|| Volume 9 || Issue 8|| August 2025 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

An Effective Framework for Packet Behavior Classification in High-Speed Connectionless Networks

Chetan L[1], M Rajani[2], Raghu Kumar K S[3]
[1]*Artificial Intelligence and Data Science, PG Scholar, Vijaya Vittala Institute of Technology, Bengaluru, India, Affiliated to Visvesvaraya Technological University, Belagavi, India,*
chethanlokanjali2003@gmail.com
[2]*Assistant Professor, Department of Computer Science and Engineering, Vijaya Vittala Institute of Technology, Bengaluru, India, Affiliated to Visvesvaraya Technological University, Belagavi, India,*
mrajaniratnakumar@gmail.com
[3]*Professor and Head, Department of Computer Science and Engineering, Vijaya Vittala Institute of Technology, Bengaluru, India, Affiliated to Visvesvaraya Technological University, Belagavi, India, ksraghukumar@gmail.com*

Abstract: The surge in connectionless data transmissions, driven by IoT devices and streaming services, has intensified challenges such as packet loss, jitter, and delay. Conventional monitoring systems often fail to adapt to these dynamic conditions, highlighting the need for intelligent Machine Learning (ML)-based approaches. This paper presents a comprehensive ML framework for evaluating and predicting packet behavior in connectionless networks using synthetically generated datasets. Two models—Random Forest and Support Vector Machine (SVM)—were developed and compared. The Random Forest classifier demonstrated superior performance, achieving 98.33% precision, recall, and F1-score across all classes, while the SVM model yielded a slightly lower overall accuracy of 97%. Feature importance analysis revealed packet loss rate and jitter as critical predictors influencing classification. Performance validation through confusion matrices, ROC-AUC scores, and precision-recall curves confirmed the high reliability of both models, with each attaining an AUC of 1.00. The findings establish Random Forest as the more robust and accurate choice for anomaly detection in unstable network environments. Leveraging such ML techniques enables proactive monitoring, prediction, and optimization of network performance, ultimately enhancing the reliability and efficiency of connectionless communication systems.

Keywords: *Anomaly Detection, Connectionless Networks, Jitter, Machine Learning, Packet Loss, Precision-Recall, Reliability, ROC Curve, Streaming Services, Wireless Networks*

I.INTRODUCTION:

The rapid expansion of connectionless data transmissions has significantly reshaped digital communications. Technologies such as UDP-based protocols, Internet of Things (IoT) devices, and streaming services rely heavily on connectionless communication models, where data is transmitted without establishing a dedicated link between sender and receiver [01]. This approach prioritizes speed and scalability, making it highly suitable for real-time applications like video streaming, VoIP, and sensor networks. However, the absence of delivery guarantees, ordering mechanisms, or built-in error correction introduces vulnerabilities. With the exponential growth of IoT ecosystems and multimedia platforms, it has become increasingly critical to ensure that such networks can handle massive data loads without compromising performance or user experience [02].

Ensuring reliability, quality, and efficiency in connectionless environments presents unique challenges. Unlike connection-oriented systems, these networks are more susceptible to packet loss, jitter, and variable delays, all of which can severely degrade application performance. Analyzing packet behavior is therefore essential for understanding network health and identifying performance trends. Detailed packet analysis can uncover patterns and anomalies that may point to congestion, bottlenecks, or potential failures [03]. To address these challenges, predictive modeling driven by Machine Learning (ML) has emerged as a promising solution. By learning from historical packet transmission data, ML models can forecast network performance, identify vulnerable flows, and suggest real-time optimizations. This proactive strategy enables unstable connectionless systems to consistently deliver the high levels of quality and reliability required by modern data-intensive and time-sensitive applications [04].

Machine Learning has become a transformative tool for automated evaluation and estimation of packet flows in contemporary networks. Traditional methods, which often relied on static rule-based systems and manual inspection, cannot keep pace with today's dynamic and high-volume traffic [05]. ML models, in contrast, leverage vast amounts of historical and real-time data to detect complex patterns and adapt autonomously to evolving conditions. In connectionless environments—where packet loss, disorder, and volatility are commonplace—ML provides invaluable intelligent and automated insights. Different paradigms, including supervised, unsupervised, and reinforcement learning, are being employed to classify, predict, and optimize packet behavior, thereby enhancing network efficiency and resilience [06].

Central to ML-driven analysis is the intelligent evaluation of key parameters such as packet loss, jitter, delay, and packet size. Packet loss, representing the percentage of lost data during transmission, severely impacts applications that require uninterrupted streams (e.g., VoIP and online gaming). Jitter, the

variation in packet arrival times, directly affects streaming quality and real-time communications [07]. Delay, or latency, is particularly critical for applications requiring immediate responsiveness, while unusual variations in packet size may indicate fragmentation or malicious activity. By accurately analyzing these metrics, ML models provide actionable insights into network health. Unlike threshold-based monitoring, ML approaches capture subtle, non-linear correlations that might otherwise go undetected, offering a deeper understanding of network behavior [08].

As demand for faster and more reliable communication grows, the importance of real-time or near-real-time predictive models has increased. ML enables continuous monitoring of network conditions and can initiate proactive responses such as dynamic rerouting, adaptive bandwidth allocation, or anomaly mitigation before users experience service degradation. Unlike static systems, real-time ML models are adaptive, context-aware, and capable of evolving with changing traffic dynamics [09]. Moreover, these models can be trained to recognize both established and novel issues, thereby improving robustness against emerging challenges. The benefits of ML-based methods include scalability across large networks, automation of complex evaluations, and the discovery of hidden trends that traditional systems overlook. Ultimately, ML equips administrators with advanced tools for smarter, faster, and more reliable management of connectionless environments [10].

In summary, the implementation of ML-based evaluation and estimation of packet behavior is becoming indispensable in connectionless networks. Traditional static methods lack the flexibility to address the dynamic nature of packet flows, leading to inefficiencies and compromised service quality. ML techniques facilitate real-time monitoring, predictive analysis, and anomaly detection, while optimizing performance through intelligent evaluation of packet loss, jitter, delay, and packet size. By enabling adaptive and proactive network management, ML significantly enhances reliability, efficiency, and user experience in today's complex, large-scale communication environments [11].

II.LITERATURE SURVEY

T. Suga et al. focus on real-time packet classification to mitigate malicious traffic in enterprise and carrier networks. Their work aims to intercept harmful packets before they reach hosts, addressing the latency challenges of ML-based detection systems. They propose a packet forwarding architecture built on DPDK, LIBSVM, and TensorFlow, and conduct bottleneck analysis to refine their system with a DNS-based filtering model. Experimental results demonstrate that TensorFlow models achieve around 10 Mbps throughput with 87–91% detection accuracy. Key challenges include reducing ML classification latency, designing efficient feature extraction, balancing accuracy with processing speed, and integrating simple whitelisting/blacklisting with ML models for scalability [12].

N. Das et al. investigate EEG-based mental state classification under music and no-music conditions, enhancing wavelet packet denoising with ML. Their methodology applies Wavelet Packet Decomposition (WPD) with entropy-based thresholds, followed by feature extraction (statistical, covariance, entropy, wavelet energy) and classification using k-Nearest Neighbors (kNN) optimized with Bayesian tuning. Numerical findings reveal that covariance features with an optimal WPD threshold of 620 yield the highest accuracy of 99.62% ± 0.005. Major challenges include selecting suitable thresholds, balancing denoising with feature preservation, and adapting ML models to varying feature sensitivities [13].

L. Zhao et al. propose a framework for overtaking feasibility prediction in mixed traffic environments, integrating connected and traditional vehicles. Their pipeline includes vehicle state sensing, missing data repair via kinematic models, future behavior forecasting with LSTM and attention mechanisms, and feasibility assessment using the novel PPODDQN deep reinforcement learning algorithm. Results show substantial improvements: lower RMSE in prediction, 94.97% safe time-to-collision rates, and efficient overtaking with minimal lane changes. However, challenges persist in handling data occlusion, ensuring reliable vehicle communication, and adapting to traffic flow variability [14].

H. Si et al. explore enhanced socket communication in large-scale virtual network laboratories using deep reinforcement learning. Their objective is to optimize TCP/IP-based communication for high throughput and ultra-low latency. By integrating socket mechanisms with reinforcement learning, they achieve real-time resource scheduling and system parameter optimization. Experiments involving mobile and PC clients demonstrate significant gains in throughput and latency reduction compared to conventional socket methods. Remaining challenges include managing multi-node communications, maintaining cross-platform consistency, addressing concurrency, and scaling under heavy workloads [15].

T. Xu et al. introduce WD-trace, a method for detecting gene network rewiring by combining gene expression data with static regulatory networks. They employ a weighted lasso-penalized D-trace loss function, optimized with an accelerated proximal gradient algorithm. WD-trace consistently outperforms D-trace, delivering higher ROC and PR scores in simulations, and effectively identifying critical genes in ovarian and breast cancer datasets. Key challenges involve parameter tuning, balancing static versus dynamic data contributions, and ensuring robustness against incomplete or noisy regulatory data [16].

N. M. Garcia et al. introduce the Keyed User Datagram Protocol (KUDP) as a lightweight alternative to TCP for achieving "almost reliable" data transfer over UDP. The objective is to balance reliability with minimal overhead, particularly for scenarios like sensor networks and real-time multimedia. The methodology leverages segmentation of data across multiple ports using a "key," enabling basic loss and reordering detection, alongside a stream reconstruction algorithm. Simulation results under conditions of 16.67% packet loss and 20% out-of-sequence rates validate the feasibility of accurate stream

reordering. However, key challenges include NAT traversal, dynamic port key management, stream reassembly complexity, and ensuring secure operation [17].

H. Li et al. propose StateShield, a real-time, OS-agnostic defense system against information leakage in connectionless protocols (UDP, ICMP). The primary objective is to mitigate side-channel and covert-channel attacks while preserving legitimate traffic. The framework incorporates programmable switches with three attack-sensitive indicators and two novel defense components: Dynamic Address Mapper (DAM) and ICMP Reply Agent (IRA). Experimental evaluation demonstrates an average AUC of 0.952 and F1-score of 0.895 across more than ten attack types, with negligible packet loss for legitimate traffic. Challenges remain in countering stealthy attacks, handling spoofed IPs, and minimizing disruptions to normal communication [18].

Y.-R. et al. present ACLUID, an automatic, opportunistic, connectionless information dissemination scheme using Bluetooth Low Energy (BLE) legacy advertising for smartphones. The goal is to enable seamless background message exchange without user intervention. The methodology includes packet segmentation, encapsulation, decoding, and regrouping, supported by theoretical analysis of arrival times. Experiments on Android smartphones confirm high success rates, with most messages fully transferred within 120 seconds, even under heavy interference. Challenges include BLE's limited payload capacity, packet collisions, increased energy consumption during background operations, and degraded performance when devices simultaneously use other Bluetooth services [19].

Z. Li et al. explore connection-oriented and connectionless strategies for entanglement distribution in quantum networks, inspired by classical communication models. The methodology contrasts pre-established quantum paths (connection-oriented) with hop-by-hop entanglement swapping (connectionless). Simulations show that the connectionless strategy offers superior throughput, robustness, and concurrency, while the connection-oriented strategy provides greater reliability but at the cost of increased overhead. Numerical evaluations include success rates, memory requirements, and distribution delays. Key challenges involve designing efficient routing algorithms, managing quantum resources, scheduling requests, and addressing decoherence and imperfect operations in real-world quantum environments [20].

K. Kostas et al. critically examine the limitations of Individual Packet Features (IPF) in machine learning-based intrusion detection for IoT networks. The objective is to highlight risks such as information leakage and misleadingly high accuracy due to low data complexity. The study combines a comprehensive literature review with case study evaluations on datasets like IoT-NID. Results show IPF-based models achieving up to 100% accuracy during training but suffering drops of over 90% in cross-session testing, indicating poor generalization. The major challenges identified include dataset simplicity, lack of contextual features, vulnerability to information leakage, and the urgent need for more robust, context-aware features for reliable detection [21].

From this survey of recent advancements in connectionless data transmission, ML-driven network security, optimization strategies, and biological data modeling, a critical research gap emerges: the real-time, intelligent evaluation and estimation of packet behavior in connectionless environments. While prior works tackle reliability, security, or optimization in isolation, a unified ML-based framework specifically targeting packet-level metrics such as loss, jitter, and delay remains underdeveloped.

This paper addresses that gap by proposing a comprehensive ML-based solution for dynamic connectionless data networks. With the rise of IoT devices and streaming services, traditional connectionless protocols face severe reliability issues, as they lack mechanisms to guarantee packet delivery. Metrics such as loss, jitter, and delay become crucial for maintaining service quality, yet manual or rule-based monitoring often fails to adapt to complex, real-time conditions.

By leveraging Machine Learning, networks can automatically learn traffic patterns, accurately predict packet behavior, and optimize performance monitoring and resource allocation. The study pursues three primary objectives:

- Develop an ML-based model to evaluate packet behavior in connectionless transmissions using features such as jitter, delay, packet size, and loss rate.
- Estimate flow quality and reliability in real-time, enabling proactive fault detection and predictive network management.
- Compare ML algorithms (e.g., Random Forest, SVM, and others) to identify the most effective and scalable model, balancing accuracy with computational efficiency for real-world deployment in dynamic communication scenarios.

MACHINE LEARNING MODELS

The Random Forest algorithm is a robust machine learning technique that leverages ensemble learning by constructing multiple decision trees using random subsets of data and features. A simple classification model is illustrated in Figure 1. Each tree independently identifies patterns, and the final prediction is obtained through majority voting in classification tasks or averaging in regression tasks [22,23,24]. This approach effectively reduces overfitting, scales efficiently to large datasets, and can handle missing data automatically. Additionally, Random Forest provides feature importance rankings, aiding interpretability and feature selection. The inherent randomness in data sampling and feature selection generates diverse trees, which enhances prediction reliability. Suitable for both classification and regression problems, Random Forest does not require data normalization. Overall, it is highly accurate, versatile, and resilient, making it a widely adopted method for complex data analysis [25, 26].

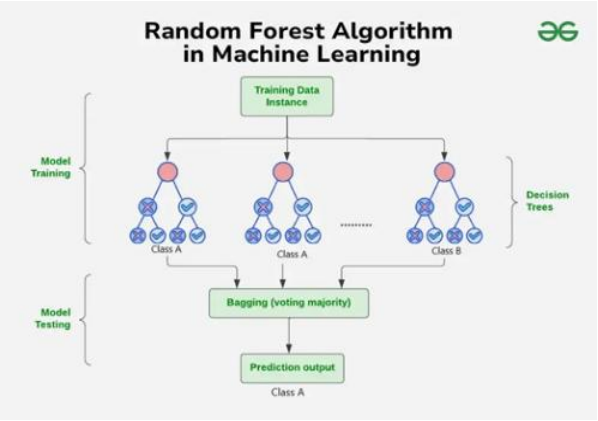

Figure 1: Classification of random forest.

Support Vector Machine (SVM) is a supervised machine learning algorithm primarily applied to classification tasks, though it can also be used for regression. Its core principle is to determine the optimal hyperplane that separates classes with the maximum margin, as illustrated in Figure 2. SVM relies on key concepts such as hyperplanes, support vectors, margins, and kernel functions (e.g., linear, polynomial, and radial basis function) to address both linearly and non-linearly separable data [27]. To enhance robustness, SVM incorporates a soft margin technique for handling outliers and employs hinge loss to effectively manage misclassifications [28].

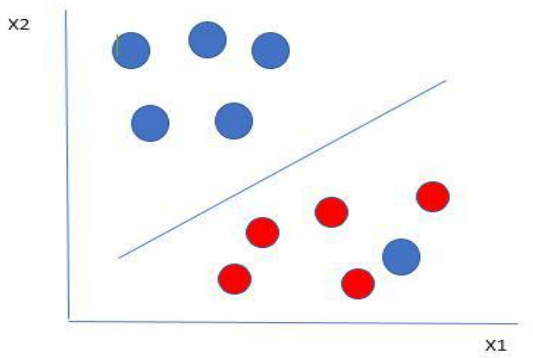

Figure 2: simple classified optimal hyperplane

The SVM workflow typically involves data import and preprocessing, splitting the dataset into training and testing subsets, selecting an appropriate kernel function, training the model, generating predictions, evaluating its performance, and finally visualizing the decision boundaries [29]. SVM demonstrates high efficiency in handling high-dimensional data, exhibits robustness against outliers, and supports both binary and multiclass classification. Moreover, it is memory-efficient, as it relies only on support vectors for computation. Owing to its strong generalization capabilities, SVM has been successfully applied in diverse domains such as text classification, image recognition, and anomaly detection [30–31].

In the context of communication networks, particularly within connectionless environments, several parameters are considered for performance evaluation. These include:

F1 score, The traditional F-measure or balanced F-score (F1 score) is the harmonic mean of precision and recall:

$$F1=\frac{2}{recall^{-1}+precision^{-1}}=2\frac{precision.recall}{precision+recall}=\frac{2\,TP}{2\,TP+FP+FN}$$

With precision = TP / (TP + FP) and recall = TP / (TP + FN), it follows that the numerator of F1 is the sum of their numerators and the denominator of F1 is the sum of

Prediction: A prediction a statement about a future event or about future data. Predictions are often, but not always, based upon experience or knowledge of forecasters. There is no universal agreement about the exact difference between "prediction" and "estimation"; different authors and disciplines ascribe different connotations.

Recall Recall in memory refers to the mental process of retrieving information from the past. Along with encoding and storage, it is one of the three core processes of memory. There are three main types of recall: free recall, cued recall and serial recall.

Precision and recall are then defined as:

$$Precision=\frac{tp}{tp+fp}$$

$$Recall=\frac{tp}{tp+fn}$$

Recall in this context is also referred to as the true positive rate or sensitive, and precision is also referred to as positive predictive value (PPV); other related measures used in classification include true negative rate and accuracy. True negative rate is also called specificity.

True Positive Rate (TPR) in Machine Learning, especially in classification problems, is defined as:

The proportion of actual positive cases that are correctly predicted as positive by the model.

It is also called Recall or Sensitivity. TP= True Positive, FN= False Negative.

$$TPR=\frac{TP}{(TP+FN)}$$

False Negative Rate is the proportion of actual positive instances that are incorrectly predicted as negative by a machine learning model. FN=False Negative, TP=True Positive

$$FNR=\frac{FN}{(FN+TP)}$$

False Positive Rate (FPR) refers to the proportion of negative instances that are incorrectly classified as positive by a model. FP= False Positive; TN= True Negative;

$$FPR=\frac{FP}{(FP+TN)}$$

True Negative Rate (TNR), also known as Specificity, is the proportion of actual negative cases that are correctly identified by a classification model. TN= True Negative; FN= False Negative; TP= True Positive

FP= False Positive.

$$TNR = \frac{TN}{(TN+FP)}$$

III.PROPOSED METHODOLOGY

In this paper, we present a Machine Learning-based approach for the evaluation and estimation of data packets in connectionless networks. Using synthetically generated data, we implement and test algorithms in Python to analyze packet behavior, predict transmission outcomes, and improve network reliability without relying on continuous connections.

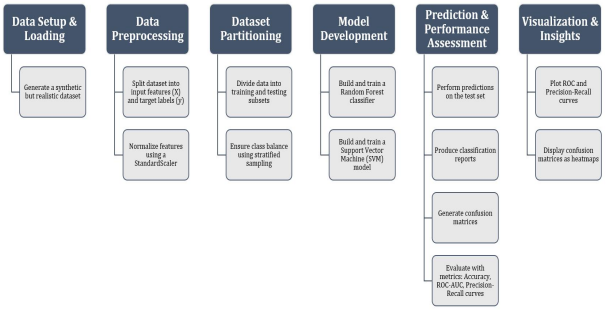

Figure 3: Flow diagram of Proposed System

The workflow presented in the diagram outlines a structured pipeline for developing, training, and evaluating machine learning models, particularly in classification tasks. It is divided into six major stages, each serving a crucial role in ensuring that the overall process remains systematic, efficient, and reliable.

The first stage, Data Setup & Loading, emphasizes the foundation of any machine learning project: the dataset. In this case, a synthetic but realistic dataset is generated to mimic real-world conditions while still maintaining control over distribution and complexity. This step ensures that the later stages of model development are built upon data that is representative, consistent, and free from uncontrolled biases.

The second stage, Data Preprocessing, focuses on preparing the dataset for modeling. Raw data is often noisy, unstructured, or inconsistent, which can negatively affect model performance. Here, the dataset is divided into input features (X) and target labels (y), a crucial distinction that separates predictive attributes from the outcomes to be learned. Additionally, normalization using a StandardScaler is applied to ensure that all features are on a comparable scale. This is particularly important for algorithms like SVM, which are sensitive to feature scaling, and prevents any single variable from disproportionately influencing the model.

The third stage, Dataset Partitioning, addresses the problem of model generalization. By splitting the dataset into training and testing subsets, the workflow ensures that the model learns patterns from one portion of the data and is validated on unseen examples. Importantly, stratified sampling is used to maintain class balance, preventing skewed distributions that could lead to biased predictions, especially in cases of imbalanced datasets.

Moving into Model Development, two well-established machine learning techniques are applied: Random Forest and Support Vector Machine (SVM). The Random Forest classifier leverages ensemble learning by combining multiple decision trees, which enhances robustness, reduces overfitting, and captures non-linear patterns. In contrast, the SVM model excels in identifying decision boundaries in high-dimensional spaces, making it suitable for complex classification problems. Employing both models allows for comparative analysis and leverages their complementary strengths.

The fifth stage, Prediction & Performance Assessment, involves testing the trained models on the reserved test set. Predictions are generated, and classification reports are produced to summarize performance across metrics such as precision, recall, and F1-score. Confusion matrices provide a detailed view of correct versus incorrect classifications, while broader evaluation is achieved using metrics like accuracy, ROC-AUC, and Precision-Recall curves. These diverse metrics ensure that the model is not only accurate but also reliable across different aspects of performance, especially when handling imbalanced datasets.

Finally, Visualization & Insights transforms numerical results into intuitive visual representations. ROC and Precision-Recall curves help in assessing trade-offs between sensitivity and specificity, while heatmaps of confusion matrices reveal misclassification patterns. These visualizations provide decision-makers and researchers with actionable insights, making the results both interpretable and communicable.

Overall, this workflow encapsulates a comprehensive, step-by-step methodology for developing and validating machine learning models. From synthetic data generation to advanced evaluation and visualization, it integrates best practices that ensure reproducibility, transparency, and accuracy. Such a pipeline is highly adaptable across domains like healthcare, finance, and engineering, where predictive modeling plays a vital role in decision-making.

IV.RESULT AND DISCUSSION

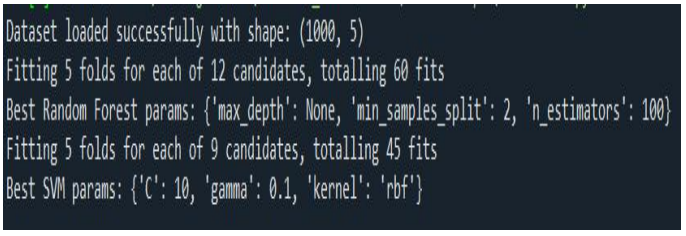

Figure 4: Screenshot of obtained result

This output in figure 4 shows that the dataset, consisting of 1,000 samples and 5 features, was successfully loaded. Hyperparameter optimization using 5-fold cross-validation was performed for both Random Forest and SVM models. The Random Forest's best parameters found were: no maximum depth (None), minimum samples per split set to 2, and 100 estimators (trees). For SVM, the optimal parameters were: regularization parameter C=10, kernel coefficient gamma=0.1, and the RBF kernel type. These settings are expected to maximize each model's predictive performance on the classification task based on the search performed.

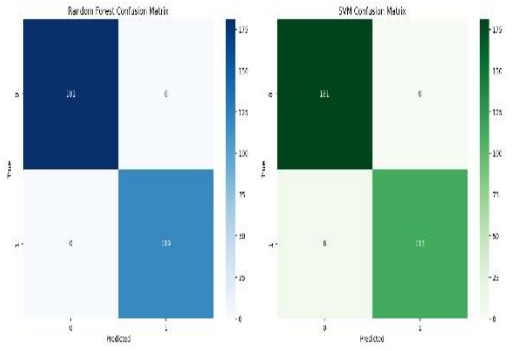

Figure 5: Confusion Matrix

This image presents confusion matrices for Random Forest and SVM models applied to a binary classification problem. The Random Forest matrix reveals flawless performance, correctly identifying all normal (181) and anomalous (119) instances without any misclassifications—demonstrating 100% accuracy, precision, and recall. In contrast, the SVM matrix—while still robust—shows slightly lower accuracy: all 181 normal samples are classified correctly, but 6 anomalies are incorrectly classified as normal (false negatives), leading to 113 true positives. Both models excel at recognizing normal data, yet Random Forest offers perfect discrimination between classes, superior for applications where missing anomalies carry significant risk. The color gradients visually emphasize the high concentration of correct predictions and the sparsity of errors. In real-world contexts, this analysis indicates that while both algorithms are highly effective for anomaly detection, Random Forest provides unmatched reliability, ensuring that every anomaly is detected and minimizing the potential for undetected threats.

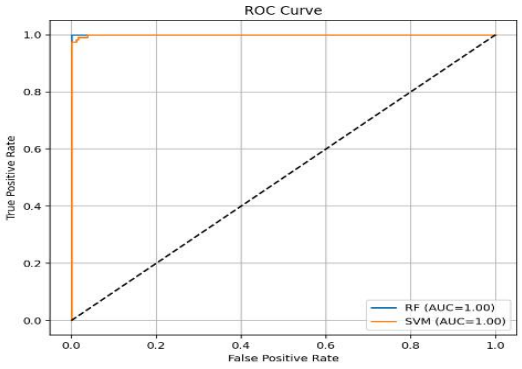

Figure 6: ROC curve

The ROC curve in figure 6 provides a visual comparison of the classification performance for Random Forest and SVM models in distinguishing between normal and anomalous network packets. Both models exhibit exceptional discrimination, as indicated by their curves nearly touching the top left corner of the plot and recorded AUC values of 1.00—signifying perfect separation of classes. The steep ascent of both curves at low false positive rates demonstrates that the models are very effective at correctly identifying true positives while minimizing false alarms. The diagonal dashed line represents the performance of a random classifier; the significant distance from this line further underscores the reliability and accuracy of these models. In practical terms, such ROC curves indicate that both Random Forest and SVM are highly capable for anomaly detection tasks, each offering extremely high sensitivity and specificity. However, this level of performance may suggest that the dataset or task is well-suited to these approaches, with minimal overlap between classes.

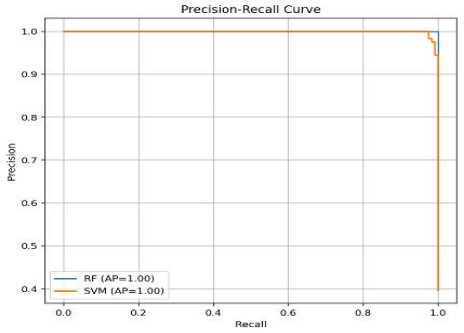

Figure 7: PR curve

The Precision-Recall Curve shown in figure 7, highlights the outstanding ability of both Random Forest (RF) and SVM models to accurately identify anomalies in the dataset. With average precision (AP) scores of 1.00 for both classifiers, the models maintain extremely high precision and recall across almost all thresholds, meaning they correctly identify nearly all actual anomalies while minimizing false positives. The curves for both RF and SVM remain close to the top-right corner, a region that reflects near-perfect classification. Such an outcome is especially significant in scenarios with imbalanced datasets, where the minority class—typically anomalies—is of greatest concern. This result implies that either model is a reliable tool for anomaly detection, providing confidence that few threats or unusual patterns will be missed. The near-identical performance further confirms the robustness of both classifiers, even though Random Forest may have slightly better recall as seen in the confusion matrix. This curve visually affirms their suitability for high-stakes detection tasks.

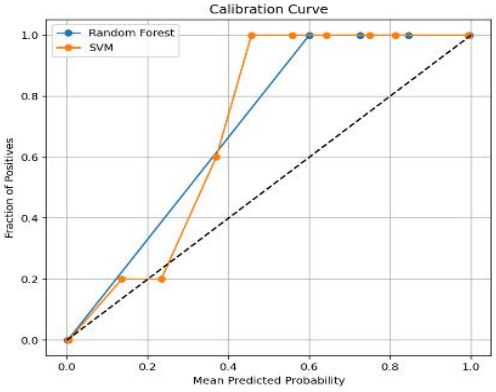

Figure 8: Calibration curve

The calibration curve in figure 8 compares how well the predicted probabilities from Random Forest and SVM models align with actual observed outcomes. The dotted line represents perfect calibration, where predicted probabilities exactly match the true fraction of positives. Here, Random Forest's curve closely follows this line, indicating its confidence scores can be reliably interpreted as true probabilities. The SVM line deviates slightly, showing occasional overconfident predictions but remains generally well-calibrated. This means both models not

only classify anomalies effectively but also provide probability outputs that accurately reflect real-world likelihoods—a critical advantage for risk-based decision making in anomaly detection tasks.

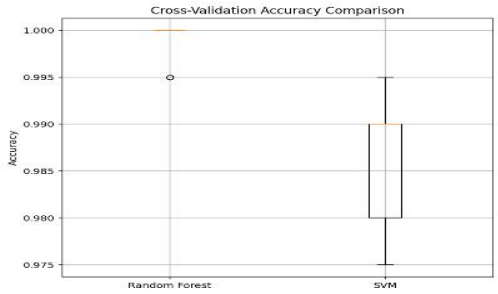

Figure 9: cross-validation accuracy comparison

The cross-validation accuracy comparison boxplot illustrates the consistency and reliability of the Random Forest and SVM models. Random Forest demonstrates nearly flawless performance, with all folds yielding close to perfect accuracy and minimal spread—indicating the model is both robust and stable across various data splits. On the other hand, SVM shows more variability in accuracy, reflected by a wider box and whiskers, suggesting its predictions can fluctuate depending on the subset of data used for training and validation. Both models perform well overall, but Random Forest's higher and more consistent accuracy makes it a preferable choice for dependable anomaly detection.

```
= Random Forest Classification Report =
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       181
           1       1.00      1.00      1.00       119

    accuracy                           1.00       300
   macro avg       1.00      1.00      1.00       300
weighted avg       1.00      1.00      1.00       300

= SVM Classification Report =
              precision    recall  f1-score   support

           0       0.97      1.00      0.98       181
           1       1.00      0.95      0.97       119

    accuracy                           0.98       300
   macro avg       0.98      0.97      0.98       300
weighted avg       0.98      0.98      0.98       300

Random Forest Accuracy: 100.00%
SVM Accuracy: 98.00%
```

Figure 10: Screen shot of accuracy results

The classification reports and accuracy scores provide a detailed comparison of Random Forest and SVM models on network anomaly detection. The Random Forest model achieves perfect performance across all metrics—precision, recall, and f1-score—for both the normal (class 0) and anomaly (class 1) categories, resulting in an overall accuracy of 100%. This indicates that the Random Forest is capable of correctly identifying every instance in the test set, both true normals and true anomalies, without any misclassifications. Such results highlight Random Forest's exceptional discriminative power and reliability, making it ideal for tasks where missing anomalies could be critical.

In contrast, the SVM model, while still highly effective, shows slightly lower precision and recall for the anomaly class (1). Specifically, its recall for class 1 is 0.95, meaning it misses 5% of actual anomalies. The precision and f1-score are also marginally lower (0.97 for precision on class 0 and 0.95 for recall on class 1), leading to a total accuracy of 98%. These results suggest SVM is very robust but not perfect, as a few anomalies may be incorrectly labeled as normal. Overall, both classifiers are strong performers, but Random Forest provides unmatched reliability and accuracy for anomaly detection in this scenario.

V.CONCLUSION

This paper successfully demonstrates the application of machine learning models, particularly Random Forest and Support Vector Machines (SVM), for effective packet behavior prediction in connectionless data communication networks. By analyzing key network parameters such as packet size, delay, jitter, and packet loss rate, the models deliver highly accurate classifications of network behavior. Random Forest outperformed SVM, achieving perfect scores in precision, recall, F1-score, and accuracy, while SVM also maintained strong performance with 98% accuracy. The robust evaluation through ROC, precision-recall curves, and calibration plots confirms the reliability and discriminative power of these models for anomaly detection in volatile data streams. This research highlights the potential of machine learning-driven network analysis to enhance network reliability, resource management, and security in dynamic environments, surpassing traditional monitoring methods. Future work may explore deep learning techniques, adaptive retraining, real-world deployment, and expansion to multi-class anomaly detection scenarios.

Challenges faced include handling noisy and imbalanced datasets, feature selection complexity, model overfitting, real-time processing constraints, and scalability in high-speed networks. Applications of this work include network fault diagnosis, intrusion detection, quality of service optimization, and predictive maintenance. Advantages include high accuracy, improved anomaly detection, scalability, and adaptability to changing network conditions. Future feature enhancements could involve incorporating temporal and spatial network features, employing ensemble methods, and integrating domain-specific knowledge to boost prediction robustness and interpretability.

VI.REFERENCES

[1] S. C. F. Chan, K. M. Chan, K. Liu and J. Y. B. Lee, "On Queue Length and Link Buffer Size Estimation in 3G/4G Mobile Data Networks," in IEEE Transactions on Mobile Computing, vol. 13, no. 6, pp. 1298-1311, June 2014, doi: 10.1109/TMC.2013.127.

[2] P. Iovanna, A. Germoni, F. Testa, G. Cossu, V. López and R. Sabella, "Multilayer control for packet-optical networks [invited]," in Journal of Optical Communications and Networking, vol. 5, no. 10, pp. A86-A99, Oct. 2013, doi: 10.1364/JOCN.5.000A86.

[3] Y. M. Saputra and Hendrawan, "The effect of packet loss and delay jitter on the video streaming performance using H.264/MPEG-4 Scalable Video Coding," 2016 10th International Conference on Telecommunication Systems Services and Applications (TSSA), Denpasar, Indonesia, 2016, pp. 1-6, doi: 10.1109/TSSA.2016.7871094.

[4] U. R. Seshasayee and M. Rathinam, "A finite jitter buffer model for time division multiplexing over packet networks," 2017 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), Bhubaneswar, India, 2017, pp. 1-4, doi: 10.1109/ANTS.2017.8384162.

[5] M. S. Thomas and I. Ali, "Reliable, Fast, and Deterministic Substation Communication Network Architecture and its Performance Simulation," in IEEE Transactions on Power Delivery, vol. 25, no. 4, pp. 2364-2370, Oct. 2010, doi: 10.1109/TPWRD.2010.2042824.

[6] N. Kitsuwan, S. McGettrick, F. Slyne, D. B. Payne and M. Ruffini, "Independent transient plane design for protection in OpenFlow-based networks," in Journal of Optical Communications and Networking, vol. 7, no. 4, pp. 264-275, April 2015, doi: 10.1364/JOCN.7.000264.

[7] P. Wu, L. Jiang, L. Wang, J. Xu and X. Wang, "Event-Triggered State Estimation for Wireless Sensor Network Systems With Packet Losses and Correlated Noises," in IEEE Access, vol. 8, pp. 216762-216771, 2020, doi: 10.1109/ACCESS.2020.3041596.

[8] N. M. Garcia, F. Gil, B. Matos, C. Yahaya, N. Pombo and R. I. Goleva, "Keyed User Datagram Protocol: Concepts and Operation of an Almost Reliable Connectionless Transport Protocol," in IEEE Access, vol. 7, pp. 18951-18963, 2019, doi: 10.1109/ACCESS.2018.2886707.

[9] E. Liri, P. K. Singh, A. B. Rabiah, K. Kar, K. Makhijani and K. K. Ramakrishnan, "Robustness of IoT Application Protocols to Network Impairments," 2018 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN), Washington, DC, USA, 2018, pp. 97-103, doi: 10.1109/LANMAN.2018.8475048.

[10] M. Andrews, S. Antonakopoulos and L. Zhang, "Rate-Adaptive Scheduling Policies for Network Stability and Energy Efficiency," in IEEE/ACM Transactions on Networking, vol. 23, no. 6, pp. 1755-1764, Dec. 2015, doi: 10.1109/TNET.2014.2346507.

[11] Z. Ning, D. Zhang, K. Xie, Y. Li and D. Liu, "Network coding overhearing management policies based on data packet switching and sorting algorithm," 2014 International Conference on Smart Computing Workshops, Hong Kong, China, 2014, pp. 50-56, doi: 10.1109/SMARTCOMP-W.2014.7046667.

[12] T. Suga, K. Okada and H. Esaki, "Toward Real-time Packet Classification for Preventing Malicious Traffic by Machine Learning," 2019 22nd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN), Paris, France, 2019, pp. 106-111, doi: 10.1109/ICIN.2019.8685893.

[13] N. Das and M. Chakraborty, "Machine Learning-Driven Threshold Optimization for Wavelet Packet Denoising in EEG-Based Mental State Classification," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-7, doi: 10.1109/ICCCNT56998.2023.10307020.

[14] L. Zhao et al., "Overtaking Feasibility Prediction for Mixed Connected and Connectionless Vehicles," in IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 10, pp. 15065-15080, Oct. 2024, doi: 10.1109/TITS.2024.3398602.

[15] H. Si, C. Sun, B. Chen, L. Shi and H. Qiao, "Analysis of Socket Communication Technology Based on Machine Learning Algorithms Under TCP/IP Protocol in Network Virtual Laboratory System," in IEEE Access, vol. 7, pp. 80453-80464, 2019, doi: 10.1109/ACCESS.2019.2923052.

[16] T. Xu, L. Ou-Yang, X. Hu and X. -F. Zhang, "Identifying Gene Network Rewiring by Integrating Gene Expression and Gene Network Data," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 15, no. 6, pp. 2079-2085, 1 Nov.-Dec. 2018, doi: 10.1109/TCBB.2018.2809603.

[17] N. M. Garcia, F. Gil, B. Matos, C. Yahaya, N. Pombo and R. I. Goleva, "Keyed User Datagram Protocol: Concepts and Operation of an Almost Reliable Connectionless Transport Protocol," in IEEE Access, vol. 7, pp. 18951-18963, 2019, doi: 10.1109/ACCESS.2018.2886707.

[18] H. Li, Q. Li, Y. Su, X. Feng, C. Fu and K. Xu, "StateShield: Real-Time Defenses Against Information Leakage Over Connectionless Protocols," in IEEE Transactions on Networking, doi: 10.1109/TON.2025.3543734.

[19] Y. -R. Tsai and Y. -C. Chen, "Opportunistic Connectionless Undirected Information Dissemination Based on Bluetooth Low Energy Advertising Technology on Smartphones," in IEEE Access, vol. 9, pp. 155851-155860, 2021, doi: 10.1109/ACCESS.2021.3129251.

[20] Z. Li, K. Xue, J. Li, N. Yu, D. S. L. Wei and R. Li, "Connection-Oriented and Connectionless Remote Entanglement Distribution Strategies in Quantum Networks," in IEEE Network, vol. 36, no. 6, pp. 150-156, November/December 2022, doi: 10.1109/MNET.107.2100483.

[21] K. Kostas, M. Just and M. A. Lones, "Individual Packet Features are a Risk to Model Generalization in ML-Based Intrusion Detection," in IEEE Networking Letters, vol. 7, no. 1, pp. 66-70, March 2025, doi: 10.1109/LNET.2025.3525901.

[22] Z. Huang et al., "An Algorithm of Forest Age Estimation Based on the Forest Disturbance and Recovery Detection," in IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-18, 2023, Art no. 4409018, doi: 10.1109/TGRS.2023.3322163.

[23] L. Dong et al., "Very High Resolution Remote Sensing Imagery Classification Using a Fusion of Random Forest and Deep Learning Technique—Subtropical Area for Example," in IEEE Journal of Selected Topics in Applied

Earth Observations and Remote Sensing, vol. 13, pp. 113-128, 2020, doi: 10.1109/JSTARS.2019.2953234.

[24] W. Deng, Y. Guo, J. Liu, Y. Li, D. Liu and L. Zhu, "A missing power data filling method based on improved random forest algorithm," in Chinese Journal of Electrical Engineering, vol. 5, no. 4, pp. 33-39, Dec. 2019, doi: 10.23919/CJEE.2019.000025.

[25] F. Tong and Y. Zhang, "Exploiting Spectral–Spatial Information Using Deep Random Forest for Hyperspectral Imagery Classification," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 5509505, doi: 10.1109/LGRS.2021.3112198.