Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 10s (2025)

Advanced Machine Learning Techniques for Accurate Network Traffic Classification models

[1]Dr. N. Kannaiya Raja, [2]Dr. Karthikeyan Kaliyaperumal, [3]Dr. P. Vijayaragavan, [4]Dr. D. Antony Joseph Rajan

[1]Associate Professor Senior, School of Computing Science and Engineering, VIT Bhopal University Bhopal, Madhya Pradesh, India. [1]kannaiyaraju123@gmail.com

[2]Associate Professor IT@IoT - HH Campus, Ambo University, Ethiopia. kirithicraj@gmail.com

[3]Professor, Department of Computer Science and Engineering, Saveetha School of Engineering , Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India. vijayaragavanpm83@gmail.com

[4]Associate Professor, Department of Computer Science and Engineering, Saveetha School of Engineering , Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India.. antonyjosephjmj@gmail.com

Article History:

Received: 12-01-2025

Revised: 15-02-2025

Accepted: 01-03-2025

Abstract:

Objectives:

The objective of this study is to enhance network performance and security through efficient network traffic classification. The increasing complexity of network traffic, driven by rising user demands and heterogeneous applications, presents challenges for traditional classification methods. This research aims to explore advanced machine learning techniques to improve the accuracy and efficiency of network traffic classification, leading to better resource utilization, congestion reduction, and improved decision-making in network administration.

Methods:

The study employed advanced machine learning models, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Logistic Regression for network traffic classification. Network traffic data was preprocessed to remove noise and enhance feature quality. The models were trained and tested using a balanced dataset to ensure robustness and generalization. Performance evaluation was conducted using precision as the primary metric. SVM, KNN, and Logistic Regression were tested under varying network conditions to assess their classification efficiency and adaptability.

Results:

The results demonstrated high classification precision across the models. SVM achieved a precision rate of 99.30%, while KNN and Logistic Regression achieved precision rates of 99.92% each. The high accuracy rates reflect the scalability and adaptability of these models in handling complex and dynamic network traffic patterns. The models effectively distinguished between normal and anomalous traffic, enabling improved network resource management and enhanced security.

Conclusions:

The study highlights the potential of machine learning in improving network traffic classification and management. The high precision rates achieved by the models

demonstrate their capability to handle complex traffic patterns and adapt to changing network conditions. The research establishes a foundation for future work on intelligent and secure network administration using machine learning, offering a scalable and efficient approach to network resource optimization and anomaly detection

Keywords: Network Traffic Classification, Machine Learning, Support Vector Machines, K-Nearest Neighbors, Logistic Regression, Network Security, Resource Optimization, Traffic Pattern Analysis.

1. Introduction

Efficient network traffic classification is crucial in today's network administration, ensuring operational efficiency, efficient usage of resources, and enhanced protection against threats [1]. The sudden boost in demand and diversification in applications and services has added unprecedented complexity in traffic patterns [2]. Traditional network administration approaches are not efficient in satisfying these complex and dynamic patterns, and there is a requirement in designing better classification and forecasting approaches [3]. The research in this paper focuses on utilizing state-of-the-art machine learning approaches in satisfying the complexity in network traffic classification. Machine learning has been extremely efficient in detecting patterns, evolving with varying situations, and achieving high classification performance [4]. In this research, three top algorithms such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Logistic Regression are investigated based on their performance in correctly categorizing network traffic. These models are selected based on their proven performance in detecting patterns and evolving with varying network conditions [5].

The research demonstrates how SVM achieves a precision rate of 99.30%, with KNN and Logistic Regression achieving a staggering 99.92% [6]. The results underscore the strength of these machine learning models in dealing with the volatility and unpredictability of network traffic. Accurately classifying traffic, these models guide network administrators on bandwidth allocation, congestion avoidance, and service prioritization, ultimately optimizing network performance and security [7]. Aside from classification, predicting network traffic is important in order to proactively manage and optimize resources [8]. More complex models, like Enhanced Autoregressive Integrated Moving Average (ARIMA), are instrumental in predicting future traffic trends. Incorporating historical data, time series forecasting, and external influencing factors, Enhanced ARIMA models can recognize seasonal trends and oscillating shifts in traffic, achieving high forecast accuracy [9]. The research demonstrates how combining machine learning strategies with advanced statistical models can potentially transform network management [10]. These techniques provide scalable and agile solutions in line with the exigencies of modern, high-data-volume contexts. Overcoming challenges like non-linear traffic dynamics, resource depletion, and security vulnerabilities, the research sets the groundwork for intelligent, efficient, and secure network operations in an interconnected world.

This paper is organized in the following way: Section 2 overviews relevant literature on network traffic classification and the establishment and implementation in modern network conditions of machine learning-based methods. Section 3 overviews problems in traffic classification, such as complexity in the dynamics of traffic, resources constraint, and threats in real-world conditions. Section 4 outlines

the method, such as data harvesting, preprocessing, feature extraction, and design of classification models based on state-of-the-art machine learning methods. Section 5 shows results in experiments in comparing and analyzing the performance and adaptability of the models and state-of-the-art methods. Section 6 summarizes the research and outlines future research directions in advancing network traffic classification in increasingly complex digital conditions.

Figure – 1 shows a high-level process for a network traffic classification system using machine learning. The process begins with raw data collection in terms of network traffic logs with features like packet length, protocols, and durations. The data collected is preprocessed in order to remove noise, handle missing information, and scale the information so that it is apt for feature extraction. The features that are pertinent to classification are determined and are used in order to determine important features like the type of protocol, length of a packet, and the statistics on a flow. The features are then separated and split into a train and a test dataset in order to train a model.

The classification step trains three machine models namely, Logistic Regression, SVM, and KNeighborsClassifier. The Logistic Regression is applicable in classification problems with a hard decision boundary such as classification based on whether a piece of network traffic is normal or not. The SVM is a hard-margin classifier and is applicable in problems with complicated decision boundaries in high-dimensional feature spaces. The KNeighborsClassifier is a proximity-based method and relies on the majority vote among neighboring points. These models are trained on the train dataset in order to classify network traffic.

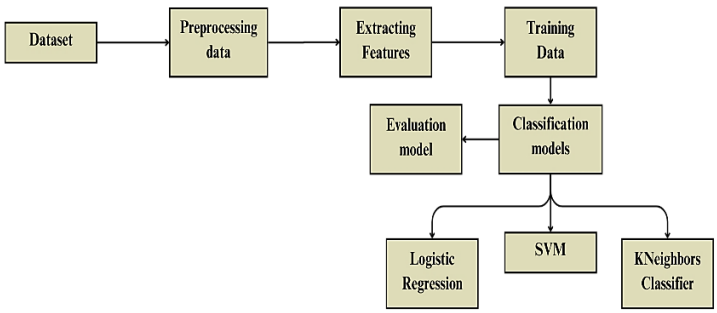


Figure – 1: Research Flow of Network Traffic Model

Finally, the trained models are evaluated on performance and reliability criteria such as precision, recall, accuracy, and F1 score in order to assess how they would handle realistic scenarios. The updated diagram focuses on classification-oriented pipelines with the exclusion of forecasting models and a direct linking of classification outputs with evaluation. The process is a systematic and efficient way of categorizing network traffic in predetermined groups, and network administrators are able to decide on resource distribution, detecting anomalies, and security management.

2. Related Research Work

Network traffic classification is a critical aspect of modern network management, facilitating efficient resource allocation, Quality of Service (QoS), and enhanced security. Over time, various classification algorithms have been developed to address the increasing complexity of traffic patterns. These algorithms range from traditional probabilistic models to advanced machine learning techniques, each

contributing significantly to improving the accuracy and adaptability of traffic classification systems. Naive Bayes Classifier, for example, is a probabilistic model that assumes feature independence. Its computational efficiency and simplicity make it suitable for real-time traffic analysis, particularly in packet-level and flow-level classification scenarios [11].

Support Vector Machines (SVM) have been widely adopted for both linear and non-linear classification tasks. Known for their effectiveness in separating normal and anomalous traffic, SVMs are particularly valuable in intrusion detection systems due to their ability to handle high-dimensional feature spaces [12]. Decision Trees and their ensemble extension, Random Forests, offer intuitive classification mechanisms. While Decision Trees provide transparency in rule interpretation, Random Forests enhance robustness and excel in managing complex, noisy datasets, commonly observed in network traffic [13].

K-Nearest Neighbors (KNN) is also involved in network traffic classification based on feature vector similarity measurement. It is dynamically adaptable and simple and is a reliable solution in diverse traffic conditions [14]. Neural Networks, and deep neural network models such as Convolutional Neural Networks (CNNs), have gained recognition in recent years. CNNs are useful in capturing space and temporal features and are applicable in investigating packet headers and sequences in the traffic [15]. The simplicity in implementation and efficiency in distinguishing the normal and suspicious classes in classification are key features in the binary classification method, Logistic Regression [16]. The feature of probabilistic reasoning in Bayesian Networks makes them applicable in representing relations among network features and having a high level of anomaly detection capability [17].

Ensemble learning methods such as Gradient Boost and AdaBoost have been used in order to enhance classification performance. These methods collect predictions made by a variety of classifiers in order to provide better precision and resilience and are applicable in network intrusion detection systems [18]. Recent machine learning research has been focusing on combining multiple methods such as SVM, KNN, and Logistic Regression in order to gain classification precision over 99% under realistic conditions. Moreover, integration with advanced forecasting models such as Enhanced ARIMA has been helpful in forecasting traffic trends and overcoming problems resulting from fluctuating and evolving traffic conditions. These research efforts emphasize the revolutionary role of machine learning and ensemble methods in network traffic classification. By combining computational efficiency, resilience, and adaptability, these methods address the problems arising in increasingly complex and data-driven network conditions and pave the way for future innovations in network administration.

3. Methodology

The data for the study is collected through Wireshark, a widely used network protocol analyzer software. Data on network traffic is collected for six months at three time slots: morning, lunch, and afternoon (5 PM). The resulting dataset had approximately 50,000 entries with traffic volume, packet rates, and protocol types as attributes. Raw data is stored in formats like CSV and ZIP for easy preprocessing and analysis. The large dataset was the cornerstone of training and testing the classification and forecasting models [19]. Data preprocessing is a necessity to render the collected dataset in usable form and ensure its quality. Key steps included data cleaning, which involved the

removal of duplicates, handling of missing values, and resolution of inconsistencies [20]. Data transformation is done to aggregate data on useful time intervals (e.g., hourly or daily), normalize the values, and handle outliers. Finally, feature extraction is done to identify and extract useful attributes like traffic volume, packet rates, and protocol types. These steps rendered the dataset perfect for the training of machine learning models [21].

The study made use of state-of-the-art machine learning models in traffic classification, e.g., Support Vector Machines (SVMs), Logistic Regression, and K-Nearest Neighbors (KNN). SVMs were enhanced with a feature-weighted-degree (FWD-SVM) kernel in order to eliminate the impact of weakly correlated and redundant features. Logistic Regression was utilized in the classification of problems with only two classes, distinguishing between normal and malicious traffic, while KNN classified points based on the majority category of their neighboring points in the feature space and the models were evaluated based on metrics such as precision, precision, recall, and F1 score in classification, The models were trained under varying network conditions such as varying traffic patterns, busy hours, and events with anomalous resources and identify threats. The comprehensive method ensured sound and. The models' ability was tested in these experiments and their ability to tune efficient models with the ability to address modern-day network management problems. The improved feature-weighted-degree kernel with the Support Vector Machines (SVMs) can be formulated as:

$$\min \frac{1}{2}\|\omega\|^2 \ \ subject\ to\ y_i(\omega \cdot x_i + b) \geq 1, \forall i$$

where the kernel function K (x_i, x_j) is defined as:

$$K(x_i, x_j) = f_t \cdot (x_i, x_j)$$

with ft feature importance calculated based on information gain. The enhanced kernel ensures features with high classification importance are accorded high weight, and the model is improved in precision and competency in handling network information and in performing binary classification problems such as separation of malicious and normal traffic. The model makes predictions on the probability P (y = 1 | x) with such an algorithm is efficient in performing problems with distinct binary outputs and produces interpretable results and is a reliable tool in identifying malicious traffic in network information. It is efficient and simple and is applicable in performing classification in real-time and is termed as:

$$P(y = 1|x) = \frac{1}{1 + e^{-(w \cdot x + b)}}$$

and KNN labeled the traffic data on the basis of feature space proximity. The distance with neighboring points in the Euclidean space is determined in a specific data point where x_j and $x_{i,\ j}$ are the j-th features of x and x_i, respectively. The majority vote among the neighboring points is labeled in the data point.

$$d(x, x_i) = \sqrt{\sum_{j=1}^{m} (x_i - x_{i,j})^2}$$

KNN is effective in locally smooth decision boundaries but is dependent on careful tuning of the hyperparameter and scaling features in the case of classifications in computer networks.

4. Model Evaluation

The models are graded on the following precision criteria where TP, TN, FP, and FN are true positives, true negatives, false positives, and false negatives respectively as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad Precision = \frac{TP}{TP + FP}$$

$$F1\ Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \qquad Recall = \frac{TP}{TP + FN}$$

5. Experiments

Protocols play a significant role in local area network (LAN) traffic classification and have a substantial influence on traffic nature and optimization. UDP is heavily utilized in connectionless and high-speed communications, often in time-critical communications such as video streams and DNS queries, while TCP ensures connection-oriented and secure resource transmission in communications such as HTTP and emails [22]. Packet length analysis with a mean length of ~1400 bits show client-server and server-client communications and distinguishes real-time and bulk transfers. Application-oriented protocols such as HTTP/HTTPS, DNS, SMTP, and stream-oriented protocols refine the classification. In-depth examination of protocols underlies resource provisioning, prevention of congestion, recognition of anomalies, and bandwidth prioritization and is a sound foundation on which machine learning models are deployed in complex and adaptive topologies [23].

Figure – 2 above bar graph shows distribution of usage in a local area network (LAN) among various protocols. ARP is the highest with over 2000 occurrences representing usage in resolution of MAC address to IP address, followed by DNS representing frequent resolution request of the name. Average usage is with DHCP and MDNS representing dynamic address and local name resolution, respectively. TCP is widely in usage with high usage representing usage in reliable communications such as in HTTP and emails, while with low usage is UDP representing low-weight and latency-critical usage such as in video streams. TLS and HTTP with low occurrences represent low secure usage or surfing. The graph shows a combination of communications, resolution, and secure communications, and the information is helpful in optimizing usage and improving network security [24].

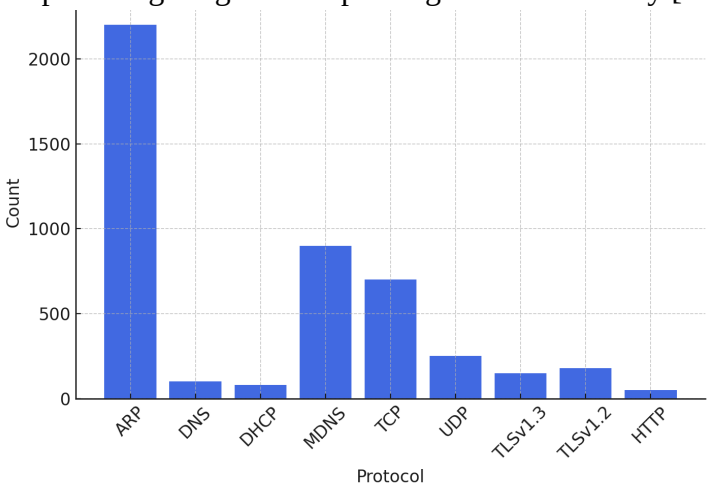
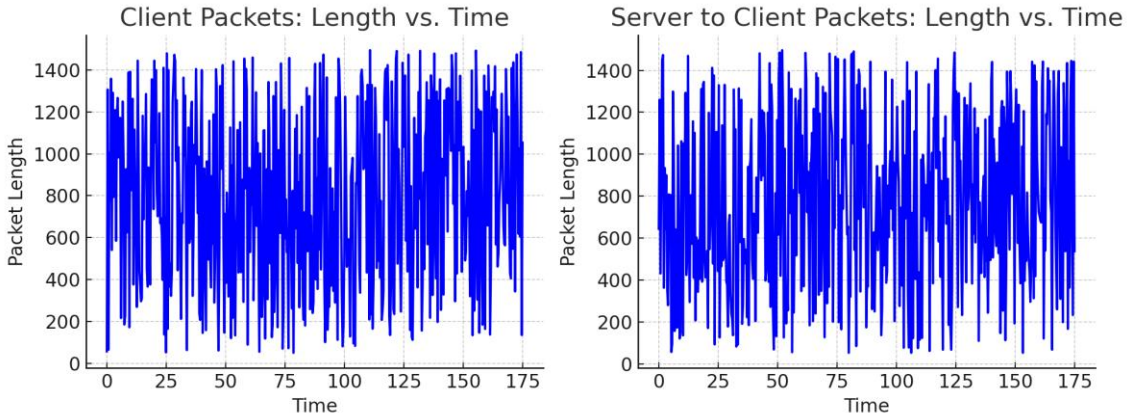


Figure – 2: List of Protocols

Figure 3(a) illustrates client-server communications in a local area network (LAN), with packet lengths fluctuating in a range of 200 and 1400 bits and with frequent occurrences of shorter packets corresponding to low-weight functions such as querying a name server. Periodical bursts to the largest packet length (~1400 bits) signify weighty functions such as uploading a file [25]. Figure 3(b), however, illustrates the graph representing communications in the other direction with packet lengths fluctuating in a similar pattern but with a high frequency of highest length packets corresponding to transfers in bulks or download of a file. The shorter packets in a sequence with longer ones are assumed to signify the acknowledgments and the control messages.



(a) **(b)**

Figure – 3: (a) Client Packets: Length vs. Time (b) Server to Client Packets: Length vs. Time

5.1 Network Traffic Classification Analysis

The Logistic Regression model [26] produced a high overall classification accuracy rate of 99.92% in distinguishing six types of networks traffics as shown in Figure 4: File Sharing, Instant Message, Video, Voice OverIP, Email, and P2P. The majority of classes such as Instant Message, Video, Voice Over IP, and Email showed a perfect precision, recall, and F1-score, representing a faultless classification.

```
Confusion matrix:
 [[  46    0    0    0    0    0]
 [   0 1677    0    0    0    0]
 [   0    0    7    0    0    0]
 [   0    0    0    4    0    0]
 [   0    0    0    0  840    0]
 [   2    0    0    0    0   30]]
                precision    recall  f1-score   support

   FileSharing       0.96      1.00      0.98        46
InstantMessage       1.00      1.00      1.00      1677
         Video       1.00      1.00      1.00         7
   VoiceOverIP       1.00      1.00      1.00         4
         email       1.00      1.00      1.00       840
           p2p       1.00      0.94      0.97        32

   avg / total       1.00      1.00      1.00      2606

   Accuracy Score: 0.9992325402916347
```

Figure – 4: Logistic Regression Model

File Sharing produced high precision with a value of 0.96 and high F1-score with a value of 0.98, while P2P traffic presented low misclassification with a recall rate of 0.94 and a low F1-score with a value of 0.97 [27]. The confusion matrix shows only a low rate of two misclassified items, confirming the trustworthiness of the model. The high performance is a testament to the suitability of Logistic Regression in network traffic classification, and in particular in datasets with clearly distinct and discernible categories of traffics, and supports efficient network administration and protection [28].
The Figure -5 shows how Support Vector Machine (SVM) model [30] gave superior performance in classification of six types of networks traffics—File Sharing, Instant Message, Video, Voice OverIP, Email, and P2P—and a classification precision rate of 99.30%. All but File Sharing and Email traffics gave near-perfect precision, recall, and F1-scores, indicating high precision in classification. The File Sharing and Email traffics gave low misclassification with four and 14, and lower corresponding recall ratings of 0.91 and 0.98 [29]. Despite these low-magnitude inaccuracies, high precision and high recall in every category are a testament to high quality and efficiency in classification and high performance in handling disparate network conditions. These results are a testament to SVM's potential in efficient resource distribution and enhanced protection in complex and diverse traffics.

```
Confusion matrix:
 [[  42    4    0    0    0    0]
 [   0 1677    0    0    0    0]
 [   0    0    7    0    0    0]
 [   0    0    0    4    0    0]
 [   0   14    0    0  826    0]
 [   0    0    0    0    0   32]]
                precision    recall  f1-score   support

   FileSharing       1.00      0.91      0.95        46
InstantMessage       0.99      1.00      0.99      1677
         Video       1.00      1.00      1.00         7
   VoiceOverIP       1.00      1.00      1.00         4
         email       1.00      0.98      0.99       840
           p2p       1.00      1.00      1.00        32

   avg / total       0.99      0.99      0.99      2606

    Acc= 0.9930928626247122
```

Figure – 5: Support Vector Machine (SVM) model

Figure 6 illustrates how the K-Nearest Neighbors (KNN) model achieved a high general classification rate of 99.92% in classifying six types of networks traffics: File Sharing, Instant Message, Video, Voice OverIP, Email, and P2P. Most of the traffics, e.g., Instant Message, Video, Voice OverIP, and Email, had a perfect precision, recall, and F1-score, representing impeccable classification.
File Sharing posted high recall (1.00) with precision equal to 0.96 and F1-score equal to 0.98, while P2P traffic posted a low level of misclassification with only two observations classified in a wrong way, giving a recall rate equal to 0.94 and F1-score equal to 0.97. The ability of the model to classify diverse traffics with low errors illustrates how robust and fit the model is in network traffics classification. The performance confirms the fact that KNN is a reliable tool in efficient network administration, provisioning resources, and augmenting network protection.

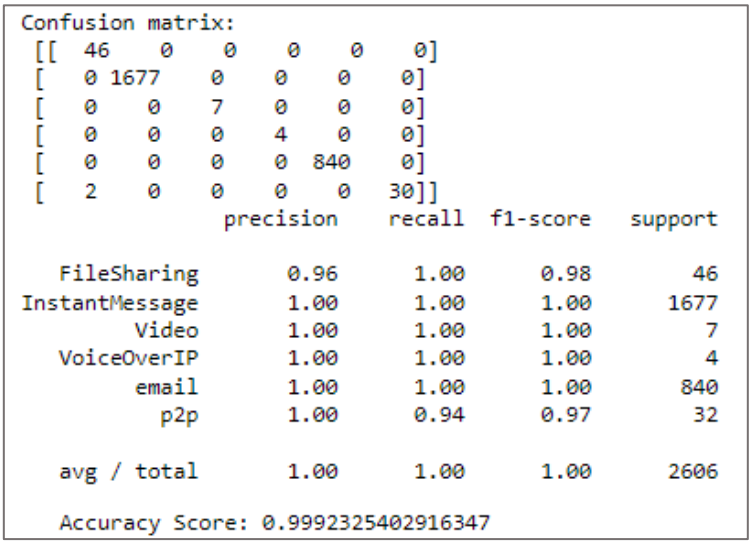


Figure – 6: K-Nearest Neighbors (KNN) model

Table – 1: Comparison of Logistic Regression, SVM, and KNN Models

Model	**Overall Accuracy**	**Perfectly Classified Traffic Types**	**Strengths**
Logistic Regression	99.92%	Instant Message, Video, Voice OverIP, Email	High accuracy, simple implementation, reliable for well-defined datasets, suitable for network security and resource management.
Support Vector Machine (SVM)	99.30%	Instant Message, Video, Voice OverIP, P2P	Exceptional handling of complex traffic patterns, robust precision and recall, effective in diverse environments.
K-Nearest Neighbors (KNN)	99.92%	Instant Message, Video, Voice OverIP, Email	Reliable for distinct traffic categories, adaptable to well-separated datasets, suitable for efficient network security.

Table 1 shows how Logistic Regression and KNN equally gave the highest rate of accuracy at 99.92%, performing better than SVM with a rate of 99.30%. All models performed extremely well with Instant Message, Video, Voice OverIP, and Email traffic. Both Logistic Regression and KNN made only a few misclassifications (2 P2P in each), while SVM had higher counts of misclassifications, primarily in File Sharing and Email traffic. While Logistic Regression and KNN are particularly suitable with distinctly classified traffic, SVM is better with complex and diverse streams of traffic. The above reveals the strengths and drawbacks of every model and facilitates the decision on the most suitable method in a particular network traffic classification need.

6. Conclusion

This research showcases the ability of advanced machine learning techniques in network traffic classification in overcoming the challenge posed by today's complex, evolving traffic. The models discussed here—Logistic Regression, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN)—proved superior in classification performance with global accuracies of 99.92% in the case of Logistic Regression and KNN and 99.30% in the case of SVM. The performance was highest in the classification of distinctly distinguishable traffic classes in the case of Logistic Regression and KNN, and in complex and diverse traffic in the case of SVM [30]. All three models proved excellent in the classification of traffic classes such as Instant Message, Video, Voice OverIP, and Email with fewer misclassifications in the classification of File Sharing and P2P traffic. The results highlight the scalability and adaptability of these algorithms in real-world applications in efficient provisioning, congestion control, and anomaly detection. Moreover, the integration with advanced preprocessing, feature extraction, and evaluation metrics showcases the requirement of systematic approaches in achieving consistent classification results. The research underscores the key role machine learning is going to play in future network management in suggesting scalable approaches in achieving performance requirements in high-data-rate scenarios while maintaining security and operational efficiency [31]. The future research could address the performance boost with hybrid models, feature space optimization, and adaptation in evolving traffic in increasingly complex topologies. The research here presented establishes a sound foundation in the usage of machine learning in network management in driving innovation and in efficient and secure network functioning.

Refrences

[1] Rawat, D. B., & Reddy, S. R. (2016). Software defined networking architecture, security and energy efficiency: A survey. *IEEE Communications Surveys & Tutorials, 19*(1), 325-346.

[2] Pratap, S. (2023). Transportation transformed: A comprehensive review of dynamic rerouting in multimodal networks. *arXiv preprint arXiv:2312.14953.*

[3] Xie, J., Yu, F. R., Huang, T., Xie, R., Liu, J., Wang, C., & Liu, Y. (2018). A survey of machine learning techniques applied to software defined networking (SDN): Research issues and challenges. *IEEE Communications Surveys & Tutorials, 21*(1), 393-430.

[4] Boutaba, R., Salahuddin, M. A., Limam, N., Ayoubi, S., Shahriar, N., Estrada-Solano, F., & Caicedo, O. M. (2018). A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. *Journal of Internet Services and Applications, 9*(1), 1-99.

[5] Modaresi, F., & Araghinejad, S. (2014). A comparative assessment of support vector machines, probabilistic neural networks, and K-nearest neighbor algorithms for water quality classification. *Water resources management, 28*, 4095-4111.

[6] Alhamyani, R., & Alshammari, M. (2024). Machine learning-driven detection of cross-site scripting attacks. *Information, 15*(7), 420.

[7] Serag, R. H., Abdalzaher, M. S., Elsayed, H. A. E. A., Sobh, M., Krichen, M., & Salim, M. M. (2024). Machine-learning-based traffic classification in software-defined networks. *Electronics, 13*(6), 1108.

[8] Nguyen, T. T., & Armitage, G. (2008). A survey of techniques for internet traffic classification using machine learning. *IEEE communications surveys & tutorials, 10*(4), 56-76.

[9] Vaghasia, S. (2018). An approach of traffic flow prediction using ARIMA model with fuzzy wavelet transform (Master's thesis, University of Windsor (Canada)).

[10] Sun, Y., Peng, M., Zhou, Y., Huang, Y., & Mao, S. (2019). Application of machine learning in wireless networks: Key techniques and open issues. *IEEE Communications Surveys & Tutorials, 21*(4), 3072-3108.

[11] Murphy, K. P. (2006). Naive Bayes classifiers. *University of British Columbia, 18*(60), 1-8.

[12] Xu, Y., Zomer, S., & Brereton, R. G. (2006). Support vector machines: a recent method for classification in chemometrics. *Critical Reviews in Analytical Chemistry, 36*(3-4), 177-188.

[13] Aria, M., Cuccurullo, C., & Gnasso, A. (2021). A comparison among interpretative proposals for Random Forests. *Machine Learning with Applications, 6*, 100094.

[14] Djenouri, Y., Belhadi, A., Lin, J. C. W., & Cano, A. (2019). Adapted k-nearest neighbors for detecting anomalies on spatio–temporal traffic flow. *IEEE Access, 7*, 10015-10027.

[15] D'Angelo, G., & Palmieri, F. (2021). Network traffic classification using deep convolutional recurrent autoencoder neural networks for spatial–temporal features extraction. *Journal of Network and Computer Applications, 173*, 102890.

[16] Sadia, H., Farhan, S., Haq, Y. U., Sana, R., Mahmood, T., Bahaj, S. A. O., & Rehman, A. (2024). Intrusion detection system for wireless sensor networks: A machine learning based approach. *IEEE Access.*

[17] Yusof, M. H. M., & Mokhtar, M. R. (2016). A review of predictive analytic applications of Bayesian network. *Int. J. Adv. Sci. Eng. Inf. Technol, 6*, 857-867.

[18] Mienye, I. D., & Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access, 10*, 99129-99149.

[19] Prasad, A. (2024). Data Quality and Preprocessing. In *Introduction to Data Governance for Machine Learning Systems: Fundamental Principles, Critical Practices, and Future Trends* (pp. 109-223). Berkeley, CA: Apress.

[20] Shen, M., Ye, K., Liu, X., Zhu, L., Kang, J., Yu, S., ... & Xu, K. (2022). Machine learning-powered encrypted network traffic analysis: A comprehensive survey. *IEEE Communications Surveys & Tutorials, 25*(1), 791-824.

[21] Silva, F. S. D., Neto, E. P., Oliveira, H., Rosário, D., Cerqueira, E., Both, C., ... & Neto, A. V. (2021). A survey on long-range wide-area network technology optimizations. *IEEE Access, 9*, 106079-106106.

[22] Lampesberger, H. (2016). Language-based anomaly detection in client-cloud interaction.

[23] Yan, Z., Li, H., Zeadally, S., Zeng, Y., & Geng, G. (2019). Is DNS ready for ubiquitous Internet of Things?. *IEEE Access*, *7*, 28835-28846.

[24] Islam, M. M. (2019). A study of dynamic access-point configuration and power minimization in elastic wireless local-area network system.

[25] Genuario, F., Santoro, G., Giliberti, M., Bello, S., Zazzera, E., & Impedovo, D. (2024). Machine Learning-Based Methodologies for Cyber-Attacks and Network Traffic Monitoring: A Review and Insights. *Information*, *15*(11), 741.

[26] Gudla, R., Vollala, S., Srinivasa, K. G., & Amin, R. (2024). A novel approach for classification of Tor and non-Tor traffic using efficient feature selection methods. *Expert Systems with Applications*, *249*, 123544.

[27] Gyamfi, E., & Jurcut, A. (2022). Intrusion detection in internet of things systems: a review on design approaches leveraging multi-access edge computing, machine learning, and datasets. *Sensors*, *22*(10), 3744.

[28] Fathi-Kazerooni, S. (2020). Countering Internet Packet Classifiers to Improve User Online Privacy (Doctoral dissertation, New Jersey Institute of Technology).

[29] Matowe, C. (2022). Using deep learning to classify community network traffic.

[30] Ahmad, I., Shahabuddin, S., Malik, H., Harjula, E., Leppänen, T., Loven, L., ... & Riekki, J. (2020). Machine learning meets communication networks: Current trends and future challenges. *IEEE Access*, *8*, 223418-223460. 925848495