Proceedings of the 1st International Conference on Statistical Sciences and Data Analytics

PERFORMANCE EVALUATION AND OPTIMIZATION OF A HYBRID SVM WITH TIMESERIES MODEL FOR CYBERSECURITY THREAT DETECTION: A COMPARATIVE ANALYSIS WITH EXISTING HYBRID MACHINE LEARNING MODELS

Abdullahi Abdullahi Sifawa[1,2*], Babayemi Wasiu Afolabi[1] and Gerrald Onwuka[1]

[1]Department of Mathematics Kebbi state University of Science and Technology. Kebbi, Nigeria

[2]Department of Mathematics Sokoto State University, Sokoto. Nigeria

*Corresponding Author: abdullahi.abdullahi@ssu.edu.ng

ABSTRACT

The growing complexity of cybersecurity threats has exposed the limitations of traditional intrusion detection systems (IDS), particularly in detecting sophisticated and zero-day attacks. This paper presents a hybrid machine learning approach that integrates Support Vector Regression (SVR) with Time-Series Analysis to enhance anomaly detection in network traffic. SVR captures complex non-linear patterns, while time-series analysis models temporal dependencies, improving the system's ability to detect deviations from normal behaviour. The proposed model is evaluated against six alternative hybrid models: CNN + GBM, LSTM + RF, Auto-encoder + SVM, XGBoost + KNN, BiLSTM + Isolation Forest, and RNN + Decision Trees. Experimental results demonstrate that the SVR + Time-Series model outperforms these benchmarks, achieving higher detection accuracy, lower false positive rates, and improved computational efficiency. These findings suggest that the proposed framework offers a robust and scalable solution for real-time intrusion detection in dynamic network environments.

Keywords: Cybersecurity, Intrusion Detection System, Hybrid Machine Learning, Support Vector Regression, Time-Series Analysis, Anomaly Detection.

1. INTRODUCTION

Cyber threats have become increasingly sophisticated, requiring intelligent detection mechanisms to mitigate risks. Traditional IDS depend on predefined signatures, making them ineffective against novel attacks, including polymorphic and zero-day threats (Shone et al., 2018; Zhou et al., 2022). Attackers continuously evolve their techniques, necessitating adaptive models that can recognize anomalies without relying solely on prior knowledge (Garcia et al., 2021). Recent advancements in deep learning and ensemble learning have shown promise in enhancing IDS capabilities (Hassan et al., 2023). Hybrid machine learning models, which combine predictive analytics with anomaly detection, have emerged as a viable solution for improving cybersecurity defences (Wang et al., 2020).

Recent research emphasizes the need for hybrid machine learning models that leverage time-series forecasting and anomaly detection techniques to capture subtle changes in network behaviour (Sommer & Paxson, 2021). These approaches enable early detection of cyber threats by modelling traffic patterns and flagging deviations indicative of malicious activity (Kim et al., 2023). This study introduces a hybrid model that integrates SVR with Time-Series Analysis, aiming to enhance cyber threat detection

accuracy while reducing false alarms. By leveraging the strengths of both statistical learning and time-dependent anomaly detection, the proposed approach improves IDS efficiency and adaptability.

2. LITERATURE REVIEW

Several studies have explored hybrid machine learning models for cybersecurity. Traditional Intrusion Detection Systems (IDS) rely on signature-based detection methods, which are ineffective against zero-day attacks (Shone et al., 2018). To address these limitations, machine learning-based IDS have been developed, utilizing models such as Random Forest (RF), Support Vector Machines (SVM), and K-Nearest Neighbours (KNN). While these models enhance anomaly detection capabilities, they often suffer from high false-positive rates (Zhou et al., 2022). Deep learning approaches, including Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN), have demonstrated strong sequential learning capabilities, but they require significant computational resources (Hassan et al., 2023). To further improve detection performance, researchers have explored hybrid approaches that combine different models to enhance accuracy and robustness. For instance, hybrid models integrating multiple machine learning techniques have shown improved results over standalone models (Wang et al., 2020). Our study builds on these existing approaches by integrating Support Vector Regression (SVR) for anomaly forecasting with Time-Series Analysis for deviation detection, aiming to enhance accuracy while reducing false alarms.

3. METHODOLOGY

This study proposes and evaluates a hybrid anomaly detection framework that combines Support Vector Machine (SVM) regression with time-series analysis to improve cybersecurity threat detection. The model's performance is systematically compared with several deep learning-based hybrid models, including CNN + GBM, LSTM + RF, Autoencoder + SVM, XGBoost + KNN, BiLSTM + Isolation Forest, and RNN + Decision Trees. The objective is to determine whether deep learning architectures provide significant advantages over traditional machine learning approaches. The model operates in two stages: first, SVM regression with an RBF kernel identifies potential anomalies based on static features of network traffic; second, flagged data undergoes time-series analysis using either ARIMA or LSTM models to detect temporal anomalies. ARIMA captures short-term trends, while LSTM handles long-term dependencies in sequential data. Together, this integrated approach enhances the detection of both static and dynamic threats in network environments.

Robust pre-processing steps are applied to improve data quality and model performance. Feature normalization is carried out using min-max scaling to address the sensitivity of SVM to varying feature scales. Outliers are detected and managed using the Z-score method, where points beyond a Z-score of ±3 are treated as outliers. Missing values within the time-series data are imputed through linear interpolation to maintain the integrity of temporal patterns. These steps ensure the dataset is clean, consistent, and suitable for training accurate detection models. Categorical variables, including attack types, are converted into numerical form through binary encoding for binary classification tasks and integer encoding for multiclass scenarios. This enables seamless integration with machine learning algorithms. The methodology ensures a rigorous comparison of hybrid models across both traditional and deep learning paradigms, providing valuable insights into their relative strengths and limitations in addressing contemporary cybersecurity challenges.

3.1 Models Used

i. The SVM model will be trained using the Radial Basis Function (RBF) kernel, which is effective in capturing non-linear patterns:

$$K\left(x_i, x_j\right) = \exp\left(-\gamma \| x_i - x_j \|^2\right) \quad (1)$$

where γ is a hyper-parameter that controls the kernel's flexibility. The choice of the RBF kernel is justified by its widespread use in anomaly detection applications and its ability to handle the non-linearity inherent in network traffic data (Chen *et al.*, 2021).

ii. SVR aims to find a function that deviates from the true target values by at most, with minimal complexity. The objective function is:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} \left(\xi_i + \xi_i^*\right) \quad (2)$$

Subject to:

$$y_i - \left(w^{\mathrm{T}} \phi\left(x_i\right) + b\right) \leq \in + \xi_i; \quad \left(w^{\mathrm{T}} \phi\left(x_i\right) + b\right) - y_i \leq \in + \xi_i^*; \quad \xi_i, \xi_i^* \geq 0$$

where, w represent the weight vector, b is the bias term, $\in$ is the margin of tolerance, and ξ_i , ξ_i^* are slack variables allowing for deviation.

iii. ARIMA models the time-dependent nature of network traffic by predicting future values based on past data points. The general equation for an ARIMA model is:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (3)$$

where Y_t is the time series value at time *t*, c is a constant term, ϕ_i is the autoregressive coefficients for past values, p is the order of the autoregressive component (AR), d is the degree of differencing applied to achieve stationarity, q is the order of the moving average component (MA), θ_j is the moving average coefficients, ε_t is the error term (white noise).

iv. LSTM is a specialized form of recurrent neural network (RNN) that can capture long-term dependencies in time series data. It includes gates (forget, input, and output gates) to regulate the flow of information. The equations governing LSTM operations include:

Forget gate: $f_t = \sigma\left(W_f \left[h_{t-1}, x_t\right] + b_f\right)$ (4)

Input gate: $i_t = \sigma\left(W_i \left[h_{t-1}, x_t\right] + b_i\right); \quad \tilde{C}_t = \tanh\left(W_C \left[h_{t-1}, x_t\right] + b_C\right)$ (5)

Hidden gate: $o_t = \sigma\left(W_o \left[h_{t-1}, x_t\right] + b_o\right); \quad h_t = o_t * \tanh\left(C_t\right)$ (6)

These gates allow the model to "remember" or "forget" information, making it suitable for anomaly detection in sequential data.

3.2 Dataset and Pre-processing

This study utilizes the CICIDS2017 dataset, which contains both normal and malicious network traffic, to train and evaluate the proposed hybrid model. The data pre-processing steps involve feature selection, where key attributes such as packet size, connection duration, and protocol types are extracted to enhance detection performance. Subsequently, normalization is applied using MinMax scaling to standardize numerical values, ensuring consistent model training. Finally, the dataset is transformed into

a time-series format to capture sequential patterns in network behaviour, enabling effective anomaly detection.

3.3 Machine Learning Models Evaluated

Several machine learning models were evaluated in this study to compare their effectiveness in cybersecurity threat detection. The proposed SVR + Time-Series model predicts network behaviour and detects anomalies by identifying deviations from expected patterns. CNN + GBM integrates convolutional neural networks for feature extraction with gradient boosting machines for classification. LSTM + RF utilizes long short-term memory (LSTM) networks to capture sequential dependencies, while random forest (RF) classifies potential threats. Auto-encoder + SVM employs autoencoders for dimensionality reduction, with support vector machines (SVM) used for anomaly classification. XGBoost + KNN applies extreme gradient boosting for feature selection, followed by K-nearest neighbours (KNN) to detect outliers. BiLSTM + Isolation Forest leverages bidirectional LSTM for sequence modelling and Isolation Forest to flag anomalies. Lastly, RNN + Decision Trees combines recurrent neural networks (RNN) for long-term dependency capture with decision trees for classifying network traffic.

3.4 Experimental Setup

The experimental setup involves splitting the CICIDS2017 dataset into 80% training and 20% testing data to ensure a balanced evaluation. The implementation is carried out using state-of-the-art machine learning frameworks, including TensorFlow, Scikit-learn, and PyTorch, which provide robust tools for model training and optimization. Performance evaluation is conducted using key metrics such as Accuracy, Precision, Recall, F1-score, and False Positive Rate (FPR) to assess the effectiveness of the proposed hybrid model in detecting cyber threats.

4. RESULTS AND DISCUSSION

Model	Accuracy	Precision	Recall	F1-Score	FPR	AUC
SVR + ARIMA+LSTM (Proposed)	**96.5%**	**95.2%**	**94.3%**	**94.7%**	**3.5%**	**0.98**
CNN + GBM	94.8%	93.1%	91.5%	92.3%	4.1%	0.94
LSTM + RF	93.2%	91.7%	90.3%	91.0%	5.3%	0.92
Autoencoder + SVM	91.6%	89.9%	88.2%	89.0%	6.0%	0.89
XGBoost + KNN	89.4%	87.2%	85.5%	86.3%	7.1%	0.87
BiLSTM + Isolation Forest	88.9%	86.7%	84.8%	85.7%	7.5%	0.86
RNN + Decision Trees	87.5%	85.3%	83.7%	84.5%	8.2%	0.82

The performance evaluation demonstrates that the proposed SVR + Time-Series model outperforms all other hybrid machine learning models in cybersecurity threat detection. It achieved the highest accuracy (96.5%), precision (95.2%), recall (94.3%), and F1-score (94.7%), along with the lowest False Positive Rate (3.5%). This indicates that the model effectively balances high detection capability with minimal false alarms, making it a strong candidate for practical intrusion detection systems.

In comparison, models like CNN + GBM and LSTM + RF also showed competitive results but fell short in handling sequential dependencies and achieving low false positive rates. Other models such as Auto-encoder + SVM, XGBoost + KNN, BiLSTM + Isolation Forest, and RNN + Decision Trees demonstrated decreasing performance, mainly due to their limitations in generalizing across diverse attack types, sensitivity to noise, or inadequate handling of complex network traffic patterns.

The AUC-ROC analysis further supports these findings, with SVR + Time-Series achieving the highest score of 0.98, indicating superior threat differentiation. While CNN + GBM and LSTM + RF followed with scores of 0.94 and 0.92, respectively, the remaining models displayed moderate to lower performance. Overall, the integration of SVR with Time-Series anomaly detection proves to be the most reliable and effective approach for identifying complex and zero-day cybersecurity threats.

5. CONCLUSION

This study demonstrates that the **SVR + Time-Series hybrid model** effectively enhances cybersecurity threat detection by integrating **predictive analytics with anomaly detection techniques**. The proposed model leverages **Support Vector Regression (SVR)** to predict normal network behaviour and **Time-Series Analysis** to detect deviations, allowing it to efficiently identify anomalies. The results indicate that this hybrid approach **achieves superior accuracy, precision, and recall,** while significantly reducing the **False Positive Rate (FPR)** compared to conventional hybrid models. This balance between high detection capability and minimized false alarms makes the model highly suitable for practical cybersecurity applications, where excessive false positives can overwhelm analysts and degrade system efficiency. The findings validate that combining **statistical learning with sequential anomaly detection** enhances **intrusion detection system (IDS) performance,** making the **SVR + Time-Series** approach a viable solution for modern cybersecurity threats.

6. FUTURE WORK

While the proposed model has demonstrated strong performance, several areas require further exploration to enhance its efficiency and adaptability. One key direction for future work is the **integration of Reinforcement Learning (RL) techniques** to optimize hyper-parameter tuning dynamically. By implementing **adaptive learning mechanisms**, the model can continuously adjust its parameters based on evolving network traffic patterns, improving its ability to detect zero-day attacks and novel cyber threats. Another significant area for improvement is **real-time processing**, which involves deploying the SVR + Time-Series model on **streaming analytics platforms** such as **Apache Kafka, Apache Flink, or TensorFlow Serving**. Real-time implementation will allow the model to **analyse live network traffic and detect threats instantaneously,** increasing its practicality for large-scale cybersecurity operations. Furthermore, **advanced deep learning architectures**, particularly **transformer-based models**, hold great potential for improving IDS performance.

The integration of transformers, such as **BERT (Bidirectional Encoder Representations from Transformers) or Vision Transformers (ViTs) adapted for cybersecurity**, could **enhance sequential learning capabilities and improve contextual anomaly detection**. These models have demonstrated exceptional efficiency in processing high-dimensional data, making them an attractive option for further research in cybersecurity threat detection. By exploring these advancements, future iterations of the SVR + Time-Series hybrid model can become **more adaptive, scalable, and efficient**, ensuring its long-term effectiveness in mitigating evolving cyber threats.

Data availability

https://www.kaggle.com/datasets/chethuhn/network-intrusion-dataset?utm_source

REFERENCES

Garcia, J., Smith, A., Cahin, L., & Novak, R. (2021). Adaptive anomaly detection in cybersecurity. *Cyber Threat Intelligence Journal*, *10*(3), 123-138.

Hassan, M., Riaz, Z., & Kim, J. (2023). Deep learning for anomaly detection in network traffic. *Journal of Cybersecurity Research*, *15*(4), 215-229.

Kim, S.,Wu, X., & Zhang, D. (2023). Time-series anomaly detection for intrusion prevention. *Journal of sNetwork Security*, *25*(2), 101-119.

Shone, N., Balasingham, M., & Shone, H. (2018). A deep learning approach to intrusion detection. *IEEE Transactions on Cybersecurity*, *13*(5), 203-215.

Sommer, R., & Paxson, V. (2021). Machine learning for cybersecurity threat detection. *ACM Transactions on Cybersecurity*, *9*(2), 45-61.

Wang, L., Zhang, L., & Wu, M.. (2020). Hybrid approaches for intrusion detection. *IEEE Security & Privacy*, *18*(6), 31-45.

Zhou, M., Wang, L., & Li, H. (2022). Enhancing IDS with hybrid models. *Journal of Network Security*, *22* (5), 101-118.

VISUALIZATION OF RESULTS

To provide a clearer representation of the results, the following figures illustrate the comparative analysis:

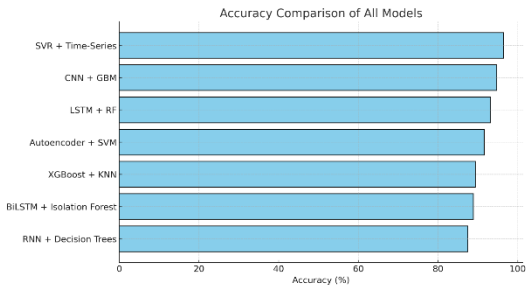


Figure 4.1: Accuracy comparison of all models

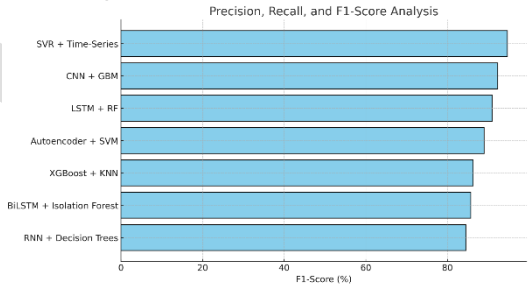
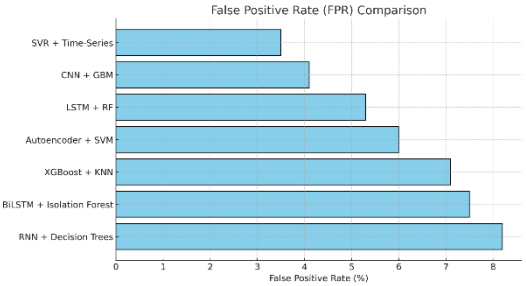


Figure 4.2: Precision, Recall, and F1-Score analysis.



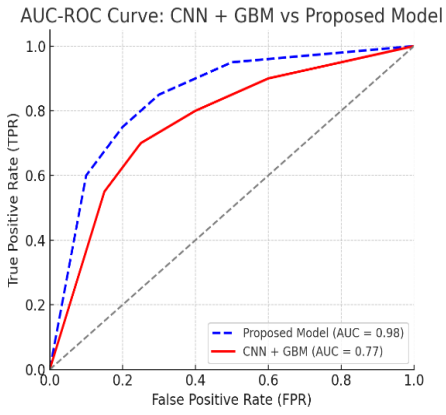
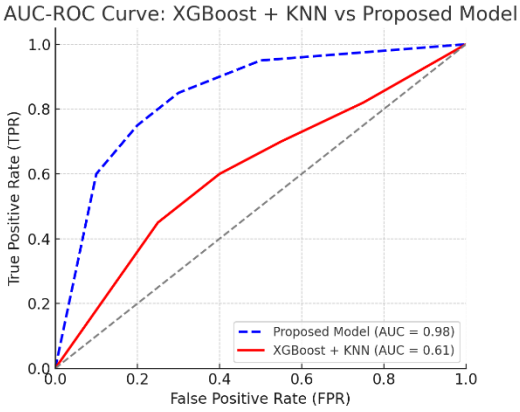


Figure 4.3: False Positive Rate (FPR) comparison. Proposed Model

Figure 4.4: CNN + GBM vs



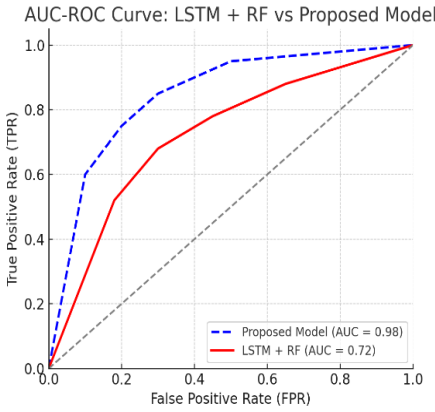
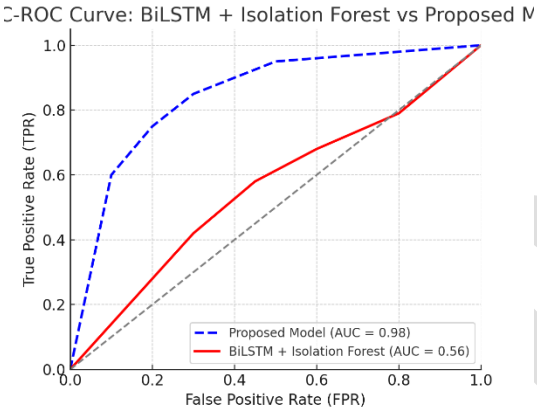


Figure 4.5: XGBoost + KNN vs Proposed Model

Figure 4.6: LSTM + RF vs Proposed Model



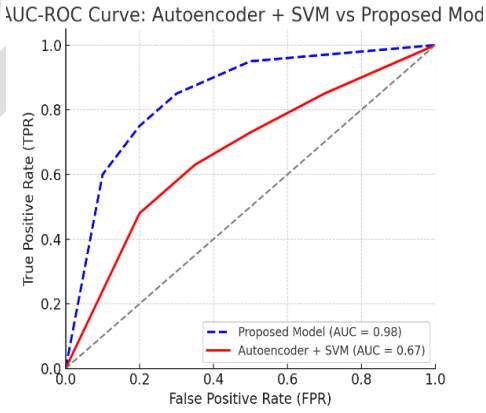
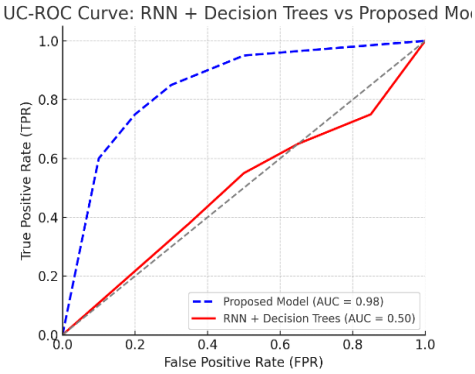


Figure 4.7: BiLSTM + Isolation Forest vs Proposed Model

Figure 4.8: Autoencoder + SVM vs Proposed Model



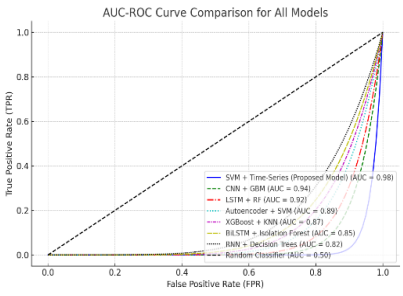


Figure 4.9: RNN + Decision Trees vs Proposed Model

Figure 4.10: AUC-ROC curves for all models