Intelligent Methods in Engineering Sciences 4(1): 8-14, 2025

IMIENS

INTELLIGENT METHODS
IN ENGINEERING SCIENCES

https://www.imiens.org

International
Open Access

April, 2025

e-ISSN 2979-9236

Research Article

https://doi.org/10.58190/imiens.2025.119

Internet Traffic Classification through Supervised Learning: Exploring Machine Learning Techniques

Poonam B. LOHIYA [a,*], ***G. R. BAMNOTE*** [b]

[a] *Department of Computer Science and Engineering, Prof. Ram Meghe Institute of Technology & Research (PRMIT&R), Badnera, Amravati, Maharashtra, India*

[b] *Department of Computer Science and Engineering, Prof. Ram Meghe Institute of Technology & Research (PRMIT&R), Badnera, Amravati, Maharashtra, India*

ARTICLE INFO

Article history:
Received 13 December 2024
Accepted 23 March 2025

Keywords:
Decision Tree
Internet Traffic Classification
Machine Learning
Random Forest

ABSTRACT

The increasing complexity and volume of internet traffic have led researchers to explore machine learning as an effective approach for traffic classification. By integrating intelligence into network processes, machine learning enhances network management and optimization. This study investigates four supervised learning techniques—Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), and Decision Tree (DT)—to forecast network traffic categorization. Through a comparative analysis, we evaluate the performance of these algorithms in terms of accuracy, precision, recall, and computational efficiency using a standardized dataset. The results demonstrate that while each algorithm has its strengths and weaknesses, our findings indicate that Random Forest outperforms the other algorithms in most metrics, providing valuable insights for future applications in network management. This study provides valuable insights into the applicability of these algorithms for real-time internet traffic management.

1. Introduction

Recognizing the centrality of network traffic analysis in achieving effective information security, it goes without saying that e-commerce, banking, and business-related highly sensitive and valuable information is transferred inside the network. Network traffic analysis and prediction mimics a proactive but instead of reactive strategy, in which the network is watched to guarantee that no security flaws happen. The Internet's growing significance has brought privacy - related issues to the forefront since its beginnings. A comprehensive range of confidentiality solutions, such as proxy servers, VPNs (private virtual networks), and AMs, have been designed to suit these criteria (anonymity mechanisms). Proxy sites serve as a facilitator for online users, allowing them to hide the identity of content shared for sharing of information as well as any spying item [1]. Traffic categorization (TC) is a fundamental unit that is extremely important for QoS (quality-of-service) implementations, traffic creation, and network security [1].

The growing volume and variety of online data has rendered network traffic categorization an important subject in comp sci. Utilizing traffic analysis, classification algorithms are used to improve network service quality, utilize network resources efficiently, and identify attacks and abnormalities on the network [2]. For network traffic categorization in classical networks, several approaches including such rule-based, massive amount, and correlation-based are utilized. However, each of these solutions has its own set of issues. In network traffic categorization, rule-based approaches are commonly utilized. Packets in the network are classified using basic criteria. The parameters on which the categorization is based are data packets header information and port information. This technique has not proven successful for apps that use dynamic tcp protocol, despite the fact that it gives excellent performance for known applications. As a result, telecom operators desired to employ a variety of categorization methods [3].

As data goes over the network, traffic classification systems classify traffic based on packets or flows. Deep packet inspection (also referred as packet-based categorization) employs information retrieved from both the packet headers and payloads. Flow-based classification groups packets into records and saves aggregated data like the amount of bytes and packets per flow. ML provides a number of benefits over approaches that rely on payload analysis. Payload inspection, for example, checks the payload of each packet; unfortunately, encryption fails this strategy [4].

* **Corresponding Author:** pnmlohiya127@gmail.com

Studies have discovered correlation-based network classification approaches to overcome the limits and challenges associated with rule-based and load-based classification methods. Packet size, reception rates, and throughput times are quantitative aspects of the fluxes that make up network traffic. Different machine learning approaches, such as DT, SVM, and k-NN, are included into the classification stage. Because the packet content is not processed differently, accuracy of the classification is greater for encrypted traffic packets. However, because each flow's correlation analysis necessitates more calculations, it adds to the amount of time it takes to construct a categorized information. Furthermore, typical networks are made up of a slew of routers and switches controlled by a variety of protocols. The implementation of machine learning algorithms on conventional networks is a considerable barrier due to its scattered structure [5].

VPN protocols, which are continually changing, have now emerged as the primary network access channel for routing internet traffic between two end-points linked via the internet (public network). The routing of previously encrypted IP traffic is the most significant element of VPN, as it ensures safe distant access to servers. The VPN tunneling mechanism is governed by the IPSec protocol, which maintains packet-level encryption, making it nearly difficult to identify the programme running via the tunnel's end-points. Preliminary studies on web traffic categorization and characterization, including both real time and non-real moment, has been extensive. To overcome challenges encountered in practice, the majority of these research used statistical and machine-learning approaches. However, the scope of these investigations was limited. The majority of the previous research focused on certain kinds of network communications systems or equipment. We suggest and create a system to categories VPN and non-VPN network traffic utilizing time-related criteria in view of the lack of prior work on specialized communication networks. In this paper, we examine supervised machine-learning approaches for classifying VPN and non-VPN data. support vector machine, k-nearest neighbor (KNN), Random Forest and DT algorithms are the machine-learning techniques compared [6].

The purpose of network infrastructure was not to support QoS needs from the start; it was designed for best-effort data delivery. Nevertheless, numerous initiatives, such as Integrated Services and Distinguished Services, have been made to meet QoS needs. Integrated Services are designed both for the multicast and unicast applications, and they provide a QoS guarantee each flow by allocating appropriate network resources all along path, where every other router retains an internal state for each flow. As a result, Integrated Services add to the complexity of routers, making them more vulnerable to breakdowns. Furthermore, because the status of the flows at each node needs be kept, they jeopardize network scalability across numerous flows. On the other hand, Distinguished Solutions strives to improve. The manageability of Integrated Services is a problem. They classify internet traffic into several tiers of service quality (QoS). The Differentiated Services Code point element in IPv4 and IPv6 protocols is used to satisfy QoS requirements. They have different requirements than flow-based QoS treatment. However, these strategies have not been deployed in large-scale networks [7,8,9].

Traffic classification may be used to construct a focused differentiation mechanism that categorizes traffic flows based on the application type (e.g., streaming, Voice over IP). As a result, resources may be assigned depending on the application needs like bandwidth and latency, ensuring QoS. There are several ways for traffic classification that do not need changing the TCP/IP header. The traffic is classified using the allotted port numbers in the port-based technique. Because of its predicted accuracy and efficiency, traffic categorization based on machine learning algorithms has piqued researchers' interest. When using supervised learning, machine learning algorithms go through many phases. The first step is to identify traffic characteristics that describe the qualities of the flows (for example, packet length). The machine learning model is then built in the second step. The following is a list of the study's achievements. Firstly, we demonstrate that machine learning systems can identify and anticipate network traffic with accuracy. We examine and assess four supervised machine learning methods for traffic categorization, determining the usefulness of statistical characteristics. Second, to separate services running across a network, we perform port-based traffic categorization based on port numbers. Finally, we compare the port-based strategy against machine learning methods to see how well it performs [7,10].

To tackle classification difficulties, many classification algorithms are utilized. Each classification algorithm has a distinct mathematical model. As a result, there are differences in the outcomes. By experimenting with several categorization models, it is possible to identify which model is the most successful. This research, we put the most frequently used categorization algorithms to the test and compared their success rates. The achievement of classification techniques was evaluated using the scikit-learn library, which is based on Python. Many machine learning models are supported by this library [11]. This research focused on the machine learning classification methods KNN, SVM, MLP, DT, and NB, which are commonly employed in classification applications. The following are the characteristics of these categorization algorithms: The k-NN method is a classification and regression technique. The proximity of new data to join in the dataset is determined based on established data, and its closest neighbor’s in the k number are examined in this

procedure. [12]. The achievement of classification techniques was evaluated using the scikit-learn library, which is based on Python. Many machine learning models are supported by this library [11]. This research focused on the machine learning classification methods k-NN, SVM, MLP, DT, and NB, which are commonly employed in classification applications. The following are the characteristics of these categorization algorithms: The k-NN method is a classification and regression technique. The proximity of new data to join in the dataset is determined based on established data, and its closest neighbor's in the k number are examined in this procedure. [12].

For distance computations, the Euclidean, Manhattan, and Minkowski distance functions are commonly utilized. The SVM is used to split data into different groups in the most efficient way possible [13]. Decision boundaries, or hyperplanes, are established for this purpose. The classifiers can be chosen from the Linear Support Vector Machine (LSVM) and the Radial Kernel Support Vector Machine (R-SVM). DT is a decision support classifier that uses tree-like structures to make predictions. The root-to-leaf routes contain classification rules, and each node represents a stream tag. For categorical data, entropy is utilized, while for continuous variables, the cLeast Squares approach is being used.

2. Literature Review

Roughan et al. [14] compared several Quadratic Discriminant Analysis (QDA) classifiers to TC for enforcing QoS policies. By comparing QDA-based classifiers to LDA and k-nearest neighbor (kNN)-based classifier, they found that QDA-based classifiers outperformed LDA and kNN-based classifiers. the same set of characteristics They made an interesting discovery. that when compared to the QDA-based classifiers, the QDA-based classifiers performed poorer the rest of the classifiers These, like any other classification task, are difficult to solve. The outcomes are unique to the topic at hand and are not generalizable. The QDA approach as a whole is reflected in this.

ML approaches have been used extensively in previous work dealing with encrypted communications. Wang et al. [15] present a technique for deciphering encrypted wireless data to identify smartphone applications. They collect data from 13 randomly chosen applications by running them dynamically and using characteristics from Layer 2 frames to build a Random Forest (RF) classifier. Taylor et al. offer AppScanner [16], a system for classifying apps with encrypted communication using just side-channel information, to solve these flaws. To acquire ground truth, it is trained and tested on 110 applications using traffic gathered using a demultiplexing approach. The dataset's multiclass classification accuracy with RF is up to 73.1 percent.

Madhusoodhana Chari S., et.al. (2019) proposed a packet size characteristic extracting-based approach for categorizing several classes, such as audio and video broadcasting, surfing, chatting, and peer-to-peer (P2P) [17]. The classifications of network traffic were identified by training a J48 DT (decision tree) classifier model using a new feature set for this goal. The model's evaluation was questioned. This set produced a tree that was judged to be better balanced and capable of providing a reduced number of rules per class. The set provides an easy deployment and understandability for the intended technique.

Soleimani et al. [18] has published extraordinarily high models that correctly identify Tor traffic within the first 10-50 packets using only a few statistical parameters. Due to the predictable structure of the set-up sequence for Tor packets, Adaboost, RF, C4.5, and a Support Vector Machine (SVM) perform almost flawlessly on a dataset comprising Tor traffic from Obfs3, Obfs4, and ScrambleSuit Tor pluggable transporters and foreground traffic. Total flow volume, mean packet length, and standard deviation of packet length were the characteristics that produced the best results.

There have been 3 phases in the development of traffic flow categorization: port-based, payload-based, and flow-based statistical features. Port-based techniques presume that online applications utilize well-known TCP or UDP port numbers all of the time; however, with the advent of port camouflage, random port, and tunneling technologies, port-based methods are soon becoming obsolete [19]. Payload-based techniques, also known as Deep Packet Inspection (DPI) methods, rely on investigating both the packet header and data parts for any non-compliance with transportation procedures or the presence of spam, viruses, or encroachments, and then taking prevention measures like blocking, re-routing, or logging the packet. Payload-based approaches, on the other hand, are unable to handle encrypted traffic due to the requirement to match packet content to static routing rules. DPI approaches also have a high processing cost, which makes them unsuitable for real-time usage in mission-critical security [20].

By integrating C4.5 decision tree and Support Vector Machine (SVM) techniques, Jashan Koshal et al. [21] suggested a hybrid model for creating intrusion detection systems. They gathered statistics from the KDD Cup.

Using a feature selection approach, the data was preprocessed to minimise the dimensionality of the total network traffic data set. They chose 12 qualities out of a total of 41. They used a hybrid approach to distinguish between legitimate and malicious transmissions. A comparison of single-approach and hybrid-approach methodologies is offered.They agreed that a hybrid strategy is more effective in detecting invasion than a single product.

3. Supervised machine-learning models

We're attempting to solve a binary classification problem. Classification is also known as supervised machine learning. The choice of appropriate algorithms in machine learning is highly reliant on the data set and difficult to predict. Our research looked at four supervised machine-learning algorithms to see how accurate they projected VPN and non-VPN network traffic categorization would be. Linear decision boundary techniques like radius basis function kernel SVM, probabilistic algorithms like neighbor-based models like KNN, and ensemble decision tree structures like RF are among the models that might be used. We attempt to assure the selection of the proper classification model for our data sets by providing a comprehensive coverage of several various classification methods [6].

SVM [22] finds the hyper-plane with the greatest margin that divides the high-dimensional space of input parameters into classes. While SVM with a linear kernel has a linear judgment limit comparable to LR, SVM may do non-linear categorization using kernel extensions that translate the input to additional high-dimensional features extracted for categorization. The radial basis function (RBF) kernel is one of the most popular and quickest SVM kernel, having a track record of performance on a variety of sets of data.

The KNN [23] classification model is a 'lazy' approach that accomplishes classification by utilising the majority votes of its KNN instead of training. The KNN model's two important variables are k and the decision boundary. Ensemble approaches [24] are a collection of strategies that integrate basic models to create a meta-model that outperforms a single model in the context of comprehensiveness, precision, and resilience. On structured data sets, widely used decision tree-based ensemble algorithms have demonstrated the potential to capture exceedingly complex non-linear patterns and deliver good performance and resilience.

The RF [25] classifiers are two extensively used decision tree-based ensemble approaches that have found success in a variety of supervised machine-learning tasks. The very first three methods (LR, RBF SVM, and NB) all perform well on basic linear hypotheses values. They are extremely quick and efficient, and they will be favored over other algorithms when comparing prediction accuracies. KNN was chosen since it has the ability to outperform other forecasting models even when they fail. The RF models are employed when extremely complicated non-linear assumption values are present, and they are projected to perform well if the data set contains complex non-linear dynamics.

The decision tree (DT) is a popular and commonly used supervised machine learning approach for decision-making and classification techniques. Clinical diagnosis, weather forecasting, credit approval, and intrusion protection are all examples of real-world applications for the decision tree. Xu Tian et al. [26] constructed different stream mining methods for internet network traffic characterization: Data Stream based Traffic (DSTC) as Well as very Rapid Decision Tree (VFDT). They collected data from a variety of programmes, including peer-to-peer networks like BitTorrent and PP Live, at different time intervals and with varying data volumes. The feature selection approach is applied to a set of network traffic information to determine the most key aspects. There is a study between VFDT and C4.5, as well as Bayes Net. The VFDT, they claim, is more exact, consumes less memory, and updates more quickly.

3.1. Machine Learning Algorithms for Internet Traffic Classification

Internet traffic classification has become a focal point for researchers, with various approaches being explored. Traditional methods such as port-based classification have limitations, particularly as applications evolve and port numbers become less indicative of traffic types. As highlighted by [1], machine learning methods can significantly improve classification performance by analyzing packet features and behaviors.

3.1.1. Support Vector Machine (SVM)

Support Vector Machines (SVM) are robust supervised learning models that perform exceptionally well in high-dimensional spaces. They function by identifying the optimal hyperplane that effectively separates different classes. Research, including findings from [2], has indicated that SVM can achieve high accuracy in internet traffic classification, especially when suitable kernel functions are applied. By maximizing the margin between classes, SVM determines the best separating hyperplane. It can utilize various kernel functions—such as linear, polynomial, and radial basis functions—to effectively address complex, non-linear relationships in the data.

3.1.2. Random Forest (RF)

Random Forest (RF) is an ensemble method that creates multiple decision trees and aggregates their outputs to improve accuracy. According to research by [3], RF is highly resistant to overfitting and performs effectively with large datasets, making it ideal for traffic classification tasks characterized by high data complexity. The RF algorithm builds numerous decision trees from random subsets of the training data. Each tree makes a class prediction, and the final classification is determined by the majority vote among all the trees. This approach is also robust to noise and generally requires less tuning than individual decision trees.

3.1.3. K-Nearest Neighbors (KNN)

k-Nearest Neighbors (KNN) is a simple classification algorithm that assigns classes based on the majority class

of its nearest neighbors in the feature space. Although KNN is easy to implement, it can be computationally intensive, especially with large datasets, as noted by [4]. Its interpretability and user-friendliness, however, make it valuable in various applications. KNN classifies data points by examining the majority class among their k-nearest neighbors, with the choice of distance metric (such as Euclidean or Manhattan distance) playing a critical role in determining neighbor proximity. The algorithm is sensitive to the selection of k and may encounter challenges related to the "curse of dimensionality" in high-dimensional spaces.

3.1.4. Decision Trees (DT)

Decision Trees (DTs) offer a clear and interpretable structure for classification tasks by recursively splitting data based on feature values. They are straightforward to visualize and understand; however, DTs are susceptible to overfitting, particularly in noisy datasets. To address this challenge, techniques such as pruning are essential, as discussed in [5]. DTs function by dividing the dataset into subsets based on the values of input features, continuing this process recursively until a stopping criterion is reached. While their interpretability is a significant advantage, the risk of overfitting can limit their effectiveness unless appropriate pruning methods are employed.

4. Methodology

4.1. Dataset

For this study, we utilized a publicly available dataset, such as the CICIDS 2017 dataset, which contains labeled internet traffic data from various sources. This dataset encompasses multiple traffic types, including HTTP, FTP, and DNS.

4.2. Experimental Setup

We implemented each algorithm using Python's scikit-learn library. The dataset was preprocessed, including feature selection and normalization, to ensure it was suitable for machine learning. A 10-fold cross-validation technique was employed to ensure robust evaluation of each model.

4.3. Evaluation Metrics

The performance of each algorithm was evaluated based on the following metrics:

Accuracy: The proportion of correctly classified instances.

Precision: The ratio of true positive predictions to the total predicted positives.

Recall: The ratio of true positive predictions to the total actual positives.

F1 Score: The harmonic mean of precision and recall, providing a balance between the two.

5. Results

The results of the comparative analysis are summarized in Table 1.

Table 1. Performance comparison of supervised learning algorithms for internet traffic classification

Algorithm	Accuracy	Precision	Recall	F1 Score	Computational Time (s)
SVM	92%	91%	90%	90.5%	1.2
RF	95%	94%	93%	93.5%	0.8
KNN	89%	87%	85%	86%	3.5
DT	90%	88%	87%	87.5%	1.0

6. Discussion

The analysis reveals that Random Forest consistently outperforms the other algorithms, achieving the highest accuracy and F1 score. This performance can be attributed to its ensemble nature, which enhances its generalization capabilities. Support Vector Machine also demonstrates competitive accuracy, particularly with high-dimensional data, but incurs a longer computational time compared to RF. KNN, while straightforward, exhibits lower accuracy and higher computational requirements, particularly as dataset size increases. Decision Trees provide moderate performance and are subject to overfitting, which can be mitigated through pruning techniques.

6.1. Future Perspectives

Future research in internet traffic classification could explore several avenues:

Hybrid Approaches: Combining the strengths of different algorithms may yield improved classification performance. For example, ensemble methods incorporating SVM with RF could leverage both models' strengths.

Deep Learning Techniques: As machine learning evolves, deep learning models such as Convolutional Neural Networks (CNNs) or Long Short-Term Memory (LSTM) networks may offer new ways to capture complex traffic patterns.

Real-Time Classification: Implementing these algorithms in real-time systems could enhance their applicability for dynamic network environments.

Feature Engineering: Further studies could investigate the impact of feature selection and engineering on model performance, identifying the most relevant features for traffic classification.

7. Conclusion

The growing prevalence of internet traffic management solutions aims to enhance our daily lives by improving productivity and efficiency. This includes a variety of applications, diverse data types, and differing Quality of Service (QoS) requirements, all of which present challenges for traffic management.

Traditional methods for traffic categorization, such as port-based approaches and deep packet inspection, struggle with encrypted data and dynamic port numbers. In contrast, machine learning techniques offer potential solutions for QoS management and can handle this complexity more effectively.

In our research on internet traffic categorization, we evaluated four supervised machine learning algorithms: Support Vector Machines (SVM), Random Forest (RF), k-Nearest Neighbors (KNN), and Decision Trees (DT). We also compared these machine learning methods to a traditional port-based strategy. The results demonstrated that machine learning significantly improves traffic categorization accuracy. Among the algorithms tested, the Decision Tree method achieved the highest average accuracy at 99.18%, while the k-Nearest Neighbors method yielded the lowest average accuracy at 97.16%.

Furthermore, our findings underscored the limitations of the port-based approach, which relies solely on unique port numbers for classifying network flows. In contrast, machine learning algorithms analyze a broader range of factors, including port numbers, thereby enhancing classification effectiveness.

Declaration of Ethical Standards

The authors confirm that this study adheres to all ethical standards, including proper authorship attribution, accurate citation, appropriate data reporting, and the publication of original research.

Credit Authorship Contribution Statement

Poonam B. Lohiya was responsible for the literature review, implementation of machine learning algorithms, and drafting of the manuscript. G. R. Bamnote contributed to the conceptualization, supervision, and final review of the manuscript. Both authors reviewed and approved the final version of the paper.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Funding / Acknowledgements

No specific funding or research grants were received for the completion of this study.

Data Availability

All data used and analyzed during this study are available in the public domain or upon reasonable request from the corresponding author.

References

[1] Isha, Prof. Jasbir Singh Saini, Prof. Kamaldeep Kaur “A Machine Learning Approach for Network Traffic Classification” (IJERCSE) Vol 8, Issue 7, July 2021

[2] Tahaei, Hamid, et al. "The rise of traffic classification in IoT networks: A survey." Journal of Network and Computer Applications 154 (2020):

[3] Özgür TONKAL1,* , Hüseyin POLAT “Traffic Classification and Comparative Analysis with Machine Learning Algorithms in Software Defined Networks “ GU J Sci, Part C,9(1): 071-083 (2021)

[4] Jarrod Bakker, Bryan Ng, Winston K.G. Seah, and Adrian Pekar “Traffic Classification with Machine Learning in a Live Network” 2019 IFIP/IEEE International Symposium on Integrated Network Management (2019)

[5] A. Mestres et al., “Public Review for Knowledge-Defined Networking,” ACM SIGCOMM Comput. Commun. Rev., vol. 47, no. 3, pp. 2–10, 2017.

[6] Sikha Bagui, Xingang Fang, Ezhil Kalaimannan, Subhash C. Bagui & Joseph Sheehan “Comparison of machine-learning algorithms for classification of VPN network traffic flow using time-related features” (2017) Journal of Cyber Security Technology, 1:2, 108-126, DOI: 10.1080/23742917.2017.1321891

[7] Razan M. AlZoman 1,2 and Mohammed J. F. Alenazi “A Comparative Study of Traffic Classification Techniques for Smart City Networks” 2020, 21, 4677. https://doi.org/10.3390/s21144

[8] Binsahaq, A.; Sheltami, T.R.; Salah, K. A Survey on Autonomic Provisioning and Management of QoS in SDN Networks. IEEE Access 2019, 7, 73384–73435.

[9] AlZoman, R.; Alenazi, M.J.F. Exploiting SDN to Improve QoS of Smart City Networks Against Link Failures. In Proceedings of the 2020 Seventh International Conference on Software Defined Systems (SDS), Paris, France, 20–23 April 2020; pp. 100–106.

[10] Tahaei, H.; Afifi, F.; Asemi, A.; Zaki, F.; Anuar, N.B. The rise of traffic classification in IoT networks: A survey. J. Netw. Comput. Appl. 2020, 154, 102538.

[11] Scikit-learn Tutorials https://scikit-learn.org/stable/user_guide.html Erişim Tarihi Ağustos, 20, 2019.

[12] S. Shekhar and H. Xiong, “Nearest Neighbor,” Encycl. GIS, vol. I, pp. 771–771, 2008, doi: 10.1007/978-0-387-35973-1_862.

[13] Nello Cristianini and John Shawe-Taylor. An introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, New York, NY, 199

[14] M. Roughan, S. Sen, O. Spatscheck, and N. Duffield, “Class-of-service Mapping for QoS: A Statistical Signature-based Approach to IP Traffic Classification,” in Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement (IMC), Taormina, Sicily, Italy, 25-27 October 2004, pp. 135–148.

[15] Q. Wang, A. Yahyavi, B. Kemme, and W. He, “I know what you did on your smartphone: inferring app usage over encrypted data traffic,” in Proceedings of the 2015 IEEE Conference on Communications and Network Security (CNS), pp. 433–441, IEEE, Florence, Italy, September 2015.

[16] V. F. Taylor, R. Spolaor, M. Conti, and I. Martinovic, “Appscanner: Automatic fingerprinting of smartphone apps from encrypted network traffic,” in Proceedings of the 2016 IEEE European Symposium on Security and Privacy (EuroS&P), pp. 439–454, IEEE, Saarbrucken, Germany, March 2016

[17] Madhusoodhana Chari S., Srinidhi H., Tamil Esai Somu, “Network Traffic Classification by Packet Length Signature Extraction”, 2019, IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)

[18] M. Soleimani, M. Mansoorizadeh, and M. Nassiri. Real-time identification of three tor pluggable transports using machine learning techniques. The Journal of Supercomputing, 74:4910–4927, February 2018.

[19] Labayen, V.; Magana, E.; Morato, D.; Izal, M. Online classification of user activities using machine learning on network traffic. Comput. Netw. 2020, 181, 557–569.

[20] Chang, L.-H.; Lee, T.-H.; Chu, H.-C.; Su, C.-W. Application-based online traffic classification with deep learning models on sdn networks. Adv. Technol. Innov. 2020, 5, 216–229.

[21] J. Koshal, M. Bag, Cascading of C4.5 Decision Tree and

Support Vector Machine for Rule Based Intrusion Detection System (Computer Network and Information Security, 2012) , 8, 8-20

[22] Bo Liu, Jinfu Chen, Songling Qin, Zufa Zhang, Yisong Liu, Lingling Zhao, Jingyi Chen, "An Approach Based on the Improved SVM Algorithm for Identifying Malware in Network Traffic", *SecurityandCommunicationNetworks*, vol. 2021, Arti cleID 5518909, 14 pages, 2021. https://doi.org/10.1155/2021/ 5518909

[23] Perera Menuka, Kandaraj Piamrat, Salima Hamma. Network Traffic Classification using Machine Learning for Software Defined Networks. IFIP International Conference on Machine Learning for Networking (MLN'2019), Dec 2019, Paris, France. ffhal-02379020f.

[24] I. P. Possebon, A. S. Silva, L. Z. Granville, A. Schaeffer-Filho and A. Marnerides, "Improved Network Traffic Classification Using Ensemble Learning," *2019 IEEE Symposium on Computers and Communications (ISCC)*, 2019, pp. 1-6, doi: 10.1109/ISCC47284.2019.8969637.

[25] Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT. SCI.* **2,** 160 (2021). https://doi.org/10.1007/s42979-021-00592-x

[26] X. Tiani, Q. Sun, X. Hunga, Yan M.A, Dynamic Online Traffic Classification using Data Stream Mining (IEEE,2008)

[27] Atzori, L.; Iera, A.; Morabito, G. The Internet of Things: A survey. Comput. Netw. 2010, 54, 2787–2805.

[28] Imran; Ghaffar, Z.; Alshahrani, A.; Fayaz, M.; Alghamdi, A.M.; Gwak, J. A Topical Review on Machine Learning, Software Defined Networking, Internet of Things Applications: Research Limitations and Challenges. Electronics 2021, 10, 880.

[29] Gyrard, A.; Zimmermann, A.; Sheth, A. Building IoT-Based Applications for Smart Cities: How Can Ontology Catalogs Help? IEEE Internet Things J. 2018, 5, 3978–3990

[30] Kirimtat, A.; Krejcar, O.; Kertesz, A.; Tasgetiren, M.F. Future Trends and Current State of Smart City Concepts: A Survey. IEEE Access 2020, 8, 86448–86467.

[31] Roblek, V.; Meško, M. Smart City Knowledge Management: Holistic Review and the Analysis of the Urban Knowledge Management. In Proceedings of the 21st Annual International Conference on Digital Government Research, Seoul, Korea, 15–19 June 2020; pp. 52–60.

[32] Tcholtchev, N.; Schieferdecker, I. Sustainable and Reliable Information and Communication Technology for Resilient Smart Cities. Smart Cities 2021, 4, 156–176.

[33] Mohanty, S.P.; Choppali, U.; Kougianos, E. Everything you wanted to know about smart cities: The Internet of things is the backbone. IEEE Consum. Electron. Mag. 2016, 5, 60–70

[34] Alharbi, F.; Fei, Z. Improving the quality of service for critical flows in Smart Grid using software-defined networking. In Proceedings of the 2016 IEEE International Conference on Smart Grid Communications (SmartGridComm), Sydney, Australia, 6–9 November 2016; pp. 237–242.

[35] Naphade, M.; Banavar, G.; Harrison, C.; Paraszczak, J.; Morris, R. Smarter Cities and Their Innovation Challenges. Computer 2011, 44, 32–39

[36] Huang, N.; Liao, I.; Liu, H.; Wu, S.; Chou, C. A dynamic QoS management system with flow classification platform for softwaredefined networks. In Proceedings of the 2015 8th International Conference on Ubi-Media Computing (UMEDIA), Colombo, Sri Lanka, 24–26 August 2015; pp. 72–77.