Leveraging Spatiotemporal Pattern for Cyberattack Detection in Distribution Systems

Paradesi Subba Rao, Vemu Raghavendrasharma, Vadde Harikrishna, Degulapadu Mondla Varun Tej, Pasupuleti Sangeeth Kumar and Pula Sandeep Kumar
Department of Computer Science and Engineering, Santhiram Engineering College,
Nandyal-518501, Andhra Pradesh, India

Keywords: Machine Learning, Spatiotemporal Analysis, Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), Cybersecurity, Power Distribution Networks, Cyberattack Detection, Random Forests.

Abstract: The growing digitalization of power systems where networks are used for the distribution of electricity has also made them susceptible to emerging sophisticated cyber threats, which necessitates the need for advanced detection systems. Traditional security approaches have often failed to capture the complex spatiotemporal dynamics of cyberattacks, leading to delayed or false negative identification of a threat. Here we propose a machine learning based cyber security system which leverages LSTM networks and CNNs to extract spatial features and detect temporal behaviour patterns, and uses Random Forest for making decisions. The proposed approach is able to enhance detection accuracy and reduce the false positive rates by using spatiotemporal data as compared to traditional techniques.

1 INTRODUCTION

The growing complexity and interdependence of modern distribution networks have rendered them susceptible to cyberattacks. Traditional techniques for cyberattack detection often involve rule-based or static analysis systems which may not suffice against advanced, adaptable attackers. Yet it allows a more flexible and effective way to detect cyberattacks using spatiotemporal pattern in data. But spatiotemporal patterns can help uncover structure and behaviour patterns in data that may indicate normal or abnormal activity describing spatial and temporal dependencies. Patterns or abnormalities indicating cyberattacks can be detected by looking at the patterns. Machine learning algorithms do fairly well at spatiotemporal patterns, especially Random Forests, Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks.

1.1 Networks Using Long Short-Term Memory (LSTM)

Sequence prediction tasks. They can learn long-term dependencies and temporal patterns better than other models. For example, LSTM networks could be used to detect such anomalies that do not fit expected patterns in times of cyberattacks through modelling the network usage over time.

1.2 Convolutional Neural Networks (CNN)

CNNs are primarily employed in the processing and analysis of image and other spatial data but can also be applied to time-series data treating it as a sequence of spatial patterns. This enables CNNs to discover local correlations within data which may be useful in detecting spatial anomalies in network traffic. Combining spatial and temporal patterns enables better cyberattack detection through the application of CNNs and LSTM networks together.

1.3 Random Forest

The Random Forests are an ensemble learning method that combine multiple decision trees to improve its classification performance. They perform very well on high-dimensional data, and they may deliver insights into feature importance. Random

Rao, P. S., Raghavendrasharma, V., Harikrishna, V., Tej, D. M. V., Kumar, P. S. and Kumar, P. S.
Leveraging Spatiotemporal Pattern for Cyberattack Detection in Distribution Systems.
DOI: 10.5220/0013877500004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 2*, pages 89-93
ISBN: 978-989-758-777-1

Forests can be used in the detection of cyberattacks, where a large number of features of network traffic are checked to identify patterns associated with an attack.

1.4 Objectives

The main goal of this research is to create an end-to-end framework for detecting cyberattacks in distribution systems using spatiotemporal patterns. This comprises:

- Compiling and preprocessing distribution system network traffic data.
- Finding spatiotemporal patterns in the data using Random Forests, CNNs, and LSTM networks.
- Evaluating how well these algorithms detect cyberattacks in comparison.
- Evaluating the new strategy's effectiveness by contrasting it with traditional techniques.

This study will aim to provide a better and more dependable method for identifying cyberattacks on distribution systems by utilizing the advantages of LSTM networks, CNNs, and Random Forest.

This will strengthen the security and dependability of these infrastructures overall.

2 LITERATURE REVIEW

2.1 Distribution Systems Are Classified by Machine Learning

As smart grids become more interconnected, there is a rise in cyberattacks on electricity distribution networks. Conventional rule-based security systems have trouble identifying complex, dynamic assaults. By examining both temporal trends and spatial correlations in power system data, spatiotemporal machine learning models offer a viable method to improve cyberattack detection.

2.2 How Cyberattack Can Be Detected in Systems for Distribution

System for distributing power cyberattacks can be effectively detected by applying machine learning techniques that examine patterns of network traffic, voltage fluctuations, and system activity. The integration of Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs), and Random Forest (RF) strengthens the identification of attacks by understanding temporal and spatial dependencies in system data.

- **Data Collection & Preprocessing Sources:** SCADA logs, smart meter readings, network activity, and system event logs.
- **Preprocessing:** Data normalization, feature extraction, as well as noise reduction for clean inputs during model training.
- **Cyberattack Detection Methodology:** CNN for the Extraction of Spatial Features, CNNs work well for spatial correlation identification in system states the convolutional layers learn normal vs. abnormal system behaviour patterns and are therefore effective at identifying localized cyber intrusions.
- **LSTM for Recognizing Temporal Patterns:** LSTMs are made to capture sequential dependencies inside time-series data and are therefore well-suited to identify attacks that unfold over time. They examine trends in power fluctuations, communication delays, or anomalous command sequences in the distribution network.
- **Random Forest for Decision Making:** Random Forest classifier acts as the last decision layer, combining features learned by CNNs and LSTMs. It achieves strong classification through the combination of multiple decision trees that diminish overfitting and enhance detection accuracy.

3 METHODOLOGY

3.1 Theoretical Structure

It starts with a raw data set such as a CSV file that requires essential data pre-processing. This includes rescaling the data so that each variable/feature has a specific range (typically between −1 and 1), a data loading system, and creating sequential structures for time-series models. Exocytic Next Module Future Extraction This is where the pre-processed data must be analysed to identify relevant features. This includes drawing out temporal features (representations of trends and fluctuations with time), spatial features, and static or location-based data. These extracted variables form the basis of strong prediction models. Figure 1 show the Evolution of Spatiotemporal patterns in Distribution systems.

It is where the extracted attributes are sent into the Model Training module to train three different models based on machine learning: Random Forest, Long Short-Term Memory network (LSTM), and

Convolutional Neural Network (CNN). The chosen models were selected due to their ability to process complex and sequential data effectively. This Module analyses the performance of each trained model. The CNN, LSTM, and Random forest models are compared using the best measures to do so. From the performances of the models, the best performing model is selected to become the last deployed model. This systematic process guarantees that we adopt the most feasible model for the task.

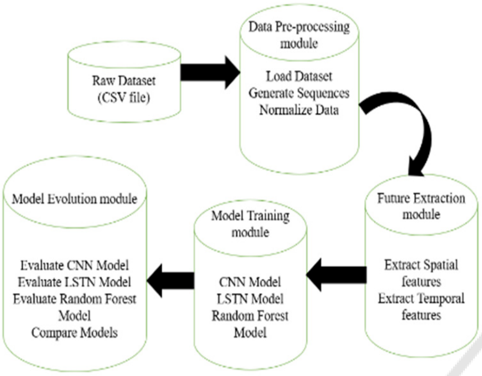


Figure 1: Evolution of Spatiotemporal Patterns in Distribution Systems.

3.2 The Datasets Data that Has Been Utilized

The data used for detection of cyberattacks in distribution systems must have both normal operating data and examples of cyber intrusions, recording key spatiotemporal features of system behaviour. Sources of data are usually smart meter measurements, and network traffic data, which give real-time information on power system operation, voltage fluctuations, and communication patterns among system elements. Furthermore, testbeds like Power World Simulator, IEEE test systems can be used to simulate real-world threats like False Data Injection (FDI), Denial of Service (DoS) attacks to create cyberattack scenarios. The dataset contains a number of imperative features such as time-series power measurements (voltage, current, and frequency variations), network traffic characteristics (packet size, source address, destination address, and transmission delays publicly accessible datasets like the ICS/SCADA cybersecurity dataset, NSL-KDD, UNSW-NB15, and CICIDS2017 is suitable for training machine learning models, Moreover, simulated attack data over IEEE 33-Bus or 118-Bus test cases.

The dataset is pre-processed before model training, comprising cleaning of data to deal with missing values, feature scaling to meet the requirements of CNN and LSTM models, and data labelling to identify normal and attack cases. A well-organized dataset allows the discussed machine learning paradigm CNN to extract spatial features, LSTM for temporal examination, and Random Forest for classification efficiently detect cyber threats on distribution systems to improve grid resilience and security.

3.3 This Approach Is to Extract Features Described

Feature extraction is crucial step in the process of using spatiotemporal patterns for cyberattack detection in distribution systems. Taking valuable characteristics out of raw data enables us to convert the data into a suitable format for use by machine learning algorithms. Here, we detail the method for feature extraction utilized in our work, which comprises the utilization of Networks using Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and Random Forests.

Through the utilization of the advantages of LSTM networks, CNNs, and Random Forests, our feature extraction method guarantees that those features most pertinent and significant to cyberattack detection in distribution systems are utilized.

4 RESULTS AND EVALUATION

The Random Forest model successfully classifies 186 out of 186 examples, demonstrating good performance in categorizing class '0'. However, it only correctly classifies 14 out of 14 examples with zero misclassifications as '0', demonstrating low performance for class '1'. This indicates that class '1' may have low recall but excellent precision.

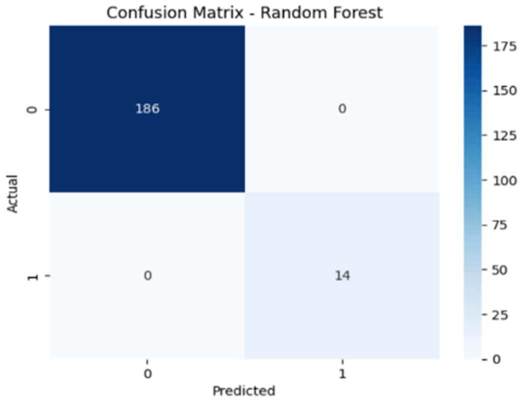
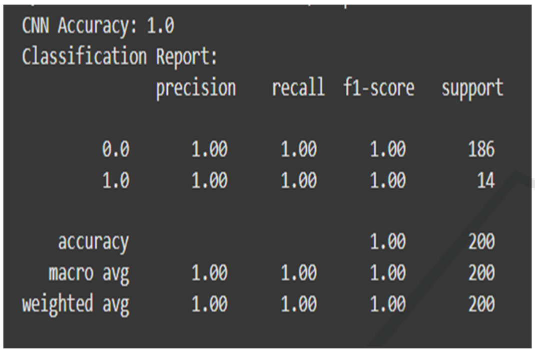


Figure 2: Confusion Matrix – Random Forest.

The model has a clear bias towards class '0' prediction, suggesting that either the dataset is unbalanced or the model requires further fine-tuning to improve class '1' prediction. The model has to be improved to perform better on the minority class, although overall it has good accuracy for the majority class. Figure 2 show the Confusion Matrix – Random Forest.

An ideal classification on the given dataset is demonstrated by the Convolutional Neural Network (CNN) in the following report, which registers 100% accuracy on all measures (precision, recall, and F1-score) for both classes. Figure 3 show the Optimal CNN Classification Precision.



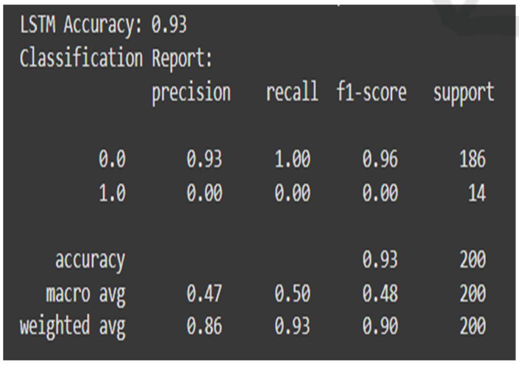
```
CNN Accuracy: 1.0
Classification Report:
              precision    recall  f1-score   support

         0.0       1.00      1.00      1.00       186
         1.0       1.00      1.00      1.00        14

    accuracy                           1.00       200
   macro avg       1.00      1.00      1.00       200
weighted avg       1.00      1.00      1.00       200
```

Figure 3: Optimal CNN Classification Precision.

The LSTM model achieves a 93% accuracy overall, but struggles with predicting class '1', showing zero precision, recall, and F1-score. Figure 4 show the LSTM Model Performance.



```
LSTM Accuracy: 0.93
Classification Report:
              precision    recall  f1-score   support

         0.0       0.93      1.00      0.96       186
         1.0       0.00      0.00      0.00        14

    accuracy                           0.93       200
   macro avg       0.47      0.50      0.48       200
weighted avg       0.86      0.93      0.90       200
```

Figure 4: LSTM Model Performance.

5 DISCUSSION

186 out of 186 correctly predicted the majority class ('0'), demonstrating the remarkable accuracy of the Random Forest model.

With all accuracy metrics (precision, recall, and F1-score) ideal for both classes, the CNN model performs exceptionally well. It shows that the CNN model has successfully identified the hidden patterns and is prepared to accurately classify both majority and minority classes.

The LSTM model has a high overall accuracy of 93%, highlighted especially by its outstanding performance on the majority class ('0'). The CNN model performs exceptionally well.

6 CONCLUSIONS

In our paper we provide a solution for cyberattack detection in machine learning-based distribution systems models. Out among the models tested, the Random Forest Classifier proved the actual and predictive analysis and also LSTM provide 0.93 accuracy. Which mainly describes how the data could be detected and also ensure significantly increase actual and prediction values.

REFERENCES

Paradesi Subba Rao,"Detecting malicious Twitter bots using machine learning" AIP Conf. Proc. 3028, 020073 (2024),https://doi.org/10.1063/5.0212693

Paradesi SubbaRao,"Morphed Image Detection using Structural Similarity Index Measure"M6 Volume 48 Issue 4 (December 2024),https://powertechjournal.com

Parumanchala Bhaskar, et al. "Machine Learning Based Predictive Model for Closed Loop Air Filtering System." *Journal of Algebraic Statistics* 13.3 (2022): 416-423.

Mahammad, Farooq Sunar, et al. "Prediction Of Covid-19 Infection Based on Lifestyle Habits Employing Random Forest Algorithm." *Journal Of Algebraic Statistics* 13.3 (2022): 40-45.

Devi, M. Sharmila, et al. "Machine Learning Based Classification and Clustering Analysis of Efficiency of Exercise Against Covid-19 Infection." *journal of algebraic statistics* 13.3 (2022): 112-117.

Parumanchala Bhaskar, et al. "Incorporating Deep Learning Techniques to Estimate the Damage of Cars During the Accidents" *AIP Conference Proceedings*. Vol. 3028. No. 1. AIP Publishing, 2024.

Mr.M.Amareswara Kumar,EFFECTIVE FEATURE ENGINEERING TECHNIQUE FOR HEART DISEASE PREDICTION WITH MACHINE LEARNING" in *International Journal of Engineering & Science Research, Volume 14, Issue 2, April-2024 with ISSN* 2277-2685.

Chaitanya, V. Lakshmi. "Machine Learning Based Predictive Model for Data Fusion Based Intruder Alert

System." *journal of algebraic statistics* 13.2 (2022): 2477-2483

Mahammad, Farooq Sunar, Karthik Balasubramanian, and T. Sudhakar Babu. "A comprehensive research on video imaging techniques." *All Open Access, Bronze* (2019).

Parumanchala Bhaskar, et al "Cloud Computing Network in Remote Sensing-Based Climate Detection Using Machine Learning Algorithms" remote sensing in earth systems sciences(springer).

Chaitanya, V. Lakshmi, and G. Vijaya Bhaskar. "Apriori vs Genetic algorithms for Identifying Frequent Item Sets." *International journal of Innovative Research &Development* 3.6 (2014): 249-254.

Suman, Jami Venkata, et al. "Leveraging natural language processing in conversational AI agents to improve healthcare security." *Conversational Artificial Intelligence* (2024): 699-711.

Mandalapu, Sharmila Devi, et al. "Rainfall prediction using machine learning." *AIP Conference Proceeding*s. Vol. 3028. No. 1. AIP Publishing, 2024.

Mahammad, Farooq Sunar, et al. "Key distribution scheme for preventing key reinstallation attack in wireless networks." *AIP Conference Proceedings*. Vol. 3028. No. 1. AIP Publishing, 2024.

Chaitanya, V. Lakshmi, et al. "Identification of traffic sign boards and voice assistance system for driving." *AIP Conference Proceedings*. Vol. 3028. No. 1. AIP Publishing, 2024

Mahammad, Farooq Sunar, and V. Madhu Viswanatham. "Performance analysis of data compression algorithms for heterogeneous architecture through parallel approach." *The Journal of Supercomputing* 76.4 (2020): 2275-2288.

Devi, M. Sharmila, et al. "Extracting and Analyzing Features in Natural Language Processing for Deep Learning with English Language." Journal of Research Publication and Reviews 4.4 (2023): 497-502.

Sunar, Mahammad Farooq, and V. Madhu Viswanatham. "A fast approach to encrypt and decrypt video streams for secure channel transmission." *World Review of Science, Technology and Sustainable Development* 14.1 (2018): 11-28.

Gowthami, V., et al. "Knowledge Based System for Immunity Improvement Against Covid-19 Infection." *journal of algebraic statistics* 13.3 (2022): 01-07.

Mahammad, Farooq Sunar, et al. "Heuristics Approach Based Expert System for Covid-19 Infection *susceptibility." journal of algebraic statistics* 13.3 (2022): 46-51.

Reddy, E. Madhusudhana, and P. Bhaskar. "Able Machine Learning Method for classifying Disease-Treatment Semantic Relations from Bio-Medical Sentences." vol 5(2018):5.https://ieeexplore.ieee.org/document/9616841