International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04746
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Leveraging Spatiotemporal Patterns for Cyber Attack Detection in Distributed System Using Machine Learning

G.MANOJ KUMAR, KANDULA HEMAPRIYA

Assistant Professor, 2MCA Final Semester, Master of Computer Applications, Sanketika Vidya Parishad Engineering College, Vishakhapatnam, Andhra Pradesh, India

Abstract

Cyber-crime is proliferating everywhere exploiting every kind of vulnerability to the computing environment. Ethical Hackers pay more attention towards assessing vulnerabilities and recommending mitigation methodologies. The development of effective techniques has been an urgent demand in the field of the cyber security community. Most techniques used in today's IDS are not able to deal with the dynamic and complex nature of cyber-attacks on computer networks. Machine learning for cyber security has become an issue of great importance recently due to the effectiveness of machine learning in cyber security issues. Machine learning techniques have been applied for major challenges in cyber security issues like intrusion detection, malware classification and detection, spam detection and phishing detection. Although machine learning cannot automate a complete cyber security system, it helps to identify cyber security threats more efficiently than other software-oriented methodologies, and thus reduces the burden on security analysts. Hence, efficient adaptive methods like various techniques of machine learning can result in higher detection rates, lower false alarm rates and reasonable computation and communication costs. Our main goal is that the task of finding attacks is fundamentally different from these other applications, making it significantly harder for the intrusion detection community to employ machine learning effectively.

IndexTerms: Cyber-crime,Machine Learning,Cyber-security,Intrusion Detection System,SpatiotemporalPatterns, Anomaly Detection,Support Vector Machine (SVM),Distributed Systems.

1. INTRODUCTION

Today, political and commercial entities are increasingly engaging in sophisticated cyberwarfare to damage, disrupt, or censor information content in computer networks. [3] In designing network protocols, there is a need to ensure reliability against intrusions of powerful attackers that can even control a fraction of parties in the network. The controlled parties can launch both passive (e.g., eavesdropping, nonparticipation) and active attacks (e.g., jamming, message dropping, corruption, and forging). Intrusion detection is the process of dynamically monitoring events occurring in a computer system or network, analyzing them for signs of possible incidents and often interdicting the unauthorized access. This is typically accomplished by automatically collecting information from a variety of systems and network sources, and then analyzing the information for possible security problems. [7] Traditional intrusion detection and prevention techniques, like firewalls, access control mechanisms, and encryptions, have several limitations in fully protecting networks and systems from increasingly sophisticated attacks like denial of service. Moreover, most systems built based on such techniques suffer from high false positive and false negative detection rates and the lack of continuously adapting to changing malicious behaviors. In the past decade, however, several Machine Learning (ML) techniques have been applied to the problem of intrusion detection with the hope of improving detection rates and adaptability. These techniques are often used to keep the attack knowledge bases up-to-date and comprehensive. In recent days, cyber-security and protection against numerous cyber-attacks are becoming a burning question. The main reason behind that is the tremendous growth of computer networks and the vast number of relevant applications used by individuals or groups for either personal or commercial use, especially after the acceptance of the Internet of Things (IoT). The cyber-attacks cause severe damage and severe financial losses in large-scale networks. The existing solutions like hardware and software firewalls, user's authentication, and data encryption methods are not sufficient to meet the challenge of upcoming demand, and unfortunately, not able to protect the computer network's several cyber-threats. [13]

1.1 Existing System

The field of cyber security is rapidly expanding, yet it remains difficult to fully grasp the extent of its impact. Malicious threats can strike anytime and anywhere, posing severe risks to individuals, organizations, and global networks. The complexity of modern internet usage and corporate systems makes it challenging for cyber security teams to maintain control and protection.[9] With the integration of Machine Learning, the cyber security landscape is evolving, enabling better analysis of massive data streams from mission-critical systems. Traditional Intrusion Detection Systems (IDS) primarily focus on north-south traffic, overlooking east-west threats that spread laterally within networks. Research indicates that only 20% of threats are caught through perimeter monitoring. IDS tools typically report suspicious activities to Security Information and Event Management (SIEM) systems, but their effectiveness depends on timely human response. Delays in identifying threats increase potential damage. Moreover, IDS cannot inspect encrypted packets, allowing attackers to bypass detection. They also frequently generate false positives, which can overwhelm analysts and obscure real threats. While tuning can reduce false alerts, it still demands constant human vigilance. Without the right personnel and policies in place, IDS tools alone cannot guarantee robust protection.[14]

1.1.1 Challenges:

High False Positives & False Negatives:

Traditional Intrusion Detection Systems (IDS) often generate excessive false alarms, making it difficult to differentiate between real threats and benign anomalies.[5]

Encrypted Traffic Detection:
IDS systems struggle to inspect encrypted packets, allowing attackers to bypass security mechanisms without being detected until it's too late.
Lateral Threat Movement:
Most IDS focus on north-south traffic (external to internal), but fail to detect threats moving laterally (east-west) within the network.
Dynamic Nature of Cyber Attacks:
The ever-evolving and sophisticated nature of cyber threats makes it challenging to develop static or rule-based systems for reliable detection.
Data Imbalance in Training Sets:
The dataset often contains a large number of normal records but very few attack instances, which hinders accurate model training and reduces detection performance.[10]
Scalability and Real-Time Processing:
Handling vast amounts of network data in real-time, especially in distributed systems, is computationally intensive and requires efficient algorithms.

1.2 Proposed system:

Machine Learning algorithms can be used to train and detect if there has been a cyber attack. As soon as the attack is detected, an email notification can be sent to the security engineers or users. Any classification algorithm can be used to categorize if it is a DoS/DDoS attack or not.[11] One example of a classification algorithm is Support Vector Machine (SVM) which is a supervised learning method that analyses data and recognizes patterns. Since we cannot control when, where or how an attack may come our way, and absolute prevention against these cannot be guaranteed yet, our best shot for now is early detection which will help mitigate the risk of irreparable damage such incidents can cause. Organizations can use existing solutions or build their own to detect cyber attacks at a very early stage to minimize the impact. Any system that requires minimal human intervention would be ideal.[17]

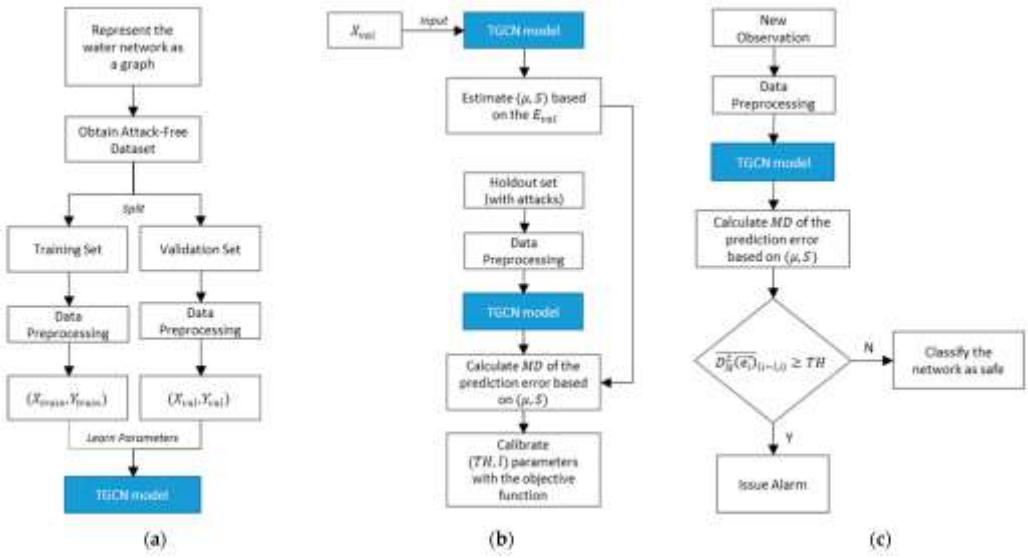

Fig: 1 Proposed Diagram

1.2.1 Advantages:

Early Detection of Attacks:
The system can detect cyber-attacks at an early stage, significantly minimizing the potential damage.
Automation with Minimal Human Intervention:
The system is designed to operate with little human input, making it ideal for real-time cyber defense scenarios.
Use of Machine Learning Algorithms:
By utilizing classification algorithms (e.g., Support Vector Machine - SVM), the system effectively distinguishes between normal and attack traffic (such as DoS/DDoS).
Proactive Alerting via Notifications:
Once an attack is detected, the system can instantly notify security engineers or users through automated email alerts.[8]
Improved Accuracy in Intrusion Detection:
Machine learning models offer improved accuracy and adaptability compared to traditional IDS, which often struggle with evolving threats

2.1 Architecture:

The purpose of the design phase is to arrange an answer of the matter such as by the necessity document. This part is that the opening moves in moving the matter domain to the answer domain. The design phase satisfies the requirements of the system. [15]The design of a system is probably the foremost crucial issue warm heartedness the standard of the software package. It's a serious impact on the later part, notably testing and maintenance. The output of this part is that the style of the document. This document is analogous to a blueprint of answer and is employed later throughout implementation, testing and maintenance. The design activity is commonly divided into 2 separate phases System Design and Detailed Design.System Design conjointly referred to as top-ranking style aims to spot the modules that ought to be within the system, the specifications of those modules, and the way them move with one another to supply the specified results.At the top of the system style all the main knowledge structures, file formats, output formats, and also the major modules within the system and their specifications square measure set. System design is that the method or art of process the design, components, modules, interfaces, and knowledge for a system to satisfy such as needs. Users will read it because the application of systems theory to development.[19]

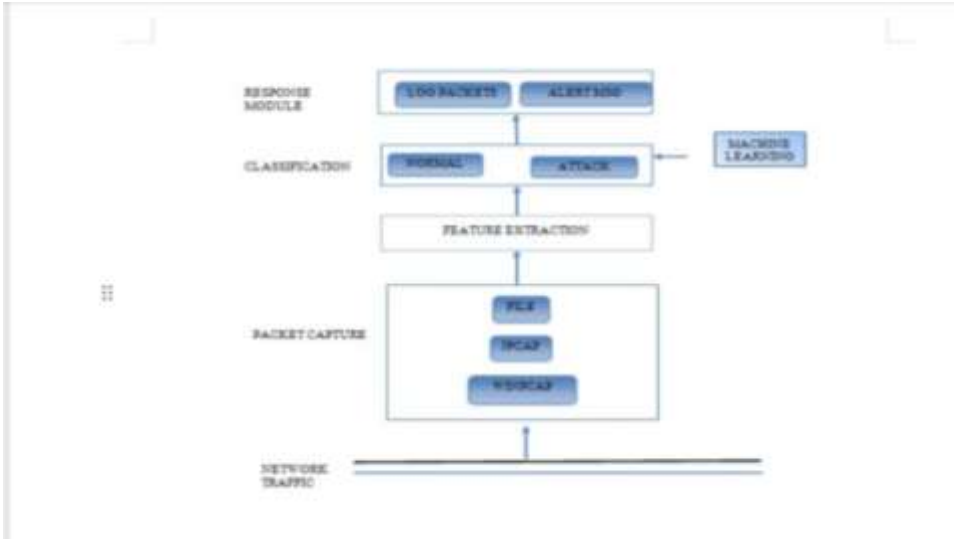

Fig:2 Architecture

UML DIAGRAMS

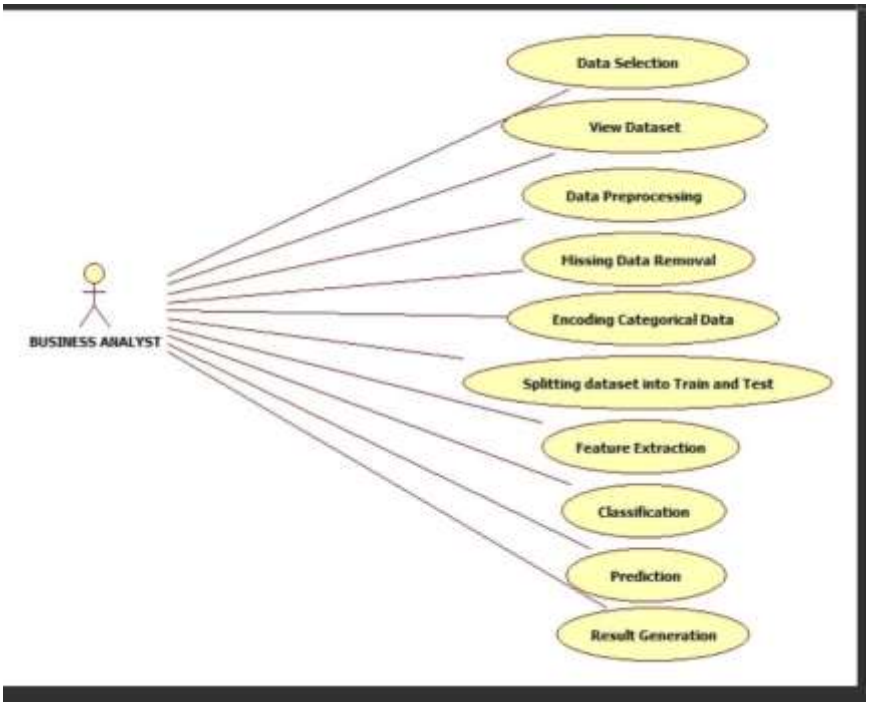

Fig:use case diagram

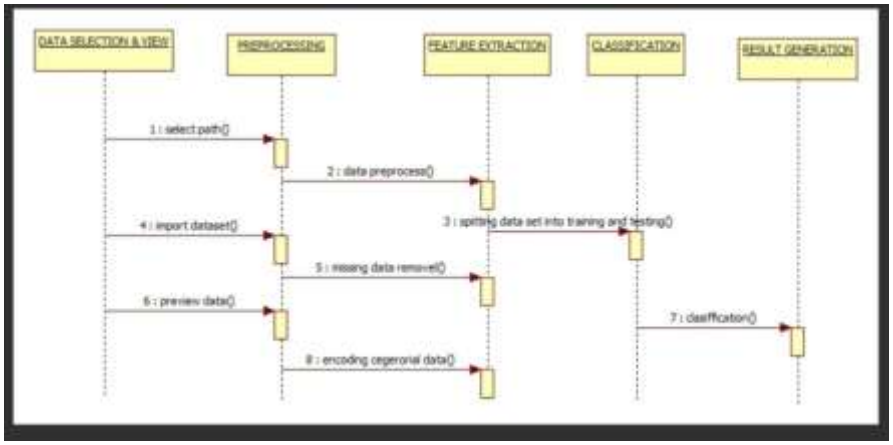

Fig: activity diagram

2.2 Algorithm:

The project "Leveraging Spatiotemporal Patterns for Cyber Attack Detection in Distribution System" utilizes a combination of advanced machine learning and deep learning algorithms to effectively detect and classify cyber-attacks, especially in imbalanced datasets typically found in critical infrastructure systems.[12]The first major algorithm employed is the AutoEncoder, a deep learning model used for feature extraction. It learns a compressed representation of the input data, which is particularly useful for handling high-dimensional datasets and reducing noise. The features extracted from the AutoEncoder are then passed to a Principal Component Analysis (PCA) module. PCA is a dimensionality reduction technique that transforms the extracted features into a lower-dimensional space, helping to retain the most significant components of the data while reducing complexity.Next, a Decision Tree classifier is applied to the reduced features.[4] This supervised learningalgorithm builds a tree-like model of decisions based on the features and classifies whether a given instance represents a normal behavior or a specific type of cyber-attack (e.g., DoS, Command Injection, etc.). The output of the Decision Tree is then further refined using a Deep Neural Network (DNN), which is trained to recognize both known and unknown attack patterns. The DNN enhances the overall detection accuracy by learning complex patterns and improving generalization.This layered approach of AutoEncoder → PCA → Decision Tree → DNN allows the system to effectively handle imbalanced datasets without relying on traditional oversampling or undersampling techniques. The combined use of these algorithms enables high detection accuracy, faster processing, and improved resilience against evolving cyber threats.[1]

2.3 Techniques:

Spatiotemporal Pattern Analysis:
The project leverages spatial and temporal features from IoT-based datasets (like SWAT) to identify how attacks evolve over time and across different devices. This technique helps detect sophisticated multi-stage attacks by analyzing data behavior patterns.

Data Preprocessing and Normalization:
Techniques like missing value imputation (replacing nulls with 0) and Min-Max Scaling (to normalize values between 0 and 1) are used to standardize the dataset and ensure compatibility across models.

Feature Extraction Using AutoEncoder:
An AutoEncoder neural network is used to learn and extract compressed latent features from high-dimensional input data. This technique is key to reducing the complexity of imbalanced datasets without losing significant information.[20]

Dimensionality Reduction Using PCA (Principal Component Analysis):
PCA is applied to the AutoEncoder-extracted features to reduce them further into a manageable number of components while preserving maximum variance. This improves training speed and reduces overfitting.

Classification Using Decision Tree:
A traditional Decision Tree classifier is used to classify the reduced feature set into different attack types. It provides initial prediction results based on feature splits.

Refinement Using Deep Neural Networks (DNN):
A DNN is used for final classification and attack attribution. It is trained on outputs from the Decision Tree, allowing the system to learn complex relationships and improve prediction accuracy.

2.4 Tools:

- **Python 3.7**:

The primary programming language used for coding the entire project. Python is chosen for its simplicity, versatility, and rich support for data science and machine learning libraries.[2]

- **TensorFlow**:

Used for building and training the AutoEncoder and Deep Neural Network (DNN) models. TensorFlow is a deep learning framework developed by Google and is ideal for high-performance computation on neural networks.

- **NumPy**:

A core scientific computing library in Python used for handling multi-dimensional arrays and performing mathematical operations during data preprocessing and transformation.

- **Pandas**:

This library is used for data manipulation and analysis, such as reading datasets, handling missing values, and preparing training and testing dataframes.

- **Scikit-learn**:

Used for implementing traditional machine learning algorithms like Decision Tree, as well as PCA (Principal Component Analysis) for dimensionality reduction and data normalization techniques such as Min-Max Scaling.

2.5 Methods:

The project utilizes a hybrid machine learning-based methodology to detect and classify cyber attacks in distribution systems. It employs an AutoEncoder to handle data imbalance and extract deep features from the SWAT dataset. These features are then reduced using Principal Component Analysis (PCA) and classified with a Decision Tree algorithm. Finally, a Deep Neural Network (DNN) is trained on the labeled outputs to enhance prediction accuracy. This multi-stage method effectively identifies various types of attacks with high precision. The development follows the Software Development Life Cycle (SDLC) umbrella model, involving stages like requirement gathering, system design, implementation, testing, and deployment, ensuring a structured and scalable approach16]

III. METHODOLOGY

3.1 Input:

The input to this project is the SWAT (Secure Water Treatment) dataset, which contains real-time sensor data from an industrial water treatment system. Each record in the dataset includes various I/O signal values representing operational behavior, along with corresponding attack labels such as Normal, DoS, NMRI, CMRI, etc. The dataset reflects both normal and attack conditions, making it suitable for training and testing the intrusion detection model. Additionally, test data without labels is used as input during the prediction phase to evaluate how well the system detects and classifies unknown attacks.[18]

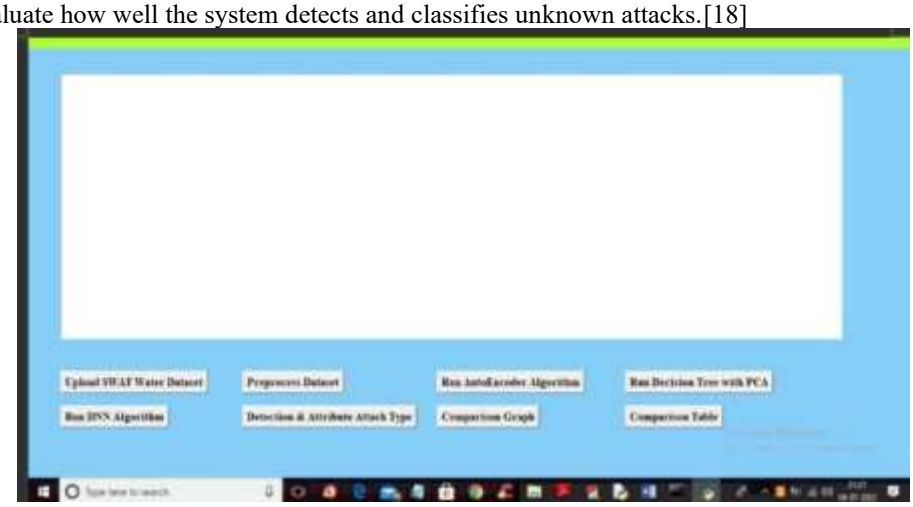

Fig : input screen

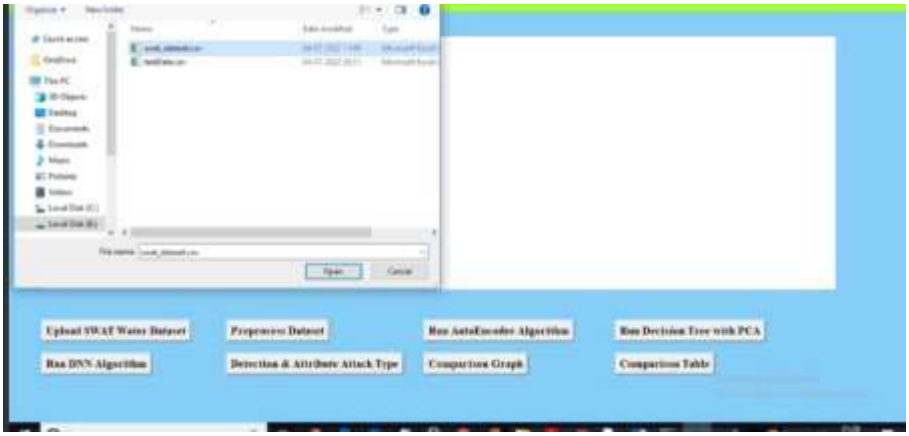

Fig : data set upload

3.2 Method of Process:

The project begins by uploading and preprocessing the SWAT dataset, which includes handling missing values and applying Min-Max normalization to scale the features. Then, an AutoEncoder deep learning model is trained to extract relevant features from the imbalanced data. These features are further refined using Principal Component Analysis (PCA) to reduce dimensionality. The transformed data is then classified using a Decision Tree algorithm to assign preliminary attack labels. Finally, a Deep Neural Network (DNN) is trained using the outputs from the Decision Tree to improve classification accuracy. The trained DNN model is then used to detect and label cyber attacks on new, unlabeled input data.[6]

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04746
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

3.3 Output:

The output of this project is the accurate detection and classification of cyber attacks present in the SWAT dataset. After processing the input data through AutoEncoder, PCA, Decision Tree, and DNN models, the system generates A list of records with their predicted attack types, such as *Normal*, *DoS*, *NMRI*, *CMRI*, etc. It also produces performance metrics like accuracy, precision, recall, and F1-score for each algorithm used. Additionally, graphical outputs such as comparison charts and confusion matrices are generated to visualize the effectiveness of different models. The final result helps identify potential threats in industrial control systems and validates the efficiency of the detection system.

Fig : detection of cyber attack

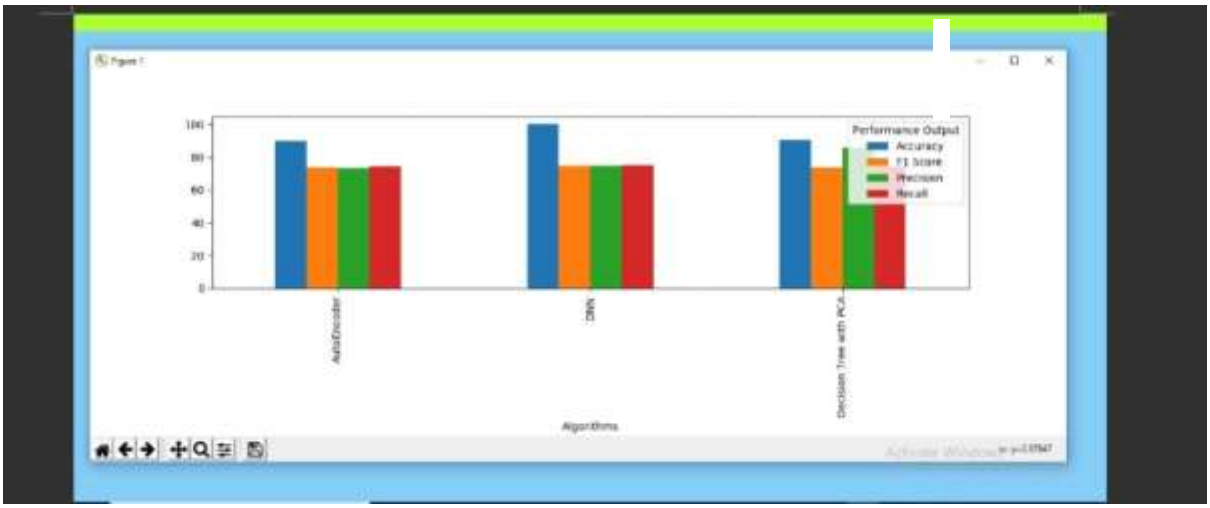

Fig:performance output

IV. RESULTS:

The project achieved a high detection accuracy of up to 99% using the Deep Neural Network (DNN) model after feature extraction through AutoEncoder and classification with Decision Tree and PCA. It successfully identified various cyber attack types such as

DoS, NMRI, CMRI, and others from the SWAT dataset, even under conditions of data imbalance. The comparison table and performance graphs clearly show that the DNN outperformed other models in terms of accuracy, precision, recall, and F1-score, proving the robustness and reliability of the proposed system for real-time cyber attack detection in industrial control environments

V.DISCUSSION:

The project discusses the limitations of traditional intrusion detection systems (IDS), particularly their inability to handle data imbalance, encrypted traffic, and false positives. It highlights how combining AutoEncoder, PCA, Decision Tree, and Deep Neural Networks (DNN) can significantly enhance detection accuracy while reducing false alarms. The discussion emphasizes the importance of early detection to prevent damage in industrial systems and the benefits of minimizing human intervention through automated classification. It also covers the significance of using real-world datasets like SWAT for training models, ensuring practical applicability in detecting both known and unknown cyber attacks effectively.

VI. CONCLUSION

The project discusses the limitations of traditional intrusion detection systems (IDS), particularly their inability to handle data imbalance, encrypted traffic, and false positives. It highlights how combining AutoEncoder, PCA, Decision Tree, and Deep Neural Networks (DNN) can significantly enhance detection accuracy while reducing false alarms. The discussion emphasizes the importance of early detection to prevent damage in industrial systems and the benefits of minimizing human intervention through automated classification. It also covers the significance of using real-world datasets like SWAT for training models, ensuring practical applicability in detecting both known and unknown cyber attacks effectively.

VII. FUTURE SCOPE:

It is not possible to develop a system that makes all the requirements of the user. User requirements keep changing as the system is being used. Some of the future enhancements that can be done to this system are:

- As the technology emerges, it is possible to upgrade the system and can be adaptable to desired environment.
- Based on the future security issues, security can be improved using emerging technologies like single sign-on.

VIII. ACKNOWLEDGEMENT:

G.manoj kumar working as an Assistant Professor in Masters of Computer Applications(MCA) in SVPEC, Visakhapatnam, Andhra Pradesh. Completed her Post graduation in Andhra University College of Engineering (AUCE). With accredited by NAAC with her areas of interest in java, Database management system.

Kandula Hemapriya is pursuing his final semester MCA in Sanketika Vidya Parishad Engineering College, accredited with A grade by NAAC, affiliated by Andhra University and approved by AICTE. With interest in Machine learning Kandula Hemapriya has taken up his PG project on LEVERGING SPATIOTEMPORAL PATTERNS FOR CYBER ATTACKS DETECTIONS IN DISTRIBUTED SYSTEM USING MACHINE LEARNING and published the paper in connection to the project under the guidance of G.MANOJ KUMAR, Assistant Professor, SVPEC.

REFERENCES

[1] Adaptive Cyber Attack Detection in Distribution Systems using Machine Learning and Spatiotemporal Patterns
https://ieeexplore.ieee.org/abstract/document/10985736
[2] Flexible Machine Learning-Based Cyberattack Detection Using Spatiotemporal Patterns for Distribution Systems
https://ieeexplore.ieee.org/abstract/document/8956077
[3] Spatiotemporal Deep Learning for Power System Applications: A Survey
https://ieeexplore.ieee.org/abstract/document/10589409
[4] Spatial-Temporal Data-Driven Model for Load Altering Attack Detection in Smart Power Distribution Networks
https://ieeexplore.ieee.org/abstract/document/10438208

[5] Detection of Cyber Attacks on Voltage Regulation in Distribution Systems Using Machine Learning
https://ieeexplore.ieee.org/abstract/document/9373306

[6] Cyber Attack Pattern Analysis Based on Geo-location and Time: A Case Study of Firewall and IDS/IPS Logs
https://jcrb.net/index.php/Journal/article/view/26

[7] Leveraging Machine Learning for Cybersecurity Resilience in Industry 4.0: Challenges and Future Directions
https://ieeexplore.ieee.org/abstract/document/10721279

[8] Proactive detection of cyber-physical grid attacks: A pre-attack phase identification and analysis using anomaly-based machine learning models
https://www.sciencedirect.com/science/article/pii/S2590005625000682

[9] An unsupervised adversarial autoencoder for cyber attack detection in power distribution grids
https://www.sciencedirect.com/science/article/abs/pii/S0378779624002955

[10] Hybrid and Spatiotemporal Detection of Cyberattack Network Traffic in Cloud Data Centers
https://ieeexplore.ieee.org/abstract/document/10433779

[11] DST-GNN: A Dynamic Spatiotemporal Graph Neural Network for Cyberattack Detection in Grid-Tied Photovoltaic Systems
https://ieeexplore.ieee.org/abstract/document/10638139

[12] Spatio-Temporal Advanced Persistent Threat Detection and Correlation for Cyber-Physical Power Systems Using Enhanced GC-LSTM
https://ieeexplore.ieee.org/abstract/document/10705117

[13] Intrusion Detection Based on Spatiotemporal Characterization of Cyberattacks
https://www.mdpi.com/2079-9292/9/3/460

[14] Deep-Learning Approach to the Detection and Localization of Cyber-Physical Attacks on Water Distribution Systems
https://ascelibrary.org/doi/abs/10.1061/(ASCE)WR.1943-5452.0000983

[15] AI-based Detection Against Cyberattacks in Cyber-Physical Distribution Systems
https://vtechworks.lib.vt.edu/items/55b87b61-5b37-4373-9787-211d2834210d

[16] Threat detection in reconfigurable Cyber–Physical Systems through Spatio-Temporal Anomaly Detection using graph attention network
https://www.sciencedirect.com/science/article/pii/S0167404825001981

[17] Enhancing Cybersecurity in smart grid: a review of machine learning approaches
https://link.springer.com/article/10.1007/s11235-025-01308-9

[18] Cyber—Physical Attack Detection in Water Distribution Systems with Temporal Graph Convolutional Neural Networks
https://www.mdpi.com/2073-4441/13/9/1247

[19] From Data to Insights: Unraveling Spatio-Temporal Patterns of Cybercrime using NLP and Deep Learning
https://discovery.ucl.ac.uk/id/eprint/10188053/

[20] A comparative analysis of various machine learning methods for anomaly detection in cyber attacks on IoT networks
https://www.sciencedirect.com/science/article/abs/pii/S2542660524001033