World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Research Article)

Explainable Artificial Intelligence and Deep Neural Network based Intrusion Detection System for Remote Sites in Oil and Gas Industry

Syed Anwarul Haque [1, *], Syed Azfarul Haque [2], Saeed M Yami [3], Panteleimon Korfiatis [4] and Vipul Thomas [5]

[1] Business System Analyst, Gas Compression Projects Department, Saudi Aramco, Al-Khobar, Saudi Arabia.
[2] Professor, Department of Physics, Jamshedpur Worker's College, Kolhan University, Jharkhand, India.
[3] Supervisor Project Engineer, Gas Compression Projects Department, Saudi Aramco, Al-Khobar, Saudi Arabia.
[4] Senior Project Engineer, Gas Compression Projects Department, Saudi Aramco, Al-Khobar, Saudi Arabia.
[5] Backbone OSP Technician, Area IT Department, Saudi Aramco, Haradh, Saudi Arabia.

World Journal of Advanced Research and Reviews, 2026, 29(01), 178-205

Publication history: Received on 24 November 2025; revised on 31 December 2025; accepted on 02 January 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.29.1.4337

Abstract

The new era of data networking involves Deep Neural Network for more efficiency and productive output. An intelligent autonomous intrusion detection system alone is not enough, when data security is important. Explainable Artificial Intelligence is based on standards and do not bypass human interference to make decision on intrusion flags. Ensuring network security is crucial and essential in Oil and Gas industries, especially for remote sites such as Wellheads, Gas Gathering Manifold and Remote Headers. Data transmission to plant and further to corporate network from remote sites is vulnerable and a target for attackers due to remoteness of the sites. These sites are mostly unmanned and remotely being monitored and controlled over network. This paper is presenting a review, analysis of integrating Explainable Artificial Intelligence (XAI) with Deep Neural Network (DNN) for Intrusion Detection System of IIoT's data transmission to corporate network in Oil and Gas industries. Here, we tried to explore and research about recent advancements in field of deep neural network-based intrusion detection system. This paper is presenting the idea to implement XAI integrated DNN for intrusion detection system in oil and Gas industries to protect any kind of cyber-attack on processed or raw data from remote sites. Many research papers have been analyzed and studied during the research and found many gaps which need to be filled when developing a smart intrusion detection system. Traditional black box transparent theory is not enough to combat cyberattacks but it needs human interface to have explanations of flags. Deep Neural Network based learning, training and testing made this research paper more accurate for intrusion detection when working with critical process data coming from IIoTs of unmanned sites of Oil and Gas industries. Weight based feature selection and attack analysis based on explanations based neural networks and pre-defined standards making this intrusion detection system best for complex networks of IIoTs.

Keywords: Deep Neural Network; Intrusion Detection Systems; Cybersecurity; Industrial Internet of Things (IIoTs); Convolutional Neural Network; Long/Short-Term Memory Neural Network.

1. Introduction

The cybersecurity in Artificial Intelligence (AI) and machine Learning (ML) era needs to be upgraded and should be based on historical patterns of worldwide threats. Attackers are aiming to target the easiest site for manipulations. Industrial Internet of Things (IIoT) is playing crucial role in remote sites of Oil and Gas industries. In recent years, there is a huge development in IIoT and connected sensing devices, making it easy for cyberattack and catastrophic damage to huge centralized systems. Therefore, an intelligent Intrusion Detection system is needed to protect from such high risk cyberthreats. Traditional intrusion detection system has many flaws such as generating overwhelming quantity of alerts and mostly signature-based detection relying on already known patterns. Cloud computing is mostly used to

[*] Corresponding author: Syed Anwarul Haque

support IIoT platforms. The ultra-modernization and digital transformation in Oil and Gas industries by extensive utilization of IIoT necessitates the development of best intrusion detection systems to counter potential cyber-threats. In recent years, many machine learning and deep learning-based intrusion detection systems have been developed including multi-Layer perceptron, convolutional neural network (CNN), recurrent neural network (RNN), Hybrid CNN-RNN, and Federated Learning, Long Short-term memory (LSTM) networks, deep autoencoders, gated recurrent units (GCU) etc.

The challenges became more severe as current AI infrastructure is facing difficult in IIoT-Cloud environment due to diverse data sources and reaching to centralized server. Advanced Deep Learning methodology is must for the cybersecurity of interconnected IIoTs in cloud domain.

Deep Neural Networks are sophisticated machine learning models inspired by the human brain's neural structure. They consist of multiple layers of interconnected nodes (neurons) that transform input data through successive nonlinear transformations to extract hierarchical patterns and make predictions

Deep Neural Network is playing pivotal role to improve response time and latency in IIoT devices applications a better energy consumption and provide a safe environment to these systems. Deep Learning models having the ability to learn patterns and accordingly can make predictions are perfect choice to study cyberattacks which are not easy to locate by traditional rule set basis.

Industrial Internet of Things (IIoT) platforms facing challenges in traditional security approaches such as scale, heterogeneity, Resource Constraints, dynamic environment factor and Novel Attack vecors. DL based IDS is having many benefits, such as:

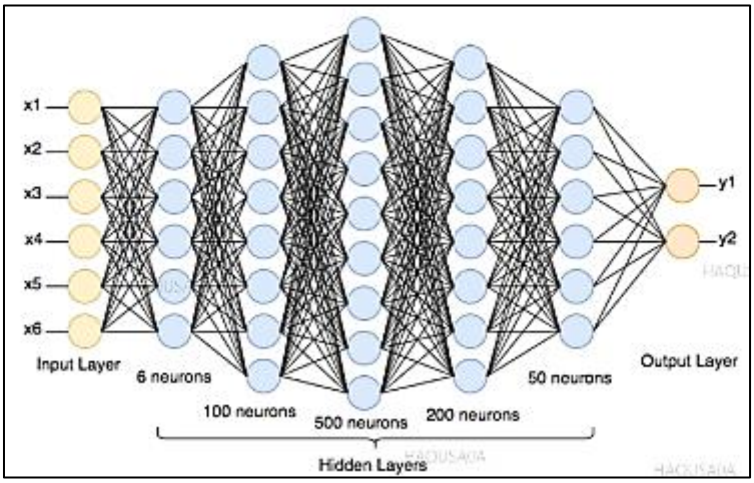
- Automatic feature learning and abstraction.
- Handling high dimensional heterogeneous data.
- Detecting sophisticated and evolving attacks.
- Temporal pattern recognition
- Scalability and Distributed Learning.
- Adaptive Learning and Concept Drift Handling
- Multi-Task Learning and Efficiency improvement

Different neural networks are being used such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), autoencoders etc.

1.1. Deep Neural Network:

Deep Neural Network is representing a paradigm shift in IoT intrusion detection system by offering:

- **Superior Detection Capabilities:** Complex patterns recognition across heterogeneous data.
- **Adaptive Intelligence:** Continuous learning from evolving threats.
- **Operational Efficiency:** Multi-task learning reduces deployment complexity.
- **Scalability:** Distributed Learning architectures for massive IoT deployments.
- **Future-Proofing:** Foundation for integrating advanced techniques such as reinforcement learning, neuromorphic computing etc.



Reference: Deep Neural Network architecture | Download Scientific Diagram

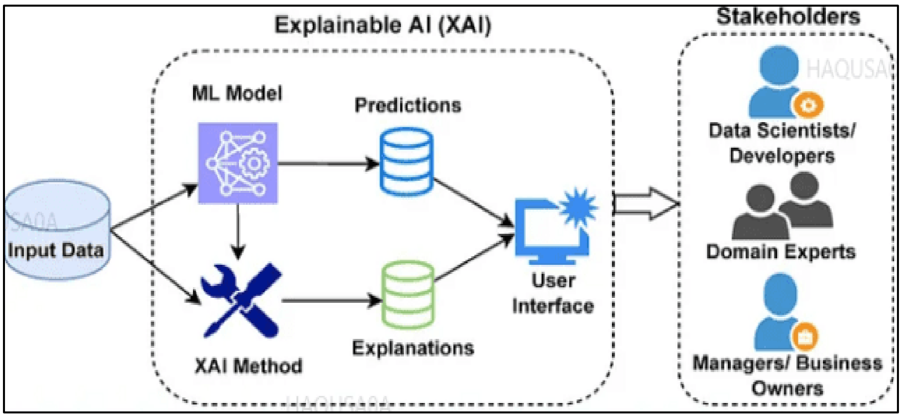
Figure 1 Deep Neural Network Architecture

As IIoT networks are growing day by day in scale and complexity, DNN based IDS transitions form a competitive advantage toa security necessity. In organizations DNN powered security system are building resilient, intelligent and adaptive IoT infrastructures capable of tomorrow's unknown attacks. DNN based IDS can be implemented in Edge side of the network initially to protect critical assets on plant side.

Deep Neural Network represent a paradigm shift in IIoT intrusion detection by providing super detection capabilities, adaptive intelligence, increasing operation efficiency and promising for future integrations.

1.2. Explainable AI

Explainable AI is creating an AI system, which can provide clear, understandable explanations for their decision and actions It helps human to interpret AI outputs by which building trust between Human and Machine.



Reference: Demystifying Explainable AI: A Beginner's Guide with Examples | by Jyoti Dabass, Ph.D. | Python in Plain English

Figure 2 Explainable AI architecture diagram

1.3. Explainable AI (XAI) Integration with Deep Neural Network (DNN) for Intrusion Detection System (IDS):

In this paper, we are presenting the idea of Explainable AI (XAI) integration with Deep Neural Network for intrusion detection in IIoT platforms. XAI is a technique that make AI/ML models decisions understandable, interpretable and transparent to human. The basic need of involving XAI is because DNNs operates as "black box", making predictions without revealing to human as why or how decisions are made.

In a security context, an alert without a reason is nuisance. XAI addresses the core operational limitations of DL-IDS.

1.3.1. Building Trust and Adoption:

Analysts must trust the system to prioritize its alerts over hundreds of others.

1.3.2. Improved Model Debugging & Refinement:

Security engineers can understand model failures (e.g. why a certain attack was missed) to improve training data or architecture.

1.3.3. Reducing False Positives:

Analysts can quickly dismiss alerts with nonsensical or benign explanations

1.3.4. Regulatory Compliance:

Each decision is documented with reasoning, clear risk scoring with competent breakdowns and producing automated reports demonstrating fair, unbiased detection.

1.3.5. Facilitating Forensic & Reporting:

Explanations provide a starting point for root cause analysis and are essential for compliance reports.

1.3.6. Accelerating Response

Explanations (Flagged because of rare SSL certificate length+ anomalous geolocation) guide immediate investigation steps.

1.3.7. Human-AI Collaboration

Saving time of security analyst by providing explanations and ranks for alerts and Analyst can learn from AI's pattern recognition.

1.3.8. Adversarial Robustness & Security

Any unusual explanation patterns flag potential adversarial attacks. XAI helps in identifying model vulnerabilities. If attack traffic generates widely different explanations for similar attacks, it indicate evasion attempt.

Integrating XAI with Deep Neural Network for IIoT intrusion detection transforms security operations from reactive monitoring to proactive and intelligent defense system. The benefits extend beyond technical improvements to encompass organizational trust, regulatory compliance and operational efficiency. XAI is not an optional requirement for IIoT platforms but it is requirement for trust and compliance. XAI is not replacing humans in decision makings. It is transparent explainable system that security teams can trust, understand and effectively collaborate with defined increasingly complex and critical systems.

2. Related Studies

Different studies have been conducted for intrusion detection system utilizing Deep Learning methodologies, K. Vinotha and Dr. P. Eswaran gave Quantum search enhances bat algorithm approach which integrates quantum-inspired search dynamics with adaptive swarm intelligence and produced an experimental results of 95.7% Intrusion detection accuracy[15]. Xue and Li produced an approach of Multiscale Convolutional Neural Network based on multi-head attention mechanism and hierarchical long short term memory network-based security intrusion detection model [5], where they utilized MCNN-MHA-HLSTM technique and adopted a multiscale convolutional neural network and combined it with head attention mechanism for improving the accuracy of intrusion detection and reducing the impact of imbalance. MCNN-MHA-HLSTM can capture the correlation and local features between security data sequence and attack data sequence. It employs multiple effective strategies, the common is data augmentation which aims to increase minority class samples by generating more diverse training samples. This can be achieved by various transformations on minority class samples, thereby, helping MCNN-MHA-HLSTM better learn and improve detection accuracy. In this approach a multiscale convolutional neural network using parallel branches. Each branch is having different depth using convolution kernels of different sizes and spans. This approach is having a multi head attention mechanism used in neural network model to focus on different parts of the input sequence. It consists multiple attention head each of which independently perform scaling dot product attention operations H times and using different query, key and value matrices each time and finally concatenates the output of all heads. The input sequence is mapped into three vectors, represented as query, key, and value information and weighted values of each position is calculated through scaled dot product attention operation for weighted summation of the input sequence. [5]. This approach is using HLSTM into two parts based on different input and output situations of modules at different levels. The hierarchical structure of HLSTM is having the capabilities to model complex data sequence across multiple time levels and learn the input data at symbol level, word level sequentially. This approach is providing 95.2% accuracy in intrusion detection. [5]. Vandana Shayka, Jaytrilok Choudhry presented a model of intrusion detection by utilizing Deep Neural Network to enhance the performance. The most effective feature from the data set is chosen through the cross-correlation method. DNN structure is used for testing and training. They reached to an accuracy of 96.23% in intrusion detection [6]. Another approach is of MS Harish, Lokesh, Lokesh S Ramanujan and others produced a method of Hybrid Deep Learning model for network intrusion detection using optimal feature fusion, [12] This approach is concentrating on creating a balanced training approach by defining an appropriate feature fusion process for deep learning. Deep Learning is considered to achieve high accuracy through an intelligent design. This model is identifying and distinguishing abnormal behavior in network traffic by utilizing hybrid deep learning methodology. The fusion process is using the RCMPA, which tune weights assigned to each characteristic, ensuring most relevant information is accentuated. The purpose of Multi-scale Dilated Deep Hybrid Network with Attention Mechanism" (MDDHN-AM) is deep learning model handling different sequential lengths of data and promptly highlighting complicated traffic patterns. [12]. MDDHN-AM is highly affective intrusion detection model that combines multi-scale detailed convolutions, hybridized temporal networks and an attention mechanism to deliver superior performance in detection its ability to adjust multipart attack pattern, prioritize critical data, and effectively learn from temporal dependencies make it a powerful tool in safeguard modern network infrastructures. The RCMPS-MDDHN-AM based ID outcomes achieved an accuracy of 96.2%. [12]. Heba Dhirar and Ali Hamad utilized software-Defined Networking approach (SDN) and increased the accuracy of Intrusion detection to 96.4% [17].Another approach from Omar Achbarou and his team is Intrusion detection system using feature selection and hybrid learning models for high performance and efficiency**.** This approach is showing the IDS system working in real time and leveraging the automatic extraction by deep learning. The model is for enhancing the accuracy and reducing the false positive rate and execution time and ensuring its ability to adapt emerging attacks in IIoT environment. [13]. This model is based on Convolutional Neural Network are well suited to capturing intricate pattern within high-dimensional data, which makes them an excellent choice for detecting anomalies in network traffic. Convolutional Neural Network (CNN) is having ability to automatically learn hierarchical feature representations from raw network data, which can enable a relatively accurate classification of network traffic as either normal or attack data. the development of the proposed model is structured into five key stages: data processing, the training-test split, feature selection, model building, evaluation. [13]. The process starts with data loading, and stored in CSV file with columns representing various network features, and a target label indicating whether the traffic consisted of normal data or an attack. This technique is converting the categorical data into numerical form. After label encoding there is a normalization of the data using minimum-maximum scaling technique. This technique adjusts the feature values to a standardized range in (0,1) to ensure each feature is contributing equally to the model and prevent bias form features with larger value range. The technique is embedded with RF, XGBoost and Light GBM producing 99.2% accuracy in intrusion detection[13] Mourad Benmalek and Abdessamad Seddiki reached to an accuracy of 99.6% by utilizing RT_IoT2022 dataset, which captures complex IoT attack scenarios. This study is utilizing particle swarm optimization, a bio metaheuristic for feature selection and optimization and reduced computational overhead and enhanced the performance [16].

3. Research Gaps and Motivation

3.1. Data Related Challenges

3.1.1. Lack of high-quality representative and current datasets:

- Obsolete Datasets

Most of the research is relying on old datasets like KDD-CUP99 or NSL-KDD, which do not reflect modern network traffic, protocols (e.g. IoT, Cloud) or sophisticated attack vectors.

- Synthetic/Lab-Generated Data

Many datasets are created in lab environments, lacking the complexity, noise and heterogeneity of real enterprise networks.

- Privacy and Sensitivity

Real network traffic contains highly sensitive information. Sharing such data for research is difficult due to privacy laws (GDPR, CCPA) and corporate policy, stifling progress.

3.1.2. Severe Class Imbalance

In real networks, malicious activity is extremely rare (<<1% of all traffic). Deep Learning models tend to become biased toward the majority class (normal traffic), leading to poor detection rates for actual attacks.

3.1.3. Lack of Labeled Data

Accurately labeling network flows or logs as specific attack types is a time-consuming, expert-level task. This scarcity of high-quality labels limits the effectiveness of supervised DL approaches.

3.1.4. Concept Drift

The statistical properties of "normal" and "malicious" traffic evolve over time (e.g. new applications, network upgrades, changing user behavior). A static DL model's performance degrades rapidly unless it is continuously retrained with fresh data.

3.2. Model-Related & Technical Challenges

3.2.1. Interpretability and Explainability (The "Black Box" Problem)

This is a critical operational hurdle. Security analysts (SOC teams) need to understand why an alert was generated to investigate, triage and respond effectively. A DL model that just outputs "anomaly score: 0.95" without explaining which features contributed to the decision is met with distrust and slows down incident response.

3.2.2. High Computational Cost

Training complex DL models (e.g., LSTM for time-series, deep autoencoders) requires significant GPU/TPU resources and time. While interface is faster, the need for frequent retraining to handle concept drift adds to operational overhead.

3.2.3. Feature Engineering Vs Representation Learning

A core promise of DL is automatic feature learning. However, for network data, raw packets (bytes) are less directly informative than for image pixels. Effective DL for Intrusion Detection System (IDS) often still requires careful preprocessing and hybrid approaches (combining handcrafted features like flow statistics with learned representations), undermining the "end to end" learning ideal.

3.2.4. Adaptability and Generalization

Models trained on one network environment often fail to generalize to another due to differences in topology, services, and normal behavior patterns. Creating a universally effective DL based IDS is extremely difficult.

3.2.5. High False Positive and False Negative Rates

Despite DL's power, tuning models to achieve an operationally acceptable balance between false alarms (which cause alert fatigue) and missed detections (which are critical security failures) remains challenging, especially with imbalanced data.

3.3. Adversarial Challenges (A Cat-and-Mouse Game):

3.3.1. Adversarial Machine Learning (AML)

Attackers can deliberately craft inputs to evade detection.

- Evasion Attacks

Manipulating attack payloads or network flow characteristics in subtle ways that are indistinguishable from normal traffic to the DL model but preserve malicious intent (e.g., adding perturbation to malware code or timing of exfiltration).

- Poisoning Attacks

Injecting carefully, crafted malicious data into the model's training set to "poison it", causing the model to learn incorrect patterns and fail during deployment.

3.3.2. Lack of Robustness

Many DL models are surprisingly fragile to these small, adversarial perturbations, making them unavailable in hostile environment where adversaries are aware of their deployment.

3.4. Operational and Deployment Challenges

3.4.1. Integration with Security Orchestration

Fitting a DL-based IDS into existing security Information and Event Management (SIEM) systems, threat intelligence platforms and SOAR (Security Orchestration, Automation and Response) Workflows in non-trivial.

3.4.2. Real-Time Processing Requirements

While DL inference can be fast, processing high-volume, high-velocity network traffic (e.g. 10/40/100 Gbps) in real-time requires optimized pipelines and hardware, which can be expensive.

3.4.3. Skill Gap

Operating, maintaining, and fine-tuning DL systems require expertise in both machine learning and cybersecurity, which is a rare combination in many security teams.

3.4.4. Trust and Accountability

Due to the “black box” nature and potential for error, security managers are hesitant to fully automate responses based on DL alerts. The lack of accountability for model errors is significant barrier.

3.5. To Mitigate the challenges, below future trends need to be implemented:

3.5.1. Hybrid Approaches:

Combining ML/DL with traditional signature-based methods and rule-based systems.

Semi-Supervised & Self-Supervised Learning:

- To reduce dependency on labelled data.
- Lifelong/Continual Learning:
- To adapt models continuously to concept drift.

Explainable AI (XAI) Integration:

- Developing techniques like SHAP or LIME specifically for network intrusion detection to build trust.

Federated Learning:

- To train models on decentralized data (e.g., across different network segments) without compromising privacy.

Focus on Production-Ready Systems:

- Research shifting from pure accuracy gains to efficiency, scalability and integration.

3.6. Explainable AI (XAI) and Deep Learning based IDS Model:

XAI and DL based IDS model is being researched here for IoTs data protection from remote areas of Oil and Gas industries. Mostly the idea of XAI and DL integration is opaque, and not clear for practical use. Here, in this paper we will discuss about the integration and implementation of the model for protecting the data at edge before reaching to SCADA server and company corporate network.

Basically, there are two kinds of XAI approach:

3.6.1. Post-Hoc, Model-Agnostic Technique:

Applied to any DL model after training

SHAP (SHapley Additive exPlanations):

It calculates the contribution of each input feature (e.g., packet size, destination port, TCP flag count) to a specific prediction.

- **Use in IDS :**

A SHAP force plot can show that connection is classified as a DDoS attack primarily because of the extremely high packet_rate (feature contribution: +0.6) and the low avg_packet_size (contribution: +3.0) despite the normal destination_port (contribution (-0.1).

LIME (Local Interpretable Model-agnostic Explanations):

Creates a simple interpretable model (like linear regression) to approximate the complex DL model's behavior locally around a single prediction.

- **Use in IDS:**

Good for explaining individual anomalous connections. The model sees this connection as a similar to SQL injection because these 5 specific byte sequences in the payload were weighted heavily.

Intrinsic or Model-Specific Techniques (Leverage the model's own architecture):

Attention Mechanism:

For sequence models (LSTMs, Transformers) Processing packet flows or system logs, attention weights reveal which parts of the sequence (e.g., which specific packets in a flow or which log entries) the model "paid attention to" when making a decision.

- **Use in IDS:**

Visually attention over a sequence of HTTP requests can pinpoint the exact malicious request in a session.

Gradient Based Methods:

Techniques like Saliency Maps or Grad-CAM (more for images) can be adapted to highlight the most influential input features by analyzing the model's gradients.

Surrogate Models

Train a globally Interpretable model (like a decision tree) to approximate the overall behavior of the black-box DL model. Less precise for individual cased but gives a global sense of model logic.

4. Contribution

This technical research paper is providing a conceptual model of Explainable Artificial Intelligence (XAI) integrated with Deep Neural Network (DNN) based Intrusion detection system, which can be utilized in Oil and Gas industries at Edge of the network to protect the data coming from IIoT sensors. Mostly Remote sites are vulnerable in Oil and Gas industries, as there is no security arrangements and mostly unmanned. These sites require strong cyber security systems to dealt with possible cyberattacks. Edge computing is a necessary tool for these sites but the data also need to be secured from any kind of intrusion or external attacks before transmitting to plant side and to corporate network.

In this paper, we studied different aspects of Deep Neural Network based Machine Learning and it integration with Explainable Artificial Intelligence (XAI) to protect output data. Feature selection method by utilizing XAI integrated DNN is been analyzed and received a report which is ready for Security analyst to review and understand threats and train the system for similar kind of patterns.

The model we proposed, is system which can be practically implemented to secure the oil and gas industries remote site locally before transmitting it to plant side network which is integrated with corporate network of the Company. Different analysis carried out to reach The integration of XAI with Deep Neural network can provide 100% intrusion detection as it is automated but interfacing with human and teaching as well for different patterns and kind of cyber-attacks.

5. Methodology and Frame Work

5.1. Step-1 Input Data from IoTs:

In proposed model, first of all, data is being collected from IoT sensors (Actuators, Transducers, Flow valves, pneumatic valves, strain gauges, thermocouples, flame detector, shutdown valves etc.) sensors installed at Remote site of Oil and Gas Industry. The data is in electrical form reaching to marshalling cabinet by instrument control wires. The collected signal is fed to hypervisor PLC which translate the data in binary sequences.

5.2. Step-2 Data Loading

Data loading refers to the systematic process of ingesting, organizing and preparing IoT data for immediate processing and analytics within Edge Computing environment. It is not just receiving bytes but making the data to consumable, reliable and ready for action. Below are basic steps of data loading:

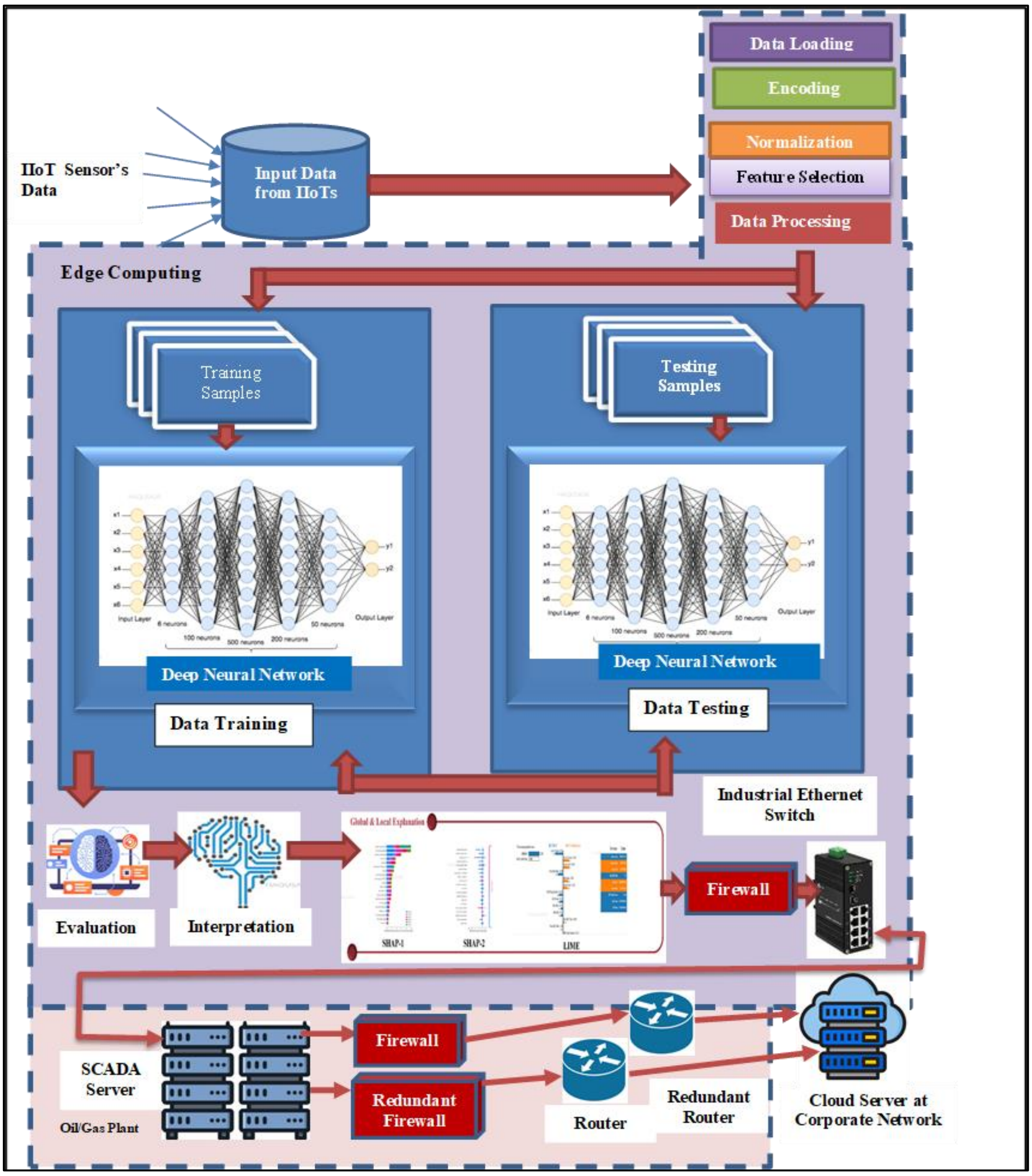


Figure 3 Proposed model of Deep Neural Network Based Intrusion Detection System for Remote sites of Oil and Gas Industries

5.3. Step-3 Data Ingestion Interface

- Protocol Handlers:
- Simultaneous support for multiple IoT protocols
- Modbus/OPC-UA
- DNP-3
- DNP Secure etc.
- Connection Management:
- Maintain persistent connections with hundreds of IoTs.
- Que Management: Handle data bursts with message queues.

5.4. Step-4 Data Buffering and Streaming

- Immediate handling of incoming data flow:
- In-Memory Buffers:
- Store microseconds of data for immediate processing.
- Streaming Pipelines:
- Creates continuous data streams for real-time analytics.
- Backpressure Handling:
- Manage what happens when data processing can't keep up with ingestion.

5.5. Step-5 Schema Management and validation

- Ensuring data quality and ingestion.
- Schema Enforcement:
- Reject malformed data immediately
- Type Conversion:
- Convert all data to consistent types (strings to numbers, etc.)
- Constraint Validation:
- Check value ranges and required fields.

5.6. Step-6 Data Parsing and Decoding

Transforming raw data into structured format.

5.7. Step-7 Metadata Attachment

- Enriching raw data with context:
- Source Metadata:
- Device ID, manufacturer, firmware version.
- Temporal Metadata:
  - Ingestion timestamp, processing latency.
- Network Metadata:
- Signal strength, connection quality.
- Geospatial Metadata:
- GPS coordinates, Location context.
- Security Metadata:
- GPS coordinates, location context

5.8. Step-8 Data Routing and Prioritization:

Directing data to appropriate processing pipelines.

5.9. Step-9 Data Encoding:

Data encoding at the edge refers to the process of transforming structured IoT data into optimized formats for storage, transmission and processing. It's about making data smaller, faster and smarter within edge constraints. Encoding of data is necessary because of below requirements:

5.9.1. Minimize size

- Reduce storage and transmission needs.
- Accelerate Processing:
- Enable faster analytics.
- Preserve Precision:
- Maintain data quality within limits.
- Enable Interoperability:
- Standardize across diverse devices.

5.10. Step-10 Data Normalization:

Data normalization at the edge is the process of transforming encoded data into consistent, comparable scales suitable for analysis, ML models and aggregation. It is about making diverse IoT data apples-to-apple comparable despite coming from different sensors, units and ranges. Normalization makes data mathematically consistent and analyzable. Below are different normalization techniques being used:

Min-Max Scaling (Normalization to (0,1):

- Data is being stored in 0 and 1 pattern as Min and Max per sensor type.
- Quantization Aware Normalization.
- Federated Learning normalization

Data normalization is the critical bridge between encoded sensor data and actionable intelligence. It transforms desperate measurements into a common mathematical language that DL models can understand and analyze effectively.

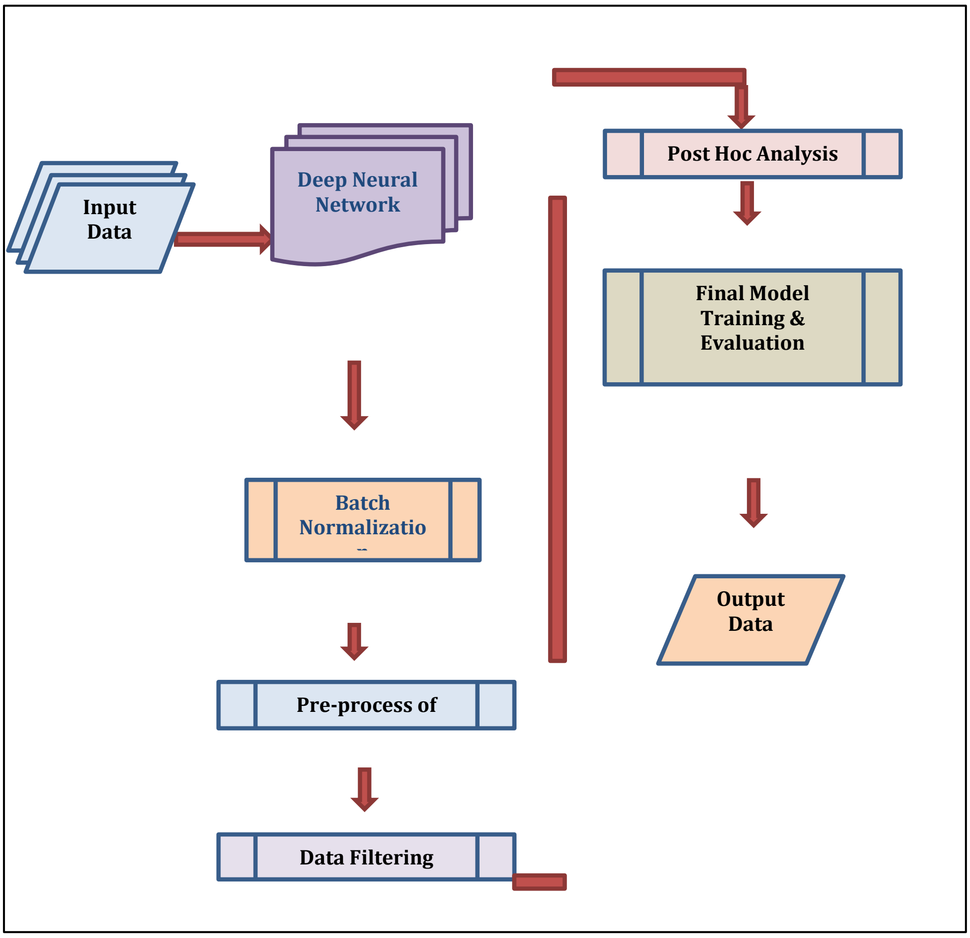


Figure 4 Feature selection process

5.11. Step 11 - Feature Selection:

Feature Selection in DL based IoT intrusion detection is the process of identifying and retaining the most relevant, informative and discriminative features from raw IoT network/sensor data while removing redundant, irrelevant or noisy features that impair detection performance and efficiency. Below are the basic functions of feature selection:

5.11.1. Computational Efficiency & Speed

Reducing input dimensionality lowers training/inference time and resource consumption, crucial for real-time or near-real-time intrusion detection.

5.11.2. Improved generalization and Reduced Overfitting:

Eliminating irrelevant redundant or noisy features and helps the model focus on meaningful patterns, improving performance on unseen attack variants.

5.11.3. Model Interpretability:

Selecting a smaller, meaningful subset of features makes it slightly easier to understand what the model is "looking at" a significant challenge in black-box DL models.

5.11.4. Data Quality Enhancement:

The process often reveals inconsistencies, outliers and redundancies in the dataset (e.g., NSL-KDD, CIC-IDS2017, UNSW-NB15).

- ⇓ **Preprocess:**
  Clean and normalize the dataset (e.g., CIC0IDS2017).
- ⇓ **Initial Filter:**
  Apply a fast filter method (e.g. Correlation + Mutual Information) to drastically reduce the feature space from, say 80 to 40 features. This removes obvious noise and redundancy.
- ⇓ **Embedded Selection with DL Model:**
  Train a DNN or CNN with L1 regularization on the first layer using the 40 features.
- ⇓ **Post-Hoc Analysis & Final Selection:**
  Use the weights from regularized layer or apply SHAP analysis to rank the 40 features. Select the top 20-25 most important features.
- ⇓ **Final Model Training & Evaluation:** Train an optimized, potentially larger DL model (without L1 constraint) on this final subset. Compare its performance and speed against a model trained on all original features.

The methodologies being used in Feature selection are below:

5.12. Filter Methods

In Filter Method Features are being selected by statistical measures:

Procedure: Compute a score for each feature and select the top-k.

5.12.1. Techniques:

- **Variance Threshold:** Remove features with very low variance (almost constant).
- **Correlation Analysis:** Remove highly correlated features (e.g., using Pearson, Spearman) . If (src_bytes and dst_bytes are perfectly correlated, one can be dropped.
- **Univariate Statistical Tests:** A NOVA F-test (for classification) identities features where mean values differ significantly across attack/normal classes.
- **Information-Theoretic Measures**: Mutual Information (MI) measures the dependency between a feature and the target table. High MI features are retained.

Advantage: Fast, Scalable, model-agnostic.

Disadvantage: Ignores feature interactions and the DL model's learning capability.

5.12.2. Wrapper Methods

Use the performance of a specific DL model to evaluate feature subsets.

Procedure: A search strategy (e.g. forward selection, backward elimination, recursive feature elimination) is used to test different feature subsets, training and evaluating the DL model each time

Example (Recursive Feature Elimination RFE):

- Train the DL model on all features.
- Rank features by importance (using gradients, weights or permutation importance).
- Remove the least important features.
- Retrain the model and repeat until the desired number of features reached.

5.12.3. Embedded Methods:

Feature selection is built in to the training process of the DL model itself. This is the most natural and popular approach for DL-IDS.

Common Techniques:

Regularization (Lasso):

Adding L1 penalty to the first layer's weights encourages many weights to become exactly zero. Features connected to zero weights are effectively unused.

Attention Mechanism

In models like Transformers or RNNs with attention, the attention weights over input features can be analyzed. Features consistently receiving low attention can be considered less important.

Differential Architecture Search:

More advanced methods where the model learns which input connections are useful.

5.12.4. Hybrid & Advanced DL-Centric Methods:

- Autoencoders for Feature Representation:

Train an unsupervised Autoencoder (AE) or Variational Autoencoder (VAE) on the data. The bottleneck layer's learned latent representation is a new reduced set of features that captures the essential information. This is feature extraction not selection but serves the same purpose of dimensionality reduction.

Convolutional Neural Networks (CNNs) on Structured Data:

Raw features are treated as a 1 D "signal". The initial convolutional layers act as automatic feature extractors, learning hierarchical patterns. The original feature set can be large, as the CNN will learn to combine them.

- Important Scoring Post-Hoc:

After training a high-performing DL model (e.g., DNN, CNN) use techniques like SHAP or gradient based Sailency Maps to explain predictions. Aggregate these explanations to rank the global importance of the original input features. This is an interpretability-driven section method.

5.13. Step -13

After feature selections, normalization of processed data is being done to standardize the data sets. Random sampling is done during the data training and testing step to preserve the class distribution the data in both sets. The data set is being thoroughly processed and help to develop a more trustworthy model.

5.14. Step- 14

At this step data is being evaluated. In data evaluation different metrices are assessed such as accuracy, precision, Recall etc. The metrices measure aspects of IDS model performance.

$$\text{Accuracy} = \frac{\text{Number of Correct Prediction}}{\text{Total Number of Prediction}}$$

Accuracy measures the how the model is predicting the class of an instance. In case of class imbalance , one class represents a small fraction of the total number of samples.

The F1 score help us to evaluate a system's overall performance and measuring trade-off between the detection of attacks and a reduction in the false positives.

$$F1\ Score = 2\ X\ \frac{\text{Precision X Recall}}{\text{Precision + Recall}}$$

Precision represents what is correct. Precision is top in intrusion detection system because a false positive can trigger an alert and wasting resources and potentially disrupting the services.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive + False Negative}}$$

Recall measures how well the model identifies all positive examples. High Recall is critical metric in intrusion detection since missing false negative will be disastrous. It is better to have falser positive than more missed attacks. It is authoritative to catch all the attacks, recall becomes the most crucial metric for evaluating the models.

5.15. Step -15

A cross-validation can be to evaluate the model performance. K fold cross validation with 5 fold models can be assessed for accuracy, consistency across different training splits.

$$CV\ Accuracy\ = \frac{1}{k} + \sum_{i=1}^{k} Accuracy_i$$

Where k = 5 in our case. Each fold involves training the model on k – 1 parts and testing it on remaining part.

Now we will see the model performance variation with data quantity in training.

$$Training\ Error = \frac{1}{n}\sum_{i=1}^{n} L\ \left(y_i, y_i^{\wedge train}\right)$$

$$Validation\ Error = \frac{1}{m}\sum_{i=1}^{m} L\ \left(y_i, y_i^{\wedge val}\right)$$

Where L is the loss function, n is the number of training samples and m is the number of validation samples.

5.16. Integration of XAI with Deep Learning:

Combining Explainable Artificial Intelligence with Deep Learning based IoT IDS after feature selection creates a transparent, trustworthy and adaptive security system. Below are the reasons for selection XAI to integrate with DL for IIoT IDS.

5.16.1. Trust and Accountability

Security Operators need explanations to trust automated alerts and take appropriate action.

5.16.2. Model Debugging & Refinement

Identify when models use wrong features or shortcuts.

5.16.3. Regulatory Compliance

GDPR, NIS2 and other regulations increasingly require Explainability in automated decision making.

5.16.4. Feature Selection Validation

XAI validates whether selected features are truly meaningful to the model's decisions.

5.16.5. Adversarial Robustness:

Understanding model vulnerabilities helps harden against attacks.

5.17. Integrated XAI-DL Training & Testing:

5.17.1. XAI-Informed Feature Selection:

Before training begins – XAI can guide final feature selection. Different methods are being used such as SHAP, LIME and Permutation importance. In SHAP values are being calculated to select simple features from baseline, which are then trained on the selected subsets and zero weight SHAP values are being removed. LIME is used for critical attack samples, verify visa LIME at the selected features are sufficient for local explanations. If not reconsider feature set. Permutation Importance is good for computing permutation after initial deep learning. If any selected feature is having zero importance, it can be safely removed. At out we receive final validated feature set with documented importance scores. Below is training procedure with XAI integration.

- ○ Standard Training Loop with added XAI Monitoring.
- ○ **Explanation-Guided Regularization**:
  - Add a loss term that penalizes inconsistent explanations across similar attack types.
  - **Implement concept alignment Loss**: Encourage model to align important features with domain expert knowledge.
- ○ **Training with Explanations:**
  - Generate adversarial examples that fool both the prediction AND the explanation.
  - Train model to be robust against these "explanation-aware" attacks.
  - Train LSTM with attention mechanism on chronologically split data.
  - Every 5 epochs generate attention heatmaps for validation samples.
  - Verify attention focuses on port scanning and communication patterns.
  - If attention shift is irrelevant features, regularization is applied.

- ○ **Testing with Comprehensive XAI Evaluation:**
  Testing goes beyond accuracy metrics to include explanation quality matrices.
- Standard Performance and Accuracy Testing
- **XAI-Specific Testing & Metrices:**
- ❖ **Explanation Faithfulness (Plausibility):**
- **Feature Ablation Test:** Systematically remove top-K important features (according to explanation). Performance should drop significantly.
- **Randomization Test:** Randomize important Vs Unimportant features. The former should cause larger performance degradation.
- **Stability:** Adding some noise changes explanations.
- ❖ **Metric:** Intra-class explanation similarity.
- ○ **Human Alignment (Domain Expert Validation):**
- Present explanations to IoT security experts.
- Questionnaire: On a scale of 1-5 does this explanation align with your understanding of why this is an attack.
- ❖ **Metric:** Expert agreement score.

Table 1 IIoT-Specific XAI Testing Scenarios

SL/No.	Test Scenario	XAI Application	Evaluation Criteria
1	Zero Attack	Generate explanations for unknown attack types. Do they highlight features similar to known attacks?	Explanation transferability across attack families.
2.	False Positive Analysis	For each false positive analyze explanation. Is the model focusing on legitimate but unusual IoT behavior.	Percentage of FPs with reasonable but incorrect explanations.
3.	Device Heterogeneity	Test same attack across different sensor. Are explanation consistent?	Cross device explanation consistency
4.	Concept Drift Detection	Monitor explanation patterns over time in simulated deployment. Sudden changes may indicate drift.	Explanation distribution shift detection.

- **Edge Case Testing:**
  - **False Positive:** Model flags firmware updated as suspicious. Explanation shows high packet size variance as reason. Analyst marks as false positive with note: "Legitimate bulk data transfer."
  - **Retraining:** Model is retrained with this sample, weighted by explanation feedback.

- **Continuous Learning with XAI Feedback Loop:** the key benefit of XAI integration with Deep Learning based IDS is the human-in-the—loop refinement. Selecting uncertain predictions (high entropy) with surprising explanations for human labelling and retraining the model by giving higher weight to samples where explanations were corrected by analysts.

6. Challenges in integrating XAI with Deep Neural Network for IDS of IIoTs:

Integrating Explainable AI (XAI) with Deep Neural Network (DNN) based intrusion detection system for IIoT environments is necessary for moving from "black box" security to "trustworthy" security. However, this integration faces several technical and structural challenges. The challenges span technical, operational and human-centered. Below are challenges of present days:

6.1. Resource Constraints

The most significant hurdle is the "explanation tax". IIoT devices are having limited CPU, memory and battery life. XAI techniques add computational and memory overhead especially for post-hoc explainers and Feature attribution methods.

6.1.1. High Complexity

XAI methods like SHAP (Shapely Additive Explanation) or LIME (Local Interpretable Model-agnostic Explanations) require thousands of model evaluation to generate a single explanation. There is often a trade-off between detection accuracy and interpretability. DNNs are valued for their high accuracy in detecting complex and novel attacks. However, this performance often comes from highly no-linear, hierarchical feature transformations that are inherently opaque. XAI methods must "reverse-engineer" these transformations in to human-understandable terms, which can be an approximation at best.

6.1.2. Real-Time Detection Requirements:

In an IDS detection must happen in milliseconds. Adding an XAI layer can introduce significant latency, making it difficult to stop an ongoing attack (like DDoS) in real time. XAI methods are often slow and iterative especially for instance-level explanations and time-series traffic analysis. Delayed explanations reduce operational usefulness. Balancing real-time detection with timely explanations is difficult.

6.1.3. Edge/Device-side XAI:

Requires ultra-lightweight DNNs and XAI techniques potentially sacrificing accuracy and explanation depth. Explanations are generated remotely, introducing latency, communication overhead and privacy concerns as raw data must be transmitted.

6.2. Accuracy-Explainability Trade-off:

There is a natural strain between how fine a model accomplishes and how simply it can be described.

6.2.1. Complexity of DNNs:

DNNs (Convolutional Neural Network (CNNs), Referral Neural Network (RNNs), Long Short-Term Memories (LSTM), are powerful because they capture high dimensional, a non-linear relationship. When XAI is being used to simplify in human readable terms, we often lose the tone that make DNN accurate at first place. Highly accurate IDS models often are the least explainable.

6.2.2. Fidelity Issues:

A "authentic" explanation must exactly reflect the model's logic. If an XAI tool provides a basic description that doesn't truly match the DNN's internal decision path, it can lead to wrong sense of security. Some XAI techniques (e.g., surrogate models like LIME) provide simple comprehensible explanations but may not faithfully represent the complex DNN's true decision-making process. Others (e.g., gradient-based methods like integrated gradients) are more faithful to the model but produce outputs (heatmaps, attribution scores) that require expert knowledge to interpret.

6.2.3. Right explanation for the Right Stakeholder:

An IIoT network administrator needs a different explanation (e.g., "This device is behaving like it's in a botnet") than a model developer (e.g., The high frequency of small TCP packets from this sensor was the primary trigger). Developing XAI that serves multiple audience is difficult.

6.2.4. Evaluation of Explanations:

There is no universal metric to judge if an XAI method's output is "correct" or "good enough ". This lack of ground truth for explanations makes it hard to validate and compare XAI techniques in this context.

6.3. Data Heterogeneity and Dynamic Environments:

IIoT networks generate huge, diverse and high rate of data streams (e.g., MQTT, CoAP, HTTP protocols). IIoT networks comprise diverse devices (sensors, cameras, valve data etc.) with different data formats, protocols and behavioral profiles. A one-size-fits for all XAI approach will not work. Explanations must be adaptable to the context of the specific device type and its normal behavior.

6.3.1. Feature Mapping:

Mapping raw network packets to "interpretable features" is difficult. A security analyst might understand "Source IP", but they may not understand why a specific weight in 1D-CNN layer triggered an alert.

6.3.2. Concept Drift

IIoT situations change continuously (new devices join, traffic patterns change). XAI models trained on static datasets may provide obsolete explanations as the primary network behavior changes.

6.3.3. Data Specificity

IIoT IDS data is often multimodal (network packets, device logs, physical sensor readings) and highly temporal. XAI methods must effectively explain decisions based on complex time-series and sequential data, which is more challenging than explaining image or text classifications.

6.3.4. Integration with Security orchestration:

Explanations need to be formatted and communicated to be consumed by Security Information and Event Management (SIEM) system, threat intelligence platforms. This requires standardization of explanation outputs which currently doesn't exist.

6.4. Adversarial Vulnerabilities

Ironically, the transparency provided by XAI can be weaponized by attackers. Attackers could use explanation outputs to perform model inversion (inferring sensitive training data), membership inference attacks or to craft more effective adversarial examples that evade detection and generate benign-looking explanations.

6.4.1. Exploiting Explanations

If an attacker recognizes why the IDS flagged their traffic (e.g., through an XAI dashboard), they can craft "antagonistic perturbations" by making minor changes to their attack traffic that bypass the specific features the DNN is looking for.

6.4.2. Actionability of Explanations:

The explanation must lead to a concrete defensive action. The modal focused on packet size and timing is less actionable than "This device is sending encrypted bursts of data to a known C&C server IP every 5 minutes, matching Botnet XYZ pattern.

6.4.3. Attacking the XAI Layer:

The XAI tool itself can be targeted. If an attacker can manipulate the explanation, they could trick a human analyst into dismissing a real alert as a "false positive".

6.5. Latency:

Delayed response to fast acting malware and the root cause is iterative nature of SHAP and LIME.

6.6. Storage:

There is limited storage at edge computing devices and not enough for large XAI libraries and its background datasets.

6.7. Robustness:

Explanations can be changed by small noise as DNNs are very sensitive to even small input changes.

6.8. Human Factor:

Intellectual overload for security analysts. Excessive technical explanation data to read, understand and make decisions.

7. Future Trend's and Research Scope Discussions

The integration of Explainable AI (XAI) with Deep Neural Network (DNN) based intrusion detection system (DNN-IDS) for IoT represents a critical frontier in cybersecurity. The future lies not just in making "black boxes" transparent, but in creating adaptive, trustworthy and collaborative security system that are intrinsic to the IoT ecosystem. There is a shifting from merely "explaining" a threat to autonomously defending against it while maintaining human oversight. The focus is moving from toward Edge-native XAI, where resource-heavy explanations are optimized for tiny IoT devices. Below are emerging trends:

7.1. Agentic and Autonomous Remediation:

The future IDSs won't just alert a human; the will use XAI to justify immediate, autonomous actions as well, such as, rerouting traffic or quarantining a compromised smart sensor. Providing a reasoning logfor the security team to review later.

7.2. Edge-XAI (TinyML Integration):

There is trend towards developing light weight XAI models (like gateways) that can run directly on IIoT edge gateways. This reduces latency and ensures that security decisions are made locally without sending sensitive data to the cloud.

7.3. Encrypted Traffic Index:

As more and more IIoT sensor and devices use end to end encryption, DNNs are being trained to detect patterns in metadata (packet length, timing) rather than payloads. Future XAI will focus on explaining these "behavioral fingerprints" rather than raw data.

7.4. Blockchain-Anchored Explanations:

To prevent attackers from tampering with security logs or XAI outputs, there is an opportunity to integrate blockchain to create immutable, transparent audit trails of why an IDS flagged a specific event.

Below are the future research opportunities :

- Developing a lightweight explainer.
- Edge optimized architectures (neuron + explanations).
- Robust XAI models that can provide enough information for a human to trust them, but not enough for an attacker to exploit the underlying logic.
- Mapping raw network features to neural language explanations.
- Explainable Federated Learning based IDS model, without sharing raw data and the explanation is also generated in a decentralized privacy-preserving way.
- Integrating data from multiple sources (network packets, devices power consumption and system logs) into a single DNN and using XAI to show how these different “modes” correlates to indicae a Zero Day attack.

8. Conclusion

In this technical paper we explored the idea to integrate Explainable Artificial Intelligence (XAI) with Deep Neural Network (DNN) for Intrusion Detection System of IIoT environment. We thoroughly investigated other research papers on same topic and came up with many loop holes and research opportunities to get 100% accuracy rate.

The model presented is showing how intrusions can be detected at unmanned and remote site of Oil and Gas industries. Edge computing with higher capacity of storage and power availability is required for 24 hours each day to avoid any kind of alteration and cyber-attacks. The remote sites are very vulnerable and require to put more focus on data security. XAI integration is introduced with DNN for human interface on decision makings and avoiding fully automated DNN based system. As many times wrong flags happen and hamper the ongoing data evaluation process.

DNN is required to test and train the system to learn the pattern of attacks. By utilization of Deep Learning methods, we can learn beyond the existing traditional attacks for which we are setting rules on firewalls.

In the new era of mass IIoT connectivity in industries, lots of care is required to receive the data and sending it to corporate networks, where each and every device is connected with network.

Reference

[1] Saad Hammood Mohammed A, Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia (UKM) Bangi, Mandeep S. Jit Singh, Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia (UKM) Bangi , Abdulmajeed Al-Jumaily, Selangor, Malaysia b Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Legane´s, Spain , Mohammad Tariqul Islam, Computer and Information Sciences Research Center (CISRC), Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia , Md. Shabiul Islam, Faculty of Engineering (FOE), Multimedia University (MMU) Cyberjaya, Selangor, Malaysia, Abdulmajeed M. Alenezi, Department of Electrical Engineering, Faculty of Engineering, Islamic University of Madinah, Madinah, Saudi Arabia, Mohamad A. Alawad, Department of Electrical Engineering, College of Engineering, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia, Muaadh A. Alsoufi, Faculty of Computing, Universiti Teknologi Malaysia (UTM), Johor Bahru, 81310, Malaysia, IG-“APSO-DNN: Deep learning intrusion detection model to detect false data injection attacks in smart grids” E-mail addresses: P117492@siswa.ukm.edu.my (S.H. Mohammed), mandeep@ukm.edu.my (M.S. Jit Singh), tariqul@ukm.edu.my (M. Tariqul Islam), shabiul. islam@mmu.edu.my (Md.S. Islam). Contents lists available at ScienceDirect Ad Hoc Networks journal homepage: www.elsevier.com/locate/adhoc https://doi.org/10.1016/j.adhoc.2025.104053 Received 22 July 2025; Received in revised form 7 October 2025; Accepted 12 October 2025 Ad Hoc Networks 180 (2026) 104053 Available online 13 October 2025 1570-8705/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

[2] Babatunde Olanrewaju-George, tjoj0107@gmail.com (B. Olanrewaju-George), Department of Engineering and Mathematics, Sheffield Hallam University, UK, Bernardi Pranggono, bernardi.pranggono@aru.ac.uk (B. Pranggono) , School of Computing and Information Science, Anglia Ruskin University, UK, "Federated learning-based intrusion detection system for the internet of things using unsupervised and supervised deep learning models", https://doi.org/10.1016/j.csa.2024.100068 , 2772-9184/© 2024 The Authors. Publishing Services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

[3] Firas Saidi, f.saidi@utb.edu.bh, College of Computer Science (CCS) AI & Metaverse Center (AIMC), Univeristy of Technology Bahrain, Building 829, road, 1213, Block: 712, Salmabad, Kingdom of Bahrain, "IDS-GPT: A Novel Deep Learning-Powered Framework for Network Traffic Intrusion Detection", 1877-0509 © 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0) Peer-review under responsibility of the scientific committee of the KES International. 10.1016/j.procs.2025.09.281.

[4] Devi Priya V.S., vsdevipriya@gmail.com, Department of CSE(Cybersecurity), School of Engineering, Dayananda Sagar University, Harohalli, Bengaluru, Karnataka, 562112, India, Sibi Chakkaravarthy Sethuraman, chakkaravarthy.sibi@vitap.ac.in, Centre of Excellence, Artificial Intelligence and Robotics (AIR), VIT-AP University, India, Centre of Excellence, Cyber Security, VIT-AP University, India d School of Computer Science and Engineering, VIT-AP University, Andhra Pradesh, 522237, , India, Muhammad Khurram Khan, mkhurram@ksu.edu.sa, Center of Excellence in Information Assurance, King Saud University, Riyadh, Saudi Arabia, "Blockchain-based Deep Learning Models for Intrusion Detection in Industrial Control Systems: Frameworks and Open Issues" , Corresponding author at: Center of Excellence in Information Assurance, King Saud University, Riyadh, Saudi Arabia. E-mail addresses: vsdevipriya@gmail.com (Devi Priya V.S.), chakkaravarthy.sibi@vitap.ac.in (S.C. Sethuraman), mkhurram@ksu.edu.sa (M.K. Khan). https://doi.org/10.1016/j.jnca.2025.104286 , Received 9 May 2024; Received in revised form 1 July 2025; Accepted 8 August 2025 Journal of Network and Computer Applications 243 (2025) 104286 Available online 1 September 2025 1084-8045/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

[5] Xue Li, 20030527@nwupl.edu.cn , School of Foreign Languages, Northwest University of Political Science and Law, Xi'an, Shaanxi, 710122, China, Yugui Zhang b a b Institute of Semiconductors, Chinese Academy of Sciences, Beijing, 100083, China, "Security application of intrusion detection model based on deep learning in english online education", https://doi.org/10.1016/j.aej.2025.03.051 , Received 6 January 2025; Received in revised form 21 February 2025; Accepted 13 March 2025 Alexandria Engineering Journal 124 (2025) 582–590 Available online 11 April 2025 1110-0168/© 2025 The Authors. Published by Elsevier B.V. on behalf of Faculty of Engineering, Alexandria University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

[6] Vandana Shayka, Jaytrilok Choudhry, Dhirendra Partap Singh, Department of Computer Science, MANIT, Bhopal, Madhya Pradesh, India, "Deep Learning based Intrusion Detection System for WSN" , 877-0509 © 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license, https://creativecommons.org/licenses/by-nc-nd/4.0 Peer-review under responsibility of the scientific committee of the International Conference on Machine Learning and Data Engineering 10.1016/j.procs.2025.04.460.

[7] Fatimah Alhayan, Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia, Nuha Alruwais, Department of Computer Science and Engineering, College of Applied Studies and Community Services, King Saud University, Saudi Arabia, P.O.Box 22459, Riyadh 11495, Saudi Arabia, Mohammad Alamgeer, Department of Information Systems, Applied College at Mahayil, King Khalid University, Saudi Arabia, Abdullah M. Alashjaee, abdullah.alashjaee@nbu.edu.sa . Department of Computer Science, College of Science, Northern Border University, Arar, Saudi Arabia, Monir Abdullah, Department of Computer Science and Artificial Intelligence, College of Computing and Information Technology, University of Bisha, Bisha 67714, Saudi Arabia, Alaa O. Khadidosf , Fouad Shoie Alallah f , Abdulrhman Alshareef, Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia, "Design of advanced intrusion detection in cybersecurity using ensemble of deep learning models with an improved beluga whale optimization algorithm", Available online 26 February 2025 1110-0168/© 2025 The Authors. Published by Elsevier B.V. on behalf of Faculty of Engineering, Alexandria University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/),

https://doi.org/10.1016/j.aej.2025.02.069 . Received in revised form 31 January 2025; Accepted 19 February 2025.

[8] Muhammad Ammar, ComSens Lab, International Graduate School of Artificial Intelligence, National Yunlin University of Science and Technology, Douliu, Yunlin 64002, Taiwan, Nadeem Javaid, Information Systems Department, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia, Abdul Khader Jilani, School of Computing and Information Science, Anglia Ruskin University, Cambridge CB11PT, UK Saudagar, Information Systems Department, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia b,∗, Imran Ahmed, School of Computing and Information Science, Anglia Ruskin University, Cambridge CB11PT, UK,"An optimized Deep and Active Learning oriented framework for intrusion detection in Internet of Sensor Things" E-mail addresses: javaidn@yuntech.edu.tw (N. Javaid), aksaudagar@imamu.edu.sa (A.K.J. Saudagar). URLs: https://www.njavaid.com, https://orcid.org/0000-0003-3777-8249 (N. Javaid). https://doi.org/10.1016/j.asej.2025.103607 Received 2 May 2025; Received in revised form 3 June 2025; Accepted 28 June 2025, Available online 15 July 2025 2090-4479/© 2025 The Author(s). Published by Elsevier B.V. on behalf of Faculty of Engineering, Ain Shams University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

[9] Mahmoud Ragab, Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, 21589, Saudi Arabia, Rayed Alakhtar, Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, 21589, Saudi Arabia, Huda Alsobhi, Information Technology Department, Faculty of Computers and Information Technology, Taif University, Taif, Saudi Arabia, Rasha Atwah, Department of Computer and Network Engineering, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia, Diaa Hamed, d Department of Software Engineering, College of Engineering, University of Business and Technology, Jeddah, Saudi Arabia e , Louai A. Maghrabi, d Department of Software Engineering, College of Engineering, University of Business and Technology, Jeddah, Saudi Arabia, Khalid Allehaibi, Computer Science Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, 21589, Saudi Arabia, "Leveraging artificial intelligence-driven deep structured learning based threat detection system for large scale communication environment", Corresponding author. E-mail addresses: mragab@kau.edu.sa (M. Ragab), ralakhtar@kau.edu.sa (R. Alakhtar), hasobhi@tu.edu.sa (H. Alsobhi), rjatwah@uj.edu.sa (R. Atwah), dzeinalabedein@kau.edu.sa (D. Hamed), l.maghrabi@ubt.edu.sa (L.A. Maghrabi), kallehaibi@kau.edu.s (K. Allehaibi). Contents lists available at ScienceDirect Journal of Radiation Research and Applied Sciences journal homepage: www.journals.elsevier.com/journal-of-radiation-research-and-applied-sciences https://doi.org/10.1016/j.jrras.2025.101828 Received 13 November 2024; Received in revised form 21 July 2025; Accepted 23 July 2025 Journal of Radiation Research and Applied Sciences 18 (2025) 101828 Available online 28 July 2025 1687-8507/© 2025 The Authors. Published by Elsevier B.V. on behalf of The Egyptian Society of Radiation Sciences and Applications. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

[10] Md. Alamgir Hossain, "Deep learning-based intrusion detection for IoT networks: a scalable and efficient approach"¸ Hossain EURASIP Journal on Information Security (2025) 2025:28 https://doi.org/10.1186/s13635-025-00202-w .

[11] Muhammad Wasim Nawaz, Department of Computer Engineering, The University of Lahore, Pakistan, Rashid Munawar† , Ahsan Mehmood, Electrical engineering department, Information Technology University, Lahore, Pakistan, Muhammad Mahboob Ur Rahman, Electrical engineering department, Information Technology University, Lahore, Pakistan, Qammer H. Abbasi, Department of Electronics and Nano Engineering, University of Glasgow, Glasgow, G12 8QQ, UK, † ∗Electrical engineering department, Information Technology University, Lahore, Pakistan, Muhammad.wasim@dce.edu.pk , mahboob.rahman@itu.edu.pk qammer.Abbasi@glasgow.ac.uk , "Multi-class Network Intrusion Detection with Class Imbalance via LSTM & SMOTE" arXiv:2310.01850v1 [cs.CR] 3 Oct 2023.

[12] Harish M.S Department of Electronics and Communication Engineering, College of Engineering Guindy, Anna University, Chennai, Tamil Nadu 600025, Lokesh S Ramanujan Computing Centre, College of Engineering Guindy, Anna University, Chennai, Tamil Nadu 600025, India , Sakthivel P, Department of Electronics and Communication Engineering, College of Engineering Guindy, Anna University, Chennai, Tamil Nadu 600025, India , Akshaya B, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai 600119, India "Hybrid deep learning model for network intrusion detection using optimal feature fusion". https://doi.org/10.1016/j.asej.2025.103904 , Received 11 August 2025; Received in revised form 6 November 2025; Accepted 18 November 2025 Ain Shams Engineering Journal 17 (2026) 103904 Available online 12

December 2025 2090-4479/© 2025 The Author(s). Published by Elsevier B.V. on behalf of Faculty of Engineering, Ain Shams University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

[13] Omar Achbarou, Cadi Ayyad University, National School of Applied Sciences, LMSC Laboratory, Marrakech, Morocco, Toufik Datsi, Ibn Zohr University, Computer Systems and Vision Laboratory, Agadir, Morocco , Outmane Bourkoukou, Cadi Ayyad University, Laboratory of computer Science and System engineering (L2IS), Marrakech, Morocco , Ahmed My El kiram, Cadi Ayyad University, Information Systems Engineering Laboratory, Marrakech, Morocco, "Enhanced intrusion detection system using feature selection and hybrid learning models for high performance and efficiency in an IOT environment", https://doi.org/10.1016/j.jer.2025.10.016 Received 13 January 2025; Received in revised form 19 October 2025; Accepted 23 October 2025 Journal of Engineering Research xxx (xxxx) xxx Available online 26 October 2025 2307-1877/© 2025 The Author(s). Published by Elsevier B.V. on behalf of Kuwait University. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

[14] Arun Silivery, Kovvuru Ram Mohan Rao, Ramana Solleti, "An Advanced Intrusion Detection Algorithm for Network Traffic Using Convolution Neural Network", February 2023, DOI:10.1109/ICECCT56650.2023.10179767,Conference: 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT), https://www.researchgate.net/publication/372563336_An_Advanced_Intrusion_Detection_Algorithm_for_Network_Traffic_Using_Convolution_Neural_Network .

[15] K. Vinotha , Dr. P. Eswaran , "Quantum Inspired Hyperparameter Optimization for Enhanced Deep Learning Based Intrusion Detection in Wireless Sensor Networks, Soft Computing" Letters (2025), doi: https://doi.org/10.1016/j.sasc.2025.200431. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

[16] Mourad Benmaleka , Abdessamed Seddikib aComputer Engineering Department, College of Engineering, Al Yamamah University, Riyadh, 11512, Saudi Arabia bEcole nationale Supérieure d'Informatique, BP 68M, Oued-Smar, Algiers, 16309, Algeria, published in Data Science and Management, https://doi.org/10.1016/j.dsm.2025.02.005 "Particle swarm optimization-enhanced machine learning and deep learning techniques for Internet of Things intrusion detection" https://doi.org/10.1016/j.dsm.2025.02.005 .

[17] Heba Dhirar , Ali Hamad Information and Communication Engineering, AL-Khwarizmi College of Engineering, University of Baghdad, Baghdad, Iraq, "Comparative evaluation of a novel IDS dataset for SDN-IoT using deep learning models against InSDN, BoT-IoT, and ToN-IoT" Corresponding author at: University of Baghdad, Baghdad, Baghdad IRAQ. E-mail address: heba.d@kecbu.uobaghdad.equ.iq (H. Dhirar). Contents lists available at ScienceDirect Measurement: Digitalization journal homepage: www.elsevier.com/locate/meadig https://doi.org/10.1016/j.meadig.2025.100015 Received 8 March 2025; Received in revised form 4 September 2025; Accepted 8 September 2025.

[18] Mimouna Abdullah Alkhonaini a , Manal Abdullah Alohali b , Mohammed Aljebreen c , Majdy M. Eltahir d , Meshari H. Alanazi e,* , Ayman Yafoz f , Raed Alsini f , Alaa O. Khadidosf "Sandpiper optimization with hybrid deep learning model for blockchain-assisted intrusion detection in iot environment", Corresponding author. E-mail address: Meshari.alanazi@nbu.edu.sa (M.H. Alanazi). a Department of Computer Science, College of Computer and Information Sciences, Prince Sultan University, Saudi Arabia b Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia c Department of Computer Science, Community College, King Saud University, P.O. Box 28095, Riyadh 11437, Saudi Arabia d Department of Information Systems, Applied College at Mahayil, King Khalid University, Saudi Arabia e Department of Computer Science, College of Sciences, Northern Border University, Arar, Saudi Arabia f Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. Contents lists available at ScienceDirect Alexandria Engineering Journal journal homepage: www.elsevier.com/locate/aej , https://doi.org/10.1016/j.aej.2024.10.032 , Received 19 August 2024; Received in revised form 19 September 2024; Accepted 7 October 2024 Alexandria Engineering Journal 112 (2025) 49–62 Available online 1 November 2024 1110-0168/© 2024 The Authors. Published by Elsevier B.V. on behalf of Faculty of Engineering, Alexandria University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

[19] Deepesh M. Dhanvijaya , Mrinai M. Dhanvijay b,∗ , Vaishali H. Kamblec, "Cyber intrusion detection using ensemble of deep learning with prediction scoring based optimized feature sets for IOT networks" Department of Electronics and Communication Engineering, National Institute of Technology, Trichy, Tiruchirappalli, India b

Department of Electronics and Telecommunication Engineering, M.E.S. College of Engineering, Pune, India c DES Pune University, Pune, India.. https://doi.org/10.1016/j.csa.2025.100088 , Received 12 July 2024; Received in revised form 14 November 2024; Accepted 20 February 2025 Available online 21 February 2025 2772-9184/© 2025 The Authors. Publishing Services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/). Corresponding author. E-mail address: mrinai.dhanvijay@mescoepune.org (M.M. Dhanvijay).

[20] Kumar, D., Mehta, M.A. (2023). An Overview of Explainable AI Methods, Forms and Frameworks. In: Mehta, M., Palade , V., Chatterjee, I. (eds) Explainable AI: Foundations, Methodologies and Applications. Intelligent Systems Reference Library, vol 232. Springer, Cham. https://doi.org/10.1007/978-3-031-12807-3_3 , Published20 October 2022 ,Publisher Name Springer, Cham, Print ISBN978-3-031-12806-6 , Online ISBN978-3-031-12807-3, eBook Packages Intelligent Technologies and Robotics Intelligent Technologies and Robotics (R0).

[21] Ben Ncir CE, Ben HajKacem MA, Alattas M. 2024. "Enhancing intrusion detection performance using explainable ensemble deep learning [PeerJ]" . PeerJ Computer Science 10:e2289 https://doi.org/10.7717/peerj-cs.2289 . https://peerj.com/computer-science/ .

[22] Usman Ahmed, School of Software, Northwestern Polytechnical University, Xian 710072, China , Zheng Jiangbin1 , Ahmad Almogren, 2 Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11633, Saudi Arabia , Sheharyar Khan, School of Software, Northwestern Polytechnical University, Xian 710072, China , Muhammad Tariq Sadiq, 3 School of Computer Science and Electronic Engineering, University of Essex, Colchester Campus, Colchester, UK, Applied Science Research Center, Applied Science Private University, Amman, Jordan, Ayman Altameem, 5 Department of Natural and Engineering Sciences, College of Applied Studies and Community Services, King Saud University, Riyadh 11543, Saudi Arabia and Ateeq Ur Rehman, School of Computing, Gachon University, Seongnam-si 13120, Republic of Korea, Correspondence: Ateeq Ur Rehman 202411144@gachon.ac.kr . "Explainable AI-based innovative hybrid ensemble model for intrusion detection". Ahmed et al. Journal of Cloud Computing (2024) 13:150 https://doi.org/10.1186/s13677-024-00712-x , Journal of Cloud Computing: Advances, Systems and Applications . Springer Open.

[23] Shakil Ibne Ahsan, Birkbeck, University of London, Malet Street, Bloomsbury, London, WC1E 7HX, UK, ∗ , Phil Legga, University of the West of England, Coldharbour Lane, Bristol, BS16 1QY, UK , S.M. Iftekharul Alam, Intel Labs, Intel Corporation, CA, USA, a University of the West of England, Coldharbour Lane, Bristol, BS16 1QY, UK , "An explainable ensemble-based intrusion detection system for software-defined vehicle ad-hoc networks" , Contents lists available at ScienceDirect Cyber Security and Applications journal homepage: http://www.keaipublishing.com/en/journals/cyber-security-and-applications , https://doi.org/10.1016/j.csa.2025.100090 Received 10 January 2024; Received in revised form 11 October 2024; Accepted 24 March 2025 Available online 1 April 2025 2772-9184/© 2025 The Authors. Publishing Services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

[24] Mohammad Naif Alatawi, " Enhancing Intrusion Detection Systems With Advanced Machine Learning Techniques: An Ensemble and Explainable Artificial Intelligence (AI) Approach" https://onlinelibrary.wiley.com/authored-by/Naif+Alatawi/Mohammed , First published: 19 January 2025, https://doi.org/10.1002/spy2.496.

[25] Hangsheng Zhang , Dongqi Han , Shangyuan Zhuang , Zhiliang Wang , Member, IEEE, Jiyan Sun, Yinlong Liu , Jiqiang Liu , and Jinsong Dong, "Explainable and Transferable Adversarial Attack for ML-Based Network Intrusion Detectors" IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, VOL. 22, NO. 5, SEPTEMBER/OCTOBER 2025. (Corresponding author: Yinlong Liu.) Hangsheng Zhang, Shangyuan Zhuang, and Yinlong Liu are with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100045, China, and also with the School of Cyber Security, University of Chinese Academy of Sciences, Beijing 101408, China (e-mail: zhanghangsheng@iie.ac.cn ; zhuangshangyuan@iie.ac.cn ; liuyinlong@iie.ac.cn). Dongqi Han is with the Faculty of Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: handongqi@bupt.edu.cn). Zhiliang Wang is with the Department of Network Sciences and Cyberspace, Tsinghua University, Beijing 100190, China (e-mail: wzl@cernet.edu.cn). Jiyan Sun is with the Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100045, China (e-mail: sunjiyan@iie.ac.cn). Jiqiang Liu is with the Faculty of Beijing Jiaotong University, Beijing 100044, China (e-mail: jqliu@bjtu.edu.cn). Jinsong Dong is with the Department of Computing, National University of Singapore (NUS), Singapore 119077 (e-mail: dcsdjs@nus.edu.sg). This article has supplementary downloadable material available at https://doi.org/10.1109/TDSC.2025.3560486 , provided by the authors. Digital Object Identifier 10.1109/TDSC.2025.3560486

[26] Syed Wali, Yasir Ali Farrukh , Irfan Khan Clean and Resilient Energy System Lab (CARES), Department of Electrical & Computer Engineering, Texas A&M University, College Station, TX, USA, "Explainable AI and Random Forest based reliable intrusion detection system" Contents lists available at ScienceDirect Computers & Security journal homepage: www.elsevier.com/locate/cose , https://doi.org/10.1016/j.cose.2025.104542 ..

[27] Said Salloum, Sajedeh Norozpour, Corresponding author email: s.a.s.salloum@edu.salford.ac.uk , "XAI-IDS: A Transparent and Interpretable Framework for Robust Cybersecurity Using Explainable Artificial Intelligence" School of Science, Engineering and Environment, University of Salford, United Kingdom, UK2Istanbul Gelişim University, Istanbul, Türkiye, SHIFRAVol. (2025), 2025, pp. 69–80ISSN: 3078-3186, https://doi.org/10.70470/SHIFRA/2025/004 .

[28] Shaikh Afnan Birahim, School of Computer Science and Engineering, University of Glasgow, G12 8QQ Glasgow, U.K. , Avijit Paul, Department of Electronics and Telecommunication Engineering, Rajshahi University of Engineering and Technology, Rajshahi 6204, Bangladesh 2 , Fahmida Rahman, Department of Computer Science and Engineering, International Islamic University Chittagong, Chittagong 4318, Bangladesh , Yamina Islam, 3Department of Computer Science and Engineering, International Islamic University Chittagong, Chittagong 4318, Bangladesh , Tonmoy Roy, 4Department of Data Analytics and Information Systems, Utah State University, Logan, UT 84322, USA 5Department of Electrical Engineering, Qatar University, Doha, Qatar 4 , Mohammad Asif Hasan, Department of Electronics and Telecommunication Engineering, Rajshahi University of Engineering and Technology, Rajshahi 6204, Bangladesh Fariha Haque, Department of Electronics and Telecommunication Engineering, Rajshahi University of Engineering and Technology, Rajshahi 6204, Bangladesh , and Muhammad E. H. Chowdhury 5,. Corresponding author: Muhammad E. H. Chowdhury (mchowdhury@qu.edu.qa), "Intrusion Detection for Wireless Sensor Network Using Particle Swarm Optimization Based Explainable Ensemble Machine Learning Approach", Received 13 December 2024, accepted 31 December 2024, date of publication 10 January 2025, date of current version 23 January 2025. Digital Object Identifier 10.1109/ACCESS.2025.3528341. IEEE Access.

[29] Harikha Manthena 1,∗, Shaghayegh Shajarian 1,∗, Jeffrey C. Kimmell 2 , Mahmoud Abdelsalam 1 , Sajad Khorsandroo 1 , and Maanak Gupta 2 , (senior member, ieee), "Explainable Artificial Intelligence (XAI) for Malware Analysis: A Survey of Techniques, Applications, and Open Challenges" published on 28 March 2025, date of current version 14 April 2025. Digital Object Identifier 10.1109/ACCESS.2025.3555926, 2025 https://creativecommons.org/licenses/by/4.0/ , IEEE Access, VOLUME 13, 2025.

[30] Sileshi Nibret Zeleke, Orchid Id: 0009 −0006 −8172 −9646, Department of Computer Science, University of Bari, Bari, Italy Amsalu Fentie Jember1, Orchid Id: 0009 −0004 −7356 −682 X , and Mario Bochicchio, Digital Health National Lab, CINI - Consorzio Interuniversitario Nazionale per, Department of Computer Science, University of Bari, Bari, Italy, Orchid Id: 0000 −0002 −9122 −6317 1 2 {sileshi.zeleke,amsalu.jember,mario.bochicchio}@uniba.it, "Integrating Explainable AI for Effective Malware Detection in Encrypted Network Traffic", arXiv:2501.05387v1 [cs.CR] 9 Jan 2025.

[31] Chathuranga Sampath Kalutharage, Edinburgh Napier University, Scotland, UK ,∗ , Xiaodong Liu, Edinburgh Napier University, Scotland, UK, Christos Chrysoulas, Heriot-Watt University, Scotland, UK , "Neurosymbolic learning and domain knowledge-driven explainable AI for enhanced IoT network attack detection and response"

[32] Osvaldo Arreche, Electrical and Computer Engineering Department, Purdue University in Indianapolis, 420 University Blvd, Indianapolis 46202, IN, USA, and Mustafa Abdallah, Computer and Information Technology Department, Purdue University in Indianapolis, 420 University Blvd, Indianapolis 46202, IN, USA, Correspondence: Mustafa Abdallah abdalla0@purdue.edu "A comparative analysis of DNN-based white-box explainable AI methods in network security", https://doi.org/10.1016/j.cose.2025.104318 Arreche and Abdallah EURASIP Journal on Information Security (2025) 2025:16 https://doi.org/10.1186/s13635-025-00201-x , EURASIP Journal on Information Security. Contents lists available at ScienceDirect Computers & Security journal homepage: www.elsevier.com/locate/cose .

[33] Nawaf Abdualaziz Almolhis1* 1*Department of Computer Science, College of Engineering and Computer Science, Jazan University Jazan, Saudi Arabia. naalmolhis@jazanu.edu.sa , https://orcid.org/0009-0004-7558-7165 , "Intrusion Detection Using Hybrid Random Forest and Attention Models and Explainable AI Visualization", Journal of Internet Services and Information Security (JISIS), volume: 15, number: 1 (February), pp. 371-384. DOI: 10.58346/JISIS.2025.I1.024 *Corresponding author: Department of Computer Science, College of Engineering and Computer Science, Jazan University Jazan, Saudi Arabia. ISSN: 2182-2069 / E-ISSN: 2182-2077.

[34] Marta Catillo ∗ , Antonio Pecchia , Umberto Villano Università degli Studi del Sannio, Pal. Bosco Lucarelli C.so Garibaldi 107, 82100 Benevento, Italy, "MultiCIDS: Anomaly-based collective intrusion detection by deep learning on IoT/CPS multivariate time series" Available online 25 January 2025 2542-6605/© 2025 The Authors.

Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/). E-mail addresses: marta.catillo@unisannio.it (M. Catillo), antonio.pecchia@unisannio.it (A. Pecchia), villano@unisannio.it (U. Villano). https://doi.org/10.1016/j.iot.2025.101519 , Received 18 September 2024; Received in revised form 13 January 2025; Accepted 20 January 2025.

[35] Waqas Ishtiaq a , Ashrafun Zannat a,b , A.H.M. Shahariar Parvez c , Md. Alamgir Hossain a,d,* , Muntasir Hasan Kanchan a,d , Muhammad Masud Tarek d a Skill Morph Research Lab., Skill Morph, Dhaka, Bangladesh b Department of Computer Science and Engineering, Bangladesh Army University of Science & Technology, Saidpur, Bangladesh c Department of Software Engineering, Daffodil International University, Dhaka, Bangladesh d Department of Computer Science and Engineering, State University of Bangladesh, Dhaka, Bangladesh, "CST-AFNet: A dual attention-based deep learning framework for intrusion detection in IoT networks" Corresponding author. Skill Morph Research Lab, Skill Morph, Dhaka, Bangladesh. E-mail addresses: waqas.ishtiaq@gmail.com (W. Ishtiaq), spzannat@baust.edu.bd (A. Zannat), shahariar.swe@diu.edu.bd (A.H.M. Shahariar Parvez), alamgir. cse14.just@gmail.com (Md. Alamgir Hossain), muntasir@sub.edu.bd (M. Hasan Kanchan), tarek@sub.edu.bd (M. Masud Tarek). Contents lists available at ScienceDirect Array journal homepage: www.sciencedirect.com/journal/array https://doi.org/10.1016/j.array.2025.100501 , Received 6 May 2025; Received in revised form 28 July 2025; Accepted 27 August 2025 Array 27 (2025) 100501 Available online 28 August 2025. 2590-0056/© 2025 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

[36] Amjad Alsirhani, Noshina Tariq, Mamoona Humayun, Ghadah Naif Alwakid, Hassan Sanaullah, "Intrusion detection in smart grids using artificial intelligence-based ensemble modelling" Published online: 25 February 2025, Cluster Computing (2025) 28:238 https://doi.org/10.1007/s10586-024-04964-9(0123456789().,-volV)(0123 .

[37] Dong, H., Kotenko, I. "Cybersecurity in the AI era: analyzing the impact of machine learning on intrusion detection." Knowl Inf Syst 67, 3915–3966 (2025). https://doi.org/10.1007/s10115-025-02366-w, published 19 February 2025.

[38] Saeid Jamshidi , , Kawser Wazed Nafi , SWAT, Polytechnique, Montréal, H3T 1J4, Quebec, Canada, Amin Nikanjam, , Huawei Distributed Scheduling and Data Engine Lab, Montréal, Quebec, Canada , , SWAT, Polytechnique, Montréal, H3T 1J4, Quebec, Canada, "Evaluating machine learning-driven intrusion detection system in IoT: Performance and energy consumption" Received 29 September 2024, Revised 25 March 2025, Accepted 31 March 2025, Available online 15 April 2025, Version of Record 17 April 2025. https://doi.org/10.1016/j.cie.2025.111103 Computers & Industrial Engineering, Volume 204, June 2025, 111103.

[39] AHWAR KHAN 1 , MD ASDAQUE HUSSAIN 2 , AND FAISAL ANWER1 1Department of Computer Science, Aligarh Muslim University, Aligarh, Uttar Pradesh 202002, India 2Faculty of Computer Studies, Arab Open University, Manama 26196, Bahrain Corresponding author: Ahwar Khan (khanahwar4@gmail.com) "Hybrid Lightweight Deep Learning-Based Intrusion Detection Approach in IoT Utilizing Feature Selection & Explainable Artificial Intelligence", Received 3 October 2025, accepted 3 November 2025, date of publication 10 November 2025, date of current version 14 November 2025. Digital Object Identifier 10.1109/ACCESS.2025.363075. 2025 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/ . IEEE Access Multidisciplinary: Rapid: Open Access.

[40] Yunfan Huang, Faculty of Engineering, Universiti Malaya, Kuala Lumpur, 50603, Wilayah Persekutuan, Malaysia, Maode Ma, College of Engineering, Qatar University, Doha, Qatar, Wong Jee Keen Raymond, Faculty of Engineering, Universiti Malaya, Kuala Lumpur, 50603, Wilayah Persekutuan, Malaysia , Chee-Onn Chow, Faculty of Engineering, Universiti Malaya, Kuala Lumpur, 50603, Wilayah Persekutuan, Malaysia, "An explainable and adaptive Internet of Things intrusion detection system supported by Large Language Model" https://doi.org/10.1016/j.engappai.2025.112911. Engineering Applications of Artificial Intelligence, Volume 163, Part 2, 1 January 2026, 112911. Elsevier.

[41] Bhawana Sharma, Manipal University Jaipur, Jaipur, Rajasthan, India, Lokesh Sharma, Manipal University Jaipur, Jaipur, Rajasthan, India, Chhagan Lal, , Department of Intelligent Systems, Cybersecurity Group, TU Delft, Netherlands , Satyabrata Roy, Manipal University Jaipur, Jaipur, Rajasthan, India, "Explainable artificial intelligence for intrusion detection in IoT networks: A deep learning based approach". Published in Expert Systems with Applications, Volume 238, Part A, 15 March 2024, 121751. https://doi.org/10.1016/j.eswa.2023.121751 , ELSEVIER.

[42] Ripal Ranpara, Faculty of Computer Applications, Marwadi University, Rajkot 360003, India , OsamahAlsalman, Department of Electrical Engineering, College of Engineering, King Saud University, Riyadh 12372, Saudi Arabia. , Om Prakash Kumar, Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India. ∗ & Shobhit K. Patel, Department of Computer Engineering, Marwadi University, Rajkot 360003, India. “A simulation-driven computational framework for adaptive energyefficient optimization in machine learning-based intrusion detection systems” email: ripal.ranpara@marwadieducation.edu.in; oalsalman@ksu.edu.sa; omprakash.kumar@manipal.edu OPEN Scientific Reports | (2025) 15:13376 | https://doi.org/10.1038/s41598-025-93254-4

[43] Vincent Zibi Mohale and Ibidun Christiana Obagbuwa* Department of Computer Science and Information Technology, Faculty of Natural and Applied Sciences, Sol Plaatje University, Kimberley, South Africa Flaminia Luccio, Ca’ Foscari University of Venice, Italy REVIEWED BY Stefano Cirillo, University of Salerno, Italy Attaullah Buriro, Ca’ Foscari University of Venice, Italy *CORRESPONDENCE Ibidun Christiana Obagbuwa ibidun.obagbuwa@spu.ac.za RECEIVED 31 October 2024 ACCEPTED 02 May 2025 PUBLISHED 22 May 2025 CITATION Mohale VZ and Obagbuwa IC (2025) Evaluating machine learning-based intrusion detection systems with explainable AI: enhancing transparency and interpretability. Front. Comput. Sci. 7:1520741. doi: 10.3389/fcomp.2025.1520741 “Evaluating machine learning-based intrusion detection systems with explainable AI: enhancing transparency and interpretability”. TYPE Original Research PUBLISHED 22 May 2025 DOI 10.3389/fcomp.2025.1520741. https://www.frontiersin.org/journals/computer-science . Hossain EURASIP Journal on Information Security (2025) 2025:28 https://doi.org/10.1186/s13635-025-00202-w.

[44] Md. Alamgir Hossain, Hossain EURASIP Journal on Information Security (2025) 2025:28 https://doi.org/10.1186/s13635-025-00202-w , “Deep learning-based intrusion detection for IoT networks: a scalable and efficient approach” EURASIP Journal on Information Security, Md. Alamgir Hossain alamgir.cse14.just@gmail.com 1 Department of Computer Science and Engineering, State University of Bangladesh, South Purbachal, Kanchan Dhaka-1461, Bangladesh. Springer Open.

[45] UsamaAhmed, Department of Artificial Intelligence, School of Systems and Technology, University of Management and Technology, Lahore 54700, Pakistan , Mohammad Nazir, 2Department of Computer Science and Information Technology, The Islamia University of Bahawalpur, Bahawalpur, Pakistan. Amna Sarwar, 3Department of Computer Science, University of Wah, Wah Cantt, Pakistan, TariqAli, Artificial Intelligence and Sensing Technologies (AIST) Research Center, University of Tabuk, Tabuk 71491, Saudi Arabia, El-Hadi M.Aggoune, Artificial Intelligence and Sensing Technologies (AIST) Research Center, University of Tabuk, Tabuk 71491, Saudi Arabia, Tariq Shahzad, Department of Computer Engineering, COMSATS University Islamabad, Sahiwal Campus, Sahiwal 57000, Pakistan & Muhammad Adnan Khan, Department of Software, Faculty of Artificial Intelligence and Software, Gachon University, Seongnam-si 13120, Republic of Korea. “Signature-based intrusion detection using machine learning and deep learning approaches empowered with fuzzy clustering” email: teshaq@ut.edu.sa ; adnan@gachon.ac.kr , Scientific Reports | (2025) 15:1726 https://doi.org/10.1038/s41598-025-85866-7 .

[46] A. Karunamurthy, Department of MCA, Sri Manakula Vinayagar Engineering College, Pondicherry, India, K.Vijayan, Department of Artificial Intelligence and Data Science, Sapthagiri NPS university, Banglore, India, Pravin R. Kshirsagar, 3Professor & Head-ETC, J D College of Engineering and Management, Nagpur, India & Kuan TakTan, Engineering Cluster, Singapore Institute of Technology, Singapore, Singapore. “An optimal federated learning-based intrusion detection for IoT environment” email: karunamurthy26@gmail.com, Scientific Reports | (2025) 15:8696 https://doi.org/10.1038/s41598-025-93501-8 . http://www.nature.com/scientificreports .

[47] C. Christy, PG and Research Department of Computer Science and Artificial Intelligence, St. Joseph’s College of Arts and Science (Autonomous), Cuddalore, Tamil Nadu, India, A. Nirmala2, A. Mary OdilyaTeena, 2Department of Computer Applications, St. Joseph’s College of Arts and Science (Autonomous), Cuddalore, Tamil Nadu, India, A. IsabellaAmali, PG and Research Department of Computer Science and Artificial Intelligence, St. Joseph’s College of Arts and Science (Autonomous), Cuddalore, Tamil Nadu, India. ,Department of Computer Applications, St. Joseph’s College of Arts and Science (Autonomous), Cuddalore, Tamil Nadu, “Machine learning based multi stage intrusion detection system and feature selection ensemble security in cloud assisted vehicular ad hoc networks” email: christypaulraj.2025@gmail.com, Scientific Reports | (2025) 15:27058 | https://doi.org/10.1038/s41598-025-96303-0 . http://www.nature.com/scientificreports .

[48] M. Sami Ataa, Eman E. Sanad & Reda A. El-khoribi, Faculty of Computers and Artificial Intelligence, Cairo University, Giza, Egypt. email: m.ataa@fci-cu.edu.eg , “Intrusion detection in software defined network using deep learning approaches” Scientific Reports | (2024) 14:29159 | https://doi.org/10.1038/s41598-024-79001-1 .

[49] Vanlalruata Hnamte∗ , Jamal Hussain, Department of Mathematics and Computer Science, Mizoram University, Tanhril, Aizawl, 796004, Mizoram, India, "Dependable intrusion detection system using deep convolutional neural network: A Novel framework and performance evaluation approach" Contents lists available at ScienceDirect Telematics and Informatics Reports journal homepage: www.elsevier.com/locate/teler , Corresponding author. E-mail address: vanlalruata.hnamte@gmail.com (V. Hnamte), https://doi.org/10.1016/j.teler.2023.100077 Received 28 April 2023; Received in revised form 5 June 2023; Accepted 4 July 2023 2772-5030/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

[50] Ahmad HIJAZI Univ.Grenoble Alpes, G-SCOP, F-38000 Grenoble, France ahd.hjz@gmail.com, EL Abed EL SAFADI Univ.Grenoble Alpes, G-SCOP, F-38000 Grenoble, France Abed.safadi@grenoble-inp.fr , Jean-Marie FLAUS Univ.Grenoble Alpes, G-SCOP, F-38000 Grenoble, France Jean-marie.Flaus@grenoble-inp.fr , "A Deep Learning Approach for Intrusion Detection System in Industry Network" paper12.pdf . www.CEUR-WS.org/vol-2343/paper12.pdf.

[51] Arun Kumar Silivery Computer Science and Engineering Osmania University Hyderabad, India. arunsilivery@osmania.ac.in , Kovvuru Ram Mohan Rao Department of Information Technology Vasavi College of Engineering , Hyderabad, India. krmrao@staff.vce.ac.in , Ramana Solleti Department of Computer Science Bhavan's Vivekananda College, Hyderabad, India ramana.cs@bhavansvc.ac.in , "An Advanced Intrusion Detection Algorithm for Network Traffic Using Convolution Neural Network", 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT) | 978-1-6654-9360-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICECCT56650.2023.1017976. Authorized licensed use limited to: OSMANIA UNIVERSITY. Downloaded on December 01,2023 at 10:05:07 UTC from IEEE Xplore. Restrictions apply.

[52] Vanlalruata Hnamte∗ , Jamal Hussain Department of Mathematics and Computer Science, Mizoram University, Tanhril, Aizawl, Mizoram 796004, India, "DCNNBiLSTM: An Efficient Hybrid Deep Learning-Based Intrusion Detection System" Contents lists available at ScienceDirect Telematics and Informatics Reports journal homepage: www.elsevier.com/locate/teler , https://doi.org/10.1016/j.teler.2023.100053 Received 17 November 2022; Received in revised form 23 February 2023; Accepted 5 March 2023 2772-5030/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

[53] Explainable AI: A Comprehensive Guide .

Author's short biography

Syed Anwarul Haque: Syed Anwarul Haque is Business System Analyst in Gas Compression Projects Department, Al Khobar Saudi Arabia. He is working in Telecom and Security field for Saudi Aramco for more than 15 years and completed many projects. He is a graduated in Electronics and Communication Engineering from RVS College of Engineering and Technology and MBA from Chandigarh University.	
Syed Azfarul Haque: Syed Azfarul Haque is Physics Professor at Jamshedpur Worker's College, Kolhan University, Chaibasa, India. He is having experience in new telecom trends and technologies and published many research papers in different International Journals. He is Doctorate in Physics from Netaji Subhash University, Jamshedpur, Jharkhand, India. His research areas are material science changes to make them usable for Electronic devices and cables.	
Saeed M Yami: Saeed M Yami is Supervisor Project Engineer with Gas Compression Projects Department. Saudi Aramco, Al Khobar Saudi Arabia. He is having very vast experience in Oil and Gas industries. His research areas are new Telecom trends, Artificial Intelligence, Machine Learning, Neural network Years of experience	`

Panteleimon Korfiatis: Panteleimon Korfiatis is Senior Project Engineer with Gas Compression Projects Department Saudi Aramco,, Al Khobar Saudi Arabia. He is having expertise in Telecommunication and Security systems and successfully completed many projects while working with Saudi Aramco. His research interest areas are Artificial Intelligence, Machine Learning, Access Control systems etc.	
Vipul Thomas: Vipul Thomas is Senior Telecom Technician and workig with Area IT Department, Saudi Aramco, Haradh for 22 years. He is having very vast knowledge of new telecom technologies and of future trend. His research interests are in Telecom Transmission, Machine Learning and Artificial Intelligence.	