένας πίνακας που συνθέτει όλες τις καταστάσεις του στρώματος δεξαμενής και Y είναι ένας πίνακας όλων των εξόδων εκπαίδευσης.

Η διαδικασία εκπαίδευσης του ESN μπορεί να εξηγηθεί μέσω των ακόλουθων βημάτων:

1. Κατά την αρχικοποίηση, δημιουργούνται τυχαία τα βάρη για τα στρώματα εισόδου και δεξαμενής (W_{in} και W_x).
2. Οδηγήστε το επόμενο διάνυσμα εισόδου $u[n + 1]$ στο στρώμα εισόδου.
3. Υπολογίστε την απόκριση του στρώματος δεξαμενής χρησιμοποιώντας την (1).
4. Αποθηκεύστε την απόκριση σε έναν πίνακα (X).
5. Επαναλάβετε τα βήματα 2-4 για όλα τα διανύσματα εισόδου.
6. Υπολογίστε τα βάρη εξόδου με βάση την κανονική εξίσωση (2).

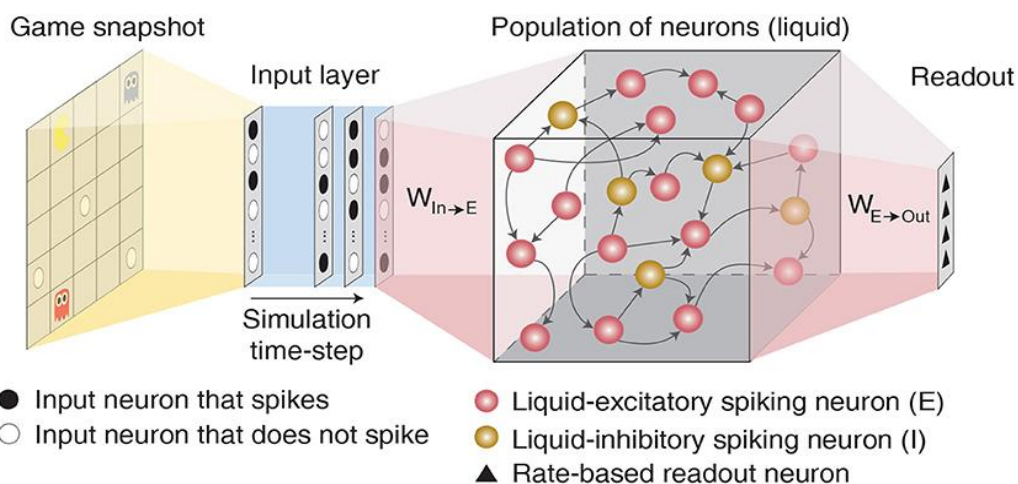
Μόλις υπολογιστούν τα βάρη του στρώματος εξόδου, το δίκτυο είναι έτοιμο και η κατάσταση του στρώματος δεξαμενής χρησιμοποιείται για τον υπολογισμό της εξόδου του δικτύου, όπως φαίνεται στην Εξίσωση (3).

$$y[n+1]=f^{out}(W_{out}x[n+1]) \quad (3)$$

όπου $y(n + 1)$ είναι η έξοδος του δικτύου, W_{out} είναι ο πίνακας βαρών στο στρώμα ανάγνωσης και f^{out} είναι η συνάρτηση ενεργοποίησης του στρώματος ανάγνωσης.

Από την άλλη πλευρά, οι Liquid State Machines (LSM) παρέχουν μια διαφορετική προσέγγιση, θεωρώντας τη δεξαμενή ως ένα "υγρό" δίκτυο νευρώνων, όπου η είσοδος ρέει μέσα και προκαλεί πολύπλοκες δυναμικές αλληλεπιδράσεις μεταξύ των νευρώνων, δημιουργώντας έναν δυναμικό χώρο καταστάσεων.[80] Η εκπαίδευση των LSM επικεντρώνεται στην προσαρμογή των συναπτικών βαρών κατά τη διάρκεια της λειτουργίας, επιτρέποντας την ανάδειξη επιθυμητής συμπεριφοράς στην απόκριση του δικτύου σε συγκεκριμένες εισόδους.[64]

Αναλύοντας περαιτέρω και αυτή την παραλλαγή, σύμφωνα με το άρθρο “[Reinforcement Learning With Low-Complexity Liquid State Machines](#)” των W Ponghiran, G Srinivasan και K Roy, η LSM αποτελείται από ένα στρώμα εισόδου που συνδέεται αραιά μέσω σταθερών συναπτικών βαρών με ένα τυχαία διασυνδεδεμένο υγρό από νευρώνες spike και ακολουθείται από ένα στρώμα ανάγνωσης, όπως απεικονίζεται στο Σχήμα 1.



ΣΧΗΜΑ 1. Απεικόνιση της αρχιτεκτονικής LSM που αποτελείται από ένα στρώμα εισόδου αραιά συνδεδεμένο μέσω σταθερών συναπτικών βάρων σε τυχαία επαναλαμβανόμενα συνδεδεμένη δεξαμενή (ή υγρό) διεγερτικών και ανασταλτικών νευρώνων αιχμής που ακολουθείται από ένα στρώμα ανάγνωσης που αποτελείται από νευρώνες βασισμένους σε τεχνητό ρυθμό.

Κάθε νευρώνας spike πυροδοτεί ένα δυναμικό δράσης που οδηγεί είτε σε διεγερτικό είτε σε ανασταλτικό αποτέλεσμα σε όλες τις θέσεις τερματισμού του. Με βάση την ορολογία που ακολουθείται στους Maass et al. (2002) και Diehl and Cook (2015), ονομάζουμε έναν νευρώνα που οδηγεί σε διεγερτικό (ανασταλτικό) αποτέλεσμα ως διεγερτικό (ανασταλτικό) νευρώνα. Το στρώμα εισόδου (συμβολίζεται με P) μοντελοποιείται ως μια ομάδα διεγερτικών νευρώνων που εκτοξεύονται με βάση την κατάσταση του περιβάλλοντος εισόδου ακολουθώντας μια διαδικασία Poisson. Οι αραιές συνδέσεις εισόδου-υγρού αρχικοποιούνται έτσι ώστε κάθε διεγερτικός νευρώνας στο υγρό να λαμβάνει συναπτικές συνδέσεις από περίπου K τυχαίους νευρώνες εισόδου. Αυτό εγγυάται την ομοιόμορφη διέγερση των νευρώνων διέγερσης του υγρού από τις εξωτερικές spike εισόδου. Τα σταθερά συναπτικά βάρη εισόδου-υγρού επιλέγονται από μια ομοιόμορφη κατανομή μεταξύ 0 και α, όπως φαίνεται στον Πίνακα 1, όπου α είναι το μέγιστο όριο που επιβάλλεται στα βάρη.

Connection type	Weight
INPUT-TO-LIQUID CONNECTIONS	
$P \rightarrow E$	[0, 0.6]
RECURRENT-LIQUID CONNECTIONS	
$E \rightarrow E$	[0, 0.05]
$E \rightarrow I$	[0, 0.25]
$I \rightarrow E$	[0, 0.3]
$I \rightarrow I$	[0, 0.01]

ΠΙΝΑΚΑΣ 1. Παράμετροι αρχικοποίησης συναπτικού βάρους για τις σταθερές συνδέσεις LSM για εκμάθηση εξισορρόπησης

Το υγρό αποτελείται από διεγερτικούς νευρώνες (συμβολίζεται με E) και ανασταλτικούς νευρώνες (συμβολίζεται με I) που συνδέονται επαναλαμβανόμενα με έναν αραιό τυχαίο τρόπο, όπως απεικονίζεται στο Σχήμα 1. Ο αριθμός των διεγερτικών νευρώνων επιλέγεται να είναι $4 \times$ ο αριθμός των ανασταλτικών νευρώνων, όπως παρατηρείται στα κυκλώματα του φλοιού (Wehr and Zador, 2003). Χρησιμοποιούμε το μοντέλο Leaky-Integrate-and-Fire (LIF) (Dayan and Abbott, 2003) για να μιμηθούμε τη δυναμική τόσο των διεγερτικών όσο και των ανασταλτικών νευρώνων αιχμής, όπως περιγράφεται από τις ακόλουθες διαφορικές εξισώσεις:

$$\frac{dV_i}{dt} = \frac{V_{rest} - V_i}{\tau} + I_i(t) \quad (1)$$

$$I_i(t) = \sum_{l \in N_p} W_{li} \cdot \delta(t - t_l) + \sum_{j \in N_E} W_{ji} \cdot \delta(t - t_j) - \sum_{k \in N_I} W_{ki} \cdot \delta(t - t_k) \quad (2)$$

όπου V_i είναι το δυναμικό μεμβράνης του i -οστού νευρώνα στο υγρό, V_{rest} είναι το δυναμικό ηρεμίας στο οποίο το V_i υποχωρεί, με χρονική σταθερά τ , απουσία ρεύματος εισόδου, και $I_i(t)$ είναι το στιγμιαίο ρεύμα που προβάλλεται στον i -οστό νευρώνα, και N_P , N_E και N_I είναι ο αριθμός των νευρώνων εισόδου, των διεγερτικών και των ανασταλτικών νευρώνων, αντίστοιχα. Το στιγμιαίο ρεύμα είναι άθροισμα τριών όρων: ρεύμα από τους νευρώνες εισόδου, ρεύμα από τους διεγερτικούς νευρώνες και ρεύμα από τους ανασταλτικούς νευρώνες. Ο πρώτος όρος ενσωματώνει το άθροισμα των προσυναπτικών αιχμών, που συμβολίζεται με $\delta(t - t_l)$ όπου t_l είναι η χρονική στιγμή των προσυναπτικών αιχμών, με τα αντίστοιχα συναπτικά βάρη (W_{li} στο 3). Ομοίως, ο δεύτερος (τρίτος) όρος ενσωματώνει το άθροισμα των προσυναπτικών ακίδων από τους διεγερτικούς (ανασταλτικούς) νευρώνες, που συμβολίζεται με $\delta(t - t_j)$ ($\delta(t - t_k)$), με τα αντίστοιχα βάρη W_{ji} (W_{ki}) στο 3. Το δυναμικό της μεμβράνης του νευρώνα ενημερώνεται με το άθροισμα των ρευμάτων εισόδου, των διεγερτικών και των αρνητικών ανασταλτικών ρευμάτων, όπως φαίνεται στο 1. Όταν το δυναμικό της μεμβράνης φτάσει ένα ορισμένο κατώφλι V_{thres} , ο νευρώνας πυροδοτεί μια ακίδα εξόδου. Στη συνέχεια, το δυναμικό της μεμβράνης επαναφέρεται στο V_{reset} και ο νευρώνας περιορίζεται από την αιχμή για μια επακόλουθη πυρίμαχη περίοδο κρατώντας το δυναμικό της μεμβράνης του σταθερό. Οι υπερπαράμετροι του μοντέλου LIF για τους διεγερτικούς και ανασταλτικούς νευρώνες παρατίθενται στον Πίνακα 2.

Parameter	Value
EXCITATORY AND INHIBITORY NEURONS	
V_{rest}	0
V_{reset}	0
V_{thres}	0.5
τ	20 ms
τ_{refrac}	1 ms
Δt (simulation time-step)	1 ms

ΠΙΝΑΚΑΣ 2. Παράμετροι μοντέλου Leaky-Integrate-and-Fire (LIF) για τους υγρούς νευρώνες.

Υπάρχουν τέσσερις τύποι επαναλαμβανόμενων συναπτικών συνδέσεων στο υγρό, δηλαδή, $E \rightarrow E$, $E \rightarrow I$, $I \rightarrow E$ και $I \rightarrow I$. Εκφράζουμε κάθε σύνδεση με τη μορφή ενός πίνακα που αρχικοποιείται να είναι αραιός και τυχαίος, γεγονός που προκαλεί την ανεξαρτησία της δυναμικής αιχμής ενός συγκεκριμένου νευρώνα από τους περισσότερους άλλους νευρώνες και διατηρεί τη διαχωριστικότητα στη δραστηριότητα αιχμής των νευρώνων. Ωστόσο, ο βαθμός αραιότητας πρέπει να συντονιστεί για να επιτευχθεί πλούσια δυναμική του δικτύου. Διαπιστώνουμε ότι η υπερβολική αραιότητα (μειωμένη συνδεσιμότητα) οδηγεί σε εξασθενημένη αλληλεπίδραση μεταξύ των νευρώνων του υγρού και καθιστά το υγρό χωρίς μνήμη. Αντίθετα, η χαμηλότερη αραιότητα (αυξημένη συνδεσιμότητα) οδηγεί σε χαοτική δραστηριότητα αιχμής, η οποία εξαλείφει τη διαχωριστικότητα στη δραστηριότητα αιχμής των νευρώνων. Αρχικοποιούμε τους πίνακες συνδεσιμότητας έτσι ώστε κάθε διεγερτικός νευρώνας να λαμβάνει περίπου C συναπτικές συνδέσεις από ανασταλτικούς νευρώνες και αντίστροφα. Η υπερπαράμετρος C ρυθμίζεται εμπειρικά, για να αποφευχθούν τα συνήθη προβλήματα χαοτικής δραστηριότητας αιχμαλώτισης που εμφανίζονται όταν (1) οι διεγερτικοί νευρώνες συνδέονται μεταξύ τους και σχηματίζουν έναν βρόχο που οδηγεί πάντα σε θετική μετατόπιση του δυναμικού της μεμβράνης και όταν (2) ένας διεγερτικός νευρώνας συνδέεται με τον εαυτό του και διεγείρεται επανειλημμένα από τη δραστηριότητά του. Συγκεκριμένα, για την πρώτη κατάσταση, έχουμε μη μηδενικά στοιχεία στον

πίνακα συνδεσιμότητας $E \rightarrow E$ (συμβολίζεται με W_{EE}) μόνο σε θέσεις όπου τα στοιχεία στο γινόμενο των πινάκων συνδεσιμότητας $E \rightarrow I$ και $I \rightarrow E$ (συμβολίζονται με W_{EI} και W_{IE} , αντίστοιχα) είναι μη μηδενικά.

Αυτό εξασφαλίζει ότι οι διεγερτικές συναπτικές συνδέσεις δημιουργούνται μόνο για εκείνους τους νευρώνες που λαμβάνουν επίσης ανασταλτικές συναπτικές συνδέσεις, γεγονός που μετριάζει την πιθανότητα συνεχούς θετικής μετατόπισης των αντίστοιχων δυναμικών μεμβράνης. Για να παρακάμψουμε τη δεύτερη κατάσταση, αναγκάζουμε τα διαγώνια στοιχεία του W_{EE} να είναι μηδενικά και εξαλείφουμε τη δυνατότητα επαναλαμβανόμενης αυτοδιέγερσης. Καθ' όλη τη διάρκεια αυτής της εργασίας, δημιουργούμε έναν πίνακα επαναλαμβανόμενης συνδεσιμότητας για υγρό με m διεγερτικούς νευρώνες και n ανασταλτικούς νευρώνες σχηματίζοντας έναν πίνακα $m \times n$ του οποίου οι τιμές αντλούνται τυχαία από μια ομοιόμορφη κατανομή μεταξύ 0 και 1. Η σύνδεση σχηματίζεται μεταξύ εκείνων των ζευγών νευρώνων όπου οι αντίστοιχες καταχωρήσεις του πίνακα είναι μικρότερες από την πιθανότητα σύνδεσης-στόχο ($= C/m$). Για παράδειγμα, θεωρήστε ένα υγρό με $m = 1.000$ διεγερτικούς και $n = 250$ ανασταλτικούς νευρώνες. Προκειμένου να δημιουργήσουμε τον πίνακα συνδεσιμότητας $E \rightarrow I$ έτσι ώστε κάθε ανασταλτικός νευρώνας να λαμβάνει συναπτική σύνδεση από έναν μόνο διεγερτικό νευρώνα ($C = 1$), σχηματίζουμε πρώτα έναν τυχαίο πίνακα $1,000 \times 250$ του οποίου οι τιμές προέρχονται από μια ομοιόμορφη κατανομή μεταξύ 0 και 1. Στη συνέχεια δημιουργούμε μια σύνδεση μεταξύ εκείνων των ζευγών νευρώνων όπου οι καταχωρήσεις του πίνακα είναι μικρότερες από 0,1% ($1/1.000$). Παρόμοια διαδικασία επαναλαμβάνεται για τη σύνδεση $I \rightarrow E$. Στη συνέχεια αρχικοποιούμε τη σύνδεση $E \rightarrow E$ με βάση το γινόμενο των W_{EI} και W_{IE} .

Ομοίως, ο πίνακας συνδεσιμότητας για $I \rightarrow I$ (συμβολίζεται με W_{II}) αρχικοποιείται με βάση το γινόμενο των W_{IE} και W_{EI} . Τα βάρη σύνδεσης αρχικοποιούνται από μια ομοιόμορφη κατανομή μεταξύ 0 και β , όπως φαίνεται στον Πίνακα 1 για διαφορετικούς επαναλαμβανόμενους πίνακες συνδεσιμότητας, εκτός αν αναφέρεται διαφορετικά. Σημειώστε ότι τα βάρη των συναπτικών συνδέσεων από ανασταλτικούς νευρώνες είναι μεγαλύτερα από εκείνα των συναπτικών συνδέσεων από διεγερτικούς νευρώνες για να ληφθεί υπόψη ο μικρότερος αριθμός ανασταλτικών νευρώνων σε σχέση με τους διεγερτικούς νευρώνες. Τα ισχυρότερα βάρη των ανασταλτικών συνδέσεων συμβάλλουν στη διασφάλιση ότι κάθε νευρώνας λαμβάνει παρόμοια ποσότητα διεγερτικών και ανασταλτικών ρευμάτων εισόδου, γεγονός που βελτιώνει τη σταθερότητα του υγρού.

Η συνδυασμένη προσέγγιση των Echo State Networks (ESN) και Liquid State Machines (LSM) αποκαλύπτει την ευελιξία των νευρωνικών δικτύων στον χειρισμό των πολυπλοκότητας και της δομής των δεδομένων. Η διεξοδική κατανόηση των παραλλαγών αυτών επιτρέπει την αξιοποίηση των επιμέρους χαρακτηριστικών τους για την αντιμετώπιση συγκεκριμένων εφαρμογών. Οι ESN είναι ιδανικές για προβλέψεις χρονοσειρών, ανίχνευση πολύπλοκων μοτίβων και ανάλυση σειριακών δεδομένων, ενώ οι LSM επιτρέπουν τη δημιουργία ενός δυναμικού χώρου καταστάσεων που είναι ιδανικός για αναγνώριση μοτίβων, εκμάθηση και απόκριση σε πολλαπλές καταστάσεις. [\[66\]\[67\]](#)

Στο σύνολο τους, αυτές οι προσεγγίσεις αναδεικνύουν τη σημασία της εξερεύνησης και της καινοτομίας στον τομέα των νευρωνικών δικτύων. [\[68\]\[70\]](#) Η ανάπτυξη παραλλαγών όπως οι ESN και LSM επιτρέπει την προσαρμογή των μοντέλων σε διαφορετικές εφαρμογές, ενισχύοντας την ικανότητά τους να αντιλαμβάνονται και να επεξεργάζονται πολυπλοκότητα χρονικών δεδομένων. [\[79\]\[81\]](#) Και οι δύο παραλλαγές αυτές προσφέρουν μια διαφορετική ματιά στον τρόπο που τα νευρωνικά δίκτυα μπορούν να αντιλαμβάνονται, να επεξεργάζονται και να

ανταποκρίνονται σε δεδομένα με χρονική δομή, ανοίγοντας τον δρόμο για ποικίλες εφαρμογές.[75][82]

Συνοψίζοντας, οι παραλλαγές του υπολογισμού δεξαμενής αντιπροσωπεύουν μια σημαντική προσέγγιση για την ανάλυση και την πρόβλεψη χρονοσειρών και σειριακών δεδομένων.[78][80] Από τις Echo State Networks με τον έλεγχο των γραμμικών τμημάτων έως τις Liquid State Machines με την εστίαση στη δυναμική διάσταση, αυτές οι παραλλαγές αντιπροσωπεύουν μια πολύτιμη προσθήκη στο φάσμα των εφαρμογών των νευρωνικών δικτύων.

5.2.3 Εφαρμογές στην Επιστήμη και τη Τεχνολογία

Ο υπολογισμός δεξαμενής, αποτελεί μια προηγμένη τεχνική που στηρίζεται στην ιδιότητα των Επαναλαμβανόμενων Νευρωνικών Δικτύων[73], έχει εξαιρετικά ευρεία εφαρμογή και συμβολίζει μια επαναστατική προσέγγιση σε διάφορους τομείς. Η αποτελεσματικότητά του έχει αναγνωριστεί σε πληθώρα πεδίων, με σημαντικές συνέπειες για την επιστήμη και την τεχνολογία.

Στον τομέα της επιστήμης, ο υπολογισμός δεξαμενής αποτελεί ένα ισχυρό εργαλείο για την ανάλυση και πρόβλεψη πολύπλοκων δυναμικών συστημάτων.[74] Ειδικότερα, στον τομέα της μετεωρολογίας, μπορεί να χρησιμοποιηθεί για την πρόβλεψη καιρικών φαινομένων[64], ενώ στην κλιματολογία μπορεί να συμβάλει στην πρόβλεψη μακροπρόθεσμων εξελίξεων. Επίσης, στην επιστήμη των υλικών, μπορεί να χρησιμοποιηθεί για την πρόβλεψη των φυσικών διαδικασιών και την ανακάλυψη νέων υλικών με επιθυμητές ιδιότητες.[70]

Από την άλλη πλευρά, στον τομέα της τεχνολογίας, ο υπολογισμός δεξαμενής είναι καθοριστικός για πολλές εφαρμογές. Στον τομέα της αναγνώρισης φωνής και κειμένου, τα RNNs μπορούν να επεξεργαστούν τα δεδομένα με χρονική δομή, βελτιώνοντας την ακρίβεια και την απόδοση των συστημάτων.[69] Στην επεξεργασία σήματος, μπορούν να απομονώσουν σημαντικά μοτίβα και χαρακτηριστικά από πολύπλοκες χρονοσειρές.[81] Στην αυτόνομη ρομποτική, τα RNNs μπορούν να επιτρέψουν σε ρομπότ να αντιλαμβάνονται και να προσαρμόζονται σε μεταβαλλόμενα περιβάλλοντα. Τέλος, στον έλεγχο διαδικασιών, μπορούν να προβλέψουν τις εξελίξεις και να λαμβάνουν αποφάσεις βάσει της χρονικής εξέλιξης των παραμέτρων.[68]

Επιπλέον, στον χώρο της ιατρικής, για παράδειγμα, τα RNNs μπορούν να αναλύσουν πολύπλοκες χρονοσειρές δεδομένων, όπως εκδηλώσεις παλμικής πίεσης ή ηλεκτροκαρδιογραφήματα, προσφέροντας πληροφορίες για παθήσεις και προβλήματα υγείας.[70]

Στον τομέα της αυτοκινητοβιομηχανίας, οι εφαρμογές του υπολογισμού δεξαμενής είναι εξίσου σημαντικές. Μπορούν να υποστηρίξουν την αυτόνομη οδήγηση, προβλέποντας τη συμπεριφορά των οχημάτων βάσει της κίνησης των γύρω οχημάτων και των τρέχουσων συνθηκών στον δρόμο.[81]

Στην καταναλωτική τεχνολογία, οι εφαρμογές των RNNs απλώνονται σε πολυάριθμους τομείς. Από την προσωποποίηση των προτάσεων σε υπηρεσίες streaming μουσικής μέχρι την πρόβλεψη των προτιμήσεων των καταναλωτών, τα RNNs μπορούν να αναδείξουν κρυφά μοτίβα στα δεδομένα και να προσφέρουν εξατομικευμένες εμπειρίες στους χρήστες.[77][76]

Ο υπολογισμός δεξαμενής, λοιπόν, αναδεικνύει τον συνδυασμό της δυνατότητας να αντιμετωπίζει τις χρονικές πτυχές δεδομένων με την ικανότητα ανίχνευσης μοτίβων και συσχετίσεων. Οι πρακτικές εφαρμογές του εκτείνονται σε ποικίλους τομείς, προσφέροντας εργαλεία για πρόβλεψη, αναγνώριση και απόφαση βάσει της χρονικής εξέλιξης των δεδομένων.[65][68][80]

5.2.4 Προκλήσεις και Μελλοντικές Εξελίξεις

Παρά την αποτελεσματικότητα που επιδεικνύει ο υπολογισμός δεξαμενής στον τομέα της μηχανικής μάθησης, δεν μπορούμε να παραβλέψουμε τις εξελίξεις που αναμένονται στο μέλλον και τις προκλήσεις που εξακολουθούν να απαιτούν λύση. Ο καθορισμός της κατάλληλης αρχιτεκτονικής δεξαμενής και η βελτιστοποίηση των υπερπαραμέτρων είναι παράγοντες που απαιτούν περαιτέρω έρευνα και ανάπτυξη. Ο διαρκής αγώνας να επιλεγούν οι κατάλληλοι αλγόριθμοι και παράμετροι για να ανταποκριθούν στις ανάγκες των συγκεκριμένων προβλημάτων αποτελεί μια διαδικασία συνεχούς εξέλιξης.

Πέρα από τον καθορισμό αυτών των τεχνικών πτυχών, μία από τις κύριες προκλήσεις που παραμένει ανοιχτή είναι η ανάπτυξη τεχνικών για την αντιμετώπιση προβλημάτων διάστασης και χρόνου.[78] Πώς μπορούμε να διατηρήσουμε αποδοτική ανάλυση όταν έχουμε να κάνουμε με δεδομένα υψηλής διάστασης ή τότε ο χρονικός παράγοντας είναι πολύ σημαντικός αλλά δυσκολεύει την επεξεργασία και την ανάλυση.[74]

Συνολικά, ο υπολογισμός δεξαμενής παρουσιάζει μία συναρπαστική προοπτική για το μέλλον. Με τη συνεχή εξέλιξη των αλγορίθμων, τη βαθύτερη κατανόηση των αρχείων τους και την αντιμετώπιση των προκλήσεων που παρουσιάζονται, αναμένεται να παραμείνει ένα ισχυρό εργαλείο για την εξέλιξη της ικανότητας των συστημάτων να αναλύουν, να μαθαίνουν και να προβλέπουν σε περίπλοκες συνθήκες.[67]

Επιπλέον, μια σημαντική πτυχή που αξίζει προσοχής είναι η ανάπτυξη πιο πολύπλοκων και διαφοροποιημένων αλγορίθμων RNN. Η συνεχής εξέλιξη της τεχνολογίας και η αύξηση της υπολογιστικής ισχύος δημιουργούν το πλαίσιο για τη δημιουργία πιο πολύπλοκων μοντέλων, τα οποία μπορούν να αντιμετωπίσουν πιο απαιτητικά προβλήματα και να αποκωδικοποιήσουν περισσότερες πτυχές των δεδομένων.

Με την εκθείαση των μεγάλων όγκων δεδομένων που συλλέγονται σε καθημερινή βάση, είναι προφανές ότι τα RNN μπορούν να αναβαθμίσουν τον ρόλο τους στην ανάλυση και την πρόβλεψη.[68] Αλλά αυτό είναι μόνο το ξεκίνημα. Η ικανότητά τους να εξαγάγουν αξία από τα δεδομένα είναι σε συνεχή ανάπτυξη. Με περαιτέρω βελτιστοποίηση των αλγορίθμων και την εφαρμογή πιο προηγμένων τεχνικών μάθησης μηχανής, οι RNN μπορούν να ενισχύσουν την ικανότητα τους να ανακαλύπτουν πρότυπα, να εντοπίζουν τις χρονικές δομές και να παρέχουν λύσεις σε ακόμη πιο πολύπλοκα προβλήματα.[78]

Επιπροσθέτως, ο υπολογισμός δεξαμενής μπορεί να συνεισφέρει σημαντικά στον τομέα της προβλεπτικής ανάλυσης.[80] Οι RNN μπορούν να αντιμετωπίσουν το πρόβλημα της πρόβλεψης μελλοντικών τάσεων με βάση τα προηγούμενα δεδομένα, κάτι που έχει εφαρμογές σε πολλούς τομείς όπως η οικονομία, η αγορά μετοχών, η μετεωρολογία και πολλοί άλλοι. Η δυνατότητα αυτή μπορεί να συμβάλει στην πρόοδο της ανθρώπινης κοινωνίας, επιτρέποντας την ανάπτυξη βιώσιμων λύσεων και την αντιμετώπιση προκλήσεων.[74]

Συμπερασματικά, ο υπολογισμός δεξαμενής αντιπροσωπεύει μια εξελιγμένη προσέγγιση στη μηχανική μάθηση, εκμεταλλευόμενος τις δυνατότητες του δυναμικού ενός νευρωνικού δικτύου.[79] Με τις πολλαπλές παραλλαγές του και τις εφαρμογές του σε διάφορους τομείς, αποτελεί μια υποσχόμενη προοπτική για την αντιμετώπιση πολύπλοκων προβλημάτων μηχανικής εκμάθησης.

5.3 Η Έννοια του Υπολογισμού Δεξαμενής

Ο υπολογισμός δεξαμενής αναδεικνύει την ιδέα της υποκείμενης δυναμικής φύσης της μνήμης στον τομέα των νευρωνικών δικτύων. Στη δομή αυτή, η δεξαμενή λειτουργεί ως μια

είδους προ-επεξεργαστική μονάδα, αντιμετωπίζοντας τα εισερχόμενα δεδομένα και προετοιμάζοντάς τα για περαιτέρω επεξεργασία. Αυτό επιτρέπει στη δεξαμενή να διατηρεί τις εσωτερικές καταστάσεις και τις δυναμικές των προηγούμενων δεδομένων, επιτρέποντας την ανίχνευση μοτίβων και την αντιμετώπιση της χρονικής εξέλιξης.

Ενδιαφέρουσα είναι και η πτυχή της τυχαιότητας στις συνδέσεις της δεξαμενής. Οι συνδέσεις αυτές δεν τροποποιούνται κατά τη διάρκεια της διαδικασίας εκπαίδευσης, και είναι τυχαία ανατιθέμενες. Αυτή η ιδιομορφία επιτρέπει τη δημιουργία ενός είδους "χαοτικού υπόβαθρου", που μπορεί να ενισχύσει την ικανότητα της δεξαμενής να ανιχνεύει μοτίβα και να αντιμετωπίζει περίπλοκες σχέσεις μεταξύ των δεδομένων.

Ο υπολογισμός δεξαμενής αναδύεται ως μια εναλλακτική προσέγγιση στον τομέα της μηχανικής μάθησης και των νευρωνικών δικτύων. Ενώ η κλασική προσέγγιση βασίζεται στον εντοπισμό και τον υπολογισμό των βαρών μεταξύ των νευρώνων, η ιδέα του υπολογισμού δεξαμενής εισάγει μια νέα δομή, εστιάζοντας στην εκμετάλλευση των δυνατοτήτων μνήμης και χρονικής αντιστάθμισης. Στην παρούσα ενότητα, θα εξεταστεί η έννοια του υπολογισμού δεξαμενής και ο τρόπος λειτουργίας της, εστιάζοντας στον ρόλο της δεξαμενής ως δυναμικού μνημητικού χώρου.

5.3.1 Βασικές Έννοιες του Υπολογισμού Δεξαμενής

5.3.1.1 Δομή της Δεξαμενής

Ο υπολογισμός δεξαμενής βασίζεται στην ιδέα της χρήσης ενός δικτύου με δομή παρόμοια με δεξαμενή.[\[63\]](#) Η δεξαμενή αποτελείται από δύο βασικά στρώματα: το επίπεδο εισόδου και τη δεξαμενή, ή αλλιώς το κρυφό επίπεδο. Το επίπεδο εισόδου δέχεται τα δεδομένα εισόδου και τα προετοιμάζει για επεξεργασία, όπως συμβαίνει σε κάθε νευρωνικό δίκτυο. Η δεξαμενή, ωστόσο, αποτελεί τον πυρήνα της προσέγγισης αυτής.

Στην περαιτέρω ανάλυση της δομής της δεξαμενής, παρατηρείται ότι η εν λόγω δομή αντιπροσωπεύει ένα από τα βασικά χαρακτηριστικά που ξεχωρίζουν τον υπολογισμό δεξαμενής από τα παραδοσιακά νευρωνικά δίκτυα.[\[69\]](#) Ενώ το επίπεδο εισόδου λειτουργεί όπως σε κάθε νευρωνικό δίκτυο, δηλαδή αναλαμβάνει την προετοιμασία των εισερχομένων δεδομένων, η δεξαμενή αποτελεί την ουσία της προσέγγισης αυτής.

Σε αντίθεση με τον τρόπο λειτουργίας των συνηθισμένων νευρωνικών δικτύων, η δεξαμενή δεν περιορίζεται σε μια σειρά απλών στρωμάτων νευρώνων.[\[65\]](#) Αντί γι' αυτό, η δεξαμενή λειτουργεί ως ένας πολυδιάστατος, δυναμικός χώρος που αποτυπώνει τις εσωτερικές καταστάσεις και τις αλληλεπιδράσεις μεταξύ των δεδομένων. Αυτή η πολυδιάστατη δομή επιτρέπει την ανίχνευση πολύπλοκων μοτίβων και την αναγνώριση της χρονικής εξέλιξης των δεδομένων, καθιστώντας τον υπολογισμό δεξαμενής ιδανικό για προβλήματα που απαιτούν ανάλυση χρονοσειρών.[\[67\]](#)

Ταυτόχρονα, η πολυδιάστατη δομή της δεξαμενής επιτρέπει την αντιμετώπιση δεδομένων που διακρίνονται από διαστάσεις και συσχετίσεις.[\[81\]](#) Αυτό είναι ιδιαίτερα σημαντικό για πολυπλοκότερα προβλήματα, όπου η ανάλυση μεμονωμένων στοιχείων δεν αρκεί για την κατανόηση της συνολικής εικόνας.[\[79\]](#)

Συνεπώς, η δομή της δεξαμενής αποτελεί τον πυρήνα της προσέγγισης του υπολογισμού δεξαμενής, προσφέροντας μια πολυδιάστατη, δυναμική χωροχρονική αναπαράσταση των δεδομένων και ενισχύοντας την ικανότητα ανίχνευσης πολύπλοκων μοτίβων και χρονικών διαδικασιών.

5.3.1.2 Μαθηματικός Φορμαλισμός

Σύμφωνα με το άρθρο “[Reservoir computing approaches to recurrent neural network training](#)” των Mantas Lukoševičius και Herbert Jaeger, έστω ότι ένα πρόβλημα ή μια εργασία στο πλαίσιο της μηχανικής μάθησης ορίζεται ως το πρόβλημα της εκμάθησης μιας λειτουργικής σχέσης μεταξύ μιας δεδομένης εισόδου $u(n) \in R^u$ και μιας επιθυμητής εξόδου $y_{target}(n) \in R^y$, όπου $n = 1, \dots, T$ και T είναι ο αριθμός των σημείων δεδομένων στο σύνολο δεδομένων εκπαίδευσης $\{(u(n), y_{target}(n))\}$. Μια μη χρονική εργασία είναι όπου τα σημεία δεδομένων είναι ανεξάρτητα μεταξύ τους και ο στόχος είναι να μάθετε μια συνάρτηση $y(n) = y(u(n))$ έτσι ώστε να ελαχιστοποιείται η $E(y, y_{target})$, όπου E είναι ένα μέτρο σφάλματος, για παράδειγμα, το κανονικοποιημένο ριζικό μέσο τετραγωνικό σφάλμα (Normalized Root-Mean-Square Error - NRMSE).

$$E(y, y_{target}) = \sqrt{\frac{(\|y(n) - y_{target}(n)\|^2)}{(\|y_{target}(n) + y_{target}(n)\|^2)}} \quad (1),$$

όπου $\|\cdot\|$, σημαίνει Ευκλείδεια απόσταση (ή νόρμα).

Μια χρονική εργασία είναι όπου τα u και y_{target} είναι σήματα σε ένα διακριτό χρονικό πεδίο $n = 1, \dots, T$ και ο στόχος είναι να μάθουμε μια συνάρτηση $y(n) = y(\dots, u(n-1), u(n))$ έτσι ώστε η $E(y, y_{target})$ να ελαχιστοποιείται. Έτσι, η διαφορά μεταξύ του χρονικού και του μη χρονικού στόχου είναι ότι η συνάρτηση $y(\bullet)$ προσπαθούμε να μάθουμε ότι έχει μνήμη στην πρώτη περίπτωση και δεν έχει μνήμη στη δεύτερη. Και στις δύο περιπτώσεις, η υποκείμενη υπόθεση είναι, φυσικά, ότι η λειτουργική εξάρτηση που προσπαθούμε να μάθουμε υπάρχει στην πραγματικότητα στα δεδομένα. Για τη χρονική περίπτωση, αυτό ορίζεται ως δεδομένα που προσκολλώνται σε ένα μοντέλο πρόσθετου θορύβου της φόρμας $y_{target}(n) = y_{target}(\dots, u(n-1), u(n)) + \theta(n)$, όπου $y_{target}(\bullet)$ είναι η σχέση που πρέπει να μαθευτεί από το $y(\bullet)$ και $\theta(n) \in R^y$ είναι ένας όρος θορύβου, ο οποίος περιορίζει την ακρίβεια εκμάθησης, δηλ. την ακρίβεια αντιστοίχισης του μαθημένου $y(n)$ με το $y_{target}(n)$.

Όποτε λέμε ότι η εργασία ή το πρόβλημα μαθαίνεται καλά, ή με καλή ακρίβεια ή ακρίβεια, εννοούμε ότι $E(y, y_{target})$ είναι μικρό. Κανονικά ένα μέρος των σημείων δεδομένων T χρησιμοποιείται για την εκπαίδευση του μοντέλου και ένα άλλο μέρος (που δεν φαίνεται κατά τη διάρκεια της εκπαίδευσης) για τη δοκιμή του. Όταν μιλάμε για σφάλματα εξόδου και απόδοση ή ακρίβεια, θα έχουμε κατά νου τα σφάλματα δοκιμών (αν δεν ορίζεται ρητά διαφορετικά). Επίσης το n , που δηλώνει τον διακριτό χρόνο, θα χρησιμοποιείται συχνά παραλείποντας το εύρος του $1 \dots T$.

Συνοψίζοντας, ο μαθηματικός φορμαλισμός της υπολογιστικής των δεξαμενών περιλαμβάνει την οριοθέτηση της δυναμικής του συστήματος, την εκπαίδευση των δεξαμενών για την προσαρμογή της δυναμικής αυτής, και την πρόβλεψη της μελλοντικής κατάστασης. Ο μαθηματικός φορμαλισμός αυτός παρέχει τη θεωρητική βάση για την κατανόηση και την εφαρμογή της υπολογιστικής των δεξαμενών σε πολλούς τομείς, όπως η πρόβλεψη, ο έλεγχος και η επεξεργασία σημάτων.

5.3.1.3 Τυχειότητα στις Συνδέσεις

Μια σημαντική πτυχή του υπολογισμού δεξαμενής είναι η τυχειότητα στις συνδέσεις της δεξαμενής. Οι συνδέσεις αυτές δεν υπόκεινται σε τροποποιήσεις κατά τη διάρκεια της εκπαίδευσης και ανατίθενται τυχαία.[64] Αυτό επιτρέπει τη δημιουργία ενός "χαοτικού υπόβαθρου", το οποίο βοηθά στην ανίχνευση πολύπλοκων μοτίβων και σχέσεων μεταξύ των δεδομένων.

Η έννοια της τυχειότητας στις συνδέσεις της δεξαμενής αναδεικνύει μια πτυχή που διαχρονικά είχε αγνοηθεί από τα παραδοσιακά νευρωνικά δίκτυα.[69] Η χρήση τυχαίων συνδέσεων μεταξύ των νευρώνων στη δεξαμενή φαίνεται αντίθετη με την ιδέα της προσεκτικής επιλογής συνδέσεων και των προσαρμογών κατά τη διάρκεια της εκπαίδευσης. Ωστόσο, αυτή η τυχειότητα δημιουργεί έναν ενδιαφέροντα μηχανισμό που βοηθά στην αντιμετώπιση πολύπλοκων προβλημάτων.

Η τυχειότητα αυτή των συνδέσεων προσδίδει στη δεξαμενή ένα είδος "στοχαστικότητας".[67] Οι νευρώνες της δεξαμενής δεν συνδέονται σύμφωνα με προκαθορισμένα μοτίβα ή κανόνες, αλλά οι συνδέσεις τους είναι ανατιθέμενες τυχαία. Αυτό δημιουργεί ένα είδος "χαοτικού υπόβαθρου" στο εσωτερικό της δεξαμενής. Σε αυτό το χαοτικό περιβάλλον, οι νευρώνες αλληλεπιδρούν μεταξύ τους με τρόπο που δεν είναι απόλυτα προβλέψιμος, και αυτό επιτρέπει στο δίκτυο να ανιχνεύει και να αντιμετωπίζει πολύπλοκα μοτίβα και σχέσεις στα δεδομένα.

Η χρήση τυχαίων συνδέσεων μεταξύ των νευρώνων επιτρέπει στη δεξαμενή να εκμεταλλευτεί την ενδιαφέρουσα ιδιότητα του χάους.[68] Το χάος δημιουργεί έναν πλούσιο χώρο καταστάσεων που μπορεί να καλύψει μεγάλη ποικιλία διαφορετικών μοτίβων και δομών. Αυτό καθιστά τη δεξαμενή ιδιαίτερα ικανή για την αντιμετώπιση πολύπλοκων προβλημάτων που περιλαμβάνουν δεδομένα με χρονική εξέλιξη και μεγάλη αβεβαιότητα.

Συνολικά, η τυχειότητα στις συνδέσεις της δεξαμενής αναδεικνύει τη σημασία της ποικιλομορφίας και της ευελιξίας στη μηχανική μάθηση. Αυτή η προσέγγιση αποδεικνύεται ιδιαίτερα χρήσιμη για την αντιμετώπιση πολύπλοκων δεδομένων και για την ανίχνευση δομών και μοτίβων που δεν είναι προφανείς με τις παραδοσιακές προσεγγίσεις.

Συνοψίζοντας, ο υπολογισμός δεξαμενής αντιπροσωπεύει μια σημαντική προσέγγιση στον χώρο της μηχανικής μάθησης. Με την εισαγωγή της έννοιας της δεξαμενής ως δυναμικού μνημητικού χώρου και της τυχειότητας στις εσωτερικές συνδέσεις, αυτή η προσέγγιση είναι σε θέση να αντιμετωπίσει προβλήματα που προκύπτουν από χρονοσειρές και μεγάλα σύνολα δεδομένων. Παρά τις ανακαλύψεις και τις εξελίξεις στον χώρο της τεχνητής νοημοσύνης, ο υπολογισμός δεξαμενής παραμένει μια σημαντική προοπτική για την επίλυση πολύπλοκων προβλημάτων που απαιτούν χρονική αντισταθμιστική ικανότητα.

5.4 Εκπαίδευση και Λειτουργία της Δεξαμενής

Η επιστημονική κοινότητα της τεχνητής νοημοσύνης και της μηχανικής μάθησης εξετάζει συνεχώς νέες προσεγγίσεις για την επίλυση προβλημάτων που απαιτούν υψηλή υπολογιστική ισχύ και προηγμένες αναλυτικές ικανότητες. Ένα από τα πεδία που έχει κεντρική σημασία σε αυτό το πλαίσιο είναι η εκπαίδευση και η λειτουργία της δεξαμενής (reservoir computing), μια προηγμένη τεχνική που επιτρέπει την αποτελεσματική επίλυση πολύπλοκων προβλημάτων μάθησης.[63]

Κατά τη διάρκεια της εκπαίδευσης, τα δεδομένα εισόδου παρουσιάζονται στη δεξαμενή, και η δυναμική απόκριση της δεξαμενής καταγράφεται. Αυτή η απόκριση στη συνέχεια τροφοδοτείται σε ένα επίπεδο ανάγνωσης, το οποίο είναι ένα εκπαιδευσιμο γραμμικό ή μη γραμμικό επίπεδο υπεύθυνο για τον χαρτογραφικό συσχετισμό της εξόδου της δεξαμενής στην επιθυμητή έξοδο. Η εκπαίδευση του επιπέδου ανάγνωσης πραγματοποιείται συνήθως χρησιμοποιώντας απλούς και αποδοτικούς αλγόριθμους μάθησης, όπως η γραμμική παλινδρόμηση ή οι μηχανές διανυσματικής στήριξης.[\[64\]\[65\]](#)

5.4.1 Η Δεξαμενή ως Μοντέλο Υπολογιστικής Μάθησης

Η δεξαμενή αποτελεί ένα εξελιγμένο παράδειγμα υπολογιστικής μάθησης που βασίζεται στην έννοια της υπολογιστικής δυναμικής. Στο πλαίσιο αυτής της προσέγγισης, η δεξαμενή αποτελεί ένα χρονικό σύστημα που είναι ικανό να αναλύει και να επεξεργάζεται πληροφορία με βάση την προηγούμενη κατάσταση του συστήματος. Σε αντίθεση με πολλές άλλες προσεγγίσεις μηχανικής μάθησης, όπου το κύριο βάρος πέφτει στην εκπαίδευση των νευρωνικών δικτύων, στη δεξαμενή η εκπαίδευση περιορίζεται συνήθως στην προσαρμογή των συνδέσμων μεταξύ των νευρωνικών μονάδων, γνωστών και ως νευρώνες της δεξαμενής.

Η δεξαμενή, επομένως, δεν απαιτεί επιβλεπόμενη εκπαίδευση με τον ίδιο τρόπο που απαιτούν άλλες προηγμένες μέθοδοι μηχανικής μάθησης, όπως τα συμβατικά νευρωνικά δίκτυα.[\[69\]](#) Αυτή η διάκριση στην εκπαίδευση της δεξαμενής οδηγεί σε μια προσέγγιση που συνδυάζει την ευελιξία της μηχανικής μάθησης με τη δυναμική ικανότητα των χρονικών συστημάτων, καθιστώντας τη δεξαμενή μια πολύ ισχυρή εργαλειοθήκη για πολύπλοκες αναλύσεις και προβλήματα υπολογιστικής μάθησης.

Ειδικότερα, οι νευρώνες της δεξαμενής διατηρούν μια εσωτερική κατάσταση, που αντιπροσωπεύει την ιστορία των εισόδων που έχουν λάβει.[\[73\]](#) Αυτή η κατάσταση είναι αυτό που διαμορφώνει την αντίδραση των νευρώνων στις νέες εισόδους και ουσιαστικά καθορίζει τον τρόπο με τον οποίο η δεξαμενή αντιμετωπίζει την πληροφορία.[\[67\]](#) Αυτό επιτρέπει στη δεξαμενή να αντιληφθεί και να διατηρήσει μακροπρόθεσμες εξαρτήσεις και πολύπλοκα μοτίβα στα δεδομένα εισόδου.

Ο ρόλος της δεξαμενής ως μοντέλου υπολογιστικής μάθησης επισημαίνει τη σημασία της για την ανάπτυξη νέων προσεγγίσεων στη μηχανική μάθηση.[\[65\]](#) Οι τεχνικές της δεξαμενής έχουν εφαρμοστεί με επιτυχία σε πολλούς τομείς, από την αναγνώριση προτύπων και την πρόβλεψη χρονοσειρών έως την επεξεργασία σημάτων και την αναγνώριση φωνής. Η ικανότητα της δεξαμενής να εξάγει πληροφορία από πολύπλοκες δομές και συσχετίσεις την καθιστά κρίσιμο εργαλείο για προβλήματα που απαιτούν αναλυτική ικανότητα, αλλά και ανεπτυγμένη υπολογιστική ισχύ.

Συνοψίζοντας, η δεξαμενή ως μοντέλο υπολογιστικής μάθησης αντιπροσωπεύει μια προηγμένη προσέγγιση που ενσωματώνει τη δυναμική των χρονικών συστημάτων για την επίλυση πολύπλοκων προβλημάτων. Μέσα από την εσωτερική της κατάσταση και τις δυναμικές των νευρώνων, η δεξαμενή αποκρυπτογραφεί και αναλύει την πληροφορία, προσφέροντας μια προοπτική που επεκτείνει τα πεδία της τεχνητής νοημοσύνης και της μηχανικής μάθησης.

5.4.2 Εκπαίδευση της Δεξαμενής

Η κύρια διαφορά της δεξαμενής από άλλες τεχνικές είναι η διαδικασία εκπαίδευσής της, η οποία βασίζεται στην έννοια της υπολογιστικής δυναμικής. Στη φάση αυτή, το σύστημα

εκπαιδεύεται να αντιδρά σε εισόδους με τρόπο που να προκαλεί εσωτερική δυναμική κατάσταση στους νευρώνες της δεξαμενής, που ονομάζεται και "εσωτερική κατάσταση" ή "διαμόρφωση της δεξαμενής".[64] Αυτή η εσωτερική κατάσταση αντιπροσωπεύει την πληροφορία που έχει αποκτηθεί από τις εισόδους και την αντίδραση του συστήματος σε αυτές.

Κατά τη διαδικασία της εκπαίδευσης, οι είσοδοι που παρέχονται στο σύστημα είναι συνήθως μια χρονική ακολουθία δεδομένων. Αυτές οι είσοδοι είναι υπεύθυνες για τη δημιουργία δυναμικών και παραμετρικών αλλαγών στο εσωτερικό του συστήματος, προκαλώντας την εμφάνιση πολύπλοκων χρονικών μοτίβων στην εσωτερική κατάσταση των νευρώνων.[67] Οι αλλαγές αυτές μπορούν να αναπαρασταθούν σαν χώρος καταστάσεων, όπου κάθε σημείο του χώρου αντιστοιχεί σε ένα συγκεκριμένο συνδυασμό των νευρώνων της δεξαμενής.[65]

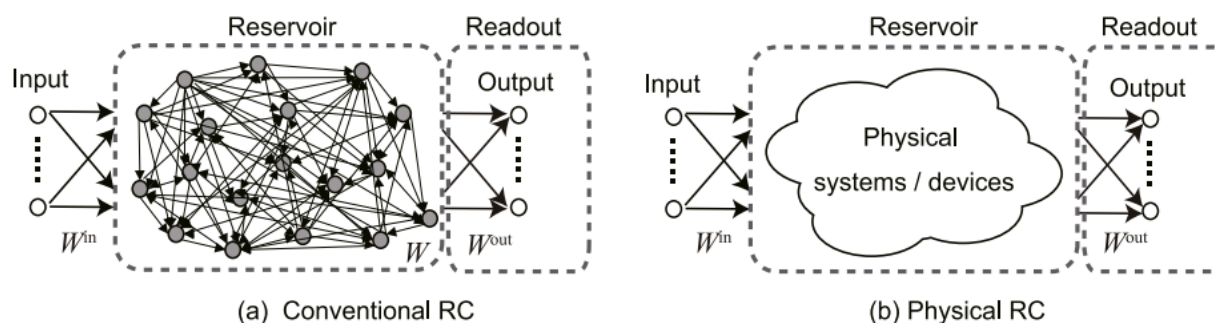
Το επιθυμητό σήμα εξόδου παράγεται ως γραμμικός συνδυασμός των σημάτων του νευρώνα από τη δεξαμενή που διεγείρεται από την είσοδο.[66] Αυτός ο γραμμικός συνδυασμός λαμβάνεται με γραμμική παλινδρόμηση, χρησιμοποιώντας το σήμα δασκάλου ως στόχο.

Μετά τη φάση της εκπαίδευσης, η εσωτερική αυτή κατάσταση χρησιμοποιείται για την παραγωγή εξόδου. Η διαδικασία αυτή συνήθως είναι μια γραμμική μετασχηματισμένη συνάρτηση της εσωτερικής κατάστασης.[69] Η παραγόμενη έξοδος μπορεί να χρησιμοποιηθεί για την αντιμετώπιση ποικίλων προβλημάτων, όπως πρόβλεψη χρονοσειρών, ταξινόμηση δεδομένων ή ακόμη και επεξεργασία σημάτων. Η εσωτερική κατάσταση που έχει αναπτυχθεί στους νευρώνες μπορεί να θεωρηθεί ως μια αναπαράσταση της πληροφορίας που παρουσιάστηκε στο σύστημα κατά τη διάρκεια της εκπαίδευσης. Αυτή η εσωτερική αναπαράσταση μπορεί να είναι υψηλής διάστασης και να περιλαμβάνει πληροφορία για σύνθετα μοτίβα, τάσεις και σχέσεις μεταξύ των δεδομένων εισόδου.

Ένα από τα πιο ενδιαφέροντα στοιχεία της δεξαμενής, είναι ότι οι νευρώνες της δεξαμενής δεν εκπαιδεύονται απευθείας για την παραγωγή συγκεκριμένης εξόδου. Αντίθετα, οι νευρώνες αντιδρούν στις εισόδους παράγοντας μια δυναμική κατάσταση που είναι ουσιαστικά ένας χώρος όπου μπορούν να αποθηκευτούν πολλά είδη πληροφορίας.[73] Αυτή η δυναμική κατάσταση χρησιμοποιείται στη συνέχεια ως είσοδος σε έναν γραμμικό μετασχηματισμό που παράγει την τελική έξοδο.[69]

Σημαντικό να σημειωθεί ότι η φάση της εκπαίδευσης αποτελεί το καίριο στάδιο για την απόκτηση της εσωτερικής κατάστασης που ανταποκρίνεται στις εισόδους. Ο σωστός σχεδιασμός των εισόδων και η διαμόρφωση της δεξαμενής είναι κρίσιμοι παράγοντες που επηρεάζουν την αποτελεσματικότητα και την απόδοση του συστήματος.

Σύμφωνα με το άρθρο “[Recent advances in physical reservoir computing: A review](#)” των G Tanaka, T Yamane, JB Héroux, κ.α., στο Reservoir Computing, τα δεδομένα εισόδου μετατρέπονται σε χωροχρονικά μοτίβα σε έναν χώρο υψηλής διάστασης από ένα RNN στη δεξαμενή. Στη συνέχεια, πραγματοποιείται ανάλυση προτύπων από τα χωροχρονικά πρότυπα στην ανάγνωση, όπως φαίνεται στο Σχήμα 1(α). Το κύριο χαρακτηριστικό της δεξαμενής είναι ότι τα βάρη εισόδου (W_{in}) και τα βάρη των αναδρομικών συνδέσεων εντός της δεξαμενής (W) δεν εκπαιδεύονται, ενώ μόνο τα βάρη ανάγνωσης (W_{out}) εκπαιδεύονται με έναν απλό αλγόριθμο μάθησης, όπως η γραμμική παλινδρόμηση. Αυτή η απλή και γρήγορη διαδικασία εκπαίδευσης καθιστά δυνατή τη δραστική μείωση του υπολογιστικού κόστους μάθησης σε σύγκριση με τα τυπικά RNN, το οποίο αποτελεί το σημαντικότερο πλεονέκτημα της δεξαμενής.



Εικ. 1. Σκελετοί Reservoir Computing όπου η δεξαμενή είναι σταθερή και μόνο τα βάρη ανάγνωσης εκπαιδεύονται. (α) Ένα συμβατικό σύστημα Reservoir Computing με δεξαμενή που βασίζεται σε RNN όπως στα ESN και LSM. (β) Ένα φυσικό σύστημα Reservoir Computing στο οποίο η δεξαμενή πραγματοποιείται χρησιμοποιώντας ένα φυσικό σύστημα ή συσκευή.

Ο ρόλος της δεξαμενής είναι να μετασχηματίζει μη γραμμικά τις ακολουθιακές εισόδους σε ένα χώρο υψηλών διαστάσεων, έτσι ώστε τα χαρακτηριστικά των εισόδων να μπορούν να αναγνωστούν αποτελεσματικά από ένα απλό αλγόριθμο μάθησης. Συνεπώς, αντί για τα RNNs, άλλα μη γραμμικά δυναμικά συστήματα μπορούν να χρησιμοποιηθούν ως δεξαμενές.

Ιδιαίτερα, η φυσική εφαρμογή του Reservoir Computing με χρήση δεξαμενών βασισμένων σε φυσικά φαινόμενα έχει προσελκύσει πρόσφατα αυξανόμενο ενδιαφέρον σε πολλούς τομείς έρευνας (Σχήμα 1(b)). Έχουν προταθεί διάφορα φυσικά συστήματα, υποστρώματα και συσκευές για την υλοποίηση του. Ένα κίνητρο για τη φυσική υλοποίηση των δεξαμενών είναι η υλοποίηση γρήγορων συσκευών επεξεργασίας πληροφοριών με χαμηλό κόστος μάθησης. Για την υλοποίηση υλικού των κανονικών RNNs, όπου απαιτείται εκπαίδευση, συχνά βασιζόμαστε σε προηγμένες τεχνολογίες νευρωνικών δικτύων και νευρομορφικών υλικών.

Αντίθετα, η φυσική υλοποίηση των δεξαμενών μπορεί να επιτευχθεί χρησιμοποιώντας διάφορα φυσικά φαινόμενα στον πραγματικό κόσμο, διότι δεν απαιτείται μηχανισμός για προσαρμοστικές αλλαγές κατά τη διάρκεια της εκπαίδευσης. Πράγματι, το φυσικό Reservoir Computing αποτελεί έναν από τους υποψήφιους των αναστρεφόμενων υπολογιστικών παραδειγμάτων βασισμένων σε νέα υλικά.

Αν και οι αρχές σχεδιασμού για το συμβατικό Reservoir Computing, όπως οι ESNs και οι LSMs, έχουν εξεταστεί εκτενώς, υφίστανται ακόμα θέματα που απαιτούν περαιτέρω εξέταση.

Συνοψίζοντας, η εκπαίδευση και η λειτουργία της δεξαμενής αντιπροσωπεύουν μια σημαντική προσέγγιση στον τομέα της τεχνητής νοημοσύνης και της μηχανικής μάθησης. Η δυνατότητά της να αντιμετωπίζει πολύπλοκες προβληματικές και η ευελιξία της στην επίλυση διαφορετικών εργασιών την καθιστούν μια ιδανική επιλογή για ποικίλους τομείς εφαρμογής. Με τη συνεχή εξέλιξη της τεχνολογίας, αναμένεται να δούμε ακόμη περισσότερες καινοτομίες σε αυτόν τον τομέα, ενισχύοντας την ικανότητα της δεξαμενής να αντιμετωπίζει προκλήσεις και να παρέχει λύσεις σε πολύπλοκα προβλήματα μάθησης.

5.5 Πλεονεκτήματα και Εφαρμογές

Ένα από τα κύρια και βασικά πλεονεκτήματα του υπολογισμού δεξαμενής είναι η απλότητα της διαδικασίας εκπαίδευσης. Επειδή μόνο το επίπεδο ανάγνωσης εκπαιδεύεται, η διαδικασία εκπαίδευσης είναι συχνά πιο γρήγορη και λιγότερο απαιτητική υπολογιστικά σε σύγκριση με τα

παραδοσιακά RNNs.[73] Επιπλέον, ο υπολογισμός δεξαμενής έχει εφαρμοστεί με επιτυχία σε μια ποικιλία εργασιών, συμπεριλαμβανομένων της πρόβλεψης χρονοσειρών, της αναγνώρισης ομιλίας, της επεξεργασίας εικόνων και των προβλημάτων ελέγχου.

5.5.1 Πλεονεκτήματα της Τεχνικής

5.5.1.1 Επεξεργασία Πολύπλοκων Προβλημάτων

Η επεξεργασία πολύπλοκων προβλημάτων αποτελεί ένα θεμελιώδες πλεονέκτημα της τεχνικής της δεξαμενής, το οποίο δίνει στην αρχιτεκτονική αυτή τη δυνατότητα να εξετάζει και να αναλύει σύνθετα και πολυδιάστατα δεδομένα. Αυτή η ιδιότητα της δεξαμενής είναι σημαντική και αναγνωρίζεται σε πολλούς τομείς, όπως η επιστήμη των δεδομένων, η τεχνητή νοημοσύνη, και η μηχανική μάθηση. Ας εμβαθύνουμε σε αυτό το θέμα:

Η δυνατότητα της δεξαμενής να αντιμετωπίζει πολύπλοκα προβλήματα πηγάζει από την ιδιαίτερη δομή της. Οι νευρώνες της δεξαμενής, που αποτελούν τον "αντανακλαστήρα" των εισερχομένων σημάτων, διατηρούν τις πληροφορίες για παρατεταμένα χρονικά διαστήματα. Αυτό τους επιτρέπει να ανιχνεύουν, να αναγνωρίζουν και να αποθηκεύουν πολύπλοκα χρονικά μοτίβα στα δεδομένα.[67] Κατά συνέπεια, η δεξαμενή είναι ιδανική για την ανάλυση και την πρόβλεψη δεδομένων που περιλαμβάνουν χρονικές διαστάσεις, όπως χρονοσειρές και σήματα.

Ένα παράδειγμα εφαρμογής είναι η πρόβλεψη των μελλοντικών τιμών σε χρονοσειρές. Με τη χρήση της δεξαμενής, μπορούμε να εκπαιδύσουμε το σύστημα να αναγνωρίζει προτερήματα και περιοδικά μοτίβα στις ιστορικές τιμές και να προβλέπει τις μελλοντικές εξελίξεις.[69] Επίσης, η δεξαμενή μπορεί να χρησιμοποιηθεί για την ανίχνευση ασυνήθιστων παρατηρήσεων σε χρονοσειρές, όπως απρόβλεπτες αλλαγές ή ανωμαλίες.[81]

Είναι σημαντικό να επισημανθεί ότι η επεξεργασία πολύπλοκων προβλημάτων δεν περιορίζεται μόνο στις χρονοσειρές. Η δυνατότητα της δεξαμενής να αναγνωρίζει μοτίβα και δομές είναι εφαρμόσιμη σε μια ποικιλία από δεδομένα, ανοίγοντας τον δρόμο για προηγμένες εφαρμογές σε πολλούς τομείς.[76]

5.5.1.2 Μη Γραμμική Επεξεργασία

Η δεξαμενή είναι ιδιαίτερα αποτελεσματική στην επεξεργασία μη γραμμικών δεδομένων, καθώς οι νευρώνες που απαρτίζουν το εσωτερικό αυτής της αρχιτεκτονικής επιδεικνύουν μια εξαιρετική ικανότητα να αναπαραστήσουν πολύπλοκες μη γραμμικές σχέσεις μεταξύ των δεδομένων εισόδου και εξόδου. Αυτό αντιπροσωπεύει ένα κρίσιμο πλεονέκτημα της δεξαμενής σε σχέση με άλλες μεθόδους μηχανικής μάθησης, ειδικά όταν αντιμετωπίζουμε προβλήματα που δεν μπορούν να περιγραφούν με απλές γραμμικές συσχετίσεις.[63]

Οι νευρώνες της δεξαμενής εμφανίζουν μια δυναμική απόκριση που προσομοιάζει τη συμπεριφορά του διακριτικού στοιχείου του νευρώνα. Αυτό σημαίνει ότι η απόκρισή τους σε ένα εισερχόμενο δείγμα δεν είναι απλώς μια σταθερή τιμή, αλλά ένα περίπλοκο χρονικό μοτίβο που εκφράζει την επεξεργασία των εισερχομένων δεδομένων. Η δυναμική αυτή απόκριση επιτρέπει στη δεξαμενή να αναπαράγει μη γραμμικές συναρτήσεις πολύ καλύτερα από τις συνηθισμένες γραμμικές μοντέλα.[64]

Η διαφορετική δομή και ο τρόπος λειτουργίας των νευρώνων της δεξαμενής επιτρέπουν τη δημιουργία πολυπλοκότερων συναρτήσεων, όπως αυτές που σχετίζονται με την αναγνώριση μοτίβων, την πρόβλεψη τάσεων και την εξαγωγή συμπερασμάτων από πολυδιάστατα και

ακανόνιστα δεδομένα. Αυτό το χαρακτηριστικό της δεξαμενής την καθιστά κατάλληλη για προβλήματα που απαιτούν αποτελεσματική αντιμετώπιση μη γραμμικών δεδομένων, όπως αναγνώριση πολύπλοκων προτύπων και πρόβλεψη διακυμάνσεων σε πολυμεταβλητά δεδομένα.[65]

5.5.1.3 Ενεργειακή Αποδοτικότητα

Μια από τις αξιοσημείωτες και αναγκαίες αρετές της δεξαμενής είναι η υψηλή της ενεργειακή αποδοτικότητα. Αυτή η χαρακτηριστική ιδιότητα της τεχνικής της δεξαμενής αποκτά ιδιαίτερη σημασία στο πλαίσιο της αυξανόμενης επιθυμίας για ανάπτυξη υψηλά αποδοτικών, ενεργειακά οικονομικών, και βιώσιμων τεχνολογιών.

Ο μηχανισμός πίσω από την ενεργειακή αποδοτικότητα της δεξαμενής βασίζεται στη δομή του νευρωνικού δικτύου. Οι νευρώνες της δεξαμενής, όπως και οι νευρώνες των ανατροφοδοτούμενων νευρωνικών δικτύων, λειτουργούν με τρόπο που επιτρέπει τη χρήση απλών αριθμητικών πράξεων για την επεξεργασία των εισόδων τους και τη δημιουργία εξόδων. Αυτό σημαίνει ότι δεν απαιτείται η εκτέλεση πολύπλοκων και ενεργοβόρων υπολογισμών για την λειτουργία της δεξαμενής.[69]

Σε συνδυασμό με την ανάγκη για χαμηλή κατανάλωση ενέργειας, αυτή η ενεργειακή αποδοτικότητα καθιστά τη δεξαμενή εξαιρετικά κατάλληλη για χρήση σε συστήματα με περιορισμένους πόρους. Τα ενσωματωμένα συστήματα, όπως αισθητήρες, αυτοκίνητα με αυτοματισμό, ή ακόμη και φορητές συσκευές, απαιτούν τεχνολογίες που μπορούν να λειτουργούν με μικρή κατανάλωση ενέργειας προκειμένου να διατηρηθεί η διάρκεια ζωής των μπαταριών ή η αυτονομία των συστημάτων.[76]

Επιπλέον, η ενεργειακή αποδοτικότητα της δεξαμενής συνδυάζεται με την ικανότητά της να αντιμετωπίζει πολύπλοκες εργασίες. Αυτό σημαίνει ότι μπορεί να επιτελέσει επεξεργασία και ανάλυση δεδομένων με χαμηλή κατανάλωση ενέργειας, επιτρέποντας την αποδοτική επίλυση προβλημάτων σε ποικίλους τομείς.[67]

Αυτή η ιδιότητα ενεργειακής αποδοτικότητας της δεξαμενής καθιστά την τεχνική αυτή ιδανική για περιβάλλοντα με περιορισμένους πόρους και για εφαρμογές που απαιτούν μεγάλη απόδοση με ελάχιστη κατανάλωση ενέργειας.

5.5.1.4 Παράλληλη Επεξεργασία

Η δυνατότητα παράλληλης επεξεργασίας αποτελεί μία από τις κεντρικές χαρακτηριστικές της δεξαμενής και ανοίγει νέες προοπτικές στην αντιμετώπιση προβλημάτων που απαιτούν την ανάλυση και την εξαγωγή πληροφοριών από πολλαπλά σετ δεδομένων ταυτόχρονα. Αυτό το χαρακτηριστικό έχει ιδιαίτερη σημασία σε περιβάλλοντα όπου υπάρχει ανάγκη για επεξεργασία πολλαπλών εισόδων και την εξαγωγή συνδυασμένων πληροφοριών.

Κατά την παράλληλη επεξεργασία, η δεξαμενή λειτουργεί ταυτόχρονα σε πολλαπλές εισόδους και επεξεργάζεται τα δεδομένα που προέρχονται από αυτές. Αυτό επιτρέπει την ανάλυση και την εξαγωγή πληροφοριών από διαφορετικά κανάλια ή αισθητήρες ταυτόχρονα, χωρίς την ανάγκη για σειριακή επεξεργασία ή διαχωρισμό των δεδομένων.

Ένα παράδειγμα είναι η επεξεργασία πολυκαναλικών σημάτων, όπως εγκεφαλικές εγγραφές (EEG) ή ηλεκτροκαρδιογραφίες (ECG), όπου κάθε κανάλι παρέχει πληροφορίες για διαφορετικές πτυχές του σήματος. Η δεξαμενή μπορεί να επεξεργαστεί όλα αυτά τα κανάλια ταυτόχρονα και να αναγνωρίσει μοτίβα, τάσεις ή ανωμαλίες που ενδέχεται να μην ήταν εμφανείς σε κάθε κανάλι ξεχωριστά.[71]

Αυτή η δυνατότητα επιτρέπει την αποδοτική αξιοποίηση των δεδομένων και την εξαγωγή πληροφοριών που δεν θα ήταν εύκολα ανιχνεύσιμες με την ανάλυση κάθε καναλιού ξεχωριστά. Επιπλέον, αυτή η δυνατότητα προσφέρει επίσης τη δυνατότητα αντιμετώπισης προβλημάτων που αφορούν πολυδιάστατα δεδομένα, όπως εικόνες και βίντεο, επιτρέποντας την ανάλυση διαφορετικών πτυχών τους συγχρόνως.^[75]

Κατά συνέπεια, η δυνατότητα παράλληλης επεξεργασίας καθιστά τη δεξαμενή ιδιαίτερα ισχυρή για προβλήματα που αφορούν πολλαπλές εισόδους και την αντλία πληροφοριών από διαφορετικές πηγές ταυτόχρονα.

Το συνολικό πλεονέκτημα της τεχνικής της δεξαμενής, παρόλα αυτά, προκύπτει από την ικανότητά της να αντιμετωπίζει πολύπλοκα και μη γραμμικά προβλήματα μάθησης, την ενεργειακή της αποδοτικότητα, τη δυνατότητα χρονικής επεξεργασίας και την ικανότητα παράλληλης επεξεργασίας. Αυτά τα χαρακτηριστικά την καθιστούν μια πολύτιμη εργαλειοθήκη για την αντιμετώπιση διαφόρων προβλημάτων σε ποικίλους τομείς, ενισχύοντας την εφαρμοσιμότητα της στην πραγματική καινοτομία και ερευνητική πρόοδο.

5.5.2 Εφαρμογές της Δεξαμενής

Οι εφαρμογές της δεξαμενής εκτείνονται σε πολλούς διαφορετικούς τομείς, εκμεταλλευόμενες την ισχυρή της ικανότητα να αντιμετωπίζει πολύπλοκες προκλήσεις που σχετίζονται με την ανάλυση και την επεξεργασία δεδομένων. Παρακάτω επεκτείνουμε τις εφαρμογές της δεξαμενής σε διάφορους τομείς:

5.5.2.1 Πρόβλεψη Χρονοσειρών

Μια από τις βασικές εφαρμογές της δεξαμενής είναι η πρόβλεψη χρονοσειρών. Αναφέρεται στη διαδικασία προβλέψεων ή εκτιμήσεων των μελλοντικών τιμών μιας μεταβλητής βάσει των παρελθοντικών της τιμών. Στο πλαίσιο της δεξαμενής, η πρόβλεψη χρονοσειρών αποτελεί μία σημαντική εφαρμογή, καθώς η δυνατότητά της να αντιμετωπίζει πολύπλοκες δομές και αλληλεξαρτήσεις στα δεδομένα την καθιστά ιδιαίτερα κατάλληλη για την ανάλυση και πρόβλεψη χρονοσειρών.^[65]

Οι εφαρμογές της πρόβλεψης χρονοσειρών είναι ευρέως διαδεδομένες σε διάφορους τομείς, προσφέροντας στρατηγικές και εργαλεία για τη λήψη αποφάσεων και την προβλεπτική ανάλυση των προκλήσεων που παρουσιάζονται.^[68] Ας επεκταθούμε περισσότερο σε διάφορους τομείς όπου η πρόβλεψη χρονοσειρών έχει σημαντική αξία:

- Οικονομικές Εξελίξεις

Στον τομέα των οικονομικών, η πρόβλεψη χρονοσειρών μπορεί να χρησιμοποιηθεί για την πρόβλεψη των τιμών στις χρηματαγορές, των επιτοκίων, του πληθωρισμού και άλλων οικονομικών δεικτών. Αυτό επιτρέπει στους επενδυτές, τις επιχειρήσεις και τους οικονομολόγους να λαμβάνουν ενημερωμένες αποφάσεις.^[63]

- Κλιματικές Μεταβολές

Η πρόβλεψη χρονοσειρών είναι χρήσιμη στον τομέα της μετεωρολογίας και της κλιματολογίας. Μπορεί να χρησιμοποιηθεί για την πρόβλεψη των μετεωρολογικών συνθηκών, της θερμοκρασίας, των καιρικών φαινομένων και άλλων περιβαλλοντικών μεταβλητών.^[80]

- Κίνηση Χρηματαγοράς

Οι επενδυτικές αποφάσεις στην κίνηση των χρηματαγορών βασίζονται στην πρόβλεψη των τιμών. Η πρόβλεψη χρονοσειρών μπορεί να βοηθήσει στην πρόβλεψη των τιμών μετοχών, εμπορευμάτων και νομισματικών ζευγαριών.[\[76\]](#)

- Διαχείριση Εφοδιαστικής Αλυσίδας

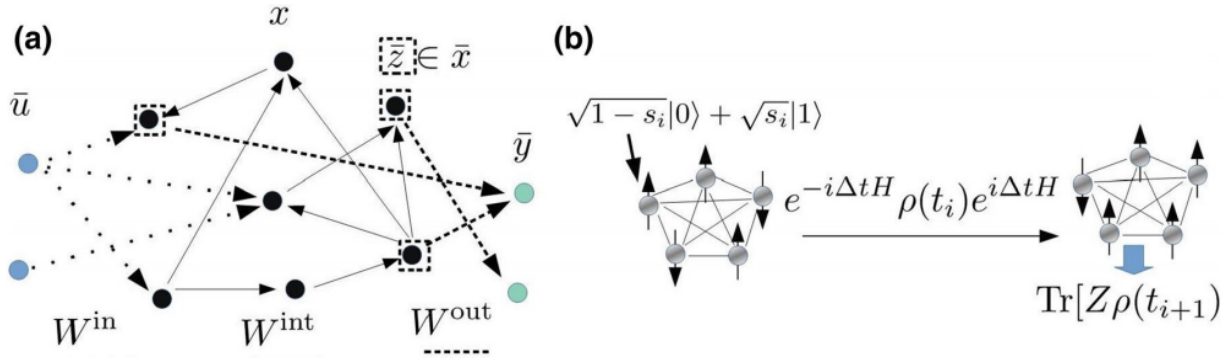
Σε βιομηχανικούς τομείς, η πρόβλεψη χρονοσειρών βοηθά στον σχεδιασμό της διαχείρισης των αποθεμάτων και της εφοδιαστικής αλυσίδας. Αυτό βοηθά στην αποτελεσματική διαχείριση των πόρων και τη μείωση των κοστών.[\[77\]](#)

Αναλύοντας περισσότερο τον τομέα αυτό, παρατίθεται ακολούθως ένα παράδειγμα μελέτης χρονοσειρών με Reservoir Computing βασισμένο στο άρθρο “[Optimizing a quantum reservoir computer for time series prediction](#)” των A. Kutvonen, K. Fujii, T. Sagawa. Οι κβαντικοί υπολογιστές και τα νευρωνικά δίκτυα υπόσχονται πολλά για το μέλλον της επεξεργασίας πληροφοριών. Στο αναφερόμενο άρθρο μελετάται ένας κβαντικός υπολογιστής δεξαμενής (QRC), ένα πλαίσιο που αξιοποιεί την κβαντική δυναμική και έχει σχεδιαστεί για τη γρήγορη και αποτελεσματική επίλυση διαχρονικών εργασιών μηχανικής μάθησης, όπως η αναγνώριση ομιλίας, η πρόβλεψη χρονοσειρών και η επεξεργασία φυσικής γλώσσας. Συγκεκριμένα, μελετάται η χωρητικότητα μνήμης και η ακρίβεια ενός κβαντικού υπολογιστή δεξαμενής που βασίζεται στο πλήρως συνδεδεμένο μοντέλο Ising με εγκάρσιο πεδίο, διερευνώντας διαφορετικές μορφές αλληλεπιδράσεων μεταξύ των σπιν και τις χρονικές κλίμακες υπολογισμού.

Ξεκινά περιγράφοντας το τυπικό πρωτόκολλο των επαναλαμβανόμενων νευρωνικών δικτύων και του υπολογισμού δεξαμενών. Δεδομένου του διακριτού διαχωρισμού του χρόνου $\{t_i\}$, οι τιμές των κόμβων εισόδου $\bar{u}(t_i) = \{u_j(t_i)\}_j^{n_u} = 0$, των κόμβων του δικτύου $\bar{x}(t_i) = \{x_j(t_i)\}_j^{n_x} = 0$ και των κόμβων εξόδου $\bar{y}(t_i) = \{y_j(t_i)\}_j^{n_y} = 0$, η εξίσωση ενημέρωσης για ένα απλό ανακυκλούμενο νευρωνικό δίκτυο δίνεται από τη σχέση

$$x_i(t_i+1) = f \left[\sum_j^{n_u} W_{x_i u_j}^{in} u_j(t_i + 1) + \sum_j^{n_u} W_{x_i x_j}^{int} x_j(t_i) \right] \quad (1)$$

όπου f είναι κάποια μη γραμμική διαφορίσιμη συνάρτηση, όπως η σιγμοειδής συνάρτηση, W^{in} , W^{int} είναι τα βάρη μετάβασης από τους κόμβους εισόδου και τους κρυφούς κόμβους και n_u , n_x και n_y είναι ο αριθμός των κόμβων εισόδου, των κρυφών κόμβων και των κόμβων εξόδου, αντίστοιχα. Ένα απλό επαναλαμβανόμενο νευρωνικό δίκτυο απεικονίζεται στο Σχήμα 1α.



(α) Κατά τον υπολογισμό της δεξαμενής τα δεδομένα εισόδου τροφοδοτούνται από τους κόμβους εισόδου $\bar{u}$ με βάρη W^{in} στους κρυφούς κόμβους του δικτύου x . Η εσωτερική αμετάβλητη δυναμική διέπεται από το W^{int} . Οι έξοδοι λαμβάνονται με στάθμιση των κόμβων ανάγνωσης z , που αντιστοιχεί στο $\langle Z_i \rangle = \text{Tr}[Z_i \rho]$ στη διάταξη μας, με τα βάρη ανάγνωσης W^{out} .

(β) Στην κβαντική δεξαμενή που υπολογίζει μια είσοδο s_i τροφοδοτείται σε ένα σπιν ορίζοντας την κατάσταση του σε $|\Psi s_i\rangle = \sqrt{1-s_i}|0\rangle + \sqrt{s_i}|1\rangle$. Μετά την εξέλιξη του ελεύθερου χρόνου, οι τιμές του κόμβου ανάγνωσης λαμβάνονται ως οι μέσοι όροι συνόλου των περιστροφών $\langle Z \rangle$

Οι έξοδοι του δικτύου δίνονται από

$$y_i(t_i+1) = \sum_j^{n_{out}} W_{y_i, z_j}^{out} z_j(t_i) \quad (2)$$

όπου υποθέτουμε ότι τα σήματα ανάγνωσης συλλέγονται από ένα υποσύνολο $\{Z_i\} \in \{x_i\}$ μεγέθους $n_{out} < n_x$ από τους κρυμμένους κόμβους. Γενικά η έξοδος $y_i(t_i+1)$ θα μπορούσε να έχει συνεισφορά από τις τελευταίες εισόδους $\bar{u}(t_i)$ και οι έξοδοι $\bar{y}(t_i)$ ή οι ίδιοι οι κόμβοι θα μπορούσαν να έχουν πιο σύνθετο σχεδιασμό.

Δεδομένων των τιμών των κόμβων εισόδου $\bar{u}(t_i)$ (ακολουθία εισόδου), ο στόχος της εκπαίδευσης είναι να βρεθεί ένα σύνολο βαρών W^{out} , που ελαχιστοποιούν το σφάλμα

$$E = \sum_i \left\| \bar{y}^{targ}(t_i) - \bar{y}(t_i) \right\|, \quad \text{όπου } \bar{y}^{targ}(t) \text{ είναι η ακολουθία-στόχος και } \|\bullet\| \text{ συμβολίζει}$$

μια διαφορίσιμη συνάρτηση απόστασης, όπως η μέση τετραγωνική απόσταση ή η διασταυρούμενη εντροπία μεταξύ κατανομών. Η συνήθης μέθοδος οπισθοδιάδοσης για τη βελτιστοποίηση συνήθως εκατομμυρίων παραμέτρων είναι η αξιολόγηση και η αριθμητική κάθοδος κατά μήκος των κλίσεων $\frac{dE}{dW}$ για όλα τα στοιχεία της W^{in} και W^{int} . Ο υπολογισμός δεξαμενής επιδιώκει τη βελτιστοποίηση της εργασίας με λιγότερη υπολογιστική ισχύ εκπαιδεύοντας τα βάρη εξόδου W^{out} μόνο. Στην περίπτωση του μέσου τετραγωνικού σφάλματος η εργασία ανάγεται στην τυπική εργασία γραμμικής παλινδρόμησης και η βέλτιστη διαμόρφωση των βαρών, $W_{opt}^{out} = \text{argmin}_W \sum_i (\bar{y}^{targ}(t_i) - \bar{y}(t_i))^2$ δίνεται από τη σχέση $W_{opt}^{out} = Z^{-1} Y$

(3) όπου $Y_{i,j} = \bar{y}_j^{targ}(t_i)$ και Z^{-1} δηλώνει το ψευδοαντίστροφο Moore-Penrose του πίνακα

$Z_{i,j} = z_j(t_i)$. Δεδομένου ότι στον υπολογισμό δεξαμενής η βελτιστοποίηση μεταξύ των δικτύων δεν είναι απαραίτητη κατά τη διάρκεια της εκπαίδευσης, ένα φυσικό σύστημα μπορεί να οδηγήσει άμεσα τη δυναμική του δικτύου.

Η μνήμη του δικτύου μετράται με την εργασία βραχυπρόθεσμης μνήμης (STM), η οποία ορίζεται ως μια εργασία για να θυμάται την είσοδο πριν από d βήματα:

$$y_d^{targ}(t_i) = s_{i-d} \quad (4)$$

Η ακρίβεια μνήμης ορίζεται ως η συσχέτιση μεταξύ της πρόβλεψης και της συνάρτησης στόχου:

$$R(d) = \text{cov}^2(y, y_d^{targ}) / \sigma^2(y) \sigma^2(y_d^{targ}), \quad (5)$$

όπου cov συμβολίζει τη συνδιακύμανση και σ^2 διακύμανση. Η χωρητικότητα μνήμης C ορίζεται ως $C = \sum_{d=1}^{d_c} R(d)$, όπου d_c είναι μια απόσταση αποκοπής, την οποία θέτουμε σε 100.

Για αποτελεσματική χρήση, η υπολογιστική δεξαμενών απαιτεί πολύπλοκη δυναμική, μεγάλα μεγέθη δικτύου και περιορισμένη δυναμική. Αυτές οι απαιτήσεις ικανοποιούνται φυσικά από αλληλεπιδρώντα κβαντομηχανικά συστήματα σπιν, των οποίων ο χώρος καταστάσεων κλιμακώνεται εκθετικά σε σχέση με τον αριθμό των σπιν και η δυναμική διέπεται από σύνθετη μοναδιαία δυναμική. Ο συμβολισμός εδώ ακολουθεί τον τυπικό συμβολισμό της κβαντομηχανικής. Η κατάσταση ενός σπιν είναι ένα δισδιάστατο μιγαδικό διάνυσμα που καλύπτεται από τις ιδιοκαταστάσεις του τελεστή Pauli Z , $\{|0\rangle, |1\rangle\}$. Ο συνολικός χώρος καταστάσεων N qubits είναι ένας χώρος τανυστικού γινομένου των δισδιάστατων επιμέρους χώρων σπιν. Η συνολική (καθαρή) κατάσταση του συστήματος αναπαρίσταται από ένα 2^N -διάστατο διάνυσμα $|\Psi\rangle$. Ένα στατιστικό μείγμα καθαρών καταστάσεων μπορεί να περιγραφεί από $2^N \times 2^N$ -διάστατο πίνακα πυκνότητας ρ . Για ένα κλειστό κβαντικό σύστημα, η χρονική εξέλιξη διέπεται από τη χρονικά ανεξάρτητη εξίσωση Schrödinger, η οποία για τον πίνακα πυκνότητας μπορεί να γραφεί ως εξής

$$\rho(t+\Delta t) = e^{-i\Delta t H} \rho(t) e^{i\Delta t H}, \quad (6)$$

όπου H είναι ένα $2^N \times 2^N$ -διάστατος ερμιτιανός πίνακας (hermitian matrix), η Χαμιλτονιανή (Hamiltonian) του συστήματος, η οποία καθορίζει τη δυναμική του συστήματος. Εδώ εξετάζεται ένα μοντέλο που έχει μελετηθεί εκτενώς, το πλήρως συνδεδεμένο μοντέλο Ising με εγκάρσιο πεδίο, του οποίου η Χαμιλτονιανή δίνεται από τη σχέση

$$H = \sum_{i,j} J_{ij} X_i X_j + h \sum_i Z_i, \quad (7)$$

όπου X_i και Z_i είναι οι τελεστές Pauli X και Z στη θέση i . Σε αυτό το μοντέλο όλα τα σπιν αλληλεπιδρούν μεταξύ τους στην κατεύθυνση x και είναι συνδεδεμένα με ένα εξωτερικό μαγνητικό πεδίο στην κατεύθυνση z .

Θεωρούμε μονοδιάστατες εισόδους $\bar{s}(t_i)=s_i$ και εξόδους $\bar{y}(t_i)=y_i$. Τη χρονική στιγμή t_i η είσοδος s_i τροφοδοτείται στο σύστημα θέτοντας την κατάσταση του πρώτου spin σε

$|\Psi_{s_i}\rangle = \sqrt{1-s_i}|0\rangle + \sqrt{s_i}|1\rangle$. Ο πίνακας πυκνότητας του συστήματος γίνεται τότε

$\rho \mapsto |\Psi_{s_i}\rangle\langle\Psi_{s_i}| \otimes T_{r_1}[\rho]$, όπου T_{r_1} συμβολίζει το ίχνος στον πρώτο βαθμό ελευθερίας του spin.

Αφού οριστεί η είσοδος, το σύστημα συνεχίζει να εξελίσσεται για χρόνο Δt . Κατά τη διάρκεια αυτού του χρόνου η δυναμική διέπεται από την εξίσωση (6) και η πληροφορία που κωδικοποιείται στο πρώτο spin θα διαδοθεί στο σύστημα. Κατά τη διάρκεια $[t_i, t_i+1]$ μετράμε τις μέσες τιμές των spin στην κατεύθυνση z , $Z_i = T_r[Z_i \rho]$, η οποία αντιστοιχεί στους κόμβους ανάγνωσης στη διάταξή μας. Αφού συγκεντρώσουμε μια επαρκή ποσότητα ζευγών σημάτων εισόδου-εξόδου, τα βάρη ανάγνωσης W^{out} εκπαιδεύονται σύμφωνα με την Εξίσωση (3) και οι αντίστοιχες έξοδοι και οι μελλοντικές προβλεπόμενες έξοδοι λαμβάνονται από την Εξίσωση (2). Ο βρόχος εισόδου-ανάγνωσης απεικονίζεται παραπάνω στο Σχήμα 1β.

Σημειώνεται ότι μια πειραματική υλοποίηση της διάταξης απαιτεί πολλαπλά αντίγραφα του ίδιου συστήματος, ώστε να καταστεί άσχετη η επίδραση της οπισθοδρόμησης των μετρήσεων. Αυτό είναι εφικτό με τη χρήση συστημάτων πυρηνικού μαγνητικού συντονισμού, όπου μπορεί να χρησιμοποιηθεί ένα τεράστιο σύνολο πανομοιότυπων μορίων.

Στο άρθρο που παρατέθηκε ως παράδειγμα, μελετήθηκε η χωρητικότητα μνήμης και η ακρίβεια ενός κβαντικού υπολογιστή-δεξαμενής που βασίζεται στο πλήρως συνδεδεμένο μοντέλο Ising με εγκάρσιο πεδίο. Παρατηρήσαμε ότι το διάστημα δεδομένων εισόδου Δt έχει μια εξαρτώμενη από το σύστημα βέλτιστη τιμή Δt_{opt} η οποία οδηγεί στη μέγιστη χωρητικότητα.

Επιπλέον, τα αποτελέσματα υποδεικνύουν ότι η απόκλιση στις δυνάμεις αλληλεπίδρασης μεταξύ των spin οδηγεί σε βελτιωμένη χωρητικότητα μνήμης μέχρι κάποιο όριο. Ενώ η μέγιστη χωρητικότητα αποτελεί ένδειξη απόδοσης, σε πρακτικές εφαρμογές μπορεί να μας ενδιαφέρει η ακρίβεια της μνήμης, η οποία γενικά βελτιώνεται με τη χρήση ενός ελαφρώς μεγαλύτερου χρονικού διαστήματος εισόδου από το Δt_{opt} . Ακόμη, υποδηλώνουν ότι η γρήγορη πτώση του ΟΤΟC είναι σημάδι της γρήγορης κωδικοποίησης πληροφοριών από τα spin εισόδου στα άλλα. Αυτές οι ρυθμίσεις παρουσιάζουν την ακριβή μνήμη με κόστος το μήκος της μνήμης. Επιπλέον, ως παράδειγμα πραγματικής εφαρμογής, προβλέψαμε τις τιμές των μετοχών χρησιμοποιώντας τον κβαντικό υπολογιστή δεξαμενής. Πιθανές μελλοντικές ερευνητικές κατευθύνσεις περιλαμβάνουν πιο προηγμένες στρατηγικές εισόδου-εξόδου δεδομένων, οι οποίες θα μπορούσαν να υλοποιήσουν π.χ. πολυδιάστατες εισόδους και εξόδους, μείγμα πολλαπλών χρονοδιαγραμμάτων ανάγνωσης και αποτελεσματικές στρατηγικές για τον καθορισμό των βέλτιστων χρονοδιαγραμμάτων για δεδομένες εργασίες.

Συνολικά, η πρόβλεψη χρονοσειρών αποτελεί ισχυρό εργαλείο για την αντιμετώπιση πολλαπλών προκλήσεων σε διάφορους τομείς. Η δεξαμενή ως αλγόριθμος μηχανικής μάθησης παρέχει τη δυνατότητα ανάλυσης πολύπλοκων μοτίβων και αλληλεξαρτήσεων στα δεδομένα, κάνοντας την κατάλληλη για ακριβείς και αξιόπιστες προβλέψεις.

5.5.2.2 Αναγνώριση Μοτίβων

Η δεξαμενή, ως αρχιτεκτονική νευρωνικού δικτύου, είναι εξαιρετικά κατάλληλη για την αναγνώριση πολύπλοκων μοτίβων σε δεδομένα. Αυτό σημαίνει ότι μπορεί να ανακαλύπτει κρυμμένες συνοχές και δομές που δεν είναι εύκολα αντιληπτές με το ανθρώπινο μάτι.

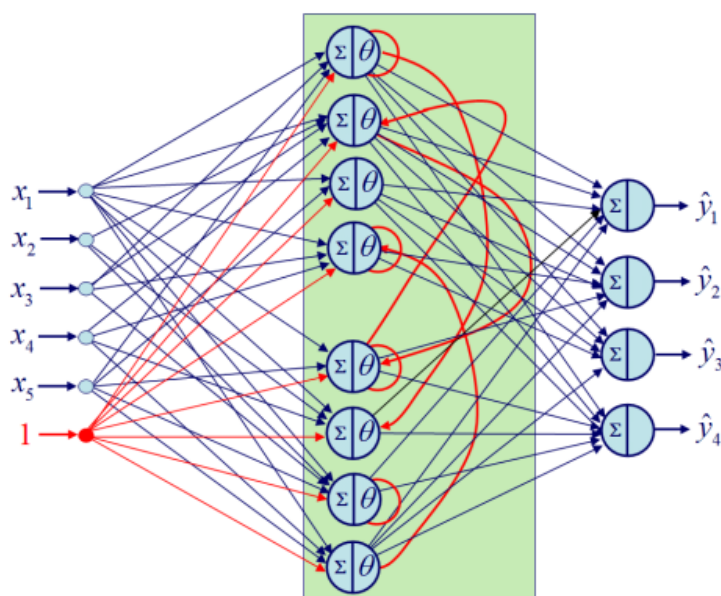
Ένα από τα σημαντικότερα παραδείγματα αναγνώρισης μοτίβων είναι η εφαρμογή της στην ανάλυση εγκεφαλικών εγγραφών, γνωστών ως ηλεκτροεγκεφαλογράφημα (EEG).^[64] Το εγκεφαλικό EEG καταγράφει την ηλεκτρική δραστηριότητα του εγκεφάλου και παρέχει δεδομένα σχετικά με τη λειτουργία του. Η δεξαμενή μπορεί να εκπαιδευτεί να αναγνωρίζει πρότυπα σε αυτά τα δεδομένα, όπως ενδείξεις για επιλεγμένες εγκεφαλικές δραστηριότητες ή ακόμη και την αναγνώριση ενδείξεων επιληπτικών κρίσεων.^[71]

Επιπλέον, στον τομέα της ιατρικής εικόνας, όπως η ραδιολογία, η δεξαμενή μπορεί να ανιχνεύσει ελλείψεις ή παρουσία συγκεκριμένων δομών σε εικόνες, όπως ακτινογραφίες, μαγνητικές τομογραφίες (MRI) και άλλες.^[77] Αυτό μπορεί να βοηθήσει τους ιατρούς να κάνουν ακριβείς διαγνώσεις και να προσδιορίζουν παθήσεις πιο αποτελεσματικά.

Εμβαθύνοντας μελετάται το άρθρο “[Reservoir computing for static pattern recognition](#)” των MJ Embrechts, LA Alexandre, JD Linton. Με βάση αυτό, παρατηρούμε ότι παρόλο που οι προηγούμενες εφαρμογές του υπολογισμού ταμιευτήρων περιορίζονταν κυρίως σε εφαρμογές ανάλυσης χρονοσειρών. Ο υπολογισμός δεξαμενής για στατική αναγνώριση προτύπων εκμεταλλεύεται τη δυναμική του στρώματος της δεξαμενής. Η συνάρτηση ενεργοποίησης εφαρμόζεται μόνο αφού η δυναμική της δεξαμενής έχει σταθεροποιηθεί. Σημειώνεται ότι μέχρι στιγμής οι εξόδοι από το στρώμα δεξαμενής προσδιορίζονται με μια μη επιβλεπόμενη διαδικασία και δεν υπάρχει ανάγκη εφαρμογής της οπισθοδιάδοσης ή της οπισθοδιάδοσης μέσω του χρόνου για το πρόβλημα της κρυφής ανάθεσης πίστωσης: δηλαδή, η δεξαμενή μπορεί να θεωρηθεί ως στάδιο μετασχηματισμού δεδομένων και η μηχανική μάθηση θα εφαρμοστεί μόνο στα μετασχηματισμένα δεδομένα. Επίσης εξαιτίας αυτού του λόγου, η μελετούμενη εργασία δεν εξετάζει συνδέσεις μεταξύ του στρώματος εισόδου και του στρώματος εξόδου (ή αντίστροφα).

Μόνο τα βάρη του στρώματος εξόδου καθορίζονται από μια διαδικασία με επίβλεψη. Επειδή το στρώμα εξόδου δεν έχει συνάρτηση ενεργοποίησης, τα βάρη αυτά μπορούν να προσδιοριστούν με ένα απλό μοντέλο γραμμικής παλινδρόμησης ή με την εφαρμογή εναλλακτικών γραμμικών τεχνικών, όπως η παλινδρόμηση κορυφογραμμής, η παλινδρόμηση κύριων συνιστωσών και η μερική παλινδρόμηση ελαχίστων τετραγώνων. Η δεξαμενή θα έχει συνήθως μερικές εκατοντάδες νευρώνες και οι διασυνδέσεις μεταξύ των μονάδων της δεξαμενής ανάδρασης θα είναι αραιές (π.χ., λιγότερο από 10 τοις εκατό συνδέσεις).

Τα βάρη του στρώματος εισόδου και των συνδέσεων της δεξαμενής επιλέγονται από μια ομοιόμορφη τυχαία κατανομή μεταξύ μηδέν και ενός μέγιστου ορίου. Αυτό το μέγιστο όριο επιλέγεται έτσι ώστε να σταθεροποιούνται οι ταλαντώσεις μετά το στρώμα δεξαμενής. Το κύριο πλεονέκτημα της υπολογιστικής δεξαμενής είναι ότι επειδή η δεξαμενή χρειάζεται μια ισχυρή προετοιμασία των βαρών και όχι μια άκρως επαναληπτική εκπαίδευση των βαρών, η υπολογιστική δεξαμενή δεν απαιτεί εκτεταμένο συντονισμό των παραμέτρων και επομένως μπορεί κατ' αρχήν να μαθαίνει από λιγότερα (ή ελπίζω μόνο από λίγα) πρότυπα. Η υπολογιστική δεξαμενή, τουλάχιστον όσον αφορά την είσοδο και τα βάρη της δεξαμενής, μπορεί επομένως να θεωρηθεί ως μια αληθοφανής μέθοδος που μοιάζει με τον εγκέφαλο: δηλαδή είναι δυνατό ή αληθοφανές ότι παρόμοιες διαδικασίες λαμβάνουν χώρα στον εγκέφαλο. Αυτή η τελευταία δήλωση πρέπει να εξεταστεί στο πλαίσιο του αλγορίθμου backpropagation. Επειδή ο αλγόριθμος backpropagation είναι ιδιαίτερα προβλεπόμενος και όχι τοπικός, είναι εξαιρετικά απίθανο ο εγκέφαλος να χρησιμοποιεί μια μέθοδο παρόμοια με τον αλγόριθμο backpropagation για την αναγνώριση προτύπων.



Σχήμα: Απλό σχήμα ενός τεχνητού νευρωνικού δικτύου που βασίζεται σε υπολογισμό δεξαμενών για στατική αναγνώριση προτύπων (με βάρη μεροληψίας στο στρώμα εισόδου). Σημειώνεται ότι η συνάρτηση μεταφοράς στο στρώμα δεξαμενής εφαρμόζεται αφού οι έξοδοι της δεξαμενής έχουν συγκλίνει σε σταθερές τιμές.

Επι της ουσίας, η δεξαμενή αναδεικνύει το δυναμικό της στο να ανακαλύπτει κρυμμένες δομές και παρουσιάζει ένα ισχυρό εργαλείο για την αναγνώριση προτύπων, τόσο σε δεδομένα υψηλής διάστασης όσο και σε περιπλοκότερες προκλήσεις αναγνώρισης μοτίβων.

5.5.2.3 Αναγνώριση Προτύπων σε Εικόνες

Η αναγνώριση προτύπων σε εικόνες αποτελεί έναν από τους πιο σημαντικούς και προκλητικούς τομείς στη μηχανική όραση και την τεχνητή νοημοσύνη. Η δεξαμενή, με την ιδιαίτερη δομή της, παρέχει έναν αποτελεσματικό τρόπο για την αναγνώριση πολύπλοκων προτύπων σε εικόνες και την εξαγωγή σημαντικών πληροφοριών από αυτές.

Η αναγνώριση αντικειμένων σε βιομηχανικές διεργασίες απαιτεί την αυτόματη αναγνώριση και παρακολούθηση των αντικειμένων κατά τη διάρκεια της παραγωγής.[77] Αυτό μπορεί να συμβάλει στην παρακολούθηση της ποιότητας, την αποτροπή σφαλμάτων και τη βελτίωση της αποτελεσματικότητας της παραγωγικής διαδικασίας.[71] Επίσης, η δυνατότητα της δεξαμενής να αναγνωρίζει πρότυπα μπορεί να εφαρμοστεί στην ανίχνευση ελαττωμάτων σε παραγωγικές γραμμές.[76] Σε περιβάλλοντα όπως η κατασκευή αυτοκινήτων ή η παραγωγή ηλεκτρονικών συσκευών, η ανίχνευση ακόμη και μικρών ελαττωμάτων είναι ζωτικής σημασίας για την ποιότητα των προϊόντων και την αποφυγή απορρίψεων.[68]

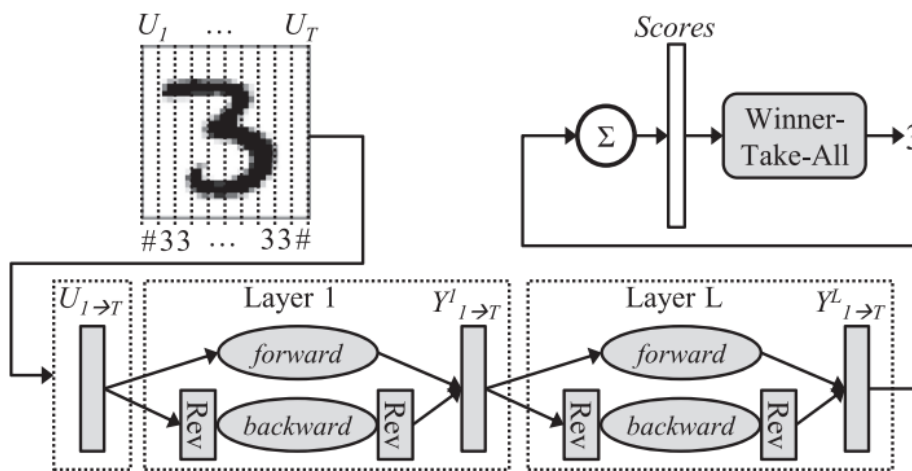
Επιπλέον, η αναγνώριση προσώπων σε εικόνες ασφαλείας αποτελεί έναν τομέα μεγάλης σημασίας, ειδικά σε περιβάλλοντα όπως αεροδρόμια, τράπεζες και κυβερνητικά κτίρια. Η δεξαμενή μπορεί να εκπαιδευτεί να αναγνωρίζει μοναδικά χαρακτηριστικά των ανθρώπινων προσώπων και να παρέχει αυξημένη ασφάλεια και παρακολούθηση σε ευαίσθητες περιοχές.

Σύμφωνα μάλιστα με το άρθρο “[On the application of reservoir computing networks for noisy image recognition](#)” των A Jalalvand, K Demuynck, W De Neve, JP Martens Σε πολλά συστήματα επεξεργασίας εικόνας που βασίζονται σε νευρωνικά δίκτυα, η είσοδος είναι ένας πίνακας εικονοστοιχείων ολόκληρης της εικόνας. Ωστόσο, για να εκμεταλλευτούμε τις ιδιότητες

δυναμικού συστήματος ενός Reservoir Computing Network (RCN), πρέπει να δημιουργήσουμε μια διαδοχική ροή εισόδου. Αυτό μπορεί να επιτευχθεί με σάρωση της εικόνας κατά στήλες (οριζόντια σάρωση) ή κατά σειρές (οριζόντια σάρωση).

Οι αναγνώσεις της δεξαμενής που θα περικλείονται στον αναγνωριστή αντιστοιχούν στα δέκα ψηφία και στο λευκό διάστημα που υπάρχει σε κάθε εικόνα ψηφίου. Με την εισαγωγή αυτού του λευκού χώρου και με το να οραματιζόμαστε μια εικόνα ως ένα ψηφίο που περιβάλλεται από λευκό χώρο, μπορούμε να επιτύχουμε ότι οι αναγνώσεις ψηφίων θα αντιδρούν κυρίως σε χαρακτηριστικά που είναι τυπικά για το ψηφίο που αντιπροσωπεύουν.

Μια τετριμμένη διαδικασία που οδηγεί στην επιθυμητή ροή εισόδου είναι η οριζόντια σάρωση: η εικόνα σαρώνεται κατά στήλες από αριστερά προς τα δεξιά και οι επόμενες στήλες (που ονομάζονται πλαίσια) αποτελούν την ακολουθία διανυσμάτων εισόδου (βλέπε Σχήμα 1).



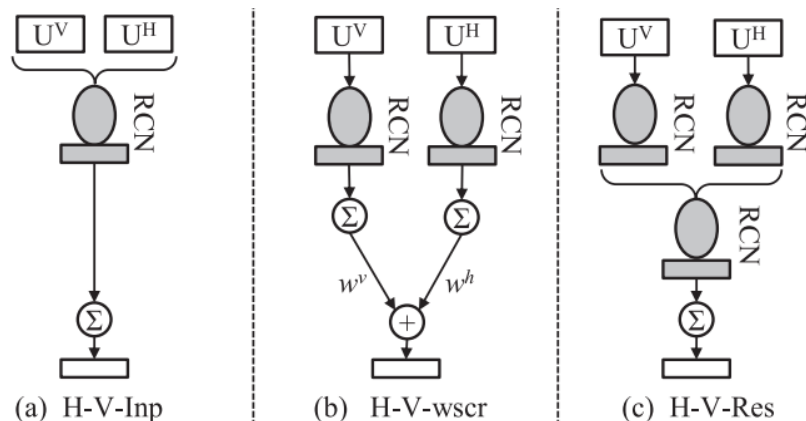
Εικ. 1. Αρχιτεκτονική ενός βαθύ αναγνωριστικού ψηφίου βασισμένου σε RCN που αξιοποιεί την αμφίδρομη επεξεργασία σε κάθε επίπεδο.

Οι βαθμολογίες των ψηφίων προκύπτουν από τη συσσώρευση των αναγνώσεων των ψηφίων σε όλη τη διάρκεια του χρόνου (το στοιχείο Σ) και ένας αλγόριθμος "ο νικητής τα παίρνει όλα" επιστρέφει το νικητήριο ψηφίο με την υψηλότερη δραστηριότητα ανάγνωσης.

Θα μπορούσε επίσης να επωφεληθεί κανείς από την αμφίδρομη επεξεργασία, πράγμα που σημαίνει ότι κάθε RCN περιέχει δύο δεξαμενές: την προς τα εμπρός δεξαμενή που επεξεργάζεται τις εισόδους $U_1 \rightarrow T$ ενώ την προς τα πίσω δεξαμενή που επεξεργάζεται τις εισόδους $U_T \rightarrow 1$. Οι εξόδοι της δεύτερης δεξαμενής αντιστρέφονται χρονικά πριν συνδυαστούν με τις εξόδους της εμπρόσθιας δεξαμενής. Ένα βαθύ (cascade) RCN προκύπτει από τη στοίβαξη πολλαπλών RCN, όπως απεικονίζεται στο Σχήμα 1. Κάθε στρώμα του βαθύ RCN είναι ένα βασικό αμφίδρομο RCN.

Τα στρώματα εκπαιδεύονται το ένα μετά το άλλο χρησιμοποιώντας τις ίδιες επιθυμητές εξόδους σε κάθε στρώμα. Το επιχείρημα για την κλιμάκωση των στρωμάτων είναι ότι το νέο στρώμα μπορεί να διορθώσει ορισμένα από τα λάθη που έγιναν από τα προηγούμενα στρώματα, επειδή προσφέρει πρόσθετη ικανότητα χρονικής μοντελοποίησης και έναν νέο εσωτερικό χώρο στον οποίο θα πραγματοποιηθεί η ταξινόμηση.

Παρατηρώντας και μια πιο σύνθετη αρχιτεκτονική, δεδομένου ότι είναι επίσης κατάλληλο για συνεχή handwriting recognition (HWR), η σάρωση οριζόντιας εικόνας φαίνεται να είναι μια προφανής επιλογή. Ωστόσο, για την αναγνώριση μεμονωμένων ψηφίων, μπορεί κανείς να εξετάσει και την κάθετη σάρωση, καθώς και μια συνδυασμένη προσέγγιση σάρωσης. Αυτές που προτείνονται στη μελέτη αυτή απεικονίζονται στο Σχήμα 2.



Σχήμα 2. Διαφορετικοί τρόποι συνδυασμού οριζόντιας (H) και κάθετης σάρωσης (V) σε ένα σύστημα: (α) παροχή στο RCN μιας γραμμής και μιας στήλης της εικόνας, (β) υπολογισμός ενός σταθμισμένου αθροίσματος των βαθμολογιών ψηφίων (συσσωρεύσεις με την πάροδο του χρόνου) που προκύπτουν από ένα H-RCN και ένα V-RCN και (γ) παροχή των εξόδων H-RCN και V-RCN σε ένα άλλο RCN και συσσώρευση των βαθμολογιών των αναγνώσεων αυτού του RCN.

→ Συνδυασμός χαρακτηριστικών εισόδου

Μια απλή στρατηγική συνδυασμού είναι να τροφοδοτείται το RCN με τη συνένωση μιας γραμμής και μιας στήλης σε κάθε χρονικό βήμα (βλέπε Σχήμα 2α). Προφανώς, η προσέγγιση αυτή προϋποθέτει τετράγωνη εικόνα, που οδηγεί σε πανομοιότυπο αριθμό πλαισίων ανά κατεύθυνση σάρωσης.

→ Σταθμισμένο άθροισμα αποτελεσμάτων

Μια άλλη στρατηγική είναι η δημιουργία δύο ανεξάρτητων παράλληλων συστημάτων: ένα που χρησιμοποιεί οριζόντια (H-RCN) και ένα που χρησιμοποιεί κάθετη σάρωση (V-RCN). Οι βαθμολογίες ψηφίων μπορούν στη συνέχεια να ληφθούν ως γραμμικός συνδυασμός των βαθμολογιών που προκύπτουν από τα δύο υποσυστήματα (βλέπε Σχήμα 2β). Το πλεονέκτημα αυτής της προσέγγισης είναι ότι μπορεί να εφαρμοστεί και σε ορθογώνιες εικόνες.

→ Συνδυασμός αναγνώσεων

Η τρίτη επιλογή είναι να παρέχονται οι συνδυασμένες αναγνώσεις του V-RCN και του H-RCN στο τελικό RCN αναγνώρισης ψηφίων (βλ. Σχ. 2γ). Προφανώς, η προσέγγιση αυτή προϋποθέτει και πάλι τετράγωνη εικόνα.

Στόχος της μελετούμενης εργασίας ήταν η διερεύνηση των δυνατοτήτων των υπολογιστικών δικτύων δεξαμενών (RCNs) στο πλαίσιο της επεξεργασίας εικόνας, με ιδιαίτερη έμφαση στη χειρόγραφη αναγνώριση ψηφίων και την αποθορυβοποίηση εικόνας.

Γενικότερα, η ικανότητα της δεξαμενής να αναγνωρίζει πολύπλοκα πρότυπα σε εικόνες έχει εφαρμογές μεγάλης σημασίας σε πολλούς τομείς και συνεισφέρει στη βελτίωση της παραγωγικότητας, της ασφάλειας και της ποιότητας των προϊόντων και των υπηρεσιών.

5.5.2.4 Πρόβλεψη και Έλεγχος Συστημάτων

Η εφαρμογή της δεξαμενής στην πρόβλεψη και τον έλεγχο δυναμικών συστημάτων ανοίγει νέες διαστάσεις στον τομέα της μηχανικής, της ρομποτικής, της πρόβλεψης κίνησης και

μεταφορών, καθώς και της περιβαλλοντικής μοντελοποίησης. Αυτή η εφαρμογή επιτρέπει την αποτελεσματική διαχείριση των δυναμικών παραμέτρων που επηρεάζουν τη συμπεριφορά των συστημάτων και την πρόβλεψη της εξέλιξής τους.

- Πρόβλεψη Δυναμικών Συστημάτων

Η δεξαμενή μπορεί να εκπαιδευτεί για να προβλέπει τη μελλοντική εξέλιξη δυναμικών συστημάτων. Αυτό έχει εφαρμογή σε πολλούς τομείς, όπως στην πρόβλεψη του καιρού, της θαλάσσιας κυκλοφορίας, της οικονομίας, και ακόμη και στην πρόβλεψη της κίνησης των οχημάτων σε αστικά περιβάλλοντα.[64] Η δυνατότητα της δεξαμενής να ανταποκριθεί σε πολύπλοκα δυναμικά μοντέλα την καθιστά ιδανική για την ανάλυση και πρόβλεψη της συμπεριφοράς τέτοιων συστημάτων.

- Έλεγχος Δυναμικών Συστημάτων

Η δεξαμενή μπορεί επίσης να χρησιμοποιηθεί για τον έλεγχο των δυναμικών συστημάτων.[68] Με τη χρήση ανάδρασης από τον εξωτερικό κόσμο, μπορεί να ρυθμίζει τις εισόδους του συστήματος έτσι ώστε να επιτυγχάνεται η επιθυμητή συμπεριφορά. Αυτό έχει εφαρμογή σε ρομποτικά συστήματα, ελέγχου διαδικασιών παραγωγής, και άλλες βιομηχανικές εφαρμογές.

- Πρόβλεψη Κίνησης και Μεταφορών

Στον τομέα των μεταφορών, η δεξαμενή μπορεί να χρησιμοποιηθεί για την πρόβλεψη της κίνησης και της κυκλοφορίας.[81] Αυτό μπορεί να συμβάλλει στον σχεδιασμό αποτελεσματικότερων συστημάτων μεταφορών, την αντιμετώπιση της κυκλοφοριακής συμφόρησης, και τη βελτίωση της ασφάλειας στους δρόμους.

- Περιβαλλοντική Μοντελοποίηση

Στον τομέα του περιβάλλοντος, η δεξαμενή μπορεί να χρησιμοποιηθεί για τη μοντελοποίηση και πρόβλεψη περιβαλλοντικών φαινομένων, όπως η κατανομή των ρύπων σε μια περιοχή[76] ή οι προβλεπόμενες κλιματικές μεταβολές[79].

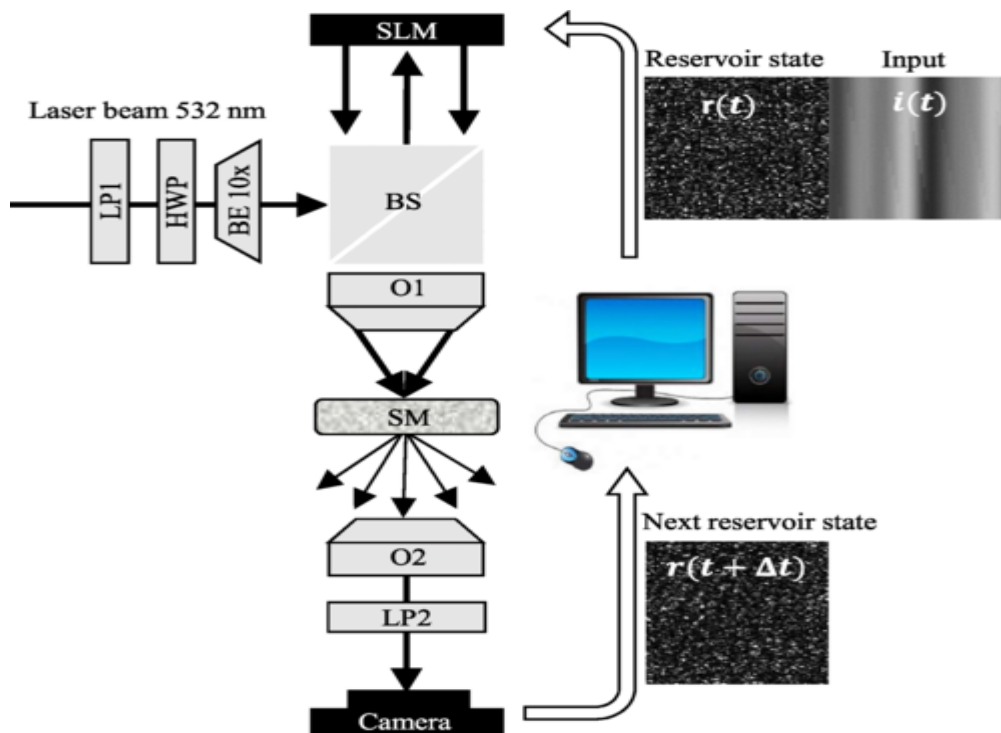
Ο συνδυασμός της δυνατότητας πρόβλεψης και ελέγχου δυναμικών συστημάτων με την ευελιξία και την αναγνώριση προτύπων της δεξαμενής ανοίγει νέες προοπτικές για τη βελτίωση της απόδοσης και της ανταγωνιστικότητας τεχνολογικών συστημάτων σε πολλούς διαφορετικούς τομείς.

Στο άρθρο “[Large-Scale Optical Reservoir Computing for Spatiotemporal Chaotic Systems Prediction](#)” των Mushegh Rafayelyan, Jonathan Dong, Yongqi Tan, κ.α. μελετάται ένα παράδειγμα που προτείνει ένα οπτικό σχήμα που εκτελεί υπολογισμό δεξαμενών σε πολύ μεγάλα δίκτυα, τα οποία είναι δυνητικά σε θέση να φιλοξενήσουν πολλά εκατομμύρια πλήρως συνδεδεμένους φωτονικούς κόμβους χάρη στις εγγενείς ιδιότητες της παραλληλίας και της επεκτασιμότητας.

Οι πειραματικές μελέτες επιβεβαιώνουν ότι, σε αντίθεση με τους συμβατικούς υπολογιστές, ο χρόνος υπολογισμού του οπτικού σχήματος εξαρτάται μόνο γραμμικά από τον αριθμό των φωτονικών κόμβων του δικτύου, ο οποίος οφείλεται σε ηλεκτρονικά γενικά έξοδα, ενώ το οπτικό μέρος του υπολογισμού παραμένει πλήρως παράλληλο και ανεξάρτητο από το μέγεθος της δεξαμενής.

Για να αποδειχθεί η επεκτασιμότητα του οπτικού σχήματος, εκτελούνται για πρώτη φορά προβλέψεις σε μεγάλα χωροχρονικά χαοτικά σύνολα δεδομένων που λαμβάνονται από την εξίσωση Kuramoto-Sivashinsky χρησιμοποιώντας οπτικές δεξαμενές με έως και 50.000 οπτικούς κόμβους. Τα αποτελέσματά είναι εξαιρετικά δύσκολα για τις συμβατικές μηχανές von Neumann και προωθούν σημαντικά την κατάσταση της τεχνολογίας των μη συμβατικών προσεγγίσεων υπολογισμού με δεξαμενές, γενικά.

Η πειραματική διάταξη για την εκτέλεση της οπτικής δεξαμενής (optical RC) φαίνεται στο Σχήμα 1 και περιγράφεται λεπτομερώς στο Παράρτημα Α. Τα βασικά οπτικά στοιχεία της διάταξης είναι οι phase-only χωρικοί διαμορφωτές φωτός (Spatial Light Modulators - SLM), το μέσο σκέδασης και η κάμερα. Οι SLM παρέχουν τόσο την κωδικοποίηση του διανύσματος εισόδου $i(t)$ διάστασης D_{in} στο και η κωδικοποίηση της επόμενης κατάστασης της δεξαμενής $r(t)$ διάστασης D_{res} (συνολική διάσταση D_{in} στο + D_{res}) στο χωρικό προφίλ φάσης του φωτός. Το μέσο σκέδασης εξασφαλίζει την τυχαία γραμμική ανάμιξή τους, η οποία ισοδυναμεί με τον γραμμικό πολλαπλασιασμό τους με μεγάλους πυκνούς τυχαίους πίνακες που αποτελούνται από ανεξάρτητες και ταυτόσημα κατανομημένες (independent and identically distributed - i.i.d.) τυχαίες μιγαδικές μεταβλητές. Τέλος, η κάμερα εκτελεί μια μη γραμμική ανάγνωση της έντασης του μιγαδικού πεδίου για την επόμενη κατάσταση της δεξαμενής $r(t + \Delta t)$, το οποίο αποστέλλεται από τον υπολογιστή πίσω στο SLM προκειμένου να εμφανιστεί με νέα είσοδο, και η διαδικασία επαναλαμβάνεται. Το άνω και το κάτω ένθετο στο Σχ.1 είναι αντίστοιχα παραδείγματα εικόνων που εμφανίζονται στο SLM και ανιχνεύονται από την κάμερα.



ΣΧΗΜΑ 1. Πειραματική διάταξη για την εκτέλεση υπολογισμού οπτικής δεξαμενής. Ο SLM λαμβάνει από τον υπολογιστή τη συνεπακόλουθη είσοδο $i(t)$ σε συνδυασμό με την κατάσταση της δεξαμενής $r(t)$ και την αποτυπώνει στο χωρικό προφίλ φάσης της ανακλώμενης δέσμης. Το μέσο σκέδασης παρέχει μια σύνθετη γραμμική ανάμειξη ολόκληρης της κωδικοποιημένης πληροφορίας. Τέλος, η κάμερα εκτελεί μια μη γραμμική ανάγνωση για την επόμενη κατάσταση της δεξαμενής $r(t + \Delta t)$, το οποίο αποστέλλεται από τον υπολογιστή πίσω στο SLM για να εμφανιστεί με νέα είσοδο, και η διαδικασία επαναλαμβάνεται. LP1 και LP2 (γραμμικοί πολωτές - linear polarizers), HWP (πλάκα μισού κύματος - half-wave plate), BE (διαστολέας δέσμης - beam expander), BS (διαχωριστής δέσμης- beam splitter), O1 και O2 (στόχοι - objectives).

Υπάρχει ένας αριθμός ρυθμιζόμενων παραμέτρων σχετικά με την κωδικοποίηση των καταστάσεων εισόδου και δεξαμενής στο SLM που περιγράφεται παρακάτω. Χωρίς απώλεια γενικότητας, υποθέτουμε ότι ο αριθμός των επιπέδων γκρι της κάμερας και του SLM είναι ίσος με 256.

Το SLM βαθμονομείται έτσι ώστε τα gray levels από 0 έως 255 να αντιστοιχούν γραμμικά σε καθυστερήσεις φάσης από 0 έως 2π . Επιπλέον, θεωρείται ότι ολόκληρο το σύνολο δεδομένων εισόδου είναι αρχικά κλιμακωμένο από 0 έως 255 και ο χρόνος λήψης της κάμερας ρυθμίζεται αρχικά ώστε να παρέχει ακόρεστες καταστάσεις δεξαμενής που κυμαίνονται και πάλι από 0 έως 255. Κατά συνέπεια, η κωδικοποίηση των καταστάσεων εισόδου και της δεξαμενής στο SLM μπορεί να περιγραφεί από $i(t) \rightarrow s_{in} i(t)$ και $r(t) \rightarrow s_{res} i(t)$, αντίστοιχα, με δύο παράγοντες κλιμάκωσης $0 \leq s_{in/res} \leq 1$. Επιπλέον, κάθε κλιμακωτή τιμή από τις καταστάσεις εισόδου και δεξαμενής μπορεί να κωδικοποιηθεί σε πολλαπλό αριθμό εικονοστοιχείων SLM που σχηματίζουν ένα μακροεικονοστοιχείο. Αυτές οι τροποποιήσεις πραγματοποιούνται στον υπολογιστή κάθε φορά πριν από την αποστολή των καταστάσεων εισόδου και δεξαμενής στο SLM, δεδομένου ότι λαμβάνουμε τις ρητές μορφές τους για τη χρονική στιγμή t μόνο στο $t - \Delta t$, εκτός από τις καταστάσεις εισόδου για εκπαίδευση, οι οποίες είναι ήδη γνωστές πριν από την έναρξη της εκπαίδευσης και μπορούν να τροποποιηθούν εκ των προτέρων. Σημειώνουμε ότι οι αναφερόμενες τροποποιήσεις απαιτούν μόνο απλό πολλαπλασιασμό με κλιμακωτές τιμές και δεν συνεπάγονται μεγάλη υπολογιστική επιβάρυνση. Ο αριθμός των εικονοστοιχείων σε ένα μακροεικονοστοιχείο συμβολίζεται με p_{in} για την κωδικοποίηση εισόδου και p_{res} για την κωδικοποίηση καταστάσεων δεξαμενής. Κατά συνέπεια, ο υπολογισμός της δεξαμενής στο οπτικό σύστημα μπορεί να περιγραφεί από την ακόλουθη αναδρομική σχέση:

$$r(t + \Delta t) = F \left[s_{res} r(t) \oplus J_{p_{res}}, s_{in} i(t) \oplus J_{p_{in}} \right] \quad (1)$$

όπου η συνάρτηση F αντιπροσωπεύει ολόκληρη την οπτική διάταξη, δηλαδή δέχεται ως δύο ορίσματα τους κωδικοποιημένους πίνακες που αντιστοιχούν στην είσοδο και στην κατάσταση της δεξαμενής, τους στέλνει στο SLM και επιστρέφει την επόμενη κατάσταση της δεξαμενής που ανιχνεύεται από την κάμερα. Το σύμβολο $\oplus$ αναφέρεται στο γινόμενο Kronecker και $J_{p_{in/res}}$

αναφέρεται στον all-ones πίνακα με $p_{in/res}$ αριθμό γραμμών και στηλών προκειμένου να διασφαλιστεί η κωδικοποίηση μακροεικονοστοιχείων του SLM. Το μαθηματικό πλαίσιο που περιγράφει το οπτικό αυτό δίκτυο είναι πολύ παρόμοιο με το συμβατικό δίκτυο δεξαμενής.

Συμπερασματικά, προτείνεται ένα οπτικό δίκτυο υπολογισμού δεξαμενών που μπορεί να εκτελέσει, για πρώτη φορά από όσο γνωρίζουμε, προβλέψεις σε μεγάλα πολυδιάστατα χαοτικά σύνολα δεδομένων. Χρησιμοποιούμε την εξίσωση Kuramoto-Sivashinsky ως παράδειγμα χωροχρονικού χαοτικού συστήματος. Τέλος, αποδεικνύεται πειραματικά ότι το μελετούμενο οπτικό δίκτυο μπορεί να κλιμακωθεί σε ένα εκατομμύριο κόμβους. Ο χρόνος υπολογισμού του αυξάνεται γραμμικά μόνο όταν αυξάνεται ο αριθμός των κόμβων, λόγω των ηλεκτρονικών επιβαρύνσεων, ενώ η ταχύτητα του οπτικού μέρους (του πολλαπλασιασμού του πίνακα) είναι ανεξάρτητη από το μέγεθος της δεξαμενής και δεν απαιτεί αποθήκευση στη μνήμη.

Τα αποτελέσματα, τα οποία είναι πολύ δύσκολο να επιτευχθούν από συμβατικές μηχανές von Neumann, ανοίγουν την προοπτική να επιτευχθούν προβλέψεις σε πολύ μεγάλα σύνολα δεδομένων πρακτικού ενδιαφέροντος, όπως η τύρβη, με υψηλή ταχύτητα και χαμηλή κατανάλωση ενέργειας

5.5.2.5 Διαχείριση Ενέργειας

Η εφαρμογή της δεξαμενής στη διαχείριση ενέργειας αποκτά μεγάλη σημασία στον σύγχρονο κόσμο, όπου η βιώσιμη και αποδοτική χρήση των πόρων είναι ζωτικής σημασίας για την προστασία του περιβάλλοντος και την εξασφάλιση της ενεργειακής ασφάλειας. Η δεξαμενή, ως αναδραστικό νευρωνικό δίκτυο, παρέχει ενδιαφέρουσες δυνατότητες για τη βελτιστοποίηση της διαχείρισης της ενέργειας σε διάφορα συστήματα. Αναλυτικότερα:

- Ηλεκτρικά Δίκτυα

Στα ηλεκτρικά δίκτυα, η δεξαμενή μπορεί να χρησιμοποιηθεί για τη βελτιστοποίηση της λειτουργίας τους. Τα ηλεκτρικά δίκτυα αντιμετωπίζουν προκλήσεις όπως η διακύμανση του φορτίου, οι αλλαγές στη γεωγραφία των πηγών παραγωγής, και η ανάγκη για αποδοτικότερη κατανομή ενέργειας. Με τη χρήση της δεξαμενής, μπορούν να αναπτυχθούν μοντέλα πρόβλεψης φορτίου, προκειμένου να προσαρμόζεται η παραγωγή ενέργειας στις διακυμάνσεις της ζήτησης.[63] Επιπλέον, η δεξαμενή μπορεί να χρησιμοποιηθεί για την πρόβλεψη διακυμάνσεων στο δίκτυο, προσφέροντας έτσι τη δυνατότητα για αποτελεσματικότερη ρύθμιση της παραγωγής.[69]

- Ενεργειακά Συστήματα

Σε ενεργειακά συστήματα όπως ανανεώσιμες πηγές ενέργειας (ηλιακά πάρκα, αιολικά πάρκα), η δεξαμενή μπορεί να χρησιμοποιηθεί για την πρόβλεψη της παραγωγής ενέργειας από αυτές τις πηγές.[76] Αυτό μπορεί να βοηθήσει στην αποτελεσματική διαχείριση της παραγωγής ενέργειας και την ομαλή ενσωμάτωση των ανανεώσιμων πηγών στο ενεργειακό σύστημα.[80]

Στον τομέα αυτό μελετώντας το άρθρο “[Reservoir Computing Approaches Applied to Energy Management in Industry](#)” των V Colla, I Matino, S Dettori, βλέπουμε ότι προτείνεται μια νέα εφαρμογή αυτής της προσέγγισης για τη μοντελοποίηση βιομηχανικών διεργασιών. Τα μοντέλα που αναπτύχθηκαν αποτελούν μέρος ενός σύνθετου συστήματος για τη βελτιστοποίηση της εκμετάλλευσης των απερίων της διεργασίας σε μια ολοκληρωμένη χαλυβουργία. Παρουσιάζονται και συζητούνται δύο μοντέλα, στα οποία εφαρμόζονται τόσο ρηχά νευρωνικά δίκτυα Echo-State (shallow Echo-State Neural Networks) όσο και βαθιά νευρωνικά δίκτυα Echo-State (Deep Echo State Neural networks).

Στη συνέχεια παρουσιάζονται ορισμένα υποδειγματικά μοντέλα βιομηχανικών διεργασιών, τα οποία δείχνουν την αποτελεσματικότητα της προτεινόμενης προσέγγισης σε βιομηχανικές εφαρμογές. Ειδικότερα, η διεργασία BOF έχει μελετηθεί σε βάθος με διάφορους στόχους μοντελοποίησης:

- Μοντέλο 1: πρόβλεψη της παραγωγής BOFG
- Μοντέλο 2: πρόβλεψη της παραγωγής ατμού με ανάκτηση της θερμότητας BOFG.

Προτείνεται ένα μοντέλο με βάση το ESN που προβλέπει την ανακτήσιμη ποσότητα BOFG (δηλαδή τον όγκο και τη θερμαντική ισχύ) και παρέχει μια ποιοτική ιδέα των ροών όγκου και της συνολικής θερμαντικής ισχύος BOFG που μπορούν να εκτοξευθούν με συχνότητα 1 λεπτού σε χρονικό ορίζοντα 2 ωρών.

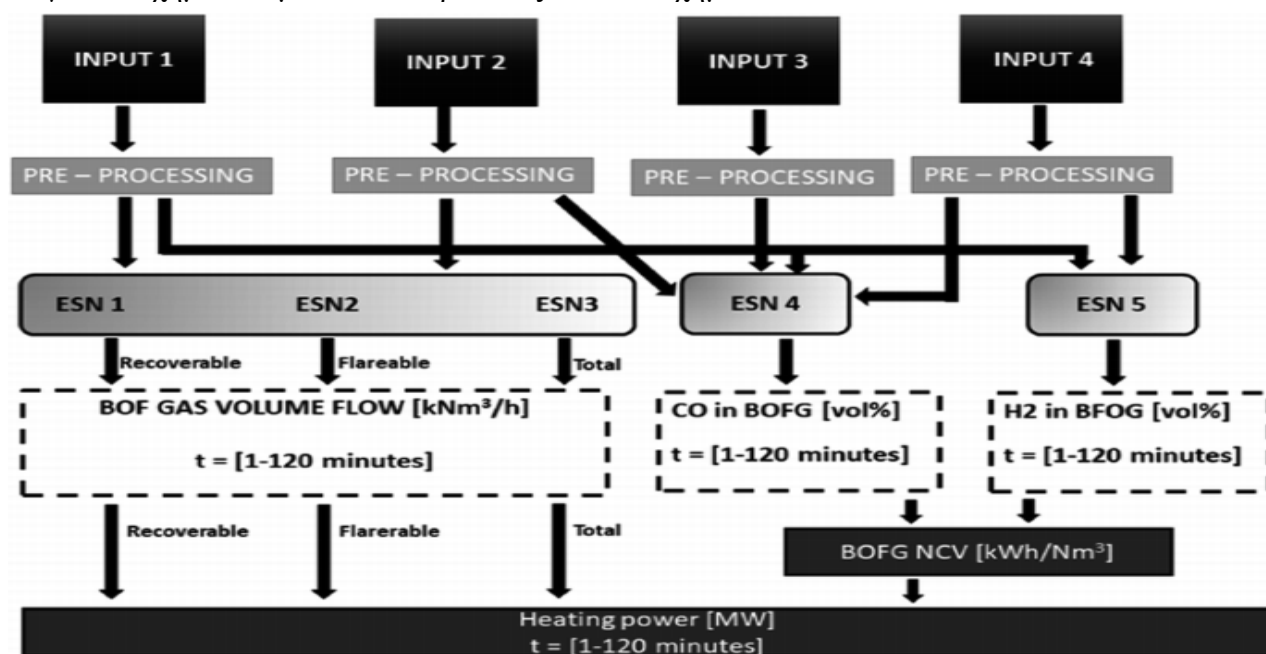
Το ποιοτικό χαρακτηριστικό της εκτίμησης των ροών όγκου και της ισχύος θέρμανσης του αναφλέξιμου και του συνολικού BOFG επηρεάζεται από την απόδοση των δεδομένων που σχετίζονται με το αέριο που καίγεται, η οποία δεν είναι συγκρίσιμη με την απόδοση των δεδομένων που σχετίζονται με το ανακτημένο αέριο και δεν επιτρέπει την εκπαίδευση του

μοντέλου με καλή απόδοση για τις εν λόγω εξόδους. Επιπλέον, προκειμένου να διασφαλιστεί η σωστή εκμετάλλευση του αερίου BOFG, η γνώση της συμπεριφοράς του ανακτήσιμου αερίου σε σχέση με το αέριο που μπορεί να καεί στον πυρσό (μέρος του αερίου που πρέπει να καεί στον πυρσό λόγω των περιορισμών της διαδικασίας και της ασφάλειας) είναι πιο ενδιαφέροντα.

Το μοντέλο περιλαμβάνει πέντε διαφορετικά ESN που πρέπει να ρυθμιστούν ξεχωριστά, προκειμένου να προβλέπουν ανεξάρτητα την ανακτήσιμη, την αναφλέξιμη και τη συνολική περιεκτικότητα του αερίου σε BOFG, CO και H_2 . Κάθε δίκτυο είναι εξειδικευμένο και, ως εκ τούτου, ο υπολογιστικός φόρτος στη φάση της εκπαίδευσης είναι μέτριος. Η πρόβλεψη της περιεκτικότητας σε CO και H_2 στο BOFG αξιοποιείται για την εκτίμηση του NCV σύμφωνα με την ακόλουθη εξίσωση:

$$NCV = \frac{(0.0789 \cdot CO + 0.0672 \cdot H_2) \cdot 1000}{22.4} \text{ kWh/Nm}^3 \quad (8)$$

Η επιτευχθείσα τιμή NCV και ο όγκος BOFG επιτρέπουν τον υπολογισμό της ισχύος θέρμανσης. Το γενικό σχήμα του μοντέλου παρουσιάζεται στο σχήμα.



Σχήμα του μοντέλου που αναπτύχθηκε για την πρόβλεψη της παραγωγής BOFG

Οι εισροές του μοντέλου, οι οποίες επιλέχθηκαν μέσω ανάλυσης VS και πλεονασμού, αναφέρονται στον πίνακα 1 με τη σχετική συστάδα και άλλες χρήσιμες πληροφορίες. Αναφέρονται κυρίως σε υλικά και οξυγόνο που τροφοδοτούνται στο BOF. Εξετάζεται επίσης η θέση της κινητής ποδιάς, η οποία είναι μια λειτουργική μεταβλητή που επηρεάζει την τροφοδοσία αέρα της διεργασίας.

Variable	Meas. unit	Input cluster	Note
Position of moveable skirt	mm	Input 1	t = 0
Blowed Oxygen Volume Flow	kNm ³ /h	Input 1	t = 0
Recovered BOFG Vol. Flow	kNm ³ /h	Input 1	t = 0
Flared BOFG Volume Flow	kNm ³ /h	Input 1	t = 0
BOF Plant Unit State	Boolean	Input 1	t = 0....120
Charged Liquid Pig iron	t	Input 1	t = 0....120
Charged Lime	t	Input 2	t = 0....120
CO content in BOFG	vol.%,	Input 3	t = 0
H ₂ content in BOFG	vol.%,	Input 4	t = 0
Recycled briquetted sinter	t	Input 4	t = 0....120

Πίνακας 1. Μεταβλητές εισόδου των ESN για την πρόβλεψη της παραγωγής BOFG (συχνότητα 1 λεπτό, t = 0 σημαίνει τρέχουσα τιμή, t = 0 ...120 σημαίνει όλες τις τιμές των επόμενων 2 ωρών).

Οι περισσότερες εισροές είναι κοινές μεταξύ των διαφόρων ESN, ενώ ορισμένες εισροές που σχετίζονται με τα φορτισμένα υλικά τροφοδοτούνται μόνο σε ορισμένα ESN. Η μεταβλητή με την ονομασία "BOF Plant Unit States" είναι ένα είδος Boolean αναπαράστασης του προγραμματισμού της διεργασίας: όταν κατέχει μοναδιαία τιμή, ο μετατροπέας βρίσκεται σε λειτουργία (έναρξη της εμφύσησης του οξυγόνου), διαφορετικά δεν πραγματοποιείται εμφύσηση οξυγόνου.

Το δεύτερο μοντέλο προβλέπει τη συνολική παραγωγή ατμού που σχετίζεται με τον προγραμματισμό της διεργασίας BOF. Το μοντέλο έχει περίοδο πρόβλεψης 2 ωρών μπροστά, ξεκινώντας από τις πιο συσχετιζόμενες μεταβλητές. Ο κατάλογος των εισροών παρουσιάζεται στον πίνακα 2. Οι μεταβλητές αυτές επηρεάζουν τη μεταφορά της θερμότητας που αξιοποιείται για την παραγωγή ατμού. Η δυναμική της παραγωγής ατμού μοντελοποιείται μέσω της χρήσης της νέας αρχιτεκτονικής που βασίζεται στο DESN.

Variable	Measurement unit	Note
Position of moveable skirt	mm	t = 0
Blowed Oxygen Volume Flow	kNm ³ /h	t = 0
Recovered BOFG Volume Flow	kNm ³ /h	t = 0
Flared BOFG Volume Flow	kNm ³ /h	t = 0
BOF Plant Unit State	Boolean	t = 0....120
Production of steam in the BOF boiler	t/h	t = 0

Πίνακας 2. Μεταβλητές εισόδου για την πρόβλεψη της παραγωγής ατμού στη διαδικασία BOG.

Πριν από την εκπαίδευση και την προσομοίωση των μοντέλων, αφαιρούνται οι ακραίες τιμές. Όλες οι παρουσιαζόμενες αρχιτεκτονικές με βάση το ESN και το DESN σχεδιάζονται σύμφωνα με μια κοινή προσέγγιση, η οποία μπορεί να συνοψιστεί ως εξής:

- ➔ Η σύνδεση των μονάδων δεξαμενής είναι χαμηλή, με αραιότητα μεταξύ 2-4%
- ➔ Οι συναρτήσεις ενεργοποίησης της δεξαμενής και της ανάγνωσης είναι, αντίστοιχα, η υπερβολική εφαπτομένη και μια γραμμική συνάρτηση
- ➔ Οι συντονισμένες υπερ-παράμετροι είναι: αριθμός νευρώνων της δεξαμενής, φασματική ακτίνα, κλιμάκωση της εισόδου και του δασκάλου (δηλαδή των στόχων της φάσης εκπαίδευσης), επίπεδο θορύβου, μετατόπιση εισόδου και δασκάλου, παράμετρος κανονικοποίησης για την παλινδρόμηση κορυφογραμμής της ανάγνωσης (φάση εκπαίδευσης). Οι αναφερόμενες υπερ-παράμετροι συντονίστηκαν ευρετικά με στόχο την

ελαχιστοποίηση του κανονικού μέσου τετραγωνικού σφάλματος ρίζας (Normal Root Mean Square Error - NRMSE) στο σύνολο δεδομένων επικύρωσης, το οποίο ορίζεται ως εξής:

$$\text{NRMSE} = 100 \frac{\sqrt{\frac{1}{N_s} \sum_{i=1}^{N_s} (\bar{y}_i - y_i)^2}}{\max(y) - \min(y)}, \text{ όπου } N_s \text{ είναι ο αριθμός των δειγμάτων του συνόλου}$$

δεδομένων, y_i και $\bar{y}_i$ είναι αντίστοιχα το i -οστό δείγμα του στόχου και της προβλεπόμενης εξόδου.

→ Ο αλγόριθμος εκπαίδευσης είναι μια offline παλινδρόμηση κορυφογραμμής της ανάγνωσης,

Όλα τα μοντέλα έχουν εκπαιδευτεί και επικυρωθεί με τη χρήση πραγματικών δεδομένων που προέρχονται από βιομηχανικές εγκαταστάσεις. Μεγάλα σύνολα δεδομένων αποκτήθηκαν με χρόνο δειγματοληψίας 1 λεπτό και χωρίστηκαν σε σύνολα εκπαίδευσης και επικύρωσης, τα οποία περιείχαν, αντίστοιχα, το 70% και το 30% των διαθέσιμων δεδομένων. Στην περίπτωση του πρώτου μοντέλου (παραγωγή BOFG), το σύνολο δεδομένων αποτελείται από 450000 δεδομένα που προέρχονται από 10 μήνες παραγωγής. Για το δεύτερο μοντέλο το σύνολο δεδομένων αποτελείται από 252000 δεδομένα που σχετίζονται με 6 μήνες παραγωγής.

Για να αξιολογηθούν οι επιδόσεις και να βρεθεί η καλύτερη λύση για τη μοντελοποίηση των διαφόρων συμπεριφορών της διαδικασίας, δοκιμάστηκαν τόσο η ρηχή όσο και η αρχιτεκτονική DESN. Το NRMSE και ο χρόνος εκπαίδευσης λαμβάνονται υπόψη για την αξιολόγηση της αποδοτικότητας.

Η χρήση DESN στο μοντέλο BOFG παρέχει σχεδόν ταυτόσημη τιμή του NRMSE. Και στις δύο περιπτώσεις, ο χρόνος εκπαίδευσης είναι λογικός, αλλά το DESN απαιτούσε σχεδόν το 75% του χρόνου εκπαίδευσης σε σχέση με τα ρηχά ESN. Από την άλλη πλευρά, ο υπολογιστικός φόρτος του DESN σε αυτή τη μελέτη περίπτωσης είναι υψηλότερος από αυτόν του ρηχού ESN. Επίσης, τα αποτελέσματα επικύρωσης για το δεύτερο μοντέλο είναι πολύ ενθαρρυντικά. Το σφάλμα είναι χαμηλό κατά τα πρώτα λεπτά της πρόβλεψης και, όπως είναι συνήθως αναμενόμενο, τείνει να αυξάνεται, καθώς θέλουμε να προβλέψουμε το φαινόμενο στο απώτερο μέλλον. Τόσο το ρηχό ESN όσο και το DESN παρουσιάζουν πολύ καλές επιδόσεις και για την πρόβλεψη 2 ωρών στο μέλλον, αλλά σε αυτή την περίπτωση το μοντέλο που βασίζεται στο DESN αποδεικνύεται η καλύτερη λύση τόσο από άποψη ακρίβειας όσο και από άποψη χρόνου υπολογισμού.

Συνολικά, η δεξαμενή συνιστά ένα πολύτιμο εργαλείο για την αντιμετώπιση των προκλήσεων της διαχείρισης ενέργειας σε σύνθετα συστήματα. Η δυνατότητα της δεξαμενής να προβλέπει και να προσαρμόζει την παραγωγή ενέργειας αποδεικνύεται αποφασιστική για την επίτευξη βιώσιμων, αποδοτικών και αξιόπιστων ενεργειακών συστημάτων.

5.5.2.6 Δεδομένα IoT και Έξυπνες Πόλεις

Η διάδραση ανάμεσα στην τεχνολογία του Διαδικτύου των Πραγμάτων (IoT) και την δεξαμενή παρέχει σημαντικές δυνατότητες για τη δημιουργία και τη λειτουργία έξυπνων πόλεων, οι οποίες αξιοποιούν τα δεδομένα που παράγονται από διάφορες συσκευές και αισθητήρες για τη

βελτίωση της ποιότητας ζωής των κατοίκων και την αποτελεσματικότερη διαχείριση των αστικών υποδομών.[69][65][76]

Οι έξυπνες πόλεις βασίζονται στη συλλογή, ανάλυση και αξιοποίηση μεγάλου όγκου δεδομένων που προέρχονται από διάφορες πηγές όπως αισθητήρες, κινητά τηλέφωνα, αυτοκίνητα, κτίρια και πολλές άλλες συσκευές που συνδέονται στο διαδίκτυο.[68][81][78] Σε αυτό το πλαίσιο, η δεξαμενή ως μοντέλο επεξεργασίας δεδομένων μπορεί να αποτελέσει ένα ισχυρό εργαλείο για την ανάλυση και την αναγνώριση μοτίβων σε αυτά τα δεδομένα, καθώς και για την πρόβλεψη των τάσεων και των αναγκών της πόλης.[63][64]

Ένα παράδειγμα εφαρμογής είναι η πρόβλεψη της κίνησης και της κυκλοφορίας σε μια έξυπνη πόλη. Με τη χρήση αισθητήρων κίνησης που τοποθετούνται σε διάφορα σημεία της πόλης, συλλέγονται δεδομένα για την κίνηση των οχημάτων. Η δεξαμενή μπορεί να εκπαιδευτεί να αναγνωρίζει τα μοτίβα της κίνησης, να αντλεί σημαντικές πληροφορίες για τις συνηθισμένες ώρες αιχμής, τις συνήθεις διαδρομές και τα μοτίβα εναλλαγής της κίνησης. Στη συνέχεια, μπορεί να χρησιμοποιηθεί για την πρόβλεψη της κίνησης σε συγκεκριμένες ώρες και περιοχές, βοηθώντας έτσι τους κατοίκους να επιλέγουν βέλτιστες διαδρομές και τρόπους μετακίνησης.[67][80]

Πιο συγκεκριμένα ένα ακόμα παράδειγμα για τον τομέα αυτό που μελετήθηκε είναι από το άρθρο “[Smart city infrastructure protection: real-time threat detection employing online reservoir computing architecture](#)” των L Gao, X Deng, W Yang. Η παρούσα εργασία παρουσιάζει την υλοποίηση μιας εξειδικευμένης αρχιτεκτονικής online υπολογισμού δεξαμενών για την προστασία των υποδομών έξυπνων πόλεων, η οποία έχει χαμηλές απαιτήσεις σε υπολογιστικούς πόρους, είναι αποτελεσματική και κατάλληλη για ανάλυση ροής δεδομένων σε πραγματικό χρόνο. Πιο συγκεκριμένα, περιγράφει την ανάπτυξη ενός δικτύου ηχητικής κατάστασης, το οποίο αποτελείται από αναλογικούς νευρώνες με αραιές τυχαίες συνδέσεις στα επίπεδα εισόδου και στη δυναμική δεξαμενή. Η εκπαίδευσή του στο επίπεδο εξόδου πραγματοποιείται με την αναδρομική μέθοδο ελαχίστων τετραγώνων. Για τη δοκιμή του προτεινόμενου μοντέλου επιλέχθηκε ένα σύνθετο σύνολο δεδομένων, το οποίο προσομοιώνει πλήρως τις ψηφιακές επιθέσεις που μπορεί να αντιμετωπίσει ο μηχανοτρονικός εξοπλισμός που χρησιμοποιείται στα έξυπνα δίκτυα ύδρευσης που βρίσκονται στο μέτωπο των έξυπνων πόλεων.

Χρησιμοποιήθηκε υπολογισμός δεξαμενής επεξεργασίας ροής σε πραγματικό χρόνο με ESN, προκειμένου να μοντελοποιηθεί το πρόβλημα της ανίχνευσης επιθέσεων σε ένα έξυπνο δίκτυο ύδρευσης. Η προσπάθεια αυτή απαιτεί τη χρήση συγκεκριμένων παραμέτρων που σχετίζονται με το πρόβλημα, δηλαδή το sliding windows (SLWi), το στρώμα εισόδου, το dynamical reservoir DR, το στρώμα ανάγνωσης, τον προσδιορισμό της προσέγγισης εκπαίδευσης του ESN, καθώς και τον προσδιορισμό της διαδικασίας αναβάθμισης των νευρώνων τόσο στο DR όσο και στο στρώμα ανάγνωσης.

Αρχικά, για τον προσδιορισμό του SLWi χρησιμοποιήθηκαν έννοιες και τεχνικές από τα δυναμικά συστήματα. Πιο συγκεκριμένα, χρησιμοποιήθηκε το μοντέλο πρόβλεψης που εκφράζεται από την ακόλουθη εξίσωση:

$$x(t + 1) = f(x(t)) \quad (1)$$

Η τιμή της ροής τη στιγμή $t + 1$ εξαρτάται από το ακόλουθο διάνυσμα που περιέχει τις μεταβλητές υστέρησης.

$$x(t) = (x(t), x(t - 1\tau), x(t - 2\tau), \dots, x(t - (n - 1)\tau)) \quad (2).$$

Υποθέτοντας ότι τα δεδομένα προέρχονται από ένα άγνωστο δυναμικό σύστημα n διαστάσεων, ο στόχος είναι να προσδιοριστεί ένα απλούστερο σύστημα από το οποίο θα

μπορούσαν να περιγραφούν τα δεδομένα αυτά. Με βάση το θεώρημα ενσωμάτωσης, ένα διάνυσμα μεταβλητών υστέρησης με αρκετά μεγάλη διάσταση m (τουλάχιστον διπλάσια από το μέγεθος του πραγματικού συστήματος) μπορεί να αναπαραστήσει επιτυχώς τακτικά καθορισμένες ροές δεδομένων με χρονολογική σειρά. Εάν τα δεδομένα είναι ανεξάρτητα, η διάσταση υστέρησης m , μπορεί να υποδηλώνει το μέγεθος σε SLWi, έτσι ώστε να μπορούν να μετασχηματιστούν σε m -διάστατα διανύσματα.

Ο μετασχηματισμός αυτός επιτυγχάνεται με την επαναληπτική δειγματοληψία των παρατηρήσεων ροής με ρυθμό τ , όπου τ δηλώνει το πολλαπλάσιο των υστερήσεων μεταξύ των δειγματοληψιών που εξασφαλίζει επαρκή αριθμό ανεξάρτητων παρατηρήσεων. Εάν η ροή έχει υπερδειγματοληψία, τότε για να αφαιρεθεί ο "θόρυβος" θεωρείται μια τιμή για το τ μεγαλύτερη από 1. Διαφορετικά, τίθεται $\tau = 1$. Έτσι, η επόμενη τιμή της μετασχηματισμένης ροής θα είναι:

$$\overline{x}_{t+1} = f(x_t, x_{t-r}, \dots, x_{t-(m-1)r}) \quad (3)$$

Επομένως, για τον ορισμό του μεγέθους του SLWi, πρέπει να αποφασίσουμε για τη διάσταση m της βύθισης και για τη χρονική καθυστέρηση τ που θα χρησιμοποιηθεί στη δειγματοληψία από τις παρατηρήσεις.

Ο ορισμός της διάστασης βύθισης m ακολούθησε την τεχνική των ψευδών πλησιέστερων γειτόνων (false nearest neighbors - FNN). Πρόκειται για έναν αλγόριθμο που επιτρέπει τον προσδιορισμό της ελάχιστης διάστασης εμβύθισης που είναι επαρκής για την αναπαράσταση ενός συστήματος. Η ιδέα είναι να διερευνηθεί πώς αλλάζουν τα γειτονικά σημεία ενός συγκεκριμένου σημείου, με βάση την αύξηση της διάστασης εμβύθισης από την τιμή m σε $m + 1$. Συγκεκριμένα, για μια διάσταση m , καταγράφονται τα σημεία που εμφανίζονται εσφαλμένα ως γειτονικά και στη συνέχεια η διάσταση αυξάνεται σε $m+1$. Όταν το ποσοστό των FNN αυξάνεται ή παραμένει αμετάβλητο, τότε επιλέγεται η m για τη διάσταση εμβύθισης. Η διαδικασία εφαρμογής του αλγορίθμου στην υπό εξέταση περίπτωση περιλαμβάνει τα ακόλουθα βήματα:

1. Δεδομένης μιας χρονοσειράς x_t , η παράμετρος m τίθεται σε 1 και στη συνέχεια σχηματίζονται m -διάστατα διανύσματα υστέρησης x_m .
2. Οι αποστάσεις μεταξύ κάθε σημείου x_i και των k πλησιέστερων γειτόνων του x_k υπολογίζονται με βάση την ακόλουθη συνάρτηση:

$$D(x_i, x_k) = \sqrt{\sum_{j=1}^m (x_{i+(j-1)\tau} + x_{k+(j-1)\tau})^2}$$

3. Η τιμή του m αυξάνεται σε $m+1$ και η τιμή ή R προσαρμόζεται με βάση την παρακάτω εξίσωση:

$$R(x_i, x_k, m + 1) = \sqrt{\sum_{j=1}^{m+1} (x_{i+(j-1)\tau} + x_{k+(j-1)\tau})^2} = \sqrt{R(x_i, x_k, m)^2 + (x_{i+m_\tau} - x_{k+m_\tau})^2}$$

όπου τα x_i και x_k είναι FNN.

Μετά την εφαρμογή του αλγορίθμου FNN, η ελάχιστη τιμή της επαρκούς διάστασης εμβύθισης για την αναπαράσταση του προτεινόμενου συστήματος ορίστηκε ίση με 14.

Αντίστοιχα, η τεχνική επαναδειγματοληψίας με βάση την αμοιβαία πληροφορία (RMI) χρησιμοποιήθηκε για τον προσδιορισμό της χρονικής υστέρησης τ . Η αμοιβαία πληροφορία $I_\tau(t)$ μπορεί να θεωρηθεί ανάλογη με τη συνάρτηση συσχέτισης, αλλά σε μη γραμμικό πλαίσιο. Στόχος

ήταν να επιτευχθεί ο κατάλληλος ρυθμός δειγματοληψίας τ , ώστε να επιτευχθεί η χαμηλότερη μέση αμοιβαία πληροφορία μεταξύ των μεταβλητών. Το πρώτο τοπικό ελάχιστο της $I_{\epsilon}(t)$ είναι ο κατάλληλος ρυθμός δειγματοληψίας τ , αν και είναι σημαντικό να διατηρηθεί μια ισορροπία μεταξύ της χαμηλής τιμής για το τ και της ανεξαρτησίας των μεταβλητών. Η ελάχιστη τιμή του $I_{\epsilon}(t)$ προσδιορίστηκε μετά από 51 υστερήσεις και επομένως η διαδικασία δειγματοληψίας στη χρονοσειρά έπρεπε να εκτελείται κάθε 51 τιμές. Έτσι, η χρονοσειρά υλοποιήθηκε για παράθυρο χρονικής υστέρησης SLWi 51 ms, χρησιμοποιώντας ένα κομμάτι ροής 14 εγγραφών. Στην πραγματικότητα, η ανάπτυξη ολόκληρου του όγκου των 946.722 εγγραφών της χρονοσειράς, απαιτούσε τη χρήση 67.623 τμημάτων ροής, με Sliding Windows 51 ms. Ως εκ τούτου, η διαδικασία αξιολόγησης του συνόλου δεδομένων ολοκληρώθηκε σε 3.448.773 ms ή σε 57,4796 λεπτά.

Η ανάπτυξη του ESN ακολούθησε μια διαδικασία εκπαίδευσης με επίβλεψη, όπου για κάθε είσοδο $u(n) \in R^N$ γνωρίζουμε το επιθυμητό αποτέλεσμα $y^{target}(n) \in R^N$. Η μεταβλητή n αντιπροσωπεύει τον διακριτό χρόνο και παίρνει τιμές στο διάστημα $[1, T]$, όπου T ο αριθμός των δεδομένων εισόδου στο σύνολο εκπαίδευσης. Μια είσοδος αντιπροσωπεύει μια εγγραφή του συνόλου δεδομένων, δηλαδή μετρήσεις από ενεργοποιητές και αισθητήρες. Σε κάθε χρονική στιγμή, το n είναι ένα διάνυσμα με 51 τιμές, όσες και οι μετρήσεις κάθε εγγραφής. Η επιθυμητή έξοδος $y^{target}(n)$ και η πραγματική έξοδος του δικτύου $y(n)$ είναι διανύσματα 2 τιμών που αντιπροσωπεύουν την κατάσταση που περιγράφεται από τις τιμές των ενεργοποιητών και των αισθητήρων, δηλαδή κανονική λειτουργία ή επίθεση. Ο σκοπός της εκπαίδευσης του δικτύου είναι να διδάξει ένα μοντέλο με έξοδο $y(n) \in R^N$. Οι τιμές του $y(n)$ θα πρέπει να είναι όσο το δυνατόν πιο κοντά στο $y^{target}(n)$, προκειμένου να επιτευχθεί το ελάχιστο σφάλμα $E(y, y^{target})$. Με αυτόν τον τρόπο, θα είναι σε θέση να γενικεύει και να ταξινομεί σωστά χρησιμοποιώντας δεδομένα που βλέπει για πρώτη φορά. Ως συνάρτηση σφάλματος χρησιμοποιήθηκε το μέσο τετραγωνικό σφάλμα (Root-mean-square error - RMSE). Επίσης, πρέπει να σημειωθεί ότι οι νευρώνες επιπέδου εισόδου είναι ουσιαστικά ανενεργοί εφόσον δεν εκτελούν υπολογισμούς. Ο μόνος σκοπός τους είναι να μεταδίδουν την είσοδο του δικτύου στο DR.

Η επόμενη εξίσωση χρησιμοποιήθηκε για τη διαδικασία αναβάθμισης των τιμών των νευρώνων στο DR:

$$\bar{x}(n) = \tanh(W^{in}[1; u(n)] + Wx(n-1)) \quad \text{όπου το } \bar{x}(n) \in R^N \text{ καθορίζει τις}$$
 αναβαθμισμένες τιμές για κάθε νευρώνα DR. Για την αναβάθμιση χρησιμοποιήθηκε η γνωστή εφαπτομενική υπερβολική συνάρτηση ($\tanh$), $u(n)$ είναι η είσοδος τη στιγμή n και η τιμή της προκατάληψης είναι ίση με 1.

Η τελική τιμή που θα λάβουν οι νευρώνες στο DR εκτιμάται από την επόμενη εξίσωση:

$$x(n) = (1 - \alpha)x(n-1) + \alpha\bar{x}(n)$$

όπου $x(n) \in R^N$ είναι ένα διάνυσμα που περιέχει τους νευρώνες του DR, b είναι η ολοκλήρωση διαρροής και $\bar{x}(n) \in R^N$ είναι η τιμή αναβάθμισης που προέκυψε από την προηγούμενη συνάρτηση 7. Η εφαπτομενική υπερβολική συνάρτηση $\tanh(*)$ εφαρμόστηκε σε κάθε τιμή των διανυσμάτων για ένα χρονικό σημείο n . Τα βάρη $W^{in} \in R^{N_x(1+N_u)}$ και $W \in R^{N_x \cdot N_x}$ αποδίδονται στην είσοδο και στη δεξαμενή αντίστοιχα. Η μεταβλητή $\alpha \in (0, 1]$ είναι ο ρυθμός διαρροής. Όταν $\alpha = 1$, μπορούμε να αποφύγουμε να έχουμε διαρροή ολοκλήρωσης στη διαδικασία ενημέρωσης των νευρώνων και έτσι το $\bar{x}(n) = x(n)$.

Προκειμένου να προστατευτούν τα δεδομένα από το θόρυβο που μπορεί να εμφανιστεί κατά τα αρχικά στάδια της διαδικασίας, καθώς τα βάρη W^{in} και W που συμβάλλουν στις τιμές του $x(n)$ αρχικοποιούνται τυχαία, οι αρχικές επιδράσεις των νευρώνων $x(n)$ απορρίπτονται. Δηλαδή, δεν χρησιμοποιούνται για την εκπαίδευση των βαρών W^{out} . Έτσι, αποφεύγουμε την αυθαίρετη προσαρμογή των τιμών $x(n)$ για την εκπαίδευση και συγκεκριμένα αποφεύγουμε τις τιμές $x(n) = 0$, οι οποίες οδηγούν σε μη φυσιολογική κατάσταση εκκίνησης του δικτύου.

Οι νευρώνες του επιπέδου εξόδου $y(n) \in R^N$ καθορίζονται από το εσωτερικό γινόμενο των βαρών εξόδου $W_{out} \in R^{N_y(1+N_u+N_x)}$ και του διανύσματος που δημιουργείται από την ένωση της τιμής κατωφλίου, των διανυσμάτων $u(n) \in R^{N_{u(n)}}$ και $x(n) \in R^N$, όπως φαίνεται στην εξίσωση παρακάτω:

$$y(n) = W^{out}[1;u(n);x(n)]$$

Αντίστοιχα, η μεταβολή των βαρών εξόδου W^{out} , με την οποία προσπαθούμε να τα μετατρέψουμε έτσι ώστε το αποτέλεσμα $y(n)$ του δικτύου να προσεγγίσει το επιθυμητό $y^{target}(n)$, δίνεται από την εξίσωση αυτή:

$$W^{out} = y^{target} X^T (X X^T + \beta I)^{-1}$$

Η μεταβλητή β είναι μια παράμετρος βελτιστοποίησης, που χρησιμοποιείται για την αποφυγή της υπερεκπαίδευσης. Η υλοποίηση του ESN έγινε με τη χρήση μιας προσέγγισης υπολογισμού δεξαμενής επεξεργασίας ροής σε πραγματικό χρόνο. Τα βάρη του δικτύου online μάθησης που αναπτύχθηκε, μεταβάλλονται ανά πάσα στιγμή, δηλαδή σε κάθε γραμμή εισόδου n των δεδομένων εκπαίδευσης.

Συγκεκριμένα, χρησιμοποιήσαμε τον αναδρομικό αλγόριθμο ελαχίστων τετραγώνων (recursive least square - RLS), ο οποίος λειτουργεί με βάση την ολοκλήρωση του ιστορικού σφαλμάτων, στους υπολογισμούς αναβάθμισης του δικτύου. Πιο συγκεκριμένα, ο αλγόριθμος RLS χρησιμοποιήθηκε για την προσαρμογή των βαρών του δικτύου W^{out} σε πραγματικό χρόνο, χρησιμοποιώντας ολισθαίνοντα παράθυρα. Ο αλγόριθμος περιέχει την παράμετρο παράγοντα λήθης λ , η οποία καθορίζει πόσο λιγότερο σημαντικό είναι εκθετικά το ιστορικό σφαλμάτων.

Για παράδειγμα, εάν η παράμετρος forgetting factor έχει την τιμή $\lambda = 1$, το ιστορικό σφάλματος έχει το ίδιο βάρος με το τρέχον σφάλμα του δικτύου. Εάν $\lambda < 1$, το ιστορικό σφάλματος επηρεάζει το δίκτυο χρονικά, δηλαδή το σφάλμα τη χρονική στιγμή n έχει μεγαλύτερη βαρύτητα από το σφάλμα τη χρονική στιγμή $n - 1$. Η συνάρτηση εκτίμησης σφάλματος του αλγορίθμου RLS περιγράφεται από την εξίσωση:

$$E(k) = \sum_{i=1}^k \lambda^{k-i} e(k)^2$$

Η παραπάνω συνάρτηση περιλαμβάνει την παράμετρο $e(k)$ η οποία σχετίζεται με τη διαφορά μεταξύ της επιθυμητής εξόδου y^{target} του δικτύου και της παραγόμενης y , για τη στιγμή k :

$$e(k) = y^{target}(k) - y(k)$$

Η συνάρτηση αναβάθμισης των βαρών RLS αλλάζει για κάθε χρονική στιγμή k :

$$W^{out}(k + 1) = W^{out}(k) + e(k)g(k)$$

όπου το $e(k)$ ορίζεται στην ακριβώς επάνω εξίσωση και $g(k)$ καθορίζεται από τη συνάρτηση DR που καθορίζει τη σημασία του ιστορικού σφάλματος στην ενημέρωση των βαρών:

$$g(k) = \frac{P(k-1)x(k)}{\lambda + x(k)^T P(k-1)x(k)},$$

όπου το $P(k-1)$ ορίζεται ως εξής:

$$P(k-1) = \lambda^{-1} P(k-1) - g(k)x(k)^T \lambda^{-1} P(k-1)$$

Πρόκειται για μια αναδρομική συνάρτηση που επιτρέπει την εξέταση του ιστορικού σφαλμάτων κατά την ενημέρωση των βαρών W^{out} . Επίσης, λ είναι ο παράγοντας λήθης και x οι νευρώνες του DR.

Η επικύρωση της απόδοσης του μοντέλου ESN είναι η καλύτερη μέθοδος για την παρακολούθηση της ικανότητας γενίκευσης της αρχιτεκτονικής του δικτύου. Η υπόθεση αυτή βασίζεται στο γεγονός ότι με την προσθήκη περισσότερων νευρώνων στο κρυφό στρώμα, το δίκτυο γίνεται ένας πολύ ευέλικτος χαρτογράφος συναρτήσεων. Αυτό με τη σειρά του αυξάνει τον κίνδυνο υπερπροσαρμογής. Το δίκτυο μπορεί να είναι σε θέση να χαρτογραφήσει τέλεια τα δεδομένα εκμάθησης, αλλά οι προβλέψεις σε δεδομένα δοκιμής μπορεί να είναι φτωχές. Η επικύρωση είναι με αυτή την έννοια όχι μόνο για να βρεθεί ο αριθμός των επαναλήψεων στη διαδικασία μάθησης, αλλά και μια πολύ σημαντική διαδικασία κατά την αξιολόγηση του αριθμού των κόμβων στα κρυφά στρώματα. Σε γενικές γραμμές, το προτεινόμενο ESN είναι ένα πολύ ευέλικτο σύστημα μοντελοποίησης.

Η σύνδεση, λοιπόν, της δεξαμενής με το IoT προσφέρει πολυδιάστατες δυνατότητες για τη βελτίωση της λειτουργίας των πόλεων, την αντιμετώπιση προκλήσεων και τη δημιουργία περιβαλλόντων που είναι πιο έξυπνα, αποδοτικά και βιώσιμα.

Κλείνοντας, ο υπολογισμός δεξαμενής προσφέρει ένα ισχυρό και καινοτόμο πλαίσιο για την εκμετάλλευση των δυνατοτήτων των αναδρομικών νευρωνικών δικτύων, αντιμετωπίζοντας παράλληλα τις προκλήσεις που σχετίζονται με την εκπαίδευσή τους. Αυτή η προσέγγιση έχει το δυναμικό να συνεχίσει να αναπτύσσεται και να βελτιώνει την απόδοσή της σε διάφορα πεδία της τεχνητής νοημοσύνης και της επεξεργασίας σημάτων. Η εξαιρετική απόδοσή του σε ποικίλα προβλήματα αποδεικνύει την αξία του στο να επιτρέπει την ανάπτυξη νέων εφαρμογών και την εξερεύνηση περαιτέρω των δυνατοτήτων της μηχανικής μάθησης.

6

Παραδείγματα Ανάλυσης

Ανωμαλιών με Reservoir

Computing

Στην ενότητα αυτή θα αναλυθούν λεπτομερώς δύο άρθρα, τα οποία παρουσιάζουν παραδείγματα εφαρμογής Reservoir Computing στον τομέα της ανάλυσης ανωμαλιών.

6.1 Reservoir Computing Meets Smart Grids: Attack Detection Using Delayed Feedback Networks

Το άρθρο “[Reservoir Computing Meets Smart Grids: Attack Detection Using Delayed Feedback Networks](#)” των Kian Hamedani, Lingjia Liu, Rachad Atat, Jinsong Wu και Yang Yi, παρουσιάζει μια νέα μέθοδο για την ανίχνευση επιθέσεων σε smart grids με γεννήτριες αιολικής ενέργειας με τη χρήση reservoir computing (RC). Το RC είναι ένα παράδειγμα ενεργειακά αποδοτικού υπολογισμού στον τομέα του νευρομορφικού υπολογισμού και η υλοποίηση των δικτύων delayed feedback networks (DFNs) του RC έχει δείξει ανώτερες επιδόσεις σε πολλές εργασίες ταξινόμησης. Ο συνδυασμός της χρονικής κωδικοποίησης, των DFN και ενός multilayer perceptron (MLP) ως στρώμα ανάγνωσης εξόδου αποδεικνύεται ότι αποδίδει βελτίωση των επιδόσεων σε σχέση με τις υπάρχουσες μεθόδους ανίχνευσης επιθέσεων, όπως τα MLP, οι support vector machines (SVM) και η συμβατική state vector estimation (SVE), όσον αφορά την ανίχνευση επιθέσεων σε smart grids. Οι προτεινόμενοι αλγόριθμοι αποδεικνύονται πιο ανθεκτικοί από τις MLP και SVE στην αντιμετώπιση διαφορετικών μεταβλητών, όπως το εύρος της επίθεσης, οι τύποι επίθεσης και ο αριθμός των παραβιασμένων μετρήσεων στα smart grids.

Η συγκομιδή ανανεώσιμης ενέργειας, ιδίως από πηγές όπως η ηλιακή και η αιολική ενέργεια, λαμβάνει σημαντική προσοχή λόγω των αυξανόμενων παγκόσμιων ενεργειακών απαιτήσεων και των πρόσφατων τεχνολογικών εξελίξεων. Οι τεχνολογίες αυτές έχουν τη δυνατότητα να τροφοδοτήσουν τα στοιχεία του smart grid, μειώνοντας το κόστος αντικατάστασης των

μπатаριών και συντήρησης των smart grids, ενώ παράλληλα συμβάλλουν σε ένα πιο πράσινο και βιώσιμο περιβάλλον. Για παράδειγμα, ένα ηλιακό πάνελ διαστάσεων 121 cm επί 53,6 cm ή μια ανεμογεννήτρια διαμέτρου 1 μέτρου υπό συνθήκες ταχύτητας ανέμου 8 m/s μπορεί να παράγει 100 Watt ηλεκτρικής ενέργειας. Ωστόσο, υπάρχουν μειονεκτήματα και προκλήσεις που σχετίζονται με τη συγκομιδή ενέργειας, καθιστώντας την πιο κατάλληλη ως συμπληρωματική πηγή ενέργειας παρά ως κύρια.

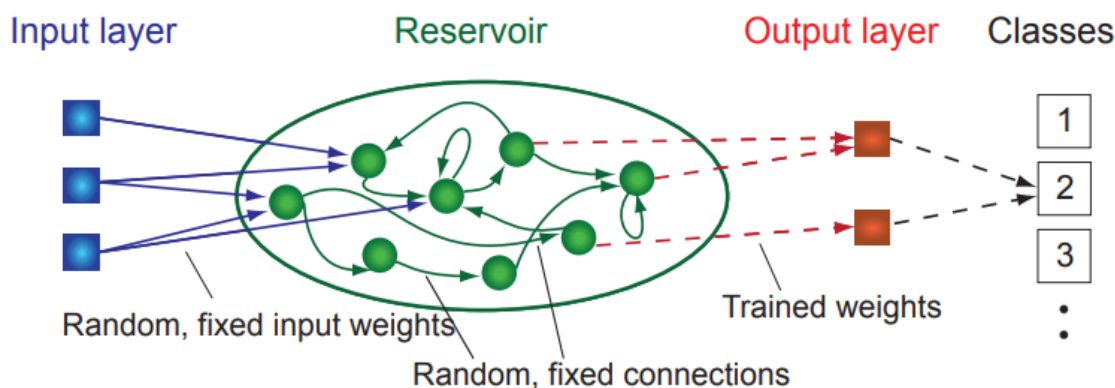
Η παρούσα εργασία επικεντρώνεται στη χρήση ανεμογεννητριών ως κύρια πηγή ηλεκτρικής ενέργειας για έξυπνα δίκτυα. Η ασφάλεια στον κυβερνοχώρο είναι ζωτικής σημασίας για τη διασφάλιση της αξιοπιστίας των smart grids, με την έγχυση ψευδών δεδομένων (false data injection - FDI) να αποτελεί μια ιδιαίτερα κρίσιμη απειλή. Η ανίχνευση επιθέσεων είναι ζωτικής σημασίας για τον μετριασμό των πιθανών ζημιών από FDI, και η εργασία διερευνά τη χρήση RNN για τον σκοπό αυτό. Τα παραδοσιακά RNN αντιμετωπίζουν δυσκολίες εκπαίδευσης, γεγονός που οδηγεί στην υιοθέτηση του RC, συγκεκριμένα των liquid state machines (LSM) και των echo state networks (ESN).

Αυτά τα συστήματα RC αποτελούνται από τρία επίπεδα: είσοδος (input), δεξαμενή (reservoir) και ανάγνωση/έξοδος (readout/output). Η δεξαμενή, που αποτελείται από τυχαία συνδεδεμένους νευρώνες, διατηρεί αμετάβλητες συνδέσεις κατά τη διάρκεια της εκπαίδευσης. Τα συστήματα RC έχουν δείξει ανώτερες επιδόσεις σε σύγκριση με τα παραδοσιακά RNN. Τα Delayed feedback networks (DFN) λειτουργούν επίσης ως συστήματα RC, προσφέροντας δυνατότητες βραχυπρόθεσμης δυναμικής μνήμης (short-term dynamic memory). Η εργασία χρησιμοποιεί interspike interval temporal encoding για τα συστήματα RC.

Η κύρια συνεισφορά της εργασίας έγκειται στην εισαγωγή των RC για την ανίχνευση επιθέσεων στα smart grids, αποδεικνύοντας την υπεροχή τους μέσω προσομοιώσεων. Ο εισαγόμενος σχεδιασμός δέχεται spike train input, προσφέρει ευρωστία έναντι των μεταβολών των επιθέσεων και χρησιμοποιεί ένα multilayer perceptron ως στρώμα ανάγνωσης για εργασίες ταξινόμησης, επιτυγχάνοντας μέσο ποσοστό ανίχνευσης επιθέσεων άνω του 99% σε προσομοιώσεις.

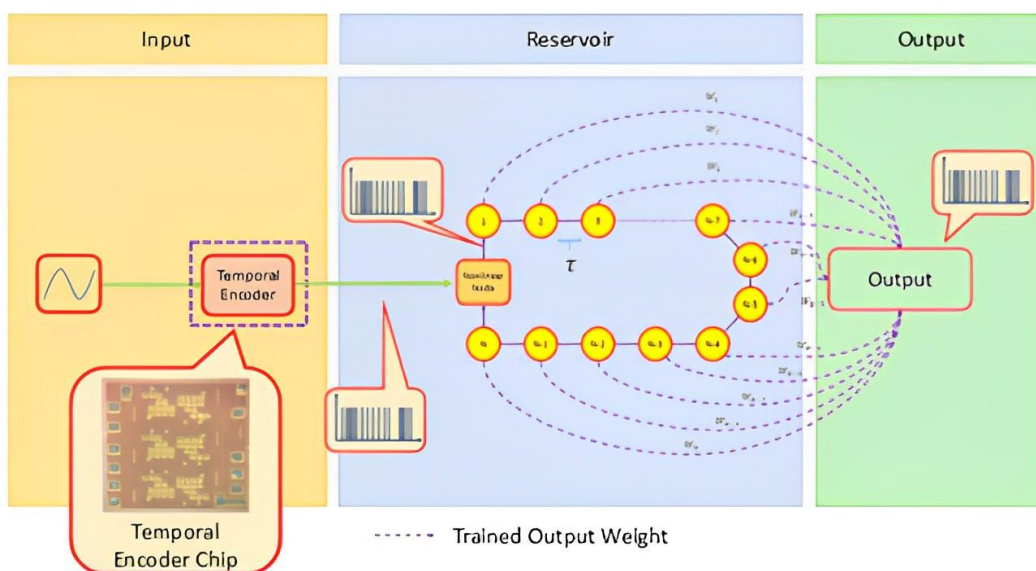
Συνοπτικά, η εργασία εξετάζει την αυξανόμενη σημασία της συγκομιδής ενέργειας από ανανεώσιμες πηγές για smart grids, τη χρήση ανεμογεννητριών για την παραγωγή ενέργειας και την εφαρμογή του RC για την ανίχνευση επιθέσεων, παρουσιάζοντας πολλά υποσχόμενα αποτελέσματα.

• REALIZING RC USING DFN



Σχήμα 1. RC System

Στο Σχήμα 1 παρουσιάζεται η δομή ενός συστήματος RC. Η μόνη διαφορά μεταξύ των παραδοσιακών μοντέλων RC, όπως το ESN και το LSM με το DFN είναι στο στρώμα της δεξαμενής. Όπως φαίνεται, το στρώμα δεξαμενής στα παραδοσιακά μοντέλα RC περιέχει νευρώνες, οι οποίοι είναι αραιά συνδεδεμένοι χρησιμοποιώντας αναδρομικές συνδέσεις, ωστόσο, στο DFN υπάρχει μόνο ένας μη γραμμικός κόμβος και η έξοδος ή η κατάσταση αυτού του μη γραμμικού κόμβου μετατοπίζεται χρονικά προκειμένου να παραχθούν οι καταστάσεις άλλων κόμβων ή εικονικών κόμβων.



Σχήμα 2. Hardware implementation of delayed feedback reservoir system.

Το Σχήμα 2 απεικονίζει τη δομή του DFN που χρησιμοποιείται στην παρούσα εργασία. Το πρώτο στρώμα είναι το στρώμα εισόδου στο οποίο εφαρμόζεται ο χρονικός κωδικοποιητής που χρησιμοποιείται. Οι λεπτομέρειες του χρονικού κωδικοποιητή εξηγούνται στην συνέχεια. Τα δεδομένα που χρησιμοποιούνται αποτελούνται από 10.000 διανύσματα των μετρήσεων που εξάγονται από το MATPOWER 5.1 [83]. Τα μισά των μετρήσεων προσβλήθηκαν από ένα τυχαίο διάνυσμα Gauss. Η διακύμανση της επίθεσης έχει οριστεί σε 0,05. Το διάνυσμα των συνδυασμένων δεδομένων που δέχθηκαν επίθεση και των δεδομένων που δεν δέχθηκαν επίθεση αποθηκεύονται και ο χρονικός κωδικοποιητής (temporal encoder) εφαρμόζεται στα δεδομένα. Για κάθε δείγμα σε αυτό το διάνυσμα, παράγεται ένα αντίστοιχο spiking train. Με αυτόν τον τρόπο, μπορούμε να μετατρέψουμε τον πίνακα μετρήσεων που εξάγεται από 57 διαύλους σε αντίστοιχο χρονικό κώδικα. Το μέγεθος του πίνακα μετρήσεων που προέρχεται από το MATPOWER είναι 137 λόγω του ότι στον ίδιο δίαυλο θα υπάρχουν πολλά μέτρα. Στο επόμενο βήμα, αυτές τα spikes εφαρμόζονται στον μη γραμμικό κόμβο του DFN. Υπάρχουν διάφορες επιλογές σχεδιασμού για τον μη γραμμικό κόμβο. Δεδομένου του ότι μας ενδιαφέρουν τα spiking neural networks, ο κόμβος εισόδου του στρώματος δεξαμενής θα είναι ένας νευρώνας LIF [84]. Τα παραγόμενα spikes θα πρέπει να μετατραπούν σε αναλογικό ρεύμα πριν εφαρμοστούν στον νευρώνα LIF. Ένα αντίστοιχο spike train θα δημιουργείται για κάθε αναλογικό ρεύμα που εφαρμόζεται στον νευρώνα LIF.

Οι καθυστερήσεις αποτελούν συχνή παρουσία σε συστήματα με δυναμικά χαρακτηριστικά, ακόμη και όταν πρόκειται για τη μετάδοση πληροφοριών μεταξύ νευρώνων στον εγκέφαλο. Για να περιγραφούν μαθηματικά τέτοια συστήματα, χρησιμοποιούνται διαφορικές εξισώσεις

καθυστερήσης (Delay differential equations). Σε αυτά τα συστήματα, η δυναμική εξαρτάται όχι μόνο από την τρέχουσα κατάσταση αλλά και από προηγούμενες καταστάσεις, εμφανίζοντας χαρακτηριστικά όπως υψηλή διαστασιμότητα (dimensionality) και βραχυπρόθεσμη μνήμη - βασικά χαρακτηριστικά που απαιτούνται για συστήματα υπολογισμού δεξαμενών (RC). Συγκριτικά με τα παραδοσιακά συστήματα RC, τα delayed feedback RC παρέχουν σχεδόν ισοδύναμες επιδόσεις.

Ένα delayed feedback RC αποτελείται από έναν μόνο μη γραμμικό κόμβο και έναν βρόχο καθυστέρησης. Τα σήματα εξόδου από αυτή την αποθήκη υποβάλλονται σε μια διαδικασία εκπαίδευσης, όπου χρησιμοποιείται ένας αλγόριθμος που στοχεύει στο να καταστήσει τον αθροισμένο συντελεστή της κατάστασης όσο το δυνατόν πιο κοντά στην επιθυμητή τιμή εξόδου. Η είσοδος εισάγεται απευθείας στον μη γραμμικό κόμβο. Για να αντισταθμιστεί η απώλεια παραλληλισμού, μια διαδικασία κάλυψης κλιμακώνει τα σήματα εισόδου, τοποθετώντας τα σε μια προσωρινή κατάσταση πριν εισέλθουν στον μη γραμμικό κόμβο. Μετά από αυτήν τη διαδικασία κάλυψης, τα σήματα μεταφέρονται στον μη γραμμικό κόμβο για μη γραμμική αντιστοίχιση. Σύμφωνα με το παραδοσιακό σύστημα ανατροφοδότησης αποθέματος, εκπαιδεύονται μόνο οι συνδέσεις βάρους εξόδου.

Εμπνευσμένη από το delayed feedback RC, προστέθηκε μια αναλογική υλοποίηση του συστήματος αυτού, η οποία είναι σε θέση να επεξεργαστεί άμεσα σήματα που βασίζονται σε spikes. Ο αναλογικός σχεδιασμός παρέχει το πλεονέκτημα της λειτουργίας σε πραγματικό χρόνο, μειώνοντας την απαιτούμενη επιφάνεια σχεδίασης και την κατανάλωση ισχύος. Στην παρούσα διάταξη, το σήμα των spikes που παράγεται από τον νευρώνα LIF μετακινείται 10 χιλιοστά του δευτερολέπτου για τη δημιουργία της κατάστασης του δεύτερου κόμβου της δεξαμενής. Αυτή η διαδικασία επαναλαμβάνεται τέσσερις φορές για τη λήψη διαφορετικών καταστάσεων. Συνήθως, η πληροφορία κωδικοποιείται πριν εισαχθεί στον μη γραμμικό κόμβο.

Συνήθως χρησιμοποιούνται δύο στρατηγικές κωδικοποίησης: κωδικοποίηση ρυθμού (rate encoding) και χρονική κωδικοποίηση (temporal encoding). Η κωδικοποίηση ρυθμού αναπαριστά την πληροφορία εισόδου με τον αριθμό των spikes, αγνοώντας άλλα χαρακτηριστικά τους. Η χρονική κωδικοποίηση, από την άλλη πλευρά, κωδικοποιεί τις πληροφορίες σε διαστήματα μεταξύ των spikes. Σε αυτόν τον σχεδιασμό, χρησιμοποιείται η χρονική κωδικοποίηση και προσαρμόζεται μια επαναληπτική δομή στον χρονικό κωδικοποιητή, όπου ο αριθμός των νευρώνων και ο αριθμός των spikes έχουν εκθετική σχέση. Αυτό μειώνει τον αριθμό των νευρώνων που απαιτούνται για να επιτευχθεί ο ίδιος αριθμός spikes. Ο χρονικός κωδικοποιητής εξασφαλίζει ότι μόνο ένας νευρώνας λειτουργεί σε δυναμική λειτουργία, μειώνοντας σημαντικά την κατανάλωση ενέργειας. Αυτός ο χρονικός κωδικοποιητής έχει κατασκευαστεί χρησιμοποιώντας μια διεργασία CMOS 180 nm με ένα σχήμα συμμετρίας για τη μεγιστοποίηση της χρήσης του χώρου του “πυρήνα” και έχει επιδείξει υψηλή ακρίβεια και χαμηλή κατανάλωση ενέργειας σε σύγκριση με άλλα σχέδια νευρώνων.

Για κάθε σετ μετρήσεων, αντλούνται πέντε διαφορετικές καταστάσεις, οι οποίες επεξεργάζονται στη συνέχεια για την εκπαίδευση ενός MLP. Το χαρακτηριστικό που χρησιμοποιείται κατά την εκπαίδευση του MLP είναι ο χρονισμός των spikes για κάθε κατάσταση σε κάθε σετ δεδομένων. Στα δεδομένα εκπαίδευσης, δεδομένου ότι το μισό από αυτά υπόκειται σε επίθεση, η αντίστοιχη ετικέτα για τα δεδομένα που επίκεινται επίθεση ορίζεται ως 1, ενώ διαφορετικά ως 0. Μετά την εκπαίδευση του MLP με τα δεδομένα εκπαίδευσης, τα δεδομένα δοκιμής χρησιμοποιούνται για να αξιολογηθεί η απόδοση του συστήματος.

• TEMPORAL ENCODER

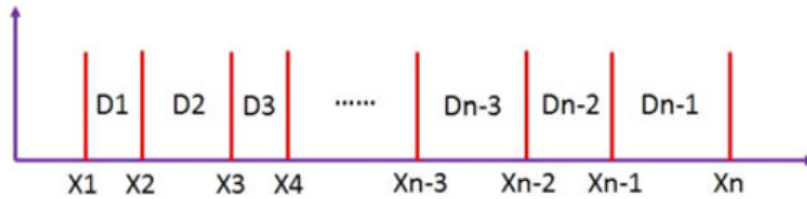
Ο κωδικοποιητής που εισήχθη στο άρθρο “Energy Efficient Spiking Temporal Encoder Design for Neuromorphic Computing Systems” [85] χρησιμεύει ως ο χρονικός κωδικοποιητής του σχεδιασμού μας. Τα αντίστοιχα interspike intervals μπορούν να εκφραστούν ως εξής:

$$D_i = f(C_i, V_i) - f(C_{i-1}, V_{i-1}) \quad (1)$$

Η συνάρτηση $f(X, Y)$ εκφράζεται ως:

$$f(C_i, V_i) = f(C_{i+1})[\beta(V_i - \gamma) + \theta] \quad (2)$$

όπου οι παράμετροι καθορίζουν τα χαρακτηριστικά του κωδικοποιητή, όπως περιόδους φόρτισης και ανθεκτικότητας. Οι C_i, V_i είναι οι παράμετροι membrane capacitance και threshold firing voltage του χρονικού κωδικοποιητή, αντίστοιχα.



Χρησιμοποιώντας τον παραπάνω χρονικό κωδικοποιητή, οποιοδήποτε δείγμα στον πίνακα μέτρησης κωδικοποιείται στις interspike interval αποστάσεις, D_i . Ο αριθμός αυτών των διαστημάτων μπορεί να ποικίλει ανάλογα με τον αριθμό των νευρώνων που χρησιμοποιούνται στον χρονικό κωδικοποιητή για κάθε δείγμα. Σε αυτό το πείραμα, για λόγους απλότητας, επιλέγουμε να χρησιμοποιήσουμε τρεις νευρώνες ($N = 3$) στον κωδικοποιητή, που οδηγεί σε τέσσερα διαφορετικά spikes, X_1 έως X_4 ή τρία διαστήματα D_1 έως D_3 που φαίνονται στο Σχήμα 3. Λόγω της ακόλουθης εξίσωσης που αναφέρεται στο άρθρο [85], υπάρχει σχέση μεταξύ του αριθμού των spikes που παράγονται και του αριθμού των νευρώνων που χρησιμοποιούνται στον χρονικό κωδικοποιητή:

$$S_N = 2^{N-1} \quad (3)$$

όπου, S_N είναι ο αριθμός των spikes που παράγονται από τον χρονικό κωδικοποιητή και N είναι ο αριθμός των νευρώνων που χρησιμοποιούνται στον κωδικοποιητή.

• SMART GRID ATTACK DETECTION FORMULATION

Τα smart grids αποτελούν έναν τρόπο για τη δημιουργία ασφαλούς και αξιόπιστης ενεργειακής μεταφοράς, που επιτρέπει τη σύνδεση μεταξύ καταναλωτών και παραγωγών ενέργειας. Για τη δημιουργία του πίνακα μετρήσεων στα smart grids, το εργαλείο MATPOWER 5.1 αποτελεί μια χρήσιμη επιλογή, καθώς προσφέρει ευελιξία στους χρήστες, επιτρέποντάς τους να ρυθμίσουν τον αριθμό των διαύλων. Στο πείραμά μας, αποφασίσαμε να χρησιμοποιήσουμε 57 διαύλους, από τους οποίους προκύπτουν 137 διαφορετικές μετρήσεις. Σημειώνεται πως η κύρια παράμετρος που επηρεάζει την ανίχνευση επιθέσεων στα smart grids είναι ο αριθμός των

μετρητών που παραβιάζονται, και όχι ο αριθμός των διαύλων. Επιλέξαμε 57 διαύλους, καθώς αυτός ο αριθμός βρίσκεται κατά προσέγγιση στο μέσον του εύρους των διαύλων που παρέχει το MATPOWER και παράγει 137 μετρητές. Αυτή η ρύθμιση παρέχει τη βάση για τη μελέτη των επιπτώσεων διαφορετικών αριθμών παραβαίνοντων μετρητών σε αυτή τη διάταξη.

Το διάνυσμα μέτρησης, το οποίο είναι η έξοδος διαφορετικών μετρητών στους διαύλους είναι z . H είναι το διάνυσμα κατάστασης, x είναι η φάση τάσης των διαύλων, και n είναι ο θόρυβος του περιβάλλοντος.

$$z = Hx + n \quad (4)$$

Όταν υπάρχει επίθεση, ένα διάνυσμα επίθεσης, a , προστίθεται στη μέτρηση. Κατά συνέπεια, η μέτρηση, $\bar{z}$, γίνεται:

$$\bar{z} = Hx + a + n \quad (5)$$

Υποθέτουμε ότι η επίθεση είναι ένα Gaussian τυχαίο διάνυσμα με 0,05 διακύμανση.

Το SVE είναι η πρώτη μέθοδος που εισήχθη για την εκτέλεση ανίχνευσης επιθέσεων για smart grids. Αυτή η μέθοδος συνίσταται στον υπολογισμό ενός υπολοίπου που δηλώνεται ως ρ . Εάν η τιμή του ρ υπερβαίνει μια προκαθορισμένη threshold value, λέγεται ότι το διάνυσμα z έχει επιτεθεί και οι μετρητές έχουν παραβιαστεί.

$$\rho = \left\| \bar{z} - H\hat{x} \right\|_2^2 \quad (6)$$

όπου $\hat{x}$ είναι το διάνυσμα που υπολογίζεται χρησιμοποιώντας τον αλγόριθμο SVE. Το $\hat{x}$ υπολογίζεται στη συνέχεια ως εξής:

$$\hat{x} = (H^T \Lambda H^{-1}) H^T \Lambda z \quad (7)$$

Η μόνη ενέργεια που πρέπει να γίνει στον αλγόριθμο SVE είναι να υπολογιστεί το $\hat{x}$, και για να γίνει αυτό, πρέπει να υπολογιστεί το Λ , όπου το Λ ορίζεται ως διαγώνιος πίνακας με τα διαγώνια στοιχεία του να είναι τα αντίστροφα της διακύμανσης των μετρήσεων. Για παράδειγμα, το jοστό διαγώνιο στοιχείο του Λ ισούται με το αντίστροφο της διακύμανσης του jοστού στοιχείου του z . Η μέθοδος SVE είναι μια πολύ απλή μέθοδος υλοποίησης, αλλά έχει πολλές ελλείψεις σε διαφορετικές καταστάσεις επίθεσης.

Στην περίπτωση που $a = HC$, η επίθεση είναι κρυφή. Αυτό σημαίνει ότι η SVE δεν είναι σε θέση να ανιχνεύσει τις κακές μετρήσεις. Στην περίπτωση των κρυφών επιθέσεων, η υπολειμματική τιμή είναι μικρότερη από το threshold και η επίθεση δεν μπορεί να ανιχνευθεί από το SVE. Για να εκτελέσει μια κρυφή επίθεση, ο κυβερνοεπιτιθέμενος πρέπει να έχει πρόσβαση σε τουλάχιστον έναν συγκεκριμένο αριθμό επιθέσεων. Έχει αποδειχθεί ότι δεν είναι δυνατόν ο επιτιθέμενος να επιλέξει οποιοδήποτε αυθαίρετο c και να το πολλαπλασιάσει με το H για να εκτελέσει την κρυφή επίθεση. Αυτό σημαίνει ότι για να πραγματοποιήσει μια κρυφή επίθεση, ο επιτιθέμενος πρέπει να έχει πρόσβαση σε τουλάχιστον k μετρήσεις, στις οποίες $k > m - n$ όπου m είναι ο αριθμός των μετρήσεων και n είναι ο αριθμός των διαύλων. Στο σύστημά μας, $m = 137$ και $n = 57$. Σε αυτή την περίπτωση, $m - n = 80$ που σημαίνει ότι με μεγάλη πιθανότητα η επίθεση θα είναι κρυφή επίθεση όταν υπάρχουν περισσότερες από 80 παραβιασμένες μετρήσεις στο δίκτυο smart grid μας. Υπό αυτό το σενάριο, ο SVE γίνεται ένας αναποτελεσματικός ανιχνευτής επιθέσεων.

• MODELING THE WIND POWER GENERATORS IN MATPOWER

Το MATPOWER είναι επίσης χρήσιμο για την μελέτη της αιολικής ενέργειας στο πλαίσιο του smart grid. Στο δίκτυο μας υπάρχουν έξι γεννήτριες ισχύος συνολικά. Μπορούμε να αντικαταστήσουμε την ισχύ που παράγεται από αυτές τις γεννήτριες με την ισχύ που προέρχεται από ανεμογεννήτριες. Ο τύπος (8) χρησιμοποιείται για να υπολογίσουμε την παραγόμενη ισχύ από μια ανεμογεννήτρια.

$$P_{avail} = 1/2 \rho A v^3 C_p \quad (8)$$

Σε αυτόν τον τύπο, χρησιμοποιούμε την πυκνότητα του αέρα (ρ), την επιφάνεια σάρωσης των πτερυγίων (A), την ταχύτητα του ανέμου (v) και τον συντελεστή ισχύος (C_p). Ο C_p έχει μέγιστη τιμή 0,59, όπως αποδείχθηκε από τον Albert Betz. Αυτό ονομάζεται όριο Betz. Στη μελέτη μας, Ορίζουμε το C_p σε 0,4 και την επιφάνεια της ανεμογεννήτριας σε 8495 m². Χρησιμοποιούμε διάφορες τιμές για την ταχύτητα του ανέμου από 0 έως 12 m/s. Οι τιμές αυτές χρησιμοποιούνται στο MATPOWER για τη δημιουργία του πίνακα H.

• SMART GRID ATTACK DETECTION USING DFN AND MLP

Το πρόβλημα των FDI μπορεί να αντιμετωπιστεί ως ένα πρόβλημα ταξινόμησης, όπου τα δεδομένα χωρίζονται σε δύο κατηγορίες: αυτά που έχουν υποστεί επίθεση και αυτά που δεν έχουν υποστεί. Αυτό το πρόβλημα δεν έχει μελετηθεί από τη σκοπιά του Reservoir Computing μέχρι στιγμής. Σε πειράματα χρησιμοποιούνται δύο διαφορετικά σύνολα δεδομένων: ένα με δεδομένα που υπέστησαν κρυφή επίθεση και ένα με δεδομένα όπου οι μετρήσεις έχουν προσβληθεί από άμεσες ή μη κρυφές επιθέσεις. Τα πειράματα εκτελούνται πάνω σε 1000 δείγματα και επαναλαμβάνονται 10 φορές. Το πρώτο βήμα είναι η κωδικοποίηση των δεδομένων χρησιμοποιώντας έναν προσωρινό κωδικοποιητή. Στη συνέχεια, τα spike trains που εξάγονται από τον κωδικοποιητή μετατρέπονται σε αναλογικά ρεύματα.

$$I^i = \sum_{t^j} K(t - t^j) H(t - t^j) \quad (9)$$

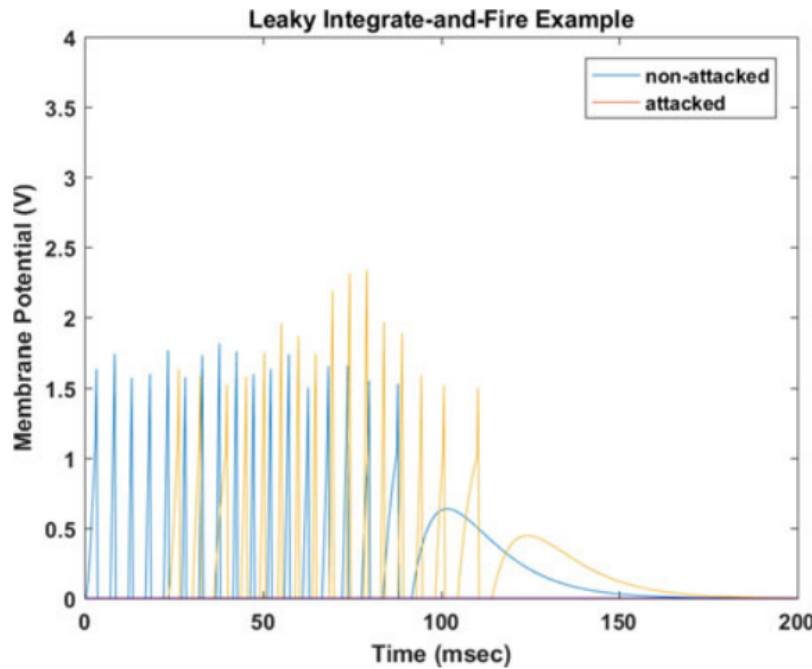
όπου H είναι η συνάρτηση Heaviside. I^i είναι το αναλογικό ρεύμα του i ου δείγματος στο z . και t^j είναι ο χρόνος εμφάνισης του j οστού spike στο αντίστοιχο spike train του i ου δείγματος που επιτυγχάνεται από τον χρονικό κωδικοποιητή και

$$K(t - t^j) = V_0 (\exp(-((t - t^j)/\tau_s)) - \exp(-((t - t^j)/\tau_f))) \quad (10)$$

όπου το τ_s έχει οριστεί στα 10 ms και το τ_f στα 2,5 ms. Οι τιμές των τ_s και τ_f πρέπει να επιλεγούν με κάποιο τρόπο ώστε $\tau_s/\tau_f = 4$. Το V_0 είναι ένας παράγοντας κανονικοποίησης για να βεβαιωθείτε ότι η μέγιστη τιμή του πυρήνα δεν υπερβαίνει το 1.

Μέχρι τώρα, έχουμε δημιουργήσει αναλογικά σήματα ρεύματος από τα (9) και (10) που αντιστοιχούν στους χρονικούς κώδικες που εξάγονται από τον προσωρινό κωδικοποιητή. Το επόμενο βήμα είναι να εισάγουμε αυτά τα ρεύματα στο DFN για τη δημιουργία αντίστοιχων καταστάσεων. Ο μη γραμμικός κόμβος του DFN είναι ένας νευρώνας LIF. Αυτά τα αναλογικά ρεύματα εφαρμόστηκαν στο DFN για τα δείγματα που είχαν υποστεί επίθεση και τα δείγματα χωρίς επίθεση.

Το αποτέλεσμα είναι ότι παράγονται τέσσερις εικονικοί κόμβοι, ως εξής: η κατάσταση του τέταρτου εικονικού κόμβου καθυστερεί κατά 40 ms σε σύγκριση με τον μη γραμμικό κόμβο. Στη συνέχεια, η κατάσταση αυτού του τέταρτου εικονικού κόμβου πολλαπλασιάζεται με ένα κέρδος ανάδρασης (feedback gain) 0,8 και προστίθεται στο νέο αναλογικό ρεύμα εισόδου. Αυτή η διαδικασία δημιουργεί πέντε διαφορετικές καταστάσεις τόσο για τα δείγματα που δέχθηκαν επίθεση όσο και για τα υπόλοιπα. Οι μέσες τιμές των καταστάσεων παρουσιάζονται στο Σχήμα 4.



Σχήμα 4. Average DFN states for attacked and nonattacked data.

Από το Σχήμα 4, παρατηρούμε ότι ο χρονισμός των spikes για τις μέσες καταστάσεις των δύο κατηγοριών δεδομένων διαφέρει σημαντικά. Φαίνεται ότι τα μέσα spikes που παράγονται από τα δείγματα που έχουν υποστεί επίθεση πιθανόν να συμβαίνουν σε σύντομα χρονικά διαστήματα, ενώ αυτά που παράγονται από δείγματα που δεν έχουν υποστεί επίθεση είναι πιθανόν να συμβαίνουν σε πιο εκτεταμένα χρονικά διαστήματα. Αυτή η διαφορά στο χρονισμό μπορεί να χρησιμοποιηθεί ως χαρακτηριστικό για τη διάκριση ανάμεσα σε αυτές τις δύο κατηγορίες. Στο επόμενο στάδιο, χρησιμοποιούμε ένα MLP και το εκπαιδεύουμε χρησιμοποιώντας αυτά τα χαρακτηριστικά.

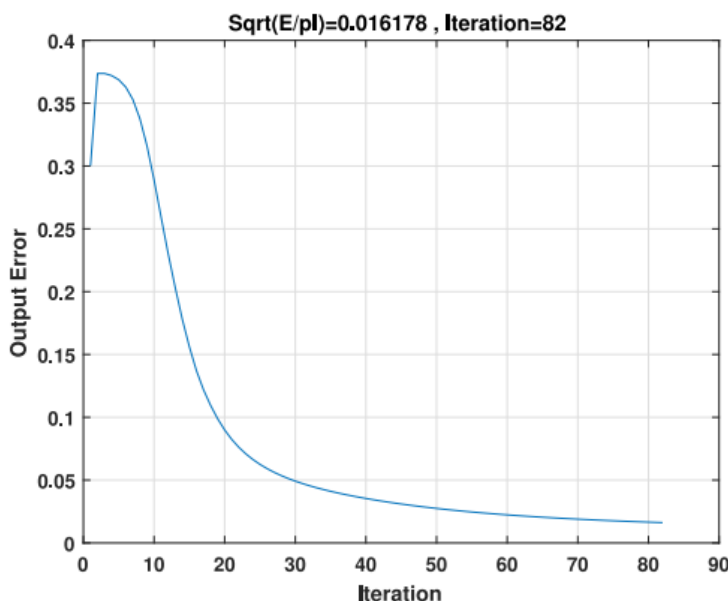
• TRAINING AN MLP WITH THE TIMING OF SPIKES

Όπως αποδεικνύεται στο άρθρο “Reservoir computing based on delay-dynamical systems” [86], το επίπεδο ανάγνωσης μπορεί να εκπαιδευτεί με έναν γραμμικό αλγόριθμο. Στον εισαγόμενο αλγόριθμο εκπαίδευσης του προαναφερόμενου άρθρου, εκχωρήθηκε ένα βάρος σε κάθε κατάσταση που εξήχθη από το DFN με τρόπο ώστε οι επιθυμητές τιμές εξόδου να μπορούν να εκτιμηθούν με το λιγότερο δυνατό σφάλμα. Η ακόλουθη έκφραση παρέχει μια καλή περίληψη του αλγορίθμου εκπαίδευσης

$$\hat{y}(k) = \sum_{i=1}^N w_i \times x[k\tau - \tau/N(N - i)] \quad (11)$$

όπου $\hat{y}(k)$ είναι η εκτιμώμενη έξοδος, w_i είναι το βάρος σύνδεσης, x είναι το διάνυσμα κατάστασης και N είναι ο αριθμός των καταστάσεων. Ο παραπάνω αλγόριθμος είναι γραμμικός και όχι επαναληπτικός, επομένως δεν θα είναι πολύ ακριβής.

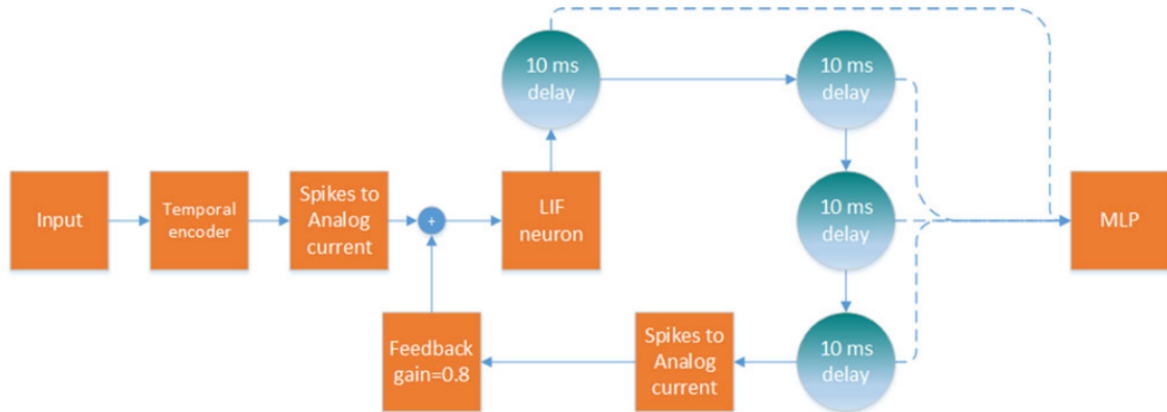
Συνεπώς, σε αυτόν τον ανιχνευτή επιθέσεων βασισμένο σε RC, χρησιμοποιούμε ένα MLP για την πρόβλεψη των αποτελεσμάτων. Η διαδικασία εκπαίδευσης του MLP βασίζεται στον αλγόριθμο backpropagation. Κάθε δείγμα που υποβάλλεται στην εκπαίδευση λαμβάνει μια ετικέτα y , όπου $y=1$ όταν το δείγμα υποβάλλεται σε επίθεση και $y=0$ όταν δεν υφίσταται επίθεση. Το χρονικό περιθώριο κατά το οποίο συμβαίνουν attacks spikes για διάφορες καταστάσεις καταγράφεται σε ένα διάνυσμα και χρησιμοποιείται ως χαρακτηριστικό κατά την εκπαίδευση του MLP. Το MLP εκπαιδεύεται με δύο κρυφά επίπεδα και ένα επίπεδο εξόδου. Κατά τη διάρκεια της εκπαίδευσης, η επιθυμητή έξοδος για τα δείγματα που δέχονται επίθεση ορίζεται σε 1, διαφορετικά ορίζεται σε 0. Όπως φαίνεται στο Σχήμα 5, το MLP εκπαιδεύεται μετά από 82 επαναλήψεις. Στη συνέχεια, τα βάρη αποθηκεύονται για χρήση στα δεδομένα δοκιμής προκειμένου να αξιολογηθεί η απόδοση του συστήματος. Τέλος, η εργασία περιλαμβάνει την εφαρμογή του αλγορίθμου SVE, του MLP και του SVM στα δείγματα για σύγκριση με την απόδοση της νέας στρατηγικής ανίχνευσης επιθέσεων που βασίζεται σε RC. Για τον ταξινομητή SVM χρησιμοποιείται η συνάρτηση Gaussian radial basis ως πυρήνας.



Σχήμα 5. Error Plot for training an MLP with DFN states spikes timings.

Όπως φαίνεται στο Σχήμα 5, το training mean square error (MSE) κατά τη διαδικασία εκπαίδευσης μειώνεται στο επίπεδο του 0,016, υποδεικνύοντας ότι το MLP είναι ικανό να διακρίνει ανάμεσα στους χρονισμούς spike καταστάσεων στα δεδομένα που έχουν υποστεί επίθεση και στα δεδομένα που δεν έχουν επιτεθεί. Ο ρυθμός εκμάθησης έχει οριστεί στο 0,01 και ο συντελεστής ορμής στο 0,5. Εάν η τιμή εξόδου υπερβαίνει το 0,5, θεωρείται ως 1, ενώ διαφορετικά θεωρείται ως 0. Υπό αυτή την προσέγγιση, η ακρίβεια της εκπαίδευσης αγγίζει σχεδόν το 100%, υποδηλώνοντας ότι σχεδόν το 100% των δειγμάτων ταξινομούνται με ακρίβεια ως επιτιθέμενα και μη επιτιθέμενα. Για να κατανοήσουμε πιο αντικειμενικά την απόδοση στην ανίχνευση, χρησιμοποιήθηκε η μέτρηση ακρίβειας όπως περιγράφεται στην εξίσωση (12). Η διαδικασία εκπαίδευσης αξιοποίησε το 50% των δειγμάτων, ενώ τα υπόλοιπα δεδομένα

αποθηκεύτηκαν για έλεγχο και επικύρωση. Τα αποτελέσματα της εφαρμογής του αλγορίθμου DFN παρουσιάζονται στη συνέχεια. Το Σχήμα 7 απεικονίζει το διάγραμμα των βημάτων του αλγορίθμου ανίχνευσης επιθέσεων που βασίζεται σε RC.



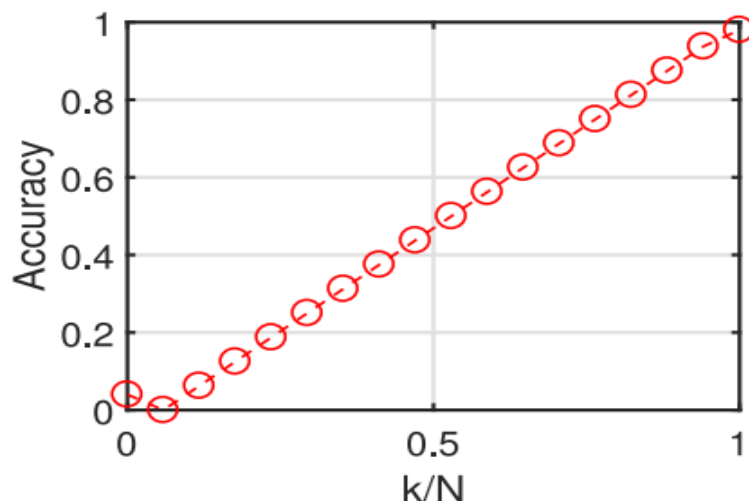
Σχήμα 7. Block diagram of the proposed DFN+MLP system for attack detection.

• State Vector Estimation

Όπως αναφέρθηκε παραπάνω, η εξίσωση (6) πρέπει να υπολογιστεί για SVE. Εάν η τιμή του ρ υπερβαίνει μια προκαθορισμένη τιμή threshold, λέγεται ότι έχει συμβεί μια επίθεση, διαφορετικά ανιχνεύεται nonattack. Αναλόγως, μπορούμε να υπολογίσουμε την τιμή του ρ όταν το διάνυσμα μέτρησης δέχεται επίθεση από το ίδιο διάνυσμα επίθεσης που αναφέρθηκε προηγουμένως. Η τιμή του ρ που επιτυγχάνεται για προσβεβλημένα διανύσματα με διαφορετικό αριθμό παραβιασμένων μετρήσεων χρησιμοποιείται για την αξιολόγηση της απόδοσης του SVE. Η μέτρηση απόδοσης που χρησιμοποιείται για την αξιολόγηση της απόδοσης ανίχνευσης είναι η ακρίβεια, η οποία ορίζεται ως

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (12)$$

όπου τα TP, TN, FP και FN αντιστοιχούν στον αριθμό των αληθινών θετικών, αληθινών αρνητικών, ψευδώς θετικών και ψευδώς αρνητικών δειγμάτων, αντίστοιχα.

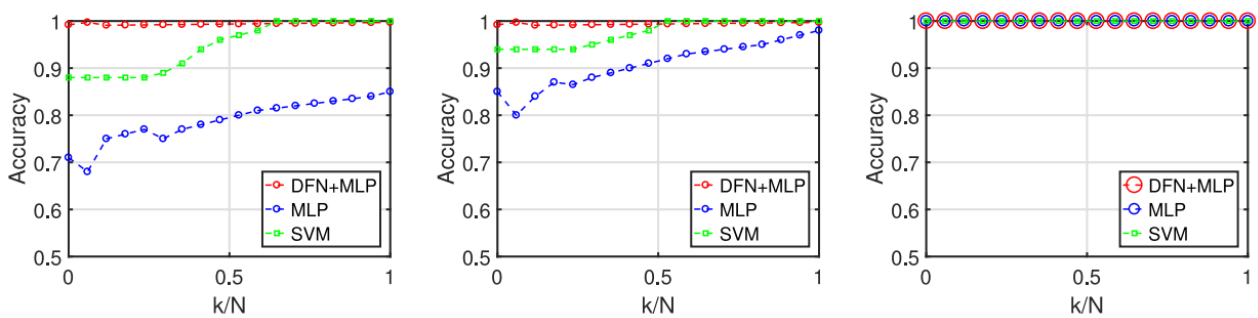


Σχήμα 6. Accuracy of the SVE.

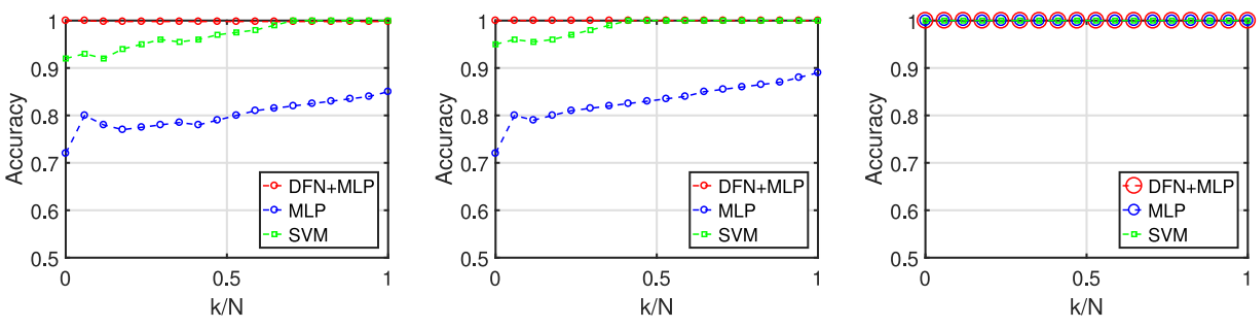
Η ανίχνευση επιθέσεων από το SVE παρουσιάζει σημαντική απόδοση, όπως φαίνεται στο Σχήμα 6. Η γενική ακρίβεια του SVE επηρεάζεται από τον αριθμό των παραβιασμένων μετρήσεων, ακόμα και όταν η επίθεση δεν είναι κρυφή. Αυτό συμβαίνει διότι η απόδοση του SVE συνδέεται σημαντικά με τη χρήση της υπολειμματικής τιμής. Όταν ο αριθμός των παραβιασμένων μετρήσεων είναι χαμηλός, παρατηρείται σημαντική μείωση της ακρίβειας του SVE. Παρακάτω, θα αποδείξουμε ότι αυτό το θέμα μπορεί να αντιμετωπιστεί πλήρως με τον εισαγόμενο ανιχνευτή επίθεσης DFN+MLP που βασίζεται σε RC.

• PERFORMANCE EVALUATION

Όπως προαναφέραμε, μόνο το 50% των δεδομένων χρησιμοποιείται για εκπαίδευση, με τα υπόλοιπα να διατηρούνται για δοκιμές και επαλήθευση. Στη συνέχεια, θα εξετάσουμε αναλυτικά την απόδοση τριών διαφορετικών αλγορίθμων: τη μέθοδο βασισμένη σε RC (DFN+MLP), το MLP και το SVM. Ο συνολικός αριθμός των δειγμάτων που χρησιμοποιήθηκαν για δοκιμή και επαλήθευση είναι 5000. Από αυτά, τα μισά υπέστησαν επίθεση, ενώ τα υπόλοιπα όχι. Στο Σχήμα 8 και το Σχήμα 9, παρουσιάζονται τα αποτελέσματα της αξιολόγησης ακρίβειας της προτεινόμενης μεθόδου για τους δύο τύπους επιθέσεων στα smart grids, ανάλογα με το μέγεθος της επίθεσης α (όπου χρησιμοποιούνται τρεις διαφορετικές τιμές: $\alpha = 0,1$, $\alpha = 1$ και $\alpha = 10$). Είναι σημαντικό να σημειωθεί ότι, επειδή το SVE δεν είναι σε θέση να ανιχνεύει κρυφές επιθέσεις, δεν συμπεριλαμβάνουμε την αξιολόγησή του στα Σχήματα 8 και 9.



Σχήμα 8. Accuracy of direct attack detection for three different methods, $\alpha = 0.1, 1, 10$.



Σχήμα 9. Accuracy of hidden attack detection for three different methods, $\alpha = 0.1, 1, 10$.

Από τα γραφήματα, είναι προφανές ότι η απόδοση τόσο του MLP όσο και του SVM επηρεάζεται σημαντικά από το μέγεθος της επίθεσης και τον αριθμό των μετρητών που προκύπτουν ως στόχοι. Σε αντίθεση με το SVE, τόσο το MLP όσο και το SVM είναι ικανά να ανιχνεύσουν κρυφές επιθέσεις. Ωστόσο, πρέπει να σημειώσουμε ότι η απόδοση τους στην ανίχνευση επιθέσεων εξαρτάται σημαντικά από

τις παραμέτρους των επιθέσεων. Για παράδειγμα, η ακρίβεια τόσο του MLP όσο και του SVM αυξάνεται όσο μεγαλώνει το μέγεθος της επίθεσης, υπονοώντας ότι μπορούν να ανιχνεύσουν επιθέσεις με υψηλή ακρίβεια όταν οι επιθέσεις είναι μεγάλες.

Ωστόσο, όταν οι επιθέσεις είναι μικρές κατά μέγεθος, τότε τα MLP και SVM αντιμετωπίζουν δυσκολίες στην ανίχνευσή τους με μικρότερη βεβαιότητα. Για παράδειγμα, όσον αφορά το MLP, η ακρίβεια του είναι 100% όταν το μέγεθος της επίθεσης είναι 10, αλλά μειώνεται στο 70% όταν το μέγεθος της επίθεσης είναι 0,1. Αυτό είναι ανεπιθυμητό στον τομέα της ανίχνευσης επιθέσεων στα smart grids, όπου το μέγεθος της επίθεσης μπορεί να ποικίλει. Αντίθετα, στη μέθοδο DFN+MLP που βασίζεται στην RC, διαπιστώνουμε ότι οι διακυμάνσεις στο μέγεθος της επίθεσης δεν επηρεάζουν την ακρίβεια σημαντικά. Συγκεκριμένα, η αλλαγή στην ακρίβεια λόγω των μεταβολών στο μέγεθος της επίθεσης είναι πολύ μικρή για την προσέγγιση που βασίζεται στην RC, και η ακρίβεια παραμένει κοντά στο 100% για όλα τα μεγέθη επίθεσης. Αυτό υποδεικνύει ότι η απόδοση στην ανίχνευση επιθέσεων με την προσέγγιση βασισμένη στην RC είναι ισχυρή ανεξαρτήτως του μεγέθους της επίθεσης. Τα Σχήματα 8 και 9 δείχνουν επίσης την ακρίβεια ως συνάρτηση του αριθμού των μετρητών που έχουν παραβιαστεί για διαφορετικές στρατηγικές ανίχνευσης επίθεσης.

Από τα γραφήματα, φαίνεται ότι η προσέγγιση που βασίζεται στο RC είναι πολύ πιο αποτελεσματική από τις μεθόδους MLP και SVM, ανάλογα με τον αριθμό των παραβιασμένων μετρητών. Επιπλέον, συγκρίνοντας τα δύο σχήματα, παρατηρούμε ότι σε αντίθεση με τις υφιστάμενες μεθόδους ανίχνευσης (SVE, MLP και SVM), η μέθοδος DFN+MLP παρέχει ομοιόμορφη απόδοση ανεξάρτητα από τη μέθοδο επίθεσης. Σε αυτήν τη μελέτη, το 50% των δειγμάτων υποβλήθηκε σε επίθεση και το υπόλοιπο όχι, διατηρώντας μια ισορροπημένη συλλογή δεδομένων. Εάν ο αριθμός των δειγμάτων που υποβλήθηκαν σε επίθεση διαφέρει σημαντικά από τα μη επιθετικά δείγματα, τότε αντιμετωπίζουμε ένα μη ισορροπημένο σύνολο δεδομένων. Σε τέτοιες περιπτώσεις, το μέτρο F1 χρησιμοποιείται για να αξιολογηθεί η απόδοση του αλγορίθμου μάθησης και μπορεί να διαχειριστεί τη μη ισορροπημένα δεδομένα.

$$F1 = (2TP) / (2TP + FP + FN) \quad (13)$$

Στην παρούσα μελέτη, παρουσιάζεται μια στρατηγική ανίχνευσης επιθέσεων που βασίζεται στο Reservoir Computing (DFN+MLP) για τα έξυπνα δίκτυα ενέργειας. Αυτή η μέθοδος αποτελείται από τρία βασικά βήματα. Στο πρώτο βήμα, κωδικοποιούμε τα μετρικά δεδομένα χρησιμοποιώντας έναν χρονικό κωδικοποιητή και μετατρέπουμε τα παραγόμενα spikes σε αντίστοιχα αναλογικά ρεύματα. Στο δεύτερο βήμα, αυτά τα αναλογικά ρεύματα εισάγονται σε ένα νευρώνα LIF και χρονικά μετατοπίζονται για τη δημιουργία καταστάσεων εικονικών κόμβων. Στο τρίτο βήμα, η έξοδος του τέταρτου εικονικού κόμβου υποβάλλεται σε πολλαπλασιασμό με έναν συντελεστή ανατροφοδότησης και προστίθεται στα νέα εισερχόμενα δεδομένα για να διατηρηθεί η αναδρομική φύση του DFN. Οι χρονικές διακυμάνσεις αυτών των καταστάσεων χρησιμοποιούνται για την εκπαίδευση ενός MLP για την ταξινόμηση. Τα αποτελέσματα των προσομοιώσεων δείχνουν ότι αυτός ο αλγόριθμος είναι σε θέση να ανιχνεύσει αξιόπιστα επιθέσεις υπό διάφορες παραμετροποιήσεις, όπως το μέγεθος και ο αριθμός των παραβιασμένων μετρητών, και να ξεπεράσει τις υπάρχουσες μεθόδους, όπως η SVE, MLP και SVM. Σημαντικό είναι επίσης να σημειωθεί ότι αυτή η μελέτη αποτελεί την πρώτη προσπάθεια επίλυσης προβλημάτων ανίχνευσης επιθέσεων σε έξυπνα δίκτυα ενέργειας χρησιμοποιώντας το Reservoir Computing. Το προτεινόμενο μοντέλο μπορεί να εφαρμοστεί σε οποιαδήποτε εργασία ταξινόμησης που υπάρχει χωροχρονική συσχέτιση μεταξύ των δειγμάτων του συνόλου δεδομένων.

6.2 *Detecting Dynamic Attacks in Smart Grids Using Reservoir Computing: A Spiking Delayed Feedback Reservoir Based Approach*

Το άρθρο “[Detecting Dynamic Attacks in Smart Grids Using Reservoir Computing: A Spiking Delayed Feedback Reservoir Based Approach](#)” των Kian Hamedani, Lingjia Liu, Shiyan Hu, Jonathan Ashdown, Jinsong Wu και Yang Yi, παρουσιάζει μία τεχνική βασισμένη σε Spiking neural network για την ανίχνευση δυναμικών κυβερνοεπιθέσεων σε ένα smart grid. Τα Spiking neural networks έχουν χρησιμοποιηθεί ευρέως για την επιβλεπόμενη αναγνώριση προτύπων διερευνώντας την υποκείμενη χωροχρονική συσχέτιση. Εν τω μεταξύ, η χωροχρονική συσχέτιση εκδηλώνεται σημαντικά μεταξύ διαφορετικών στοιχείων σε ένα smart grid, καθιστώντας το Spiking neural network έναν επιθυμητό υποψήφιο για την ανίχνευση επιθέσεων false data injection. Η επίτευξη της προτεινόμενης τεχνικής γίνεται μέσω της συνετής ενσωμάτωσης των spiking neurons με ένα ειδικό αναδρομικό νευρωνικό δίκτυο που ονομάζεται delayed feedback reservoir computing. Η κωδικοποίηση του inter-spike διαστήματος διερευνάται επίσης στη διαδικασία εκπαίδευσης με βάση τη precise-spike-driven synaptic plasticity. Τα αποτελέσματα της προσομοίωσης υποδηλώνουν ότι η εισαγόμενη μέθοδος υπερτερεί έναντι των multi-layer perceptrons και μπορεί να επιτύχει σημαντικά καλύτερες επιδόσεις σε σύγκριση με τις σύγχρονες τεχνικές. Επιπλέον, η ανάλυσή μας υποδεικνύει ότι η τιμή της καθυστέρησης στο delayed feedback reservoir θα έχει σημαντικό αντίκτυπο στη συνολική απόδοση του συστήματος.

Τα smart grids αποτελούν βελτιωμένα ηλεκτρικά δίκτυα που είναι σχεδιασμένα για έξυπνη μεταφορά και διανομή ηλεκτρικής ενέργειας με στόχο να αυξήσουν την αξιοπιστία, την ασφάλεια και την αποδοτικότητα του συστήματος. Ωστόσο, μπορεί να δημιουργήσουν νέα προβλήματα ασφάλειας αν δεν διαχειρίζονται σωστά. Οι κυβερνοεπιθέσεις και οι διαταραχές στον κυβερνοχώρο μπορούν να προέρχονται από διάφορες πηγές με διαφορετικά κίνητρα, από φάρσες έως τρομοκρατία. Λόγω του υψηλού όγκου δεδομένων που ρέει στο δίκτυο, η παρακολούθηση των πληροφοριών ασφαλείας και η ανίχνευση ανωμαλιών γίνεται δύσκολη για τους αναλυτές. Το πρόβλημα της false data injection (FDI) στα smart grids παρουσιάστηκε για πρώτη φορά στο [87].

Γενικά, υπάρχουν δύο κύριες κατηγορίες επιθέσεων FDI στα smart grids: οι single-period επιθέσεις και οι dynamic επιθέσεις. Στο πρώτο σενάριο, οι επιθέσεις πραγματοποιούνται ταυτόχρονα και ο επιτιθέμενος περιμένει μέχρι να υπάρξει ευκαιρία να εκτελέσει την επίθεση με μεγάλη πιθανότητα επιτυχίας. Οι περισσότερες από τις υπάρχουσες εργασίες για την κυβερνοασφάλεια των smart grids επικεντρώνονται σε αυτήν την περίπτωση. Στα dynamic σενάρια, ο αντίπαλος αλλοιώνει σταδιακά την κατάσταση του δικτύου προς την επιθυμητή κατάσταση, κάτι που το καθιστά δυσκολότερο να ανιχνευθεί, καθώς οι αλλαγές συμβαίνουν αργά. Μια περίληψη των υπάρχοντων αλγορίθμων ανίχνευσης FDI, μαζί με τα πλεονεκτήματα και τα μειονεκτήματά τους, περιγράφεται σε άλλη μελέτη [88].

Ο State Vector Estimation (SVE) αναδείχθηκε ως ο πρώτος αλγόριθμος που προτάθηκε για την αντιμετώπιση προβλημάτων που σχετίζονται με την αντιμετώπιση του προβλήματος FDI στα smart grids. Επιπλέον, έχουν αναπτυχθεί προσεγγίσεις βασισμένες στη μηχανική μάθηση για την ανίχνευση των FDI σε αυτά τα δίκτυα. Συγκεκριμένα, μελετήθηκαν τεχνικές που βασίζονται σε τεχνητά νευρωνικά δίκτυα (ANN), support vector machines (SVM) και k-nearest neighbor (KNN). Γενικά, οι προσεγγίσεις μηχανικής μάθησης έχουν εμφανίσει καλύτερη απόδοση σε σύγκριση με τον αλγόριθμο SVE στην ανίχνευση των FDI. Τόσο οι αλγόριθμοι μηχανικής μάθησης με επίβλεψη όσο και οι αλγόριθμοι μηχανικής μάθησης χωρίς επίβλεψη

χρησιμοποιούνται για την ανίχνευση αθέατων εγχύσεων δεδομένων σε πραγματικό χρόνο στα smart grids. Επιπλέον, σε μια άλλη εργασία μελετάται η απόδοση των SVM και KNN στην ανίχνευση false data σε ένα σύστημα IEEE-30 bus, υπό διάφορα σενάρια δεδομένων, είτε ισορροπημένων είτε μη ισορροπημένων. Τέλος, στην εργασία που αναφέρεται στο [89], αξιοποιείται μια μέθοδος που βασίζεται στο deep learning για την αποτύπωση της συμπεριφοράς των χαρακτηριστικών των επιθέσεων FDI, τα οποία χρησιμοποιούνται στη συνέχεια για την ανίχνευση των FDI σε πραγματικό χρόνο με υψηλή ακρίβεια. Δυστυχώς, αυτές οι προσεγγίσεις επικεντρώνονται μόνο στην ανίχνευση single-period επιθέσεων και δεν λαμβάνουν υπόψη τις χωροχρονικές συσχετίσεις των μετρήσεων στα smart grids. Στο προηγούμενο παράδειγμα (Ενότητα 6.1), παρουσιάστηκε μια προσέγγιση που βασίζεται στο Reservoir Computing (RC) για την ανίχνευση των FDI, εκμεταλλευόμενη τις χωροχρονικές συσχετίσεις.

Τα RC αποτελούν ένα είδος νευρωνικών δικτύων που είναι ευκολότερο να εκπαιδευτούν συγκριτικά με τα παραδοσιακά αναδρομικά νευρωνικά δίκτυα (RNN). Μοντέλα όπως το echo state network (ESN) και η liquid state machine (LSM) χρησιμοποιούν τη δομή των RNN ως δεξαμενή με μη εκπαιδευμένες εσωτερικές συνδέσεις. Ένα πιο πρόσφατο μοντέλο RC με το όνομα delayed feedback reservoir computing (DFR) χρησιμοποιεί έναν μη γραμμικό νευρωνικό κόμβο με δυναμικό βρόχο καθυστέρησης και έχει αποδειχθεί ότι προσφέρει καλύτερη απόδοση από άλλα μοντέλα RC. Επιπλέον, υπάρχουν διάφορα μαθηματικά μοντέλα για τεχνητούς νευρώνες, με τα Spiking neural networks (SNN) να αντιπροσωπεύουν ένα παράδειγμα αυτών των μοντέλων. Τα SNN χρησιμοποιούν "spikes" ως σήματα επικοινωνίας μεταξύ νευρώνων και θεωρούνται πιο βιολογικά αληθοφανή σε σύγκριση με τα παραδοσιακά τεχνητά νευρωνικά δίκτυα.

Τα SNNs έχουν ένα επιπρόσθετο πλεονέκτημα, καθώς είναι πολύ πιο αποδοτικά στην ενέργεια και κατάλληλα για υλικές υλοποιήσεις. Πολλές ερευνητικές μελέτες έχουν αποδείξει ότι τα τεχνητά SNNs εξοικονομούν ενέργεια, καθώς χρησιμοποιούν μοντέλα νευρώνων που παράγουν σήματα "spike". Ωστόσο, η προσέγγιση που ο νευρώνας χρησιμοποιεί για την κωδικοποίηση της πληροφορίας παραμένει αμφίδρομη και απαιτεί περαιτέρω έρευνα. Υπάρχουν διάφορες μέθοδοι κωδικοποίησης, οι οποίες μπορούν να χωριστούν σε δύο βασικές κατηγορίες: κωδικοποίηση με βάση το ρυθμό και χρονική κωδικοποίηση. Ενώ οι μέθοδοι κωδικοποίησης με βάση το ρυθμό ήταν πιο δημοφιλείς στο παρελθόν, πρόσφατες έρευνες έχουν αποδείξει την αποτελεσματικότητα της χρονικής κωδικοποίησης έναντι αυτών. Στη χρονική κωδικοποίηση, ο ακριβής χρόνος των σημάτων "spike" χρησιμοποιείται για την αναπαράσταση της νευρικής πληροφορίας. Πολλά πειραματικά αποτελέσματα έχουν δείξει ότι ο ακριβής χρόνος των σημάτων "spike" χρησιμοποιείται από τους νευρώνες για την αναπαράσταση και τη μετάδοση της πληροφορίας.

Το κείμενο αναφέρει ότι οι χρονικές κωδικοποιήσεις είναι αποδοτικότερες από τις κωδικοποιήσεις ρυθμού στον τομέα της νευρωνικής εκπαίδευσης. Εξετάζει τον αλγόριθμο "precise-spike-driven synaptic plasticity" (PSD) που εστιάζει στην προσαρμογή των συνάψεων μεταξύ νευρώνων χρησιμοποιώντας χρονικούς κωδικούς, προκειμένου να μάθει τα χωροχρονικά μοτίβα. Αυτό αποτελεί αντίδραση στους αλγορίθμους που χρησιμοποιούν κωδικοποίηση ρυθμού για την εκπαίδευση των νευρωνικών δικτύων. Ο PSD αποτελεί έναν κανόνα που ρυθμίζει τα βάρη των συνάψεων μεταξύ νευρώνων για να επιτρέψει την καλύτερη προσαρμογή στα χωροχρονικά μοτίβα.

Σε αυτή την εργασία, παρουσιάζεται ο αλγόριθμος PSD, ο οποίος χρησιμοποιείται για να εκπαιδεύσει τα SNN σε στόχους όπως η ανίχνευση multi-period επιθέσεων σε ένα smart grid, καθώς και η εύρεση του βέλτιστου τρόπου κωδικοποίησης των spikes. Οι dynamic επιθέσεις, αντιπροσωπεύουν μια ειδική κατηγορία επιθέσεων, όπου ο επιτιθέμενος αλλάζει σταδιακά την κατάσταση του smart grid με την πάροδο του χρόνου για να φτάσει στην επιθυμητή κατάσταση.

Το βέλτιστο σχήμα κωδικοποίησης που προκύπτει κατά την εκπαίδευση του αλγορίθμου PSD θα χρησιμοποιηθεί στην εκπαίδευση του DFR+MLP με χρήση των spikes.

• PROBLEM FORMULATION

Στο πλαίσιο των επιθέσεων στα συστήματα smart grids, υπάρχουν δύο πιθανοί στόχοι που μπορούν να εξεταστούν: οι μετρήσεις και η δομή του δικτύου smart grid. Συγκεκριμένα, η τιμή κάθε μέτρησης σε ένα smart grid εξαρτάται από την κατάσταση του συστήματος, ακολουθεί μια γραμμική συνάρτηση και επηρεάζεται από τον θόρυβο του περιβάλλοντος. Αυτό μπορεί να εκφραστεί με τον ακόλουθο τρόπο:

$$z = Hx + n \quad (1)$$

όπου, z αντιπροσωπεύει το διάνυσμα μέτρησης, που περιλαμβάνει τα πραγματικά μέρη των ροών γραμμής και των εγχύσεων διαύλου. Το H είναι μια γραμμική συνάρτηση, το x είναι το διάνυσμα κατάστασης και το n αντιπροσωπεύει τον θόρυβο του περιβάλλοντος. Σε περίπτωση επίθεσης, η εξίσωση (1) μπορεί να τροποποιηθεί ως εξής:

$$\begin{aligned} \bar{z} &= z + a, \\ \bar{z} &= Hx + n + a \end{aligned} \quad (2)$$

όπου, a αντιπροσωπεύει το διάνυσμα της επίθεσης. Αυτού του είδους η επίθεση που αναφέρεται στην εξίσωση (2) ονομάζεται παρατηρήσιμη επίθεση (observable attack). Ωστόσο, είναι δυνατόν οι αντίπαλοι να υλοποιήσουν έναν διαφορετικό τύπο επίθεσης, που ονομάζεται μη παρατηρήσιμη ή κρυφή επίθεση (unobservable or hidden attack). Αυτή η κρυφή επίθεση είναι πολύ δυσκολότερο να ανιχνευθεί, ειδικά όταν χρησιμοποιούνται παραδοσιακές μέθοδοι ανίχνευσης επιθέσεων, όπως η SVE.

Στις κρυφές επιθέσεις, θεωρείται ότι ο αντίπαλος έχει πρόσβαση στον πίνακα H , και μπορεί να κρύψει την κατάσταση που επιθυμεί στον πίνακα H για να χειραγωγήσει το διάνυσμα μέτρησης z . Οι κρυφές επιθέσεις είναι πιο δύσκολο να εντοπιστούν, καθώς δεν μπορούν να παρατηρηθούν από το κέντρο ελέγχου και συλλογής δεδομένων (Supervisory Control and Data Acquisition - SCADA) του smart grid. Στην παρούσα εργασία, εξετάζουμε τις κρυφές dynamic επιθέσεις. Η κρυφή επίθεση ορίζεται ως $a = Hc$ και η (2) αναδιατυπώνεται ως εξής,

$$\begin{aligned} \bar{z} &= Hx + n + Hc \\ \bar{z} &= H(x + c) + n \end{aligned} \quad (3)$$

όπου, c είναι η επιθυμητή κατάσταση του αντιπάλου. Με αυτόν τον τρόπο, ο επιτιθέμενος επιδιώκει να αλλοιώσει την κανονική κατάσταση του έξυπνου δικτύου με σκοπό να την κρύψει στον πίνακα H .

Οι dynamic επιθέσεις εκτελούνται με τρόπο ώστε να επηρεάσουν σταδιακά την κατάσταση ενός smart grid με τρόπο που εξυπηρετεί τους σκοπούς του επιτιθέμενου. Προκειμένου να σχεδιαστεί αυτή η επίθεση, το διάνυσμα της επίθεσης πρέπει να οριστεί ως συνάρτηση του χρόνου, διότι σε αντίθεση με την single-period που ο αντίπαλος εκτελεί την επίθεση ευκαιριακά, στις dynamic η επίθεση εκτελείται σταδιακά και με την πάροδο του χρόνου. Η εξάρτηση της από το χρόνο καθιστά το μέγεθος των διακυμάνσεων των επιθέσεων μικρότερο και είναι πιο δύσκολο να ανιχνευθούν. Στις single-period επιθέσεις οι μεταβολές του μεγέθους των επιθέσεων είναι ξαφνικές και απότομες και ανιχνεύονται ευκολότερα. Η διατύπωση της dynamic επίθεσης που χρησιμοποιείται στην παρούσα εργασία έχει ως εξής,

$$\bar{z}(t) = Hx(t) + n + a(t) \quad (4)$$

Η dynamic επίθεση $a(t)$ εξαρτάται από το χρόνο και υποθέτουμε επίσης ότι ο αντίπαλος έχει πρόσβαση στον πίνακα H . Έτσι, η επίθεση μπορεί να πραγματοποιηθεί ως κρυφή. Στις κρυφές επιθέσεις η επίθεση $a(t)$ μπορεί να εκφραστεί ως $a(t) = Hc(t)$, και το $c(t)$ ορίζεται ως εξής,

$$c(t) = A \cos(2\pi f_c t) \times N(0, 1) \quad (5)$$

όπου, A είναι το μέγεθος της επίθεσης, $\cos$ είναι η συνάρτηση συνημιτόνου, f_c είναι η συχνότητα της επίθεσης και ορίζεται σε 1 σε αυτή τη μελέτη, προκειμένου να πραγματοποιηθεί μια dynamic επίθεση που μπορεί να χειραγωγήσει σταδιακά την κατάσταση του smart grid, $N(0,1)$ είναι ένα κανονικά κατανοημένο διάνυσμα με μηδενική μέση τιμή και διακύμανση 1.

Αντικαθιστώντας την (5) στην (4), προκύπτει η ακόλουθη εξίσωση:

$$\bar{z}(t) = H(x + A \cos(2\pi f_c t) \times N(0, 1)) + n \quad (6).$$

Ο στόχος αυτής της μελέτης είναι να ανιχνεύσουμε την ύπαρξη ή την απουσία κρυφών δυναμικών επιθέσεων σε smart grids. Για να το ανιχνεύσουμε αυτό, πρέπει να έχουμε δεδομένα μετρήσεων smart grids που μπορούν να παραβιαστούν από κρυφές δυναμικές επιθέσεις. Υποθέτουμε ότι ο επιτιθέμενος δεν έχει γνώση των παραμέτρων του ανιχνευτή επιθέσεων, οι οποίες σε αυτήν την περίπτωση είναι οι παράμετροι του DFR και του MLP. Ως εκ τούτου, ο επιτιθέμενος δεν μπορεί να βελτιστοποιήσει την επίθεσή του πριν την εκτέλεσή της με σκοπό τη μείωση της πιθανότητας ανίχνευσης της επίθεσης. Ωστόσο, ο επιτιθέμενος μπορεί να προσαρμόσει το μέγεθος της επίθεσής του για να μειώσει την πιθανότητα ανίχνευσης.

Όπως θα αναφέρουμε και παρακάτω, ενώ το μέγεθος της επίθεσης είναι μικρότερο, η ανίχνευση της επίθεσης είναι πιο δύσκολη και το αντίστροφο. Συγκεκριμένα, χρησιμοποιούμε ένα δυναμικό μοντέλο επίθεσης για να θέσουμε σε κίνδυνο τις μετρήσεις του smart grid, όπως περιγράφεται στο [90]. Για να προσομοιώσουμε αυτές τις μετρήσεις στο smart grid, χρησιμοποιούμε την εργαλειοθήκη MATPOWER και το σύστημα δοκιμών IEEE 14-bus [83]. Στο περιβάλλον δοκιμών του smart grid IEEE 14-bus, υπάρχουν συνολικά 34 διαφορετικοί μετρητές που μπορούν να παραβιαστούν από τον αντίπαλο. Το μοντέλο μας υποθέτει ότι ο αντίπαλος μπορεί να έχει ποικίλη πρόσβαση στο σύστημα. Συγκεκριμένα, ο αριθμός των μετρητών που μπορεί να παραβιάσει ο αντίπαλος κυμαίνεται από 0 έως 34. Δημιουργούμε συνολικά 10.000 δείγματα για την εκπαίδευση του μοντέλου μας και 10.000 ακόμα για δοκιμή και επικύρωση, χρησιμοποιώντας την εργαλειοθήκη MATPOWER και το IEEE 14-bus test bed.

• PRECISE SPIKE DRIVEN SYNAPTIC PLASTICITY

○ Neuron Model

Υπάρχει σημαντική χωροχρονική συσχέτιση μεταξύ των διαφόρων στοιχείων ενός smart grid. Επομένως, το PSD είναι μια εξαιρετική επιλογή για την ανίχνευση αυτής της χωροχρονικής συσχέτισης και της χρήσης της για τη βελτίωση της ανίχνευσης επιθέσεων FDI. Στο πλαίσιο αυτό, αναλύουμε το νευρωνικό μοντέλο, τον κανόνα μάθησης PSD και τη διαδικασία εκπαίδευσής του. Μέχρι στιγμής, έχουν προταθεί διάφορα spiking neuron μοντέλα που επιδιώκουν να προσομοιώσουν τη συμπεριφορά των βιολογικών νευρώνων. Επιλέγονται συνήθως τα μοντέλα Leaky-Integrate-and-Fire (LIF) και Hodgkin-Huxley [91] για τεχνητούς spiking neurons. Ο νευρώνας LIF χρησιμοποιείται ευρέως λόγω της απλοτητάς του και της ευκολίας υλοποίησής του σε υλικό [85]. Στην πραγματικότητα, όταν εφαρμόζεται διεγερτικό ρεύμα σε ένα νευρώνα, αυτός αρχίζει να πυροδοτείται μόλις η τάση της μεμβράνης του υπερβεί ένα συγκεκριμένο όριο. Στον νευρώνα LIF, αυτή η σχέση μεταξύ του ερεθίσματος και της τάσης της μεμβράνης εκφράζεται ως εξής:

$$\tau_m \frac{dv_m}{dt} = - (V_m - E) + (I_{noise} + I_s)R_m \quad (7)$$

όπου V_m αντιπροσωπεύει την τάση της μεμβράνης, $\tau_m = R_m C_m$ είναι η σταθερά χρόνου του νευρώνα, και R_m και C_m αναπαριστούν την αντίσταση και τη χωρητικότητα της μεμβράνης αντίστοιχα. Επιπλέον, το E αναπαριστά την resting τάση, το I_{noise} είναι ο θόρυβος υποβάθρου και το I_s αντιπροσωπεύει το προσυναπτικό ρεύμα ή ρεύμα ερεθίσματος (presynaptic or stimulus current) [92]. Συγκεκριμένα, τα R_m ορίζονται σε 1 mega ohm και το C_m σε 10 nF.

Η εξίσωση (9) περιγράφει τον τρόπο με τον οποίο η τάση στη μεμβράνη σχετίζεται με το προσυναπτικό ρεύμα. Ενώ η εξίσωση (10) αναφέρει τη σχέση ανάμεσα στο μετασυναπτικό ρεύμα που παράγεται από ένα νευρώνα LIF και τα προσυναπτικά ρεύματα. Συγκεκριμένα, το προσυναπτικό ρεύμα ενός νευρώνα αντιπροσωπεύεται ως ένα σταθμισμένο άθροισμα των μετασυναπτικών ρευμάτων που προέρχονται από τους προσαγωγούς νευρώνες, η εξίσωση (8) δίνει τον τύπο:

$$I_s(t) = \sum_{j=1} w_i I_{PSC}^j(t) \quad (8)$$

όπου, w_i είναι το βάρος του ρεύματος από τον ιοστό προσαγωγό νευρώνα, και το I_{PSC}^i αντιστοιχεί στο μετασυναπτικό ρεύμα από τον ιοστό προσαγωγό νευρώνα. Αυτό το μετασυναπτικό ρεύμα $I_{PSC}^i(t)$ μπορεί να εκφραστεί ως εξής:

$$I_{PSC}^i(t) = \sum_{t^j} \kappa(t - t^j) H(t - t^j) \quad (9)$$

όπου, H αντιπροσωπεύει τη συνάρτηση Heaviside, t^j είναι ο χρόνος κατά τον οποίο εμφανίζεται ένα spike, και η συνάρτηση $\kappa(t)$ ορίζεται ως εξής:

$$\kappa(t - t^j) = V_0 \times \left(\exp\left(-\frac{t-t^j}{\tau_s}\right) - \exp\left(-t \frac{t-t^j}{\tau_f}\right) \right) \quad (10)$$

Στην προκειμένη περίπτωση, οι τιμές των τ_s και τ_f ορίζονται σε 10 και 2,5 αντίστοιχα. Επιπλέον, το V_0 αναφέρεται ως ο κανονικοποιητικός παράγοντας, που βοηθά στη διατήρηση του μεγέθους του πυρήνα ρεύματος κάτω από 1.

○ PSD Learning Algorithm

Ο αλγόριθμος PSD βασίζεται στον κανόνα Widrow-Huff (WH), που περιγράφεται από την ακόλουθη εξίσωση [93]:

$$\Delta w_i = \eta u_i (y_d - y_o) \quad (11)$$

Εδώ, η προτείνεται είναι ένας θετικός αριθμός και καθορίζει την ταχύτητα εκπαίδευσης. Το u_i αναπαριστά την είσοδο, το y_d είναι η επιθυμητή έξοδος, και το y_o είναι η έξοδος που παράγεται από το δίκτυο. Το Δw_i καθορίζει τον ρυθμό ενημέρωσης των βαρών. Ο κανόνας WH δεν μπορεί να εφαρμοστεί εύκολα στην εκπαίδευση των SNN, επειδή το μοντέλο WH προτάθηκε αρχικά για να χειρίζεται συνεχή σήματα. Για τα spikes, οι είσοδοι και οι έξοδοι ορίζονται απλώς ως ο χρόνος

στον οποίο εμφανίζονται τα spikes. Λόγω αυτού του λόγου, η ενημερωμένη διατύπωση που παρουσιάζεται στην (11) δεν μπορεί να εφαρμοστεί άμεσα στα SNNs.

Εφόσον κάθε νευρώνας μπορεί να πυροδοτήσει αρκετά spikes, μια συνάρτηση ώθησης μπορεί να χρησιμοποιηθεί για να περιγράψει τα χρονικά spike trains που προκαλούνται από κάθε νευρώνα. Αυτή η συνάρτηση ώθησης δίνεται ως:

$$S(t) = \sum_j \delta(t - t^j) \quad (12)$$

όπου, δ είναι η συνάρτηση Dirac ή συνάρτηση παλμού, και t^j είναι ο χρόνος κατά τον οποίο εμφανίζεται το jοστό spike.

Οι ακολουθίες εισόδου (S_{in}), εξόδου (S_{out}) και των επιθυμητών spike trains (S_{des}) μπορούν επίσης να καθοριστούν με τον ακόλουθο τρόπο [92]:

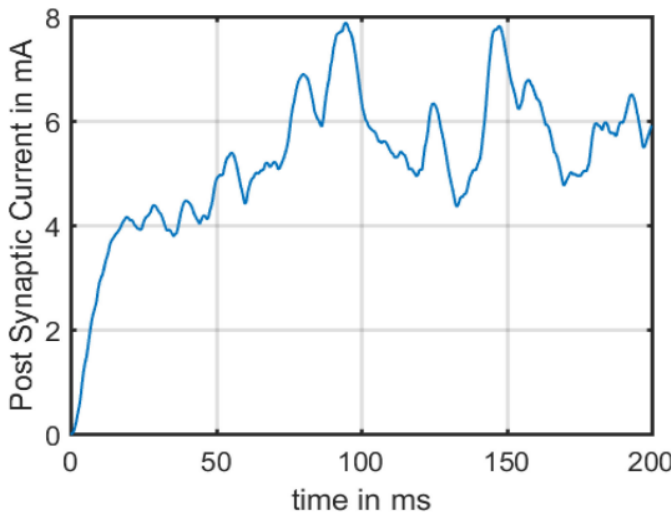
$$S_{in} = \sum_j \delta(t - t^j), \quad S_{out} = \sum_i \delta(t - t^i), \quad S_{des} = \sum_f \delta(t - t^f) \quad (13)$$

Η συνάρτηση Dirac στην είσοδο, την έξοδο και την επιθυμητή έξοδο καθιστά δύσκολη τη χρήση του παραδοσιακού κανόνα WH. Προκειμένου να επιλυθεί αυτό το ζήτημα, χρησιμοποιείται μια έννοια που ονομάζεται spike convolution. Η συνέλιξη ενός πυρήνα με διαφορετικά spike trains μετατρέπει τα spike trains σε real-valued συναρτήσεις. Με αυτόν τον τρόπο, ο κανόνας WH μπορεί να εφαρμοστεί και στα spike trains.

$$\bar{s}(t) = s(t) * \kappa(t) \quad (14)$$

όπου, $\kappa(t)$ είναι η συνάρτηση πυρήνα και προκύπτει από την (12). Με βάση αυτή την παραγωγή, το convolved spike έχει την ίδια μορφή με την I_{PSC} και η μόνη διαφορά είναι ότι η I_{PSC} είναι το σταθμισμένο άθροισμα των εισαγόμενων spike neurons.

Υπάρχει ένα spike train με την ονομασία $S_t(i)$, και μια συνελκτική έκδοση της $S_t(i)$ με την ονομασία $\bar{S}_t(i)$. Η συνελκτική έκδοση πραγματοποιείται με τη χρήση μιας συνάρτησης πυρήνα, όπως ορίζεται στην εξίσωση (10). Η $\bar{S}_t(i)$ με τη συνάρτηση πυρήνα μετατρέπει την $S_t(i)$ από ένα spike train σε ένα αναλογικό σήμα ρεύματος. Αυτή η μετατροπή είναι απαραίτητη για να επιτραπεί στους μετασυναπτικούς νευρώνες να επεξεργαστούν τις πληροφορίες που προέρχονται από τους προσυναπτικούς νευρώνες.



Σχήμα 1. Postsynaptic current after spike convolution.

Στο Σχήμα 1 παρουσιάζεται ένα μετασυναπτικό ρεύμα που παράγεται για τυχαία spikes εισόδου και τυχαία spikes εξόδου. Αυτό το αναλογικό ρεύμα είναι συνεχές και μπορεί να χρησιμοποιηθεί για την εκπαίδευση ενός SNN χρησιμοποιώντας τον κανόνα εκπαίδευσης WH. Μετά από αυτές τις τροποποιήσεις, η ενημερωμένη μορφή του κανόνα εκπαίδευσης WH εκφράζεται στην εξίσωση (15).

$$\frac{dw_i}{dt} = \eta(S_{des}(t) - S_o(t)) I_{PSC}^i(t) \quad (15)$$

Η παραπάνω εξίσωση δείχνει ότι εφαρμόζοντας τον πυρήνα που εισάγεται σε δύο διακριτά spike trains χρησιμοποιώντας την (14), μπορούμε να υπολογίσουμε την παράγωγο των βαρών που αντιστοιχεί στην ενημέρωση των βαρών σε κάθε επανάληψη.

○ ERROR FUNCTION

Η μετρική Rossum χρησιμοποιήθηκε στον αλγόριθμο PSD για να αξιολογηθεί το σφάλμα. Αυτή η μετρική υπολογίζει τον βαθμό διαφοράς μεταξύ της επιθυμητής εξόδου και της πραγματικής εξόδου που παράγεται από το δίκτυο, όπως περιγράφεται στον ακόλουθο τύπο [92]:

$$Dist = \frac{1}{\tau_s} \int_0^{+\infty} [h(t) - g(t)]^2 \quad (16)$$

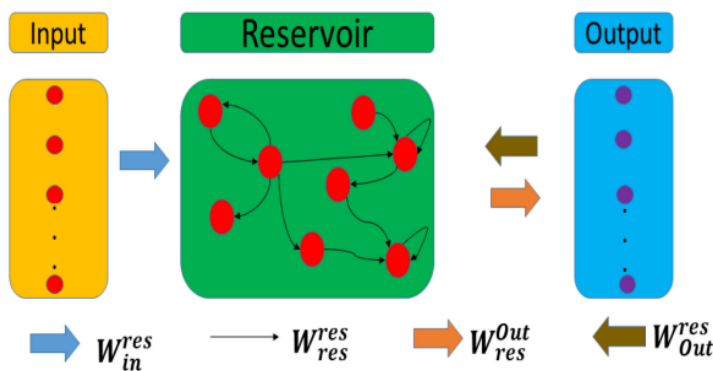
όπου, $h(t)$ αναφέρεται στη φιλτραρισμένη έκδοση της πραγματικής εξόδου, ενώ $g(t)$ αναφέρεται στη φιλτραρισμένη έκδοση της επιθυμητής εξόδου. Στο πλαίσιο της δικής μας μελέτης, θεωρούμε δύο διαφορετικά spike trains ως τις ετικέτες δύο διαφορετικών κλάσεων, δηλαδή των μετρήσεων που δέχονται επίθεση και αυτών που δεν δέχονται. Στην τρέχουσα έρευνα, η επιθυμητή έξοδος αντιστοιχεί στην ετικέτα του σχετικού εισαγόμενου σήματος. Κατά τη διαδικασία της εκπαίδευσης, η Dist δεν επηρεάζει την προσαρμογή των βαρών. Ωστόσο, λειτουργεί ως μέτρο ένδειξης ποιοτικής εκπαίδευσης, δείχνοντας πότε έχουμε επιτύχει αποτελεσματική εκπαίδευση.

● RC AND SPIKING DFR

Στο πλαίσιο αυτής της ενότητας, αρχικά παρουσιάζεται η δομή του RC, και στη συνέχεια εξηγείται ο τρόπος με τον οποίο μπορεί να τροποποιηθεί για να λειτουργεί ως spiking DFR.

○ DESIGN AND STRUCTURE

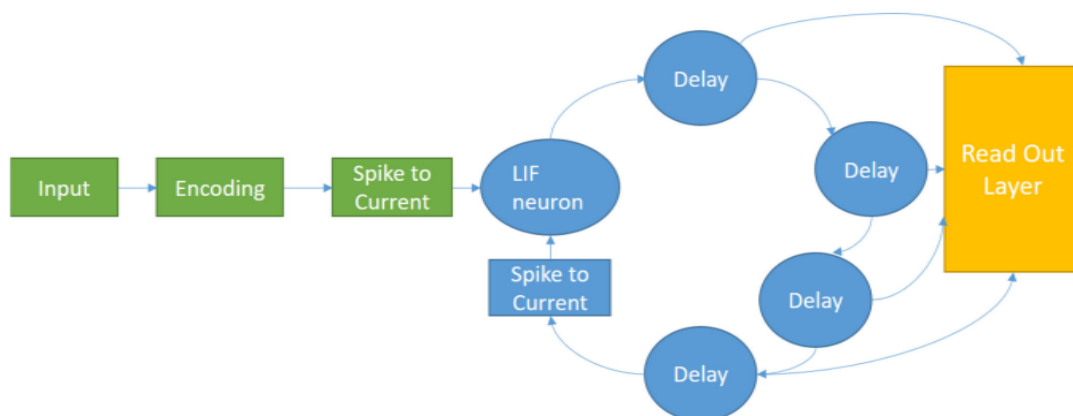
Στο παρελθόν, η λειτουργία του νευρώνα μοντελοποιούνταν ως γραμμική ολοκλήρωση πολλαπλών συναπτικών εισόδων ακολουθούμενη από μια μη γραμμικότητα threshold. Αλλά, τελευταίες ερευνητικές εξελίξεις στον τομέα της νευρωνικής information-processing υποδηλώνουν ότι η λειτουργία ενός νευρώνα είναι πολύ πιο περίπλοκη [94]. Αυτά τα δεδομένα υποδηλώνουν ότι η χωρική διάταξη και τα χαρακτηριστικά φιλτραρίσματος των βιολογικών συνάψεων και δενδριτών δημιουργούν χωροχρονικά χαρακτηριστικά που είναι αναπόφευκτα μη γραμμικά. Ακόμη, αυτές οι μη γραμμικές διαδικασίες ολοκλήρωσης στον χρόνο και τον χώρο επιτρέπουν σε ορισμένους νευρώνες να λειτουργούν σε superlinear ή sublinear καθεστώτα. Αυτές οι superlinear διεγερτικές εισροές μπορούν να μετασχηματίσουν νευρώνες σε ένα νευρωνικό δίκτυο με δύο επίπεδα που είναι ικανό να αντιμετωπίσει γραμμικά μη διαχωρίσιμες συναρτήσεις. Αυτή η δυνατότητα μη γραμμικής επεξεργασίας από έναν μόνο νευρώνα έχει προκαλέσει εκτεταμένο ενδιαφέρον στην έρευνα, όχι μόνο στον τομέα της νευροεπιστήμης, αλλά και στα τεχνητά νευρωνικά δίκτυα [95].



Σχήμα 2. Structure of reservoir computing

Το Σχήμα 2 παρουσιάζει ένα σύστημα που αποτελείται από τρία επίπεδα: επίπεδο εισόδου, δεξαμενή και εξόδου. Η δεξαμενή χρησιμοποιεί αναδρομικό νευρωνικό δίκτυο (RNN) χωρίς εκπαίδευση των συνδέσεων εντός αυτής και με τυχαίο διαμοιρασμό βαρών. Τα επίπεδα εισόδου λειτουργούν ως κλιμάκωση του εισαγόμενου σήματος και μεταφέρουν το κλιμακούμενο σήμα στη δεξαμενή, ενώ εντός της δεξαμενής, οι νευρώνες συνδέονται τυχαία, δημιουργώντας μια μη γραμμική χαρτογράφηση. Αυτή η αρχιτεκτονική βασίζεται σε ανάδραση και ανεκπαίδευτες συνδέσεις, με εκπαίδευση μόνο των εξόδων. Αυτό απλοποιεί τη διαδικασία εκπαίδευσης και απαιτεί λιγότερη υπολογιστική ισχύ και πολυπλοκότητα. Κάθε RC πρέπει να παρουσιάζει δύο βασικά χαρακτηριστικά: 1) υψηλή διαστατικότητα (high dimensionality) και 2) βραχυπρόθεσμη μνήμη (short-term memory). Το RC είναι ιδιαίτερα κατάλληλο για αναδυόμενες, μη συμβατικές περιπτώσεις υπολογισμού, αφού είναι οικονομικά αποδοτικό για εκπαίδευση και διαθέτει εξαιρετική ευελιξία στην υλοποίηση της αποθήκευσης.

Το μοντέλο DFR χρησιμοποιεί έναν μόνο νευρώνα και ένα delayed feedback για τη δημιουργία δεξαμενών με τοπολογία δακτυλίου (ring topology). Στο Σχήμα 3 παρουσιάζεται η δομή του DFR με τα εισαγόμενα spikes. Για την επεξεργασία των μετρήσεων από smart grids, πρέπει πρώτα να προσδιορίσουμε τα δεδομένα. Έπειτα, τα κωδικοποιημένα δεδομένα μετατρέπονται σε αναλογικό ρεύμα. Αυτό το ρεύμα εισέρχεται σε έναν μη γραμμικό κόμβο του DFR, όπου υπάρχει ένας νευρώνας LIF. Τα spikes που παράγονται από τον νευρώνα LIF διέρχονται από έναν βρόχο καθυστέρησης (delay loop), και για κάθε μέτρηση του smart grid παράγεται ένα spike train. Η διαδικασία αυτή επαναλαμβάνεται μέχρι να δημιουργηθούν αντίστοιχα spike trains για όλες τις μετρήσεις. Στο επόμενο στάδιο, τα interspike διαστήματα μεταξύ των spike trains χρησιμοποιούνται ως χαρακτηριστικά για την εκπαίδευση του επιπέδου ανάγνωσης, που στην περίπτωση μας είναι ένα MLP.



Σχήμα 3. Spiking delayed feedback reservoir computing.

Το MLP εκπαιδεύεται χρησιμοποιώντας χαρακτηριστικά που προέρχονται από το στρώμα δεξαμενής και έχει δύο διαφορετικές ετικέτες που αντιστοιχούν στις compromised και στις not-compromised μετρήσεις. Η εξίσωση που καθορίζει τον DFR εκφράζεται ως εξής:

$$x' = -x(t) + F(x(t - \tau), I(t), \theta) \quad (17)$$

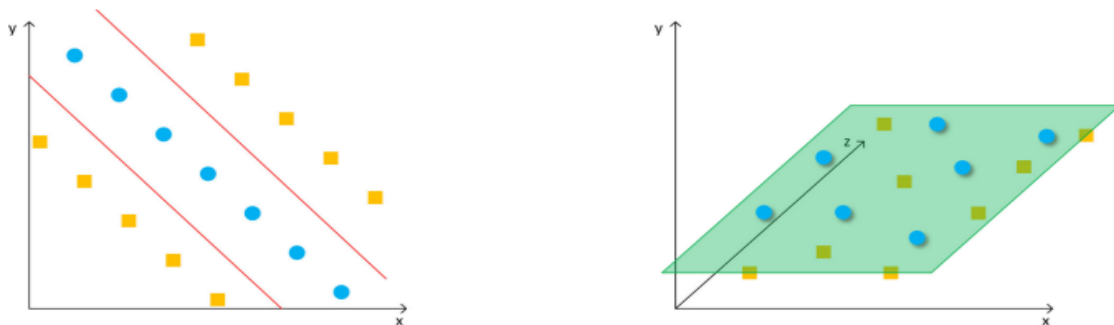
όπου το F είναι μια μη γραμμική differentiable συνάρτηση, γνωστή ως nonlinear mapping. Η τ είναι η καθυστέρηση, το $x(t)$ αναπαριστά τις καταστάσεις του DFR, και το $I(t)$ αναπαριστά το εισερχόμενο σήμα με ένα masking scheme που προστίθεται στον DFR. Στον βρόχο καθυστέρησης, ο συνολικός χρόνος καθυστέρησης, τ , διαιρείται σε N ισοδιάρκεις μονάδες καθυστέρησης, που μπορούν να εκφραστούν ως (18):

$$\tau = N\theta \quad (18),$$

όπου θ είναι το χρονικό διάστημα μεταξύ των εικονικών κόμβων. Αντίθετα με το RC, λόγω της τοπολογίας δακτυλίου του DFR, ο αριθμός των μη γραμμικών κόμβων μειώνεται σημαντικά. Η είσοδος εισάγεται απευθείας στο μη γραμμικό κόμβο και εκτελείται nonlinear mapping. Όπως και στο κλασικό RC, τα βάρη εξόδου εκπαιδεύονται ως τα μοναδικά βάρη [94].

○ HIGH DIMENSIONAL BEHAVIOR OF DFR

Ο DFR έχει αποσπάσει προσοχή στον τομέα της μηχανικής μάθησης, καθώς μπορεί να μετασχηματίσει δεδομένα σε έναν πολυδιάστατο χώρο, επιτρέποντας τη γραμμική διάκριση [94]. Αυτό επιτυγχάνεται χάρη στη χρήση της θεωρίας του χάους (chaos theory) και της ανάλυσης Lyapunov, που μετατρέπει τα δεδομένα χαμηλής διάστασης σε χώρο υψηλότερης διάστασης. Αρκετές μελέτες έχουν δείξει ότι η διάσταση Lyapunov ενός συστήματος χάους αντιστοιχεί άμεσα στην καθυστέρηση του βρόχου [96]. Στην εργασία αυτή προτείνεται η συνδυασμένη χρήση του DFR και του SNN για να εκμεταλλευτούν τα πλεονεκτήματα και των δύο μεθόδων. Η εφαρμογή του DFR στα δεδομένα βελτιώνει τη γραμμική διάκριση μεταξύ των δεδομένων που έχουν δεχθεί επίθεση και αυτών που δεν έχουν, ενώ το SNN εκμεταλλεύεται τη χωροχρονική συσχέτιση στα στοιχεία των smart grids. Το κείμενο εξετάζει επίσης τη συμπεριφορά του DFR για διάφορες καθυστερήσεις και επιλέγει την κατάλληλη καθυστέρηση για τη χρήση στις προσομοιώσεις. Όπως φαίνεται στο Σχήμα 4, τα αρχικά δεδομένα είναι δύσκολο να διαχωριστούν γραμμικά, αλλά με τον μετασχηματισμό σε χώρο υψηλότερης διάστασης γίνεται πιο εύκολο. Αυτό αποτελεί το κύριο κίνητρο για τη χρήση DFR στην εργασία αυτή.



Σχήμα 4. High dimensional mapping of data using DFR.

• TRAINING PSD FOR FDI ATTACK DETECTION

Σε αυτή την ενότητα, παρουσιάζουμε διαφορετικά συστήματα κωδικοποίησης των spikes και περιγράφουμε τον τρόπο αξιοποίησής τους για την κωδικοποίηση των μετρήσεων του smart grid. Στη συνέχεια χρησιμοποιούμε τις κωδικοποιημένες μετρήσεις για την εκπαίδευση του SNN χρησιμοποιώντας τον αλγόριθμο PSD. Κωδικοποίηση των πληροφοριών και η εκμάθηση είναι τα δύο κύρια συστατικά στοιχεία κάθε SNN [18]. Στο παρούσα εργασία, διερευνώνται τρία διαφορετικά σχήματα κωδικοποίησης για την κωδικοποίηση των μετρήσεων που παράγονται από τις προσομοιώσεις του smart grid. Δηλαδή, η latency, η latency-phase και η ISI (Inter-Spike Interval) και χρησιμοποιούνται αντίστοιχα για εκπαίδευση των SNN με τη χρήση του PSD.

○ LATENCY ENCODING

Στην περίπτωση της latency ή time to first spike (TTFS) κωδικοποίησης, ένα ερέθισμα κωδικοποιείται ως ο χρόνος, κατά τον οποίο το πρώτο spike πυροδοτείται σε σχέση με το ερέθισμα, δηλαδή η latency μεταξύ της έναρξης του ερεθίσματος και του πρώτου παρατηρούμενου spike [97]. Σύμφωνα με την εξίσωση (19), η latency υπολογίζεται ως εξής[91]:

$$Latency = t_i = t_{max} - \ln(\alpha \cdot s_i + 1) \quad (19)$$

όπου, t_{max} είναι το μέγιστο μήκος του παραθύρου κωδικοποίησης, α είναι ο παράγοντας κλιμάκωσης, και s_i αντιπροσωπεύει την ένταση κάθε εικονοστοιχείου. Στην προκειμένη περίπτωση, t_{max} ορίζεται σε 200 ms, το s_i αντικαθίσταται από μετρήσεις smart grid, και το α ορίζεται ως -1 όταν η μέτρηση είναι αρνητική και ως 1 σε διαφορετική περίπτωση. Η εξίσωση (19) προέρχεται από την παρατήρηση ότι τα ισχυρότερα ερεθίσματα οδηγούν σε πιο γρήγορη αντίδραση, ενώ τα ασθενέστερα ερεθίσματα προκαλούν μεγαλύτερη καθυστέρηση στην απόκριση των νευρώνων [98].

○ LATENCY-PHASE ENCODING

Στο [97] έχει αποδειχθεί ότι ο συνδυασμός ή η πολυπλεξία διαφορετικών προσεγγίσεων κωδικοποίησης μπορεί να αποδώσει υψηλότερη χωρητικότητα πληροφορίας. Υπάρχουν επίσης άλλες ενδείξεις που υποδεικνύουν ότι η πολυπλεξία διαφορετικών συστημάτων κωδικοποίησης συμβαίνει στον εγκέφαλο και αυτό εξηγεί γιατί ο εγκέφαλος είναι αποτελεσματικός στη διάκριση διαφορετικών αντικειμένων. Εμπνευσμένος από αυτήν την έρευνα, ο Nadasdy [99] πρότεινε ένα μηχανισμό που συνδυάζει latency με τον κώδικα φάσης (phase code). Ο σχετικός χρονισμός (relative timing) ενός spike σε σχέση με την εσωτερική subthreshold membrane potential oscillation (SMOs) είναι το χαρακτηριστικό που καθορίζει τον κώδικα φάσης ενός spike [99].

Οι SMOs μπορούν απλά να οριστούν ως εξής για τη μέτρηση του ισοστού μετρητή στο smart grid,

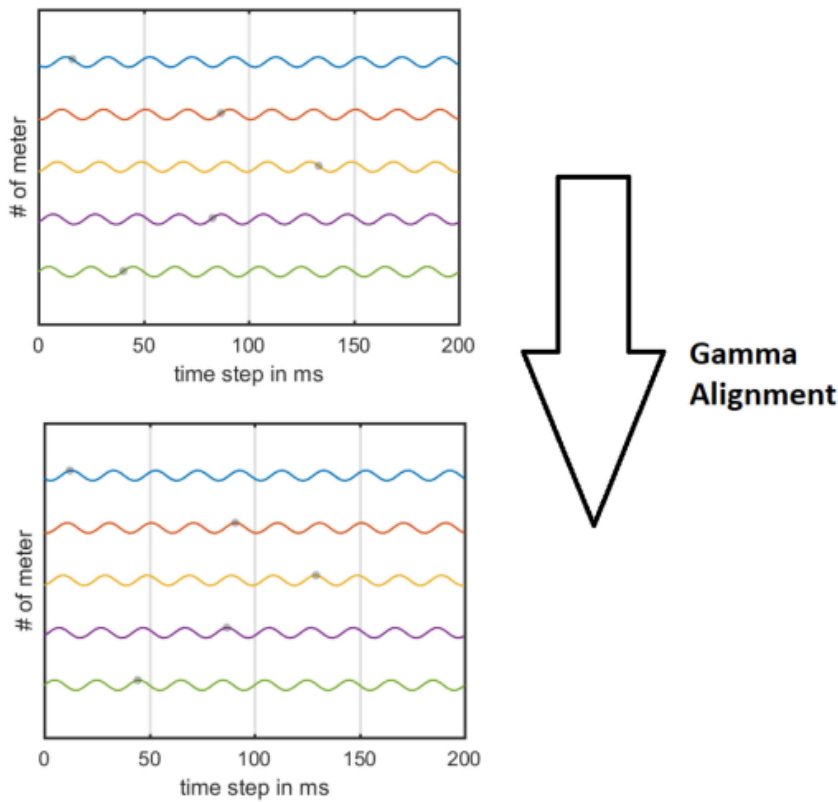
$$SMO_i = A \cos(2\pi f t + \varphi_i), \quad i = 1, 2, \dots, L \quad (20)$$

όπου, A είναι το μέγεθος του SMO και στην παρούσα εργασία ορίζεται σε 1, f είναι η συχνότητα ταλάντωσης που στην περίπτωσή μας ορίζεται σε 0,05, L είναι ο συνολικός αριθμός των μονάδων κωδικοποίησης που στην παρούσα εργασία είναι ίσος με 34 που αντιπροσωπεύει το συνολικό αριθμό των μετρητών στο σχεδιασμό του smart grid που χρησιμοποιήσαμε, φ_i είναι η αρχική φάση του i -οστού μετρητή και παρουσιάζεται ως η ακόλουθη εξίσωση,

$$\varphi_i = \varphi_0 + (i - 1)\Delta\varphi \quad (21)$$

όπου, ϕ_0 αντιστοιχεί στη φάση αναφοράς και $\Delta\phi$ είναι η διαφορά φάσης μεταξύ γειτονικών μετρητών smart grid.

Στα [91], [92] ορίζεται ότι το $\Delta\phi$ ισούται με $2\pi/N_{en}$, όπου N_{en} είναι ο συνολικός αριθμός pixels. Για την περίπτωση μας, το $\Delta\phi$ ισούται με $2\pi/L$, με L να αντιπροσωπεύει τον συνολικό αριθμό μετρητών (34 μετρητές). Η ευθυγράμμιση γάμμα (Gamma alignment) αποτελεί τη διαδικασία που συνδέει τη latency με τον κώδικα φάσης. Κατά τη διαδικασία αυτή, η latency συγχρονίζεται με τον αντίστοιχο SMO κάθε μετρητή, όπως παρουσιάζεται στο Σχήμα 5. Στην πραγματικότητα, οι latency ακολουθούν αυτή τη μέθοδο, ούτως ώστε να βρίσκονται στην κορυφή του αντίστοιχου SMO για κάθε μετρητή. Ο στόχος είναι να συνδυαστούν ο κώδικας φάσης και η latency για βελτίωση της εκπαίδευσης του SNN χρησιμοποιώντας τον κώδικα PSD.



Σχήμα 5. Aligning latency code with SMO of each meter using Gamma alignment.

○ ISI ENCODING

Το ISI αποτυπώνει το χρονικό διάστημα μεταξύ διαδοχικών spikes. δηλαδή, αντί να χρησιμοποιείται ο απόλυτος χρονισμός κάθε spike, για την κωδικοποίηση του ερεθίσματος χρησιμοποιείται η σχετική απόσταση κάθε spike σε σχέση με άλλα. Αυτό επιτυγχάνεται μέσω της παρακάτω μαθηματικής σχέσης:

$$D_i = f(C_i, V_i) - f(C_{i-1}, V_{i-1}) \quad (22)$$

όπου η συνάρτηση $f(C_i, V_i)$ ορίζεται ως εξής:

$$f(C_i, V_i) = (C_{i+1}) [\beta(V_i - \gamma) + \theta] \quad (23)$$

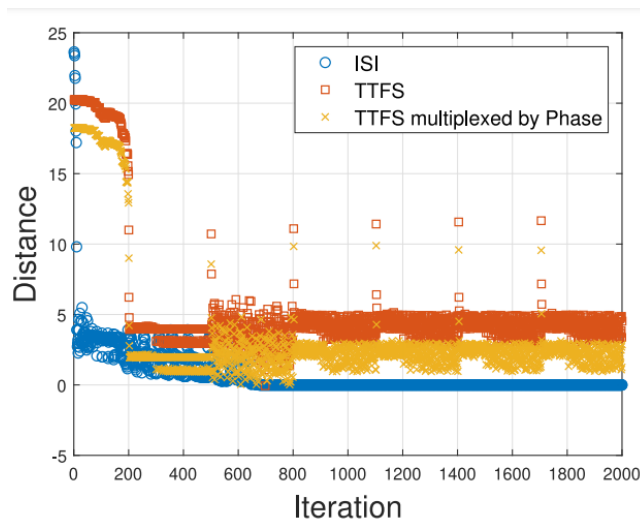
Στις παραπάνω εξισώσεις, C_i, V_i αντιστοιχούν στις παραμέτρους του νευρώνα LIF που εκτελεί την κωδικοποίηση και αντιπροσωπεύουν τη χωρητικότητα της μεμβράνης και το

threshold.[85][100] Για λόγους απλότητας, ο αριθμός των παραγόμενων spikes από τη μονάδα κωδικοποίησης ορίζεται ως 4. Αυτό σημαίνει ότι υπάρχουν συνολικά τρία interspike διαστήματα, τα οποία θα χρησιμοποιηθούν αργότερα για την εκπαίδευση του SNN με PSD (Postsynaptic Density).

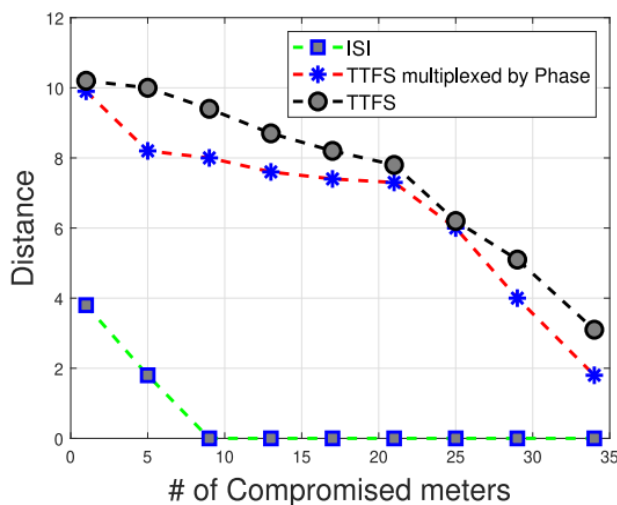
• PERFORMANCE AND ANALYSIS

○ USING SNN FOR DYNAMIC ATTACK DETECTION

Στις προσομοιώσεις μας, θεωρούμε ότι ο αριθμός των παραβιασμένων μετρητών κυμαίνεται από 0 έως 34. Στο Σχήμα 6 παρουσιάζεται το μέσο σφάλμα εκπαίδευσης για τα δεδομένα που υπέστησαν επίθεση. Τα αποτελέσματα δείχνουν ότι η πολυπλεξία TTFS ανά φάση μπορεί να βελτιώσει την απόδοση της εκπαίδευσης και ότι η πολυπλεξία διαφορετικών συστημάτων κωδικοποίησης οδηγεί σε καλύτερη ταξινόμηση. Ωστόσο, η μέθοδος ISI παραμένει ακόμα αποδοτικότερη από την πολυπλεξία TTFS ανά φάση, με το σφάλμα εκπαίδευσης του PSDS με χρήση ISI να είναι καλύτερο από άλλες μεθόδους, όπως παρατηρείται στο Σχήμα 7.



Σχήμα 6. Error of training for different encoding schemes.



Σχήμα 7. Error of training for different encoding schemes & different number of compromised meters.

Περιγράφεται μια προσομοίωση όπου χρησιμοποιούνται 10 νευρώνες σε ένα νευρωνικό δίκτυο για να αναγνωρίσουν επιθέσεις. Οι προσομοιώσεις διαρκούν 2000 επαναλήψεις, και η

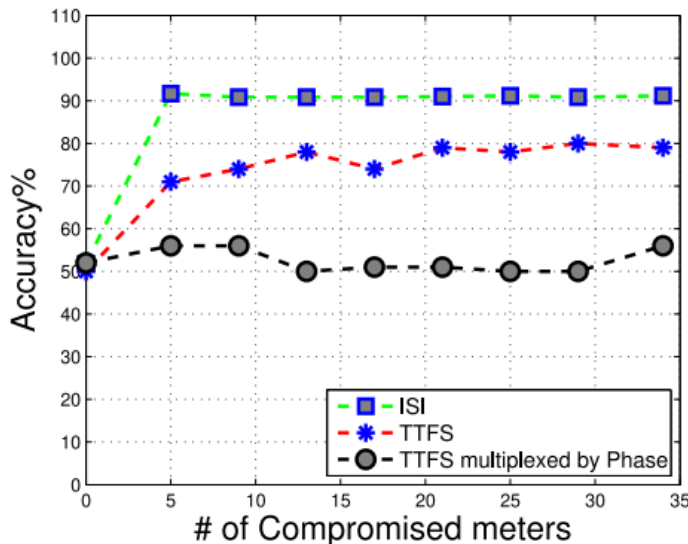
αξιολόγηση γίνεται με βάση το σφάλμα μεταξύ του παραγόμενου spike train και ενός spike train αναφοράς για κάθε επίθεση. Ο αλγόριθμος PSD χρησιμοποιείται για την εκπαίδευση του SNN. Το σφάλμα ορίζεται ως η απόκλιση μεταξύ του αναμενόμενου spike train και του παραγόμενου spike train. Κατά την επίθεση, οι νευρώνες πρέπει να εκπέμπουν συγκεκριμένα spike train στην έξοδο, ανάλογα με τον αριθμό των παραβιασμένων μετρητών. Όπως παρουσιάζεται και στο Σχήμα 7, η μελέτη δείχνει ότι όσο αυξάνεται ο αριθμός των παραβιασμένων μετρητών, το σφάλμα εκπαίδευσης μειώνεται.

Στις προσομοιώσεις μας, καθορίσαμε την ταχύτητα μάθησης σε 0,01. Οι αρχικές τιμές των βαρών προέρχονται από μια κανονική κατανομή με μέση τιμή 0,5 και διακύμανση 0,2. Σχεδιάσαμε τους νευρώνες εξόδου έτσι ώστε να εκπέμπουν τρεις παλμούς σε τρεις διαφορετικές χρονικές στιγμές για τα δείγματα κάθε κατηγορίας δεδομένων. Τα αποτελέσματα δείχνουν ότι τα SNN, ειδικά όταν εκπαιδεύονται χρησιμοποιώντας κωδικοποίηση ISI, έχουν αξιόλογη δυνατότητα να ανιχνεύουν επιθέσεις FDI. Ένα μειονέκτημα της κωδικοποίησης ISI είναι ότι προκαλεί πολλαπλούς παλμούς ανά μονάδα χρόνου, αντίθετα με την latency που προκαλεί μόνο έναν παλμό ανά μονάδα χρόνου. Αυτό δυσκολεύει υπολογιστικά την διαδικασία και αυξάνει τον χρόνο εκπαίδευσης. Η μετρική που χρησιμοποιείται για να αξιολογηθεί η απόδοση της ανίχνευσης είναι η ακρίβεια, η οποία υπολογίζεται όπως εξηγείται στην εξίσωση (24),

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (24)$$

όπου TP, TN, FP και FN αντιστοιχούν στον αριθμό των true positive, true negative, false positive και false negative πειραμάτων αντίστοιχα.

Όπως παρατηρούμε στο Σχήμα 8, η χρήση της κωδικοποίησης ISI οδηγεί σε υψηλότερο ποσοστό ανίχνευσης επιθέσεων, με βάση τη μετρική ακρίβειας. Δοκιμές που πραγματοποιούνται με δεδομένα που δεν έχουν χρησιμοποιηθεί κατά την εκπαίδευση επιβεβαιώνουν αυτήν την αποτελεσματικότητα, ενώ δείχνουν ότι η κωδικοποίηση ISI είναι ανώτερη από άλλες μεθόδους κωδικοποίησης. Συνεπώς, αν και η πολυπλεξία μπορεί να βελτιώσει την κωδικοποίηση TTFS, δεν μπορεί να ξεπεράσει την αποδοτικότητα της κωδικοποίησης ISI.



Σχήμα 8. Test results.

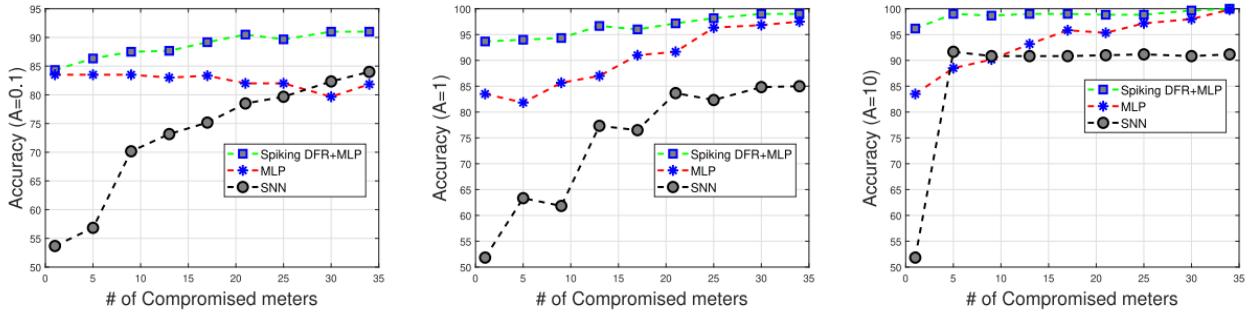
○ ATTACK DETECTION USING DFR

Σε αυτή την ενότητα χρησιμοποιείται το DFR με spikes για την ανίχνευση των dynamic επιθέσεων. Τα δεδομένα που χρησιμοποιούμε είναι μη ισορροπημένα. Αυτό σημαίνει ότι η αναλογία των εκτεθειμένων και μη εκτεθειμένων δειγμάτων δεν είναι ίση. Στο σύνολο δεδομένων μας, το 20% των δειγμάτων είναι εκτεθειμένα και το 80% δεν είναι. Οι μετρικές απόδοσης για την αξιολόγηση είναι η ακρίβεια και το F1. Η ακρίβεια ορίστηκε στην (24) και η F1 ορίζεται,

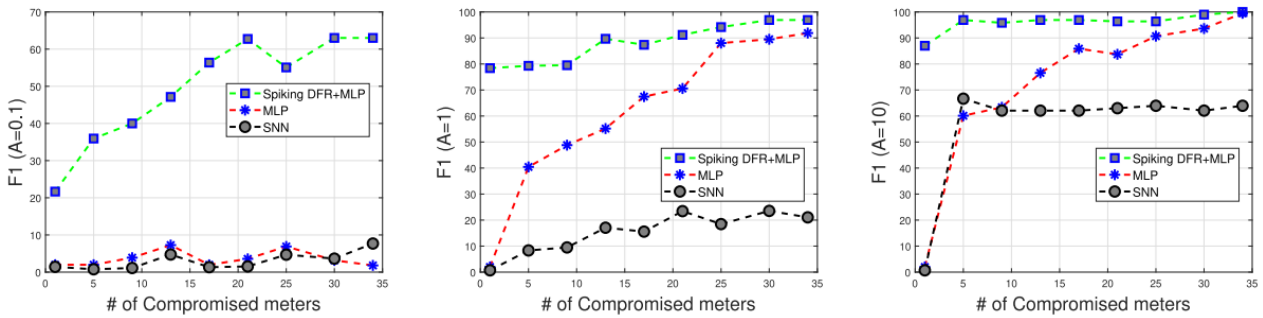
$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (25)$$

όπου $Precision = \frac{TP}{TP + FP}$, και $Recall = \frac{TP}{TP + FN}$.

Τα Σχήματα 9 και 10 παρουσιάζουν τη συνδυαστική χρήση spike νευρώνων, DFR και MLP σε ένα πείραμα, αποδεικνύοντας ότι αυτή η συνδυαστική προσέγγιση έχει υψηλότερη απόδοση σε σύγκριση με τη χρήση MLP και SNN χωριστά. Η απόδοση βελτιώνεται όσο αυξάνεται το μέγεθος της επίθεσης, λόγω της ικανότητας του DFR να προσομοιώνει RNN και ανιχνεύει χωροχρονικές συσχετίσεις. Στην αρχιτεκτονική DFR+MLP, το MLP λειτουργεί ως στρώμα ανάγνωσης του DFR. Ο μη γραμμικός κόμβος του DFR είναι ένας νευρώνας LIF που χρησιμοποιείται για τη μετατροπή της κωδικοποιημένης εισόδου σε ρεύμα, προς εφαρμογή στο νευρώνα LIF.



Σχήμα 9. Accuracy of attack detection for three different methods and magnitude of attacks, $A = 0.1, 1, 10$.



Σχήμα 10. F1 of attack detection for three different methods and magnitude of attacks, $A = 0.1, 1, 10$.

Η εξίσωση (9) χρησιμοποιείται για τη μετατροπή κωδικοποιημένης εισόδου σε ρεύμα, το οποίο εισάγεται σε ένα νευρώνα LIF, παράγοντας train spikes. Αυτά τα train spikes καθυστερούν χρονικά και μετατρέπονται ξανά σε ρεύμα, το οποίο προστίθεται στο ρεύμα της επόμενης εισόδου. Η διαδικασία αυτή επαναλαμβάνεται για όλα τα δείγματα, δημιουργώντας train spikes για κάθε είσοδο. Τα spikes αυτά έχουν διαφορετική χρονική συχνότητα και χρησιμοποιούνται ως χαρακτηριστικό γνώρισμα για την εκπαίδευση ενός MLP. Το MLP δέχεται το χρόνο εμφάνισης των spikes ως είσοδο και παράγει έξοδο 0 ή 1, ανάλογα με την κατηγορία των δεδομένων, όπου η

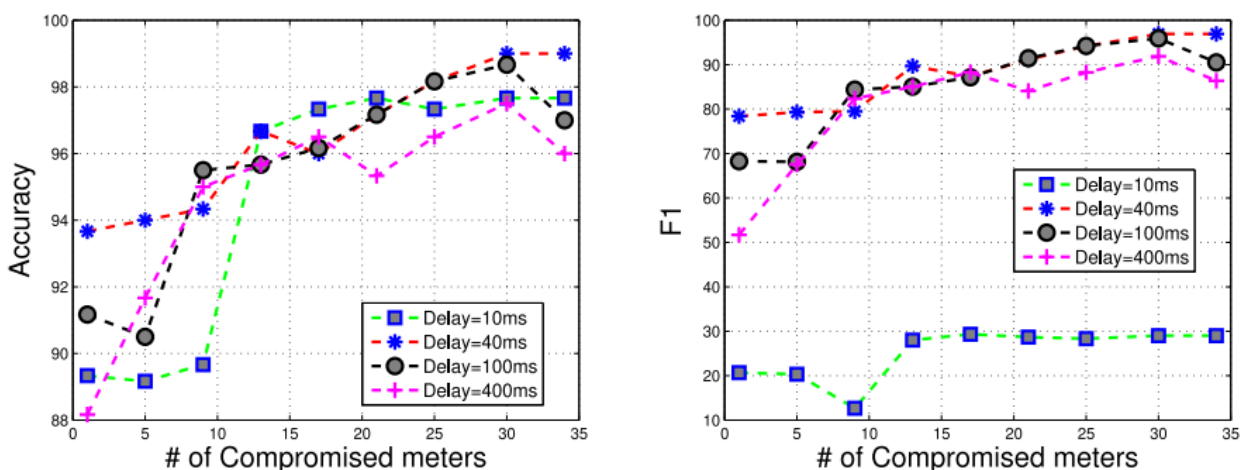
έξοδος είναι 1 για επιθέσεις και 0 για άλλα δεδομένα. Μετά την εκπαίδευση, το DFR χρησιμοποιεί τα δεδομένα που δεν έχει δει προηγουμένως για δοκιμές.

Χρησιμοποιείται ένα σύνολο δεδομένων με 10.000 δείγματα εκπαίδευσης και 10.000 δείγματα δοκιμής, όπου το 20% είναι συμβιβασμένα και το 80% δεν είναι. Οι διαφορετικές αναλογίες εκτεθειμένων και μη εκτεθειμένων δειγμάτων δημιουργούν μη ισορροπημένα δεδομένα. Η απόδοση των τριών μεθόδων (DFR+MLP, MLP και SNN), Σχήματα 9 και 10, συγκρίνεται με βάση τον αριθμό των προσβεβλημένων μετρήσεων και το μέγεθος της επίθεσης. Όσο αυξάνεται το μέγεθος της επίθεσης, η απόδοση βελτιώνεται και στα τρία μοντέλα. Το DFR+MLP με spikes επιτυγχάνει την καλύτερη απόδοση, ενώ για χαμηλά μεγέθη επιθέσεων, τόσο το MLP όσο και το SNN έχουν χαμηλά ποσοστά απόδοσης. Ωστόσο, το DFR+MLP είναι ακριβέστερο στην ανίχνευση επιθέσεων, ακόμα και για χαμηλά μεγέθη επίθεσης. Τα αποτελέσματα της προσομοίωσης καταδεικνύουν ότι η μέση ακρίβεια ανίχνευσης επιθέσεων αυξάνεται έως και 94,6% όταν συνδυάζονται οι spikes νευρώνες, ο DFR και το MLP σε μια ενιαία πλατφόρμα.

Στην απλούστερη περίπτωση με μόνο SNNs, η μέθοδος βελτιώνει τη μέση ακρίβεια κατά περίπου 17%, ενώ η βελτίωση στο F1 είναι ακόμη πιο σημαντική. Στην περίπτωση χρήσης μόνο SNNs για ανίχνευση επιθέσεων, το μέτρο F1 είναι μόλις 25%, ενώ η προτεινόμενη μέθοδος επιτυγχάνει αύξηση κατά περίπου 53%. Επιπλέον, αναφέρεται ότι οι SNNs εμφανίζουν χαμηλή απόδοση όταν το μέγεθος της επίθεσης είναι χαμηλό, ενώ η μέθοδος που παρουσιάζεται είναι ικανή να αντιμετωπίσει αυτό το πρόβλημα λόγω της χωροχρονικής συσχέτισης των δεδομένων και της χαρτογράφησης σε χώρο υψηλότερων διαστάσεων που προσφέρει.

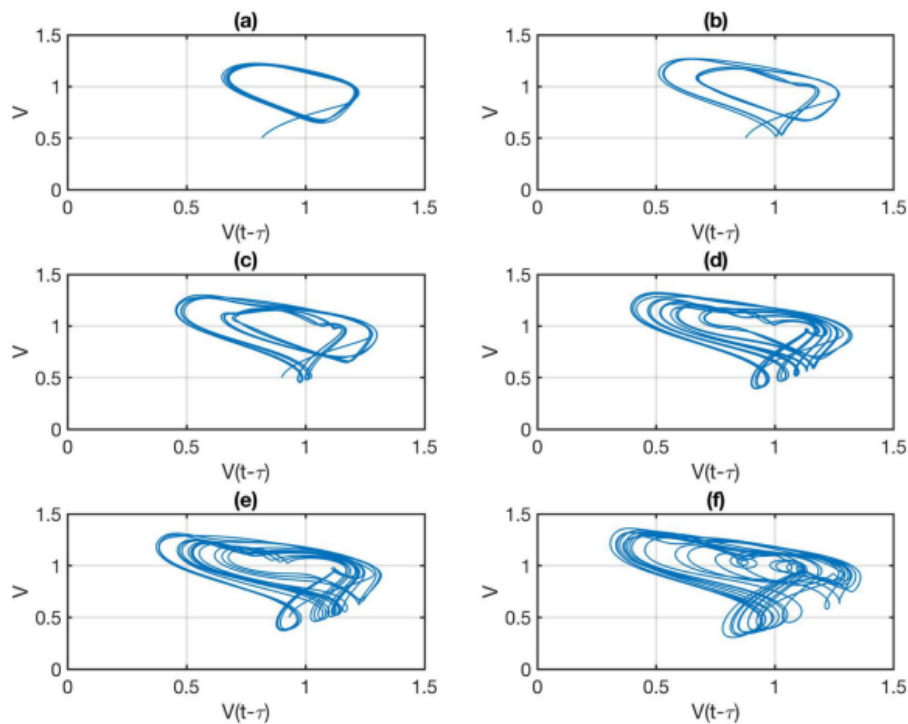
○ EFFECT OF DELAY ON THE PERFORMANCE

Στο Σχήμα 11 παρουσιάζονται τα αποτελέσματα της αξιολόγησης του DFR+MLP καθώς παραλλάσσεται η τιμή της χρονικής καθυστέρησης. Όταν η καθυστέρηση ρυθμίζεται στα 40 ms, παρατηρούμε υψηλές επιδόσεις σε όρους ακρίβειας και F1. Αντίθετα, όταν η καθυστέρηση μειώνεται στα 10 ms, παρατηρούμε χαμηλότερες επιδόσεις. Αυτό συμβαίνει επειδή απαιτείται η κατάλληλη ρύθμιση της χρονικής καθυστέρησης για να διατηρηθεί η λειτουργία του spiking DFR+MLP στα όρια του χάους.



Σχήμα 11. Effects of different values of the delay on the performance when the A = 1.

Το Σχήμα 12 αποτυπώνει τις αλλαγές στα σχήματα φάσης με την αύξηση της χρονικής καθυστέρησης. Καθώς η καθυστέρηση αυξάνεται, η δυναμική συμπεριφορά εκτείνεται από διατεταγμένη έως την άκρη του χάους (ordered to the edge of chaos) έως εντελώς χαοτική (completely chaotic). Μελέτες που αναφέρονται στο [96] υποδηλώνουν ότι, για να επιτευχθούν υψηλές επιδόσεις σε ένα δυναμικό σύστημα χωρίς διαστάσεις, η χρονική καθυστέρηση πρέπει να ρυθμιστεί με τρόπο που να επιτρέπει τη λειτουργία στα όρια του χάους. Για την περαιτέρω εξέταση της δυναμικής συμπεριφοράς, διερευνάται η λύση της εξίσωσης DDE. Η δυναμική συμπεριφορά της μη γραμμικής συνάρτησης μοντελοποιείται από την DDE με μεταβαλλόμενο χρόνο καθυστέρησης, όπως φαίνεται στο Σχήμα 12. Το σχήμα φάσης είναι μια αναπαράσταση των λύσεων που παρακολουθεί την πορεία κάθε συγκεκριμένης λύσης. Είναι ένα γραφικό εργαλείο για την απεικόνιση του τρόπου με τον οποίο οι λύσεις ενός συγκεκριμένου συστήματος διαφορικών εξισώσεων θα συμπεριφέρονταν μακροπρόθεσμα. Το σχήμα φάσης παρακολουθεί τη δυναμική συμπεριφορά και καταδεικνύει τη χαοτική ή περιοδική συμπεριφορά του συστήματος.



Σχήμα 12. Phase portraits of dynamic systems.

○ COMPARISON WITH CLASSICAL ALGORITHMS

Η SVE είναι η πρώτη μέθοδος για την ανίχνευση επιθέσεων στα smart grids [87]. Αυτή η μέθοδος βασίζεται στον υπολογισμό του υπολοίπου chi-squared που δηλώνεται ως ρ . Εάν η υπολογισμένη τιμή του ρ είναι μεγαλύτερη από μια προκαθορισμένη τιμή threshold, τότε το διάνυσμα μέτρησης z , ταξινομείται ως ένα που δεχθεί επίθεση. Η τιμή του υπολοίπου, ρ , ορίζεται ως εξής,

$$\rho = \|\bar{z} - \hat{H}\hat{x}\|_2^2 \quad (26)$$

όπου $\hat{x}$ είναι η εκτιμώμενη κατάσταση του smart grid και μπορεί να υπολογιστεί ως εξής,

$$\hat{x} = (H^T W H^{-1}) H^T W_z \quad (27)$$

όπου, W είναι ένας διαγώνιος πίνακας και τα διαγώνια στοιχεία του είναι το αντίστροφο της διακύμανσης των μετρήσεων. Το ισοτό διαγώνιο στοιχείο του W αντιστοιχεί στο αντίστροφο της διακύμανσης του ισοτού μετρητή.

Η SVE υλοποιείται πολύ εύκολα. Ωστόσο, αντιμετωπίζει ένα σημαντικό μειονέκτημα κατά την αντιμετώπιση κρυφών επιθέσεων. Στην περίπτωση κρυφών επιθέσεων, όπως θα δείξουμε και θα αποδείξουμε στην (28), η τιμή του υπολοίπου είναι πάντα μικρότερη από την τιμή threshold. Επομένως, η SVE δεν μπορεί να είναι χρήσιμη για κρυφές επιθέσεις.

$$\begin{aligned} \rho &= ||\bar{z} - H\hat{x}||_2^2 \\ &= ||z + z - H((H^T W H^{-1}) H^T W_z)||_2^2 \\ &= ||z - Hx + Hc - H((H^T W H^{-1}) H^T Hc)||_2^2 \\ &= ||z - Hx + Hc - HC||_2^2 \\ &= ||z - Hx||_2^2 \leq \tau \quad (28) \end{aligned}$$

Όπως φαίνεται στην (28), όταν η επίθεση είναι κρυφή η τιμή του υπολοίπου, ρ , είναι πάντα μικρότερη από την τιμή threshold, τ . Επομένως, η κρυφή επίθεση δεν μπορεί να ανιχνευθεί από την SVE. Δεδομένου ότι, στην παρούσα εργασία θεωρείται επίσης ότι η επίθεση είναι κρυφή, αυτό σημαίνει ότι η SVE δεν αποτελεί καλή επιλογή για εμάς.

ο COMPLEXITY ANALYSIS

Σε αυτή την ενότητα, αναλύεται η πολυπλοκότητα της προσέγγισής μας και συγκρίνεται με άλλες μεθόδους. Η υπολογιστική πολυπλοκότητα της ανίχνευσης δυναμικής επίθεσης βασίζεται στο spiking DFR και σχετίζεται με το Reservoir Computing και την ενημέρωση των βαρών του στρώματος ανάγνωσης. Στην προσέγγιση με βάση το spiking DFR, μόνο τα βάρη του στρώματος ανάγνωσης χρειάζονται εκπαίδευση, ενώ στα RNN, όλοι οι νευρώνες σε διάφορα στρώματα πρέπει να εκπαιδευτούν. Αυτό καθιστά την εκπαίδευση των RNNs πολύ πιο πολύπλοκη από τις δομές που βασίζονται σε RC. Για τη μέτρηση της πολυπλοκότητας, μπορεί να χρησιμοποιηθεί ο συνολικός αριθμός πράξεων κινητής υποδιαστολής (floating-point operations - FLOPs). Έχει αποδειχθεί στο [101] ότι η υπολογιστική πολυπλοκότητα των προσεγγίσεων μάθησης που βασίζονται σε RC αντιστοιχεί στον χρόνο εκπαίδευσης. Συνεπώς, σε αυτή την ενότητα παρουσιάζεται ο χρόνος εκπαίδευσης της προσέγγισής μας και συγκρίνεται με αυτόν των MLP και SNN. Στον ακόλουθο πίνακα, παρουσιάζονται οι χρόνοι εκπαίδευσης για τις μεθόδους spiking DFR+MLP, MLP και SNN. Οι συγκρίσεις γίνονται μεταξύ των χρόνων εκπαίδευσης αυτών των τριών αλγορίθμων, οι οποίοι συνδέονται επίσης με την υπολογιστική τους πολυπλοκότητα.

Στον πίνακα, παρατηρείται ότι ο αλγόριθμος SNN έχει την υψηλότερη υπολογιστική πολυπλοκότητα. Από το Σχήμα 6, μπορούμε να δούμε ότι ο αλγόριθμος SNN απαιτεί αρκετές επαναλήψεις για να συγκλίνει. Ο δεύτερος και τρίτος αλγόριθμος με υψηλή υπολογιστική πολυπλοκότητα είναι ο spiking DFR+MLP και ο MLP αντίστοιχα, όπως παρατηρείται στο Σχήμα 3. Ο αλγόριθμος spiking DFR+MLP διαθέτει ορισμένα δομικά στοιχεία που τον καθιστούν υπολογιστικά πιο πολύπλοκο από τον MLP. Στον αλγόριθμο που παρουσιάσαμε, υπάρχουν μπλοκ

όπως τα "spike to current" "reservoir block" και “MLP as the readout layer”, τα οποία καθιστούν τον αλγόριθμό μας πιο πολύπλοκο υπολογιστικά από ένα απλό MLP. Ωστόσο, αυτή η πρόσθετη επιβάρυνση δεν είναι τόσο σημαντική, όπως φαίνεται στον πίνακα.

COMPUTATIONAL COMPLEXITY ANALYSIS

Algorithm	Training Time
Spiking DFR+MLP	16.69 s
MLP	3.2 s
SNN	90 s

Αυτή είναι η πρώτη δημοσίευση που διερευνά τα spiking DFRs για τη δυναμική ανίχνευση κυβερνοεπιθέσεων σε ένα smart grid. Διερευνήθηκαν διαφορετικά σχήματα χρονικής κωδικοποίησης για την κωδικοποίηση των μετρήσεων που λαμβάνονται από το smart grid. Τα αποτελέσματα προσομοίωσης υποδεικνύουν ότι το ISI έχει την καλύτερη απόδοση εκπαίδευσης επειδή μεταφέρει περισσότερες πληροφορίες από άλλα σχήματα κωδικοποίησης. Από την άλλη πλευρά, το DFR μπορεί να χρησιμοποιηθεί για την αντιστοίχιση των δεδομένων σε έναν χώρο υψηλότερων διαστάσεων καθιστώντας ευκολότερη την ταξινόμηση των δεδομένων που έχουν παραβιαστεί από τα μη παραβιασμένα δεδομένα. Τα αποτελέσματα προσομοίωσης δείχνουν επίσης ότι ο συνδυασμός του DFR με το MLP μπορεί να υπερτερεί έναντι άλλων μεθόδων όπως το MLP και το SNN.

7

Σύνοψη

7.1 Επιγραμματική περίληψη

Η εργασία αυτή εμβαθύνει στη σφαίρα της ανίχνευσης ανωμαλιών στο πλαίσιο του Διαδικτύου των Πραγμάτων (IoT), ενός τομέα αιχμής με εκτεταμένες επιπτώσεις. Κοιτάζοντάς την πιο αναλυτικά:

Η εισαγωγική ενότητα θέτει τις βάσεις διαμορφώνοντας το πλαίσιο της μελέτης. Παρουσιάζει το τοπίο του IoT, τονίζοντας τη σημασία και τον αντίκτυπό του. Η δυνατότητα του IoT να αναδιαμορφώσει τους κλάδους, από την υγειονομική περίθαλψη έως τις μεταφορές, υπογραμμίζει την ανάγκη για ισχυρούς μηχανισμούς ανίχνευσης ανωμαλιών.

Αυτή η υποενότητα μεγεθύνει το κεντρικό θέμα της εργασίας - την ανίχνευση ανωμαλιών στο IoT. Όχι μόνο ορίζει την ανίχνευση ανωμαλιών, αλλά υπογραμμίζει επίσης τον κρίσιμο ρόλο της στα οικοσυστήματα του IoT. Η επισήμανση της ευαισθησίας των δικτύων IoT σε ανωμαλίες υπογραμμίζει τον επείγοντα χαρακτήρα των αποτελεσματικών μηχανισμών ανίχνευσης.

Ακολούθως, εμβαθύνει στη λεπτή ταξινόμηση και κατηγοριοποίηση των ανωμαλιών, μια καίρια πτυχή της ανίχνευσης ανωμαλιών. Με τον ορισμό των τύπων των ανωμαλιών και τον προσδιορισμό των περιοχών όπου εκδηλώνονται συνήθως, το έγγραφο παρουσιάζει μια ολοκληρωμένη προσέγγιση για την κατανόηση και την αντιμετώπιση των ανωμαλιών.

Η δεύτερη ενότητα, ονόματι “Προσεγγίσεις και Τεχνικές”, κάνει μια βαθιά βουτιά στις διάφορες μεθοδολογίες για τη μελέτη των ανωμαλιών. Αναλύοντας την επιβλεπόμενη, μη επιβλεπόμενη και ημι επιβλεπόμενη μάθηση, υπογραμμίζεται η ποικιλία των στρατηγικών που είναι διαθέσιμες στους ερευνητές. Αυτή η ευρύτητα ενισχύει τη συνάφεια του εγγράφου με ένα ευρύ φάσμα αναγνωστών με διαφορετικό τεχνικό υπόβαθρο.

Η υποενότητα “Τεχνικές” καλύπτει μια σειρά τεχνικών ανίχνευσης ανωμαλιών. Από παραδοσιακές προσεγγίσεις, όπως οι στατιστικές μέθοδοι, έως εξελίξεις αιχμής, όπως τα βαθιά νευρωνικά δίκτυα, η εργασία παρουσιάζει μια πανοραμική άποψη της εργαλειοθήκης που είναι διαθέσιμη για την ανίχνευση ανωμαλιών.

Προχωρώντας πέρα από την ανίχνευση, εμβαθύνει στο πεδίο της ανάλυσης. Διερευνά τη στατιστική ανάλυση, τη μηχανική μάθηση, τη φασματική ανάλυση και την ανάλυση συμπεριφοράς, αναδεικνύοντας την πολυδιάστατη φύση της έρευνας για την ανίχνευση ανωμαλιών.

Ένα κρίσιμο σημείο της εργασίας, η ενότητα αυτή εισάγει τα νευρωνικά δίκτυα, μια μετασχηματιστική τεχνολογία. Θέτοντας τα θεμέλια με τη δομή, τους τύπους και τις εφαρμογές των νευρωνικών δικτύων, η εργασία εφοδιάζει τους αναγνώστες με μια θεμελιώδη κατανόηση.

Στην τρίτη ενότητα, ξεκινά ένα ταξίδι στα Βαθιά Νευρωνικά Δίκτυα (DNN), ξετυλίγοντας το κουβάρι της πολυπλοκότητας και της χρησιμότητάς τους, συμπεριλαμβανομένων των θεμελιωδών αρχών, των διαδικασιών εκπαίδευσης, των προκλήσεων και των εφαρμογών τους σε διάφορους τομείς.

Στη συνέχεια, η εργασία διευρύνει τον ορίζοντα διερευνώντας τις εφαρμογές των DNNs σε τομείς όπως η εικόνα και το βίντεο, η επεξεργασία φυσικής γλώσσας, η υγεία, τα αυτόνομα οχήματα και τα κοινωνικά δίκτυα. Αυτό υπογραμμίζει την πρακτικότητα και την ευελιξία των DNNs.

Στην τέταρτη ενότητα, η εστίαση της εργασίας μετατοπίζεται τώρα στα αναδρομικά νευρωνικά δίκτυα (RNNs). Με τη διευκρίνιση των θεμελιωδών αρχών τους, του ρόλου τους στη μοντελοποίηση δεδομένων χρονοσειρών και των εξελισσόμενων τάσεων, η εργασία αναδεικνύει τη συμβολή των RNNs στην ανίχνευση ανωμαλιών.

Έπειτα η υποενότητα εμβαθύνει την κατανόηση παρουσιάζοντας συγκεκριμένες εφαρμογές των RNNs. Από την επεξεργασία φυσικής γλώσσας έως την πρόβλεψη χρονοσειρών, οι αναγνώστες εκτίθενται στον πραγματικό αντίκτυπο των RNNs.

Το τελευταίο σκέλος της εργασίας αυτής, η πέμπτη ενότητα, εισάγει το Reservoir Computing (Υπολογισμό δεξαμενής), ένα νέο παράδειγμα. Εξηγώντας τις αρχές, τις παραλλαγές, τις εφαρμογές, τις προκλήσεις και τα πλεονεκτήματά του, το έγγραφο προβάλλει μια πολλά υποσχόμενη οδό για περαιτέρω έρευνα.

Συνοψίζοντας, η εργασία αποτελεί μια ολοκληρωμένη εξερεύνηση της ανίχνευσης ανωμαλιών στο τοπίο του IoT. Διασχίζοντας διάφορες προσεγγίσεις, τεχνικές και τεχνολογίες, προσπαθώντας να εφοδιάσει τους αναγνώστες με γνώσεις αλλά και να τους εμπνεύσει να εξερευνήσουν τα αχαρτογράφητα εδάφη της ανίχνευσης ανωμαλιών και τις πολλαπλές επιπτώσεις της.

7.2 Κριτική Αξιολόγηση των ειδών Νευρωνικών Δικτύων

Τα Νευρωνικά Δίκτυα αντιπροσωπεύουν μία από τις πιο ενδιαφέρουσες και δυναμικές πτυχές της σύγχρονης επιστήμης των υπολογιστών και της τεχνητής νοημοσύνης. Μεταξύ των ποικίλων ειδών Νευρωνικών Δικτύων, τα Δίκτυα Βαθιάς Μάθησης (DNNs), τα Αναδρομικά Νευρωνικά Δίκτυα (RNNs) και ο Υπολογισμός Δεξαμενής (Reservoir Computing) αποκτούν όλο και μεγαλύτερη σημασία. Σε αυτή την εργασία, θα πραγματοποιήσουμε μια κριτική αξιολόγηση αυτών των τύπων Νευρωνικών Δικτύων, εξετάζοντας τα πλεονεκτήματα, τα μειονεκτήματα και τις εφαρμογές τους.

Τα DNNs αναδείχθηκαν ως ένα ισχυρό εργαλείο για την ανάλυση και την εξαγωγή πληροφοριών από πολύπλοκα σύνολα δεδομένων. Οι πολυεπίπεδες αρχιτεκτονικές τους επιτρέπουν την αναγνώριση προτύπων που θα ήταν δύσκολο να ανιχνευθούν με άλλες μεθόδους.

Ωστόσο, η επιτυχής εκπαίδευση DNNs απαιτεί μεγάλο όγκο δεδομένων και υψηλό υπολογιστικό κόστος.

Τα RNNs είναι ιδιαίτερα κατάλληλα για την επεξεργασία δεδομένων με χρονική σειρά, όπως χρονοσειρές και φυσική γλώσσα. Η δυνατότητά τους να διατηρούν εσωτερική κατάσταση του παρελθόντος τους τους παρέχει το πλεονέκτημα να αντιμετωπίζουν προβλήματα με δυναμικές εξελίξεις. Ωστόσο, τα προβλήματα που προκύπτουν από την εξαφάνιση ή εκθετική αύξηση των gradients (γνωστό ως πρόβλημα της εξαφάνισης/εκθετικής αύξησης του gradient) μπορούν να περιορίσουν την απόδοση των RNNs.

Ο Υπολογισμός Δεξαμενής αναδεικνύεται ως μια πρωτοποριακή προσέγγιση στα Νευρωνικά Δίκτυα, καθώς βασίζεται στην αρχή της δυναμικής ανακατασκευής ενός διακριτού συστήματος. Η δυνατότητά του να επιτελεί υπολογισμούς με υψηλή παραλληλία τον καθιστά ιδανικό για προβλήματα επεξεργασίας σημάτων και χρονοσειρών. Παρ' όλα αυτά, ο Υπολογισμός Δεξαμενής δεν είναι ακόμη ευρέως γνωστός, και οι τεχνικές του αξίζουν περαιτέρω έρευνα.

Η κριτική αξιολόγηση των τύπων Νευρωνικών Δικτύων - DNNs, RNNs και Υπολογισμού Δεξαμενής - αποκαλύπτει μια ποικιλία πλεονεκτημάτων και προκλήσεων που σχετίζονται με καθέναν από αυτούς. Ενώ τα DNNs είναι ικανά να αναγνωρίζουν πολύπλοκα πρότυπα, τα RNNs αντιμετωπίζουν προκλήσεις με την αντιμετώπιση των προβλημάτων της εξαφάνισης/εκθετικής αύξησης του gradient. Ο Υπολογισμός Δεξαμενής, παρότι υποσχόμενος, απαιτεί ακόμη επεξεργασία και περαιτέρω ανάπτυξη των τεχνικών του. Καθώς η τεχνητή νοημοσύνη συνεχίζει να εξελίσσεται, η κατανόηση και η κριτική αξιολόγηση αυτών των τύπων Νευρωνικών Δικτύων παραμένει απαραίτητη για την ανάπτυξη και την καινοτομία σε αυτόν τον τομέα.

7.3 Σύγκριση των ειδών

7.3.1 Τύπος Νευρωνικού Δικτύου: DNNs

- Δομή

Τα Deep Neural Networks (DNNs) αποτελούν την πιο συνηθισμένη αρχιτεκτονική νευρωνικού δικτύου και χαρακτηρίζονται από την πολυεπίπεδη δομή τους. Αυτά τα δίκτυα συνήθως αποτελούνται από τρία βασικά επίπεδα: το επίπεδο εισόδου, τα κρυφά επίπεδα και το επίπεδο εξόδου. Τα κρυφά επίπεδα, ενδιάμεσα ανάμεσα στην είσοδο και την έξοδο, συνήθως περιλαμβάνουν εκατοντάδες ή ακόμη και χιλιάδες νευρώνες. Αυτή η βαθιά αρχιτεκτονική επιτρέπει την ανάλυση και την αναγνώριση πολύπλοκων προτύπων στα δεδομένα.

- Εφαρμογές

Τα DNNs έχουν εφαρμογές σε ποικίλους τομείς. Είναι ιδιαίτερα αποτελεσματικά για εφαρμογές που απαιτούν εξειδικευμένη αναγνώριση προτύπων, όπως εικόνες, φυσική γλώσσα, αναγνώριση προσώπων και πολλά άλλα. Στην αναγνώριση εικόνων, μπορούν να αναγνωρίσουν αντικείμενα, πρόσωπα και σκηνές. Στη φυσική γλώσσα, μπορούν να εκτελέσουν αυτόματη μετάφραση, να δημιουργήσουν κείμενο και να αναλύσουν συναισθηματική φόρτιση. Επίσης, χρησιμοποιούνται για προβλέψεις σε χρονοσειρές δεδομένων, από την πρόβλεψη των χρηματιστηριακών αγορών μέχρι την πρόβλεψη του καιρού.

- Προκλήσεις

Μία από τις μεγαλύτερες προκλήσεις που αντιμετωπίζουν τα DNNs είναι η ανάγκη για έναν μεγάλο όγκο δεδομένων για εκπαίδευση. Αυτό σημαίνει ότι χρειάζονται υποστηρικτικά δεδομένα σε αρκετή ποσότητα και ποικιλία για να μάθουν αποτελεσματικά. Επίσης, η επιλογή του σωστού αριθμού κρυφών επιπέδων και νευρώνων μπορεί να αποτελεί πρόκληση, καθώς ανεπιθύμητες επιλογές μπορεί να οδηγήσουν σε υπερεκπαίδευση ή αδυναμία μάθησης.

Συνολικά, τα DNNs αντιπροσωπεύουν μια ισχυρή εργαλειοθήκη για την αναγνώριση προτύπων σε πολύπλοκα δεδομένα, χρησιμοποιώντας βαθιές και πολυεπίπεδες δομές νευρωνικών. Παρά τις προκλήσεις της εκπαίδευσης, η δυνατότητά τους να αντλούν πληροφορίες από τα δεδομένα τους κάνει αναπόφευκτη την εξέλιξή τους σε ένα σημαντικό εργαλείο στη μηχανική μάθηση και την τεχνητή νοημοσύνη.

7.3.2 Τύπος Νευρωνικού Δικτύου: RNNs

- Δομή

Τα Αναδρομικά Νευρωνικά Δίκτυα (RNNs) προσδίδουν μια εντυπωσιακή δυνατότητα στην αντιμετώπιση χρονικών δεδομένων, εφόσον η αναδρομική δομή τους επιτρέπει την επεξεργασία αλληλουχιών και χρονικών σειρών. Σε κάθε βήμα του χρόνου, ο κάθε νευρώνας διατηρεί μια κατάσταση που αναμεταδίδεται στο επόμενο βήμα. Αυτός ο μηχανισμός καθιστά τα RNNs ιδανικά για προβλήματα που απαιτούν κατανόηση συμπεριφοράς με βάση την χρονική σειρά.

- Εφαρμογές

Οι εφαρμογές των RNNs είναι πολύ ποικίλες και εντυπωσιακές. Αποτελούν τη βάση για προβλήματα πρόβλεψης χρονοσειρών, καθώς μπορούν να μοντελοποιήσουν και να αναγνωρίσουν πολύπλοκες χρονικές συσχετίσεις. Επίσης, είναι ιδιαίτερα χρήσιμα στην επεξεργασία φυσικής γλώσσας, όπου η σειρά των λέξεων έχει σημασία για τη σημασιολογική κατανόηση του κειμένου. Τέλος, χρησιμοποιούνται σε εφαρμογές μετάφρασης, γεννήτριας κειμένου, αναγνώρισης γραφής και άλλα.

- Προκλήσεις

Παρά τις εντυπωσιακές τους δυνατότητες, τα RNNs αντιμετωπίζουν προκλήσεις. Ένα από τα κύρια προβλήματα είναι ο "vanishing gradient", όπου η ανανεωσιμότητα των βαρών σε μεγάλο βάθος αναδρομής μπορεί να οδηγήσει σε απώλεια πληροφορίας και αδυναμία μάθησης μακρινών εξαρτήσεων. Επίσης, τα μακροπρόθεσμα προβλήματα εκπαίδευσης μπορεί να οδηγήσουν σε αστάθειες κατά την εκπαίδευση.

Τα Αναδρομικά Νευρωνικά Δίκτυα (RNNs) αντιπροσωπεύουν έναν σημαντικό τύπο νευρωνικού δικτύου με δυνατότητα αντιμετώπισης χρονικών δεδομένων. Παρά τις προκλήσεις που αντιμετωπίζουν, η ικανότητά τους να αναγνωρίζουν χρονικές συσχετίσεις και αλληλουχίες τους καθιστά αναπόσπαστο κομμάτι της επιστήμης της μηχανικής μάθησης και της επεξεργασίας φυσικής γλώσσας.

7.3.3 Τύπος Νευρωνικού Δικτύου: Reservoir Computing

- Δομή

Το Reservoir Computing αναπτύχθηκε ως μια προσέγγιση που εξερευνά μια διαφορετική αρχιτεκτονική σε σχέση με τα παραδοσιακά νευρωνικά δίκτυα. Η βασική του δομή είναι αποκεντρωμένη και περιλαμβάνει ένα "υπολογιστικό δεξαμενή", ή αλλιώς "υπολογιστική δεξαμενή", που λειτουργεί ως ενδιάμεσος χώρος επεξεργασίας δεδομένων. Αυτή η δεξαμενή δεν υπόκειται σε εκπαίδευση, αλλά διατηρεί μια σταθερή και τυχαία σύνδεση μεταξύ των νευρώνων της. Η είσοδος ενεργοποιεί τους νευρώνες του δεξαμενής, και από εκεί τα σήματα εξελίσσονται μέσω της σταθερής δομής των συνδέσεων.

- Εφαρμογές

Ο τύπος αυτός έχει ευρύ φάσμα εφαρμογών, εκτείνοντας την επιρροή του σε επιστημονικά πεδία και τεχνολογικές εφαρμογές. Στην επεξεργασία σημάτων, το Reservoir Computing μπορεί να αναγνωρίσει πολύπλοκα πρότυπα σε σήματα που αλλάζουν με το χρόνο. Επίσης, στην αναγνώριση προτύπων, μπορεί να αναδείξει συνήθως δυσδιάκριτα χαρακτηριστικά από τα δεδομένα εισόδου. Ενδιαφέροντα είναι η εφαρμογή του σε προβλήματα πρόβλεψης, καθώς μπορεί να αντλήσει πληροφορίες από προηγούμενες χρονικές στιγμές για να προβλέψει τις μελλοντικές τιμές.

- Προκλήσεις

Η διαμόρφωση του αποτελεί μια προκλητική διαδικασία, απαιτώντας προσεκτικά σχεδιασμένες συνδέσεις μεταξύ των νευρώνων και κατανόηση της επίδρασης των διαφόρων παραμέτρων. Επίσης, η επιλογή του κατάλληλου μεγέθους της δεξαμενής και η διαμόρφωση της αρχικής της κατάστασης μπορεί να επηρεάσει σημαντικά την απόδοση. Απαιτεί εξειδικευμένες γνώσεις για την αποτελεσματική χρήση αυτής της προσέγγισης.

7.4 Πλεονεκτήματα και Μειονεκτήματα

Τα νευρωνικά δίκτυα αντιπροσωπεύουν ένα σημαντικό εργαλείο στην ανίχνευση ανωμαλιών και έχουν εξελιχθεί σε διάφορους τύπους, όπως τα Deep Neural Networks (DNNs), τα Recurrent Neural Networks (RNNs) και το Reservoir Computing.

7.4.1 Τύπος Νευρωνικού Δικτύου: Deep Neural Networks (DNNs)

- Πλεονεκτήματα

- ο Ικανότητα εκμάθησης πολύπλοκων σχέσεων

Τα Deep Neural Networks (DNNs) ξεχωρίζουν στην ικανότητά τους να μάθουν και να ανακαλύψουν πολύπλοκες συσχετίσεις ανάμεσα σε δεδομένα. Αυτό τα καθιστά ιδανικά για την ανάλυση σύνθετων προβλημάτων.

- ο Εξαιρετική αποτελεσματικότητα στην επεξεργασία εικόνων, κειμένου και ήχου

Τα DNNs έχουν επιδείξει εξαιρετική απόδοση σε εργασίες όπως η αναγνώριση εικόνων, η επεξεργασία φυσικής γλώσσας και η αναγνώριση φωνής. Η δυνατότητά τους να αντιληφθούν μοτίβα σε αυτές τις μορφές δεδομένων τους καθιστά ανεκτίμητα σε εφαρμογές που συνδέονται με αυτά.

- Εφαρμογή σε πληθώρα πεδίων

Η ευελιξία των DNNs τους επιτρέπει να εφαρμοστούν σε διάφορους τομείς. Από την ανάπτυξη αυτόνομων οχημάτων μέχρι την ιατρική διάγνωση και τα κοινωνικά δίκτυα, τα DNNs είναι ιδιαίτερα πολύπλευρα.

- Μειονεκτήματα

- Υψηλή απαιτητικότητα σε υπολογιστικούς πόρους

Η εκπαίδευση των DNNs μπορεί να απαιτεί σημαντικούς υπολογιστικούς πόρους, όπως ισχυρούς επεξεργαστές ή ειδικούς επιταχυντές.

- Απαιτεί μεγάλα σύνολα δεδομένων

Για να επιτύχουν αποτελεσματική εκπαίδευση, τα DNNs απαιτούν μεγάλα σύνολα δεδομένων. Αυτό μπορεί να περιορίσει τη χρησιμότητά τους σε περιπτώσεις με περιορισμένα διαθέσιμα δεδομένα.

- Προβλήματα υπερ-προσαρμογής (overfitting)

Η υπερ-προσαρμογή μπορεί να προκαλέσει προβλήματα στη γενίκευση των αποτελεσμάτων των DNNs σε νέα δεδομένα που δεν έχουν "δει" κατά την εκπαίδευση.

7.4.2 Τύπος Νευρωνικού Δικτύου: *Recurrent Neural Networks (RNNs)*

- Πλεονεκτήματα

- Χειρισμός Χρονοσειρών

Τα RNNs είναι ιδανικά για την επεξεργασία ακολουθιακών δεδομένων όπως χρονοσειρές, καθιστώντας τους κατάλληλους για προβλήματα που αναφέρονται στον χρόνο, όπως προβλέψεις.

- Ποικιλία Εφαρμογών

Είναι κατάλληλα για προβλήματα μετάφρασης κειμένου, αναγνώρισης φωνής, και πολλά άλλα, εξαιτίας της ικανότητάς τους να αναγνωρίζουν σειρές με εσωτερική εξάρτηση.

- Μειονεκτήματα

- Πρόβλημα της Εξαφάνισης/Εκροής των Gradients

Τα παραδοσιακά RNNs αντιμετωπίζουν δυσκολίες στο να μάθουν μακροπρόθεσμες εξαρτήσεις λόγω του προβλήματος της εξαφάνισης ή της εκροής των gradients κατά την εκπαίδευση.

- Απαιτούν Επιπλέον Χρόνο για Εκπαίδευση

Λόγω της ανάγκης για την επεξεργασία ακολουθιακών δεδομένων και της προκλητικής φύσης της εκπαίδευσής τους, τα RNNs απαιτούν περισσότερο χρόνο για την εκπαίδευσή τους σε σύγκριση με άλλα τύπους δικτύων.

7.4.3 Τύπος Νευρωνικού Δικτύου: *Reservoir Computing*

- Πλεονεκτήματα

- Ευέλικτη Αρχιτεκτονική

Ένα από τα κυριότερα πλεονεκτήματα του Reservoir Computing είναι η ευελιξία της αρχιτεκτονικής του, η οποία επιτρέπει τη χρήση διαφορετικών τύπων νευρώνων σε διάφορα στάδια του δικτύου. Αυτό δίνει τη δυνατότητα για προσαρμογή του δικτύου σε διαφορετικούς τύπους εργασιών και δεδομένων.

- Απλούστερη Εκπαίδευση

Ένα άλλο ουσιαστικό πλεονέκτημα είναι ότι η εκπαίδευση του Reservoir Computing είναι συνήθως πιο απλή σε σύγκριση με άλλα πιο πολύπλοκα νευρωνικά δίκτυα. Αυτό οφείλεται στο γεγονός ότι η δεξαμενή (reservoir) δεν ανανεώνεται συνεχώς κατά τη διάρκεια της εκπαίδευσης, κάτι που μειώνει τις απαιτήσεις για πολύπλοκη παραμετροποίηση κατά την εκπαίδευση.

- Μειονεκτήματα

- Περιορισμένες Εφαρμογές

Παρά τα πλεονεκτήματα, το Reservoir Computing έχει περιορισμένες εφαρμογές σε σύγκριση με πιο πολύπλοκα νευρωνικά δίκτυα, όπως τα Deep Neural Networks (DNNs). Αυτό οφείλεται στην απλότητα της αρχιτεκτονικής του, που το περιορίζει σε ορισμένες τύπους εργασιών.

- Ενεργή Έρευνα

Αν και αναπτύσσεται δυναμικά, το Reservoir Computing είναι ακόμα υπό ενεργή έρευνα. Αυτό σημαίνει ότι η πλήρης κατανόηση της δυναμικής του και η αποτελεσματικότητά του μπορεί να είναι περιορισμένη σε ορισμένες περιπτώσεις.

Κάθε κατηγορία νευρωνικού δικτύου παρουσιάζει μια μοναδική σειρά πλεονεκτημάτων και μειονεκτημάτων, καθώς αντιμετωπίζει διαφορετικές προκλήσεις και παρέχει διαφορετικές λύσεις σε διαφορετικά προβλήματα. Η επιλογή ανάμεσα σε αυτούς τους τύπους εξαρτάται κυρίως από το είδος του προβλήματος που αντιμετωπίζετε και τις απαιτήσεις της συγκεκριμένης εφαρμογής σας.

Συνολικά, η ενδιαφέρουσα δυναμική ανάμεσα στους τρεις τύπους νευρωνικών δικτύων, δηλαδή τα Deep Neural Networks (DNNs), τα Recurrent Neural Networks (RNNs) και το Reservoir Computing, ανακαλύπτεται μέσα από μια πληθώρα παραμέτρων και παραγόντων που ενδέχεται να επηρεάσουν την τελική απόδοσή τους.

Τα DNNs, με τη βαθιά τους δομή που αποτελείται από πολλά κρυφά επίπεδα, έχουν αποδειχθεί εξαιρετικά αποτελεσματικά στην αναγνώριση πολύπλοκων προτύπων και την επεξεργασία πλούσιων δεδομένων, όπως εικόνες και ήχος. Ωστόσο, η επιτυχία τους συνδέεται

στενά με τη διαθεσιμότητα μεγάλου όγκου δεδομένων για εκπαίδευση και την τέλεια ρύθμιση των υπερπαραμέτρων.

Από την άλλη πλευρά, τα RNNs είναι κατάλληλα για δεδομένα με χρονική σειρά, όπως φυσική γλώσσα, μουσική και χρονοσειρές. Η δυνατότητά τους να διατηρούν κατάσταση μεταξύ των χρονικών βημάτων τους επιτρέπει την αντιμετώπιση μακροχρόνιων εξαρτήσεων. Ωστόσο, αντιμετωπίζουν προβλήματα όπως η "vanishing gradient" όταν η ανατροφοδότηση του σφάλλει σε μεγάλα βάθη.

Το Reservoir Computing παρουσιάζει μια διαφορετική προσέγγιση, χρησιμοποιώντας μια αποκεντρωμένη δομή με ένα υπολογιστικό δεξαμενής. Αυτό το δίκτυο παρουσιάζει πλεονεκτήματα στην αντιμετώπιση περιπλοκότερων προβλημάτων χωρίς την ανάγκη για την πλήρη ρύθμιση των υπερπαραμέτρων όπως στα DNNs.

Σε κάθε περίπτωση, η επιλογή του κατάλληλου τύπου δικτύου εξαρτάται από τη φύση των δεδομένων και τον συγκεκριμένο σκοπό της αντίστοιχης ανάλυσης στο εκάστοτε πρόβλημα. Κατά την εξέταση των διαφόρων παραμέτρων, αλγορίθμων και τεχνικών, είναι σημαντικό να ληφθούν υπόψη οι προκλήσεις και τα πλεονεκτήματα κάθε τύπου νευρωνικού δικτύου.

Βιβλιογραφία - References

- [1] [Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. \(2013\). Internet of Things \(IoT\): A vision, architectural elements, and future directions. Future Generation Computer Systems, 29\(7\), 1645-1660.](#)
- [2] [Lee, I., & Lee, K. \(2015\). The Internet of Things \(IoT\): Applications, investments, and challenges for enterprises. Business Horizons, 58\(4\), 431-440.](#)
- [3] [Kortuem, G., Kawsar, F., Fitton, D., & Sundramoorthy, V. \(2010\). Smart objects as building blocks for the Internet of Things. IEEE Internet Computing, 14\(1\), 44-51.](#)
- [4] [Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. \(2015\). Internet of Things: A survey on enabling technologies, protocols, and applications. IEEE Communications Surveys & Tutorials, 17\(4\), 2347-2376.](#)
- [5] [Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. \(2014\). Internet of things for smart cities. IEEE Internet of Things Journal, 1\(1\), 22-32.](#)
- [6] [Atzori, L., Iera, A., & Morabito, G. \(2010\). The Internet of Things: A survey. Computer Networks, 54\(15\), 2787-2805.](#)
- [7] [Li, S., Xu, L. D., & Zhao, S. \(2015\). The internet of things: A survey of topics and trends. Information Systems Frontiers, 17\(2\), 261-274.](#)
- [8] [Kortuem, G., Kawsar, F., Fitton, D., & Sundramoorthy, V. \(2010\). Smart objects as building blocks for the Internet of Things. IEEE Internet Computing, 14\(1\), 44-51.](#)
- [9] [N. Mishra and S. Pandya, "Internet of Things Applications, Security Challenges, Attacks, Intrusion Detection, and Future Visions: A Systematic Review," in IEEE Access, vol. 9, pp. 59353-59377, 2021, doi: 10.1109/ACCESS.2021.3073408.](#)
- [10] [A. A. Cook, G. Mısırlı and Z. Fan, "Anomaly Detection for IoT Time-Series Data: A Survey," in IEEE Internet of Things Journal, vol. 7, no. 7, pp. 6481-6494, July 2020, doi: 10.1109/JIOT.2019.2958185.](#)
- [11] [A. Gaddam, T. Wilkin and M. Angelova, "Anomaly Detection Models for Detecting Sensor Faults and Outliers in the IoT - A Survey," 2019 13th International Conference on Sensing Technology \(ICST\), Sydney, NSW, Australia, 2019, pp. 1-6, doi: 10.1109/ICST46873.2019.9047684.](#)
- [12] [Markou, M., & Singh, S. \(2003\). Novelty detection: a review—part 1: statistical approaches. Signal processing, 83\(12\), 2481-2497.](#)
- [13] [Markou, M., & Singh, S. \(2003\). Novelty detection: a review—part 2: neural network based approaches. Signal Processing, 83\(12\), 2499-2521.](#)
- [14] [M. Fahim and A. Sillitti, "Anomaly Detection, Analysis and Prediction Techniques in IoT Environment: A Systematic Literature Review," in IEEE Access, vol. 7, pp. 81664-81681, 2019, doi: 10.1109/ACCESS.2019.2921912.](#)
- [15] [Jamalipour, A., & Murali, S. \(2022\). A Taxonomy of Machine-Learning-Based Intrusion Detection Systems for the Internet of Things: A Survey. IEEE Internet of Things Journal, 9\(12\), 9444-9466. doi: 10.1109/JIOT.2021.3126811](#)
- [16] [Chandola, V., Banerjee, A., & Kumar, V. \(2009\). Anomaly detection: A survey. ACM computing surveys \(CSUR\), 41\(3\), 15.](#)
- [17] [Hodge, V. J., & Austin, J. \(2004\). A survey of outlier detection methodologies. Artificial Intelligence Review, 22\(2\), 85-126.](#)
- [18] [Goldstein, M., & Uchida, S. \(2016\). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PloS one, 11\(4\), e0152173.](#)
- [19] [Chalapathy, R., & Chawla, S. \(2019\). Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407.](#)

- [20] [Goldstein, M., & Uchida, S. \(2016\). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PloS one, 11\(4\), e0152173.](#)
- [21] [Patcha, A., & Park, J. M. \(2007\). An overview of anomaly detection techniques: Existing solutions and latest technological trends. Computer networks, 51\(12\), 3448-3470.](#)
- [22] [Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., Ayyash, M., Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications, IEEE Communications Surveys & Tutorials, 2015, vol. 17, no. 4, pp. 2347-2376.](#)
- [23] [Ross, Sheldon M. Introduction to probability and statistics for engineers and scientists. Academic press, 2020.](#)
- [24] [Percival, Donald B., and Andrew T. Walden. Spectral analysis for physical applications. cambridge university press, 1993.](#)
- [25] [Atzori, L., Iera, A., Morabito, G., Understanding the Internet of Things: Definition, Potentials, and Societal Role of a Fast-Evolving Paradigm, Elsevier, 2017.](#)
- [26] [Box, George EP, et al. Time series analysis: forecasting and control. John Wiley & Sons, 2015.](#)
- [27] [Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.](#)
- [28] [Bishop, Christopher M., and Nasser M. Nasrabadi. Pattern recognition and machine learning. Vol. 4. No. 4. New York: springer, 2006.](#)
- [29] [Pimentel, M. A., Clifton, D. A., Clifton, L., & Tarassenko, L. \(2014\). A review of novelty detection. Signal Processing, 99, 215-249.](#)
- [30] [Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M., Internet of Things \(IoT\): A Vision, Architectural Elements, and Future Directions, Future Generation Computer Systems, 2013, vol. 29, no. 7, pp. 1645-1660.](#)
- [31] [Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 25 \(2012\).](#)
- [32] [Graves, Alex, Greg Wayne, and Ivo Danihelka. "Neural turing machines." arXiv preprint arXiv:1410.5401 \(2014\).](#)
- [33] [Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 \(2014\).](#)
- [34] [Alpaydin, Ethem. Introduction to machine learning. MIT press, 2020.](#)
- [35] [LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 \(2015\): 436-444.](#)
- [36] [Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors." nature 323.6088 \(1986\): 533-536.](#)
- [37] [Jaeger, Herbert. "The “echo state” approach to analysing and training recurrent neural networks-with an erratum note." Bonn, Germany: German National Research Center for Information Technology GMD Technical Report 148.34 \(2001\): 13.](#)
- [38] [Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." Advances in neural information processing systems 27 \(2014\).](#)
- [39] [Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 \(2014\).](#)
- [40] [Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.](#)
- [41] [Pathak, Deepak, et al. "Context encoders: Feature learning by inpainting." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.](#)
- [42] [Greff, Klaus, et al. "LSTM: A search space odyssey." IEEE transactions on neural networks and learning systems 28.10 \(2016\): 2222-2232.](#)
- [43] [Hochreiter, Sepp, and Jürgen Schmidhuber. "Feature extraction through LOCOCODE." Neural Computation 11.3 \(1999\): 679-714.](#)

- [44] [Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." Advances in neural information processing systems 27 \(2014\).](#)
- [45] [Cho, Kyunghyun, et al. "Learning phrase representations using RNN encoder-decoder for statistical machine translation." arXiv preprint arXiv:1406.1078 \(2014\).](#)
- [46] [Schmidhuber, Jürgen. "Deep learning in neural networks: An overview." Neural networks 61 \(2015\): 85-117.](#)
- [47] [Young, Tom, et al. "Recent trends in deep learning based natural language processing." iee Computational intelligence magazine 13.3 \(2018\): 55-75.](#)
- [48] [Esteva, Andre, et al. "A guide to deep learning in healthcare." Nature medicine 25.1 \(2019\): 24-29.](#)
- [49] [Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., ... & Zhang, X. \(2016\). End to end learning for self-driving cars. arXiv preprint arXiv:1604.07316.](#)
- [50] [Santoro, Adam, et al. "Meta-learning with memory-augmented neural networks." International conference on machine learning. PMLR, 2016.](#)
- [51] [Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks." Communications of the ACM 60.6 \(2017\): 84-90.](#)
- [52] [Gehring, Jonas, et al. "Convolutional sequence to sequence learning." International conference on machine learning. PMLR, 2017.](#)
- [53] [Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 \(1997\): 1735-1780.](#)
- [54] [Boulanger-Lewandowski, Nicolas, Yoshua Bengio, and Pascal Vincent. "Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription." arXiv preprint arXiv:1206.6392 \(2012\).](#)
- [55] [Brown, Tom, et al. "Language models are few-shot learners." Advances in neural information processing systems 33 \(2020\): 1877-1901.](#)
- [56] [Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." arXiv preprint arXiv:1409.0473 \(2014\).](#)
- [57] [Graves, Alex. "Generating sequences with recurrent neural networks." arXiv preprint arXiv:1308.0850 \(2013\).](#)
- [58] [Pascanu, Razvan, Tomas Mikolov, and Yoshua Bengio. "On the difficulty of training recurrent neural networks." International conference on machine learning. Pmlr, 2013.](#)
- [59] [Maass, Wolfgang, Thomas Natschläger, and Henry Markram. "Real-time computing without stable states: A new framework for neural computation based on perturbations." Neural computation 14.11 \(2002\): 2531-2560.](#)
- [60] [Lipton, Zachary C., John Berkowitz, and Charles Elkan. "A critical review of recurrent neural networks for sequence learning." arXiv preprint arXiv:1506.00019 \(2015\).](#)
- [61] [Gers, Felix A., Jürgen Schmidhuber, and Fred Cummins. "Learning to forget: Continual prediction with LSTM." Neural computation 12.10 \(2000\): 2451-2471.](#)
- [62] [Chung, Junyoung, et al. "Empirical evaluation of gated recurrent neural networks on sequence modeling." arXiv preprint arXiv:1412.3555 \(2014\).](#)
- [63] [Lukoševičius, Mantas, and Herbert Jaeger. "Reservoir computing approaches to recurrent neural network training." Computer science review 3.3 \(2009\): 127-149.](#)
- [64] [Verstraeten, David, Benjamin Schrauwen, and Dirk Stroobandt. "Reservoir-based techniques for speech recognition." The 2006 IEEE International Joint Conference on Neural Network Proceedings. IEEE, 2006.](#)
- [65] [Gallicchio, Claudio, Alessio Micheli, and Luca Pedrelli. "Deep reservoir computing: A critical experimental analysis." Neurocomputing 268 \(2017\): 87-99.](#)
- [66] [Maass, Wolfgang, and Henry Markram. "On the computational power of circuits of spiking neurons." Journal of computer and system sciences 69.4 \(2004\): 593-616.](#)

- [67] [Laje, Rodrigo, and Dean V. Buonomano. "Robust timing and motor patterns by taming chaos in recurrent neural networks." *Nature neuroscience* 16.7 \(2013\): 925-933.](#)
- [68] [Hermans, Michiel, and Benjamin Schrauwen. "Training and analysing deep recurrent neural networks." *Advances in neural information processing systems* 26 \(2013\).](#)
- [69] [Verstraeten, David, et al. "An experimental unification of reservoir computing methods." *Neural networks* 20.3 \(2007\): 391-403.](#)
- [70] [Kudithipudi, Dhireesha, et al. "Design and analysis of a neuromemristive reservoir computing architecture for biosignal processing." *Frontiers in neuroscience* 9 \(2016\): 502.](#)
- [71] [Buteneers, Pieter, et al. "Automatic detection of epileptic seizures on the intra-cranial electroencephalogram of rats using reservoir computing." *Artificial intelligence in medicine* 53.3 \(2011\): 215-223.](#)
- [72] [Khosrozadeh, Ali, et al. "Presenting a novel page replacement algorithm based on LRU." *Journal of Basic and Applied Scientific Research* 2.10 \(2012\): 10377-10383.](#)
- [73] [Jaeger, Herbert. "Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the" echo state network" approach." \(2002\).](#)
- [74] [Hewamalage, Hansika, Christoph Bergmeir, and Kasun Bandara. "Recurrent neural networks for time series forecasting: Current status and future directions." *International Journal of Forecasting* 37.1 \(2021\): 388-427.](#)
- [75] [Jalalvand, Azarakhsh, et al. "On the application of reservoir computing networks for noisy image recognition." *Neurocomputing* 277 \(2018\): 237-248.](#)
- [76] [Gao, Lili, Xiaopeng Deng, and Weimin Yang. "Smart city infrastructure protection: real-time threat detection employing online reservoir computing architecture." *Neural Computing and Applications* \(2022\): 1-10.](#)
- [77] [Colla, Valentina, et al. "Reservoir computing approaches applied to energy management in industry." *Engineering Applications of Neural Networks: 20th International Conference, EANN 2019, Xersonisos, Crete, Greece, May 24-26, 2019, Proceedings* 20. Springer International Publishing, 2019.](#)
- [78] [Gaurav, Ankit, et al. "Reservoir computing for temporal data classification using a dynamic solid electrolyte ZnO thin film transistor." *Frontiers in Electronics* 3 \(2022\): 869013.](#)
- [79] [Tanaka, Gouhei, et al. "Recent advances in physical reservoir computing: A review." *Neural Networks* 115 \(2019\): 100-123.](#)
- [80] [Rafayelyan, Mushegh, et al. "Large-scale optical reservoir computing for spatiotemporal chaotic systems prediction." *Physical Review X* 10.4 \(2020\): 041037.](#)
- [81] [Ponghiran, Wachirawit, Gopalakrishnan Srinivasan, and Kaushik Roy. "Reinforcement learning with low-complexity liquid state machines." *Frontiers in Neuroscience* 13 \(2019\): 883.](#)
- [82] [Embrechts, Mark J., Luís A. Alexandre, and Jonathan D. Linton. "Reservoir computing for static pattern recognition." *ESANN*. 2009.](#)
- [83] [Zimmerman, Ray Daniel, Carlos Edmundo Murillo-Sánchez, and Robert John Thomas. "MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education." *IEEE Transactions on power systems* 26.1 \(2010\): 12-19.](#)
- [84] [Liu, Ying-Hui, and Xiao-Jing Wang. "Spike-frequency adaptation of a generalized leaky integrate-and-fire model neuron." *Journal of computational neuroscience* 10 \(2001\): 25-45.](#)
- [85] [Zhao, Chenyuan, et al. "Energy efficient spiking temporal encoder design for neuromorphic computing systems." *IEEE Transactions on Multi-Scale Computing Systems* 2.4 \(2016\): 265-276.](#)
- [86] [Appeltant, Lennert. "Reservoir computing based on delay-dynamical systems." *These de Doctorat, Vrije Universiteit Brussel/Universitat de les Illes Balears* \(2012\).](#)
- [87] [Liu, Yao, Peng Ning, and Michael K. Reiter. "False data injection attacks against state estimation in electric power grids." *ACM Transactions on Information and System Security \(TISSEC\)* 14.1 \(2011\): 1-33.](#)
- [88] [Cui, Shuguang, et al. "Coordinated data-injection attack and detection in the smart grid: A detailed look at enriching detection solutions." *IEEE Signal Processing Magazine* 29.5 \(2012\): 106-115.](#)

- [89] [He, Youbiao, Gihan J. Mendis, and Jin Wei. "Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism." IEEE Transactions on Smart Grid 8.5 \(2017\): 2505-2516.](#)
- [90] [Esmalifalak, Mohammad, et al. "Detecting stealthy false data injection using machine learning in smart grid." IEEE Systems Journal 11.3 \(2014\): 1644-1652.](#)
- [91] [Hu, Jun, et al. "A spike-timing-based integrated model for pattern recognition." Neural computation 25.2 \(2013\): 450-472.](#)
- [92] [Yu, Qiang, et al. "Precise-spike-driven synaptic plasticity: Learning hetero-association of spatiotemporal spike patterns." Plos one 8.11 \(2013\): e78318.](#)
- [93] [Mohammed, Ammar, et al. "Span: Spike pattern association neuron for learning spatio-temporal spike patterns." International journal of neural systems 22.04 \(2012\): 1250012.](#)
- [94] [Bai, Kangjun, et al. "Enabling a new era of brain-inspired computing: energy-efficient spiking neural network with ring topology." Proceedings of the 55th Annual Design Automation Conference. 2018.](#)
- [95] [Giles, C. Lee, and Tom Maxwell. "Learning, invariance, and generalization in high-order neural networks." Applied optics 26.23 \(1987\): 4972-4978.](#)
- [96] [Hegger, Rainer, et al. "Identifying and modeling delay feedback systems." Physical review letters 81.3 \(1998\): 558.](#)
- [97] [Panzeri, Stefano, et al. "Sensory neural codes using multiplexed temporal scales." Trends in neurosciences 33.3 \(2010\): 111-120.](#)
- [98] [Gawne, Timothy J., Troels W. Kjaer, and Barry J. Richmond. "Latency: another potential code for feature binding in striate cortex." Journal of neurophysiology 76.2 \(1996\): 1356-1360.](#)
- [99] [Nadasdy, Zoltan. "Information encoding and reconstruction from the phase of action potentials." Frontiers in systems neuroscience 3 \(2009\): 499.](#)
- [100] [Zhao, Chenyuan, et al. "Spike-time-dependent encoding for neuromorphic processors." ACM Journal on Emerging Technologies in Computing Systems \(JETC\) 12.3 \(2015\): 1-21.](#)
- [101] [Mosleh, Somayeh Susanna, et al. "Brain-inspired wireless communications: Where reservoir computing meets MIMO-OFDM." IEEE transactions on neural networks and learning systems 29.10 \(2017\): 4694-4708.](#)