A Hybrid Intelligent System for IP Traffic Classification

Muhana Magboul Ali Muslam, Senior, IEEE
Department of Information Technology-College of Computer and Information Sciences,
Imam Mohammad Ibn Saud Islamic University, P.O. Box 5701, Riyadh 11432, Saudi Arabia

***Abstract*—The classification of IP traffic is important for many reasons, including network management and security, quality of service (QoS) monitoring and provisioning, and high hardware utilisation. Recently, many machine learning-based IP traffic classifiers have been developed. Unfortunately, most of them need to be trained on large datasets and thus require a long training time and significant computational power. In this paper, I investigate this problem and, as a solution, present a hybrid system, which I call the ISITC, that combines the random forest (RF) and XGBoost (XGB) machine learning techniques with the support vector classifier (SVC) as the final estimator, the stacking classifier. This design leads to the development of a model that performs the classification of IP traffic and internet applications efficiently and with high accuracy. I evaluate the performance of the ISITC and various IP traffic classifiers, including neural network (NN), RF, decision tree (DT), and XGB classifiers and SVCs. The experimental results show that the ISITC provides the best IP traffic classification, with an accuracy of 96.7, and outperforms the other IP traffic classifiers: the NN classifier has an accuracy of 59, the RF classifier has an accuracy of 88.5, the DT classifier has an accuracy of 90.5, the XGB classifier has an accuracy of 89.8, and the SVC has an accuracy of 64.8.**

***Keywords*—Internet application classification; IP traffic classification; machine learning; machine learning techniques; stacking classifier**

I. INTRODUCTION

IP traffic classification is crucial to network management and security [1], quality of service (QoS) monitoring and provisioning [2][3], and better hardware utilisation [4]. However, the emergence of encryption and encapsulation [5] is making this a difficult task. Traditional methods such as port- and payload-based identification and deep packet inspection (DPI) are becoming increasingly ineffective due to dynamic port numbers and encryption [6].

Machine learning (ML) techniques, especially decision tree, C4.5, and random forest algorithms, have shown promise in this area [2][6]. These techniques can be used to develop real-time classification systems, with the Bayesian network being particularly effective [7]. However, these machine learning-based IP traffic classifiers need to be trained on a large dataset in order to be able to perform classification with high accuracy. Training on a large dataset is time consuming; it is not always possible to prepare large datasets and use them for the on-flight training of classifiers, and more computational resources are required. Moreover, many machine learning models do not achieve high accuracy when trained on small datasets [8]. Therefore, solutions to this problem are needed. The limitations of small datasets in achieving machine learning models with high accuracy may be due to the need for better methods [8] and the promotion of a data-centric approach to improving model performance [9]. In this research, I attempt to answer the question of how combined machine learning algorithms can be used to improve the accuracy of IP traffic classifiers with small datasets (containing only the most frequent features).

To respond to this challenge, in this paper, I investigate how to develop an effective IP traffic classifier that can be trained on a small dataset. The main objectives of this research are (1) to analyse and evaluate the performance of neural network, random forest, decision tree, XGBoost, and support vector classifiers for IP traffic classification, (2) to develop a hybrid traffic classification system that can be trained on small datasets and used to classify IP traffic with minimum latency and high accuracy, and (3) to compare the performance of individual and hybrid IP traffic classifiers in terms of accuracy. I conclude that an intelligent hybrid system (combining different machine learning methods) can efficiently and effectively classify IP traffic when trained on a small dataset, as combining the strengths of different machine learning models can increase the ability to capture different patterns in datasets that individual classifiers might miss. The proposed system combines the random forest (RF) technique and the XGBoost (XGB) technique with the support vector classifier (SVC) in a way that maximises the possibility of achieving high performance in the IP traffic classifier using a small number of data. The proposed solution is called the Intelligent System for IP Traffic Classification (the ISITC). Efficient IP traffic classifiers such as the ISITC can result in the prioritisation of bandwidth for critical services, the improvement of network performance, a reduction in the need for expensive and computationally intensive manual traffic monitoring tools, and the possibility of faster monitoring, which can lead to anomaly detection and, thus, improve security.

The main contributions in this paper are: (1) the investigation of different machine learning models used for IP traffic classification, (2) the introduction of an IP traffic classifier that can be easily implemented in networks without requiring high-performance computing and (3) depends on a small dataset with few and general features to classify IP traffic, and (4) a comparative analysis of widely used machine learning models in the field of IP traffic classification.

The remainder of the paper is organised as follows: Section II gives an overview of IP traffic classifiers based on individual machine learning models and IP traffic classifiers based on an ensemble. Section III presents the proposed solution, the intelligent system for IP traffic classification (the ISITC).

Section IV presents the results for the IP traffic classifiers, while Section V discusses the performance evaluation of the IP traffic classifiers. Section VI concludes this paper.

II. Review of IP Traffic Classifiers Based on Individual Machine Learning Models and IP Traffic Classifiers Based on an Ensemble

In this section, I examine IP traffic classifiers based on individual machine learning models and IP traffic classifiers based on an ensemble. Many machine learning algorithms that have been used to classify IP traffic have achieved varying degrees of accuracy. For example, in a previous study, the Bayesian network and C4.5 achieved 94% accuracy, but this dropped to 88% for smaller datasets [10]. Furthermore, [2] showed that the size of the dataset significantly influences the classification performance.

The random forest (RF) classifier is often used as an example of a single classifier [1][11][12]. In [1], the authors used random forest (RF), decision tree (DT), support vector machine (SVM), K-nearest neighbour (KNN), and naive Bayes (NB) classifiers to classify IP traffic. The RF classifier achieved the best accuracy, at around 87%. Random forest (RF) and convolutional neural network (CNN) classifier are used to classify the most common applications [11], and the models (RF and CNN) in this work were trained with datasets comprising more than 2 million samples. In [12], the authors focused on using random forest to study application-based traffic classification in an enterprise network. They collected traffic data in an enterprise network using OpenFlow in SDN. Then, the proposed classifiers were used to classify traffic flows in eight applications, namely: YouTube, Vimeo, Facebook, LinkedIn, Skype, BitTorrent, Web browsing (HTTP), and Dropbox. However, the proposed method is limited by the data provided by OpenFlow.

In [13], the authors combined DPI and machine learning to classify network traffic. They first identified the traffic as far as possible using the DPI module and then used the machine learning module to identify the unidentified traffic. Although the classification accuracy of this model was more than 98%, the privacy of the traffic successfully identified via DPI was compromised and there was an additional delay in classifying the unidentified traffic using DPI.

Decision trees (DTs) were used in [14], [15], and [16]. In [14], the authors determined whether the traffic flow was an elephant flow or a mouse flow. In [15], a DT was used to detect traffic among the top 40 applications in the Google Play Store. In [16], the authors used both decision tree and k-NN classifiers. They used two different datasets: one to classify the IP traffic among the top 37 apps in the Google Play Store and the other to classify the IP traffic among 45 apps.

An SVM was used in [17] to classify IP traffic to one of eight applications (PPlive, TVAnts, SopCast, Joost, Edonkey, BitTorrent, Skype, and DNS) based on Netflow records. In [18], a deep neural network (DNN) was used to classify IP traffic to 1 of 200 mobile applications.

Due to the advantages of ensemble methods, which have shown promising results in internet traffic classification, many such methods have been developed [19][20][21][22]. In one study [19], an ensemble of SVMs with different kernels, extra-tree-based feature selection, and majority voting was presented and achieved better results than single-kernel methods.

Moreover, in another study [20], ensemble learning was combined with co-training techniques to address weak adaptability, limited accuracy of data flow, and the need for large labelled training sets. Xu et al. [21] presented an ensemble method using three neural networks as the base model and weight tuning, achieving an accuracy of 96.38% for the payload of the transportation layer of packets. In [22], an ensemble classifier for IP traffic was proposed for imbalanced but not small datasets.

Although there are many methods of IP traffic classification, they need to be trained on large datasets, so there is still a need for efficient machine learning methods that can be trained on small datasets and achieve a good performance. Therefore, this study investigates this problem and proposes a technique that combines different techniques simply and efficiently to provide a solution that takes into account important factors such as the time required to train the model and the size of the training datasets (in terms of samples and features) while maintaining good performance.

III. An Intelligent System for IP Traffic Classification

This section provides an overview of the proposed solution, an intelligent IP network traffic classification system called the ISITC, and how it works.

A. An Overview of the ISITC

The ISITC is a hybrid intelligent system for IP network traffic classification that uses a combination of XGBoost (XGB) and a random forest (RF) with a support vector classifier (SVC) as the final estimator to efficiently classify network traffic into different application classes.

B. How does the ISITC Work?

During the training of the ISITC, the training data (80% of the total dataset) are divided into three folds. In each iteration, the XGB and RF are trained on two foldings and proceed to perform classification based on the remaining folding. The classifications performed by the XGB and RF are used as features to train the SVC. For each training example, a new feature set consisting of the classifications of the XGB and RF classifiers is obtained. Then, the SVC is trained on these meta-features. It is important to note that the target labels remain the same but the input features are now the classifications of the XGB and RF.

When the XGB and RF classifiers are trained on the entire training dataset, both perform classifications on the test dataset (20% of the entire dataset). These classifications performed by the XGB and RF classifiers on the test dataset are used to create the test meta-features that are used by the SVC to perform the final classifications on the test dataset. Fig. 1 shows the ISITC elements and their interactions during the training and testing processes.

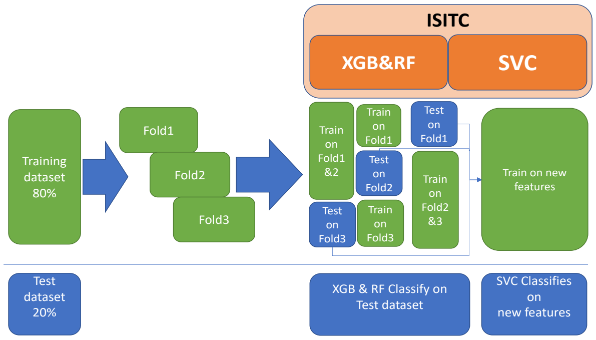

Fig. 1. The architecture of the Intelligent System for IP Traffic Classification (the ISITC).

To ensure that both the hyperparameters of the XGB and RF classifiers and the stacking ensemble are optimised, I use stacking with cross-validation (cv=3) and GridSearchCV in this solution, the ISITC. This method utilises the advantages of grid search to tune the hyperparameters and stacking for the composition of the XGB and RF classifiers. Fig. 2 shows the procedures for loading and splitting the dataset, determining the classifier parameters, the training and testing process, and the stacking process with 3-fold cross-validation for the XGB and random forest classifiers with SVC as the last estimator.

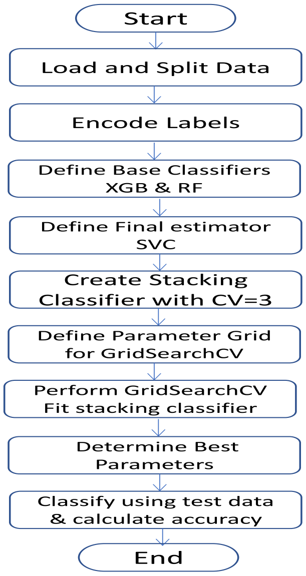
Start
Load and Split Data
Encode Labels
Define Base Classifiers XGB & RF
Define Final estimator SVC
Create Stacking Classifier with CV=3
Define Parameter Grid for GridSearchCV
Perform GridSearchCV Fit stacking classifier
Determine Best Parameters
Classify using test data & calculate accuracy
End

Fig. 2. Training and testing procedures for the Intelligent System for IP Traffic Classification (the ISITC).

The figure above shows the key steps in stacking classifiers with 3-fold cross-validation and optimising their performance using grid search.

IV. Results of the IP Traffic Classifiers

This section presents the dataset and the experimental setup, as well as the results for the IP traffic classifiers and the statistical analysis.

A. *Dataset and Experimental Setup*

The dataset used in this paper is part of the "IP Network Traffic Flows Labeled with 75 Apps" dataset [23]. The small dataset used comprises only 2172 samples and five applications. Table I shows the characteristics of the datasets and the experimental setup.

TABLE I. Dataset and Experimental Setup Under Which IP Traffic Classifiers Are Examined

Parameter	***Value***
Split ratio	80% training, 20% testing
# of samples	2172
# of classes	5
# of instances in class 0	565
# of instances in class 1	439
# of instances in class 2	418
# of instances in class 3	405
# of instances in class 4	345
Data scaling	StandardScaler
Combination Method	Stacking
Parameter tuning	GridSearchCV, 3 folds
Validation	cross-validation using cross_val_score
Statistical tests	t-test using ttest_ind

B. *The Results of the Different IP Traffic Classifiers*

In this section, the accuracy results for different IP traffic classifiers using different machine learning models are presented. In particular, the accuracy results for the neural network, random forest, decision tree, and XGBoost classifiers, the SVC, and the ISITC are presented. The experimental results show that the NN classifier has an accuracy of 59, the RF classifier has an accuracy of 88.5, the DT classifier has an accuracy of 90.5, the XGB classifier has an accuracy of 89.8, the SVC has an accuracy of 64.8, and the ISITC has an accuracy of 96.7. The accuracy of each classifier is shown in Fig. 3.

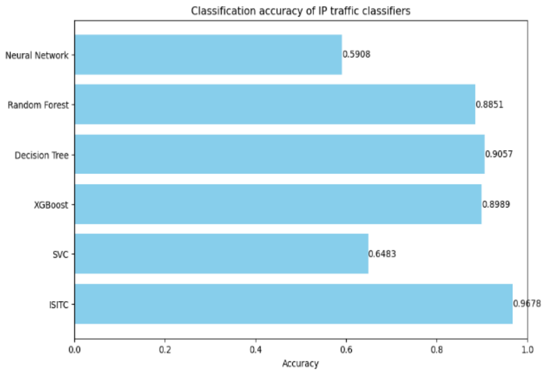

Fig. 3. Classification accuracy of IP traffic classifiers.

To determine the optimal settings for each classifier when using a small dataset, I also evaluated the performance of these classifiers, including neural network (NN), random forest (RF), decision tree (DT), and XGBoost (XGB) classifiers, an SVC, and the ISITC (a stacked model using the RF and XGB, with the SVC as the final estimator). The accuracy results for each IP traffic classifier with different hyperparameters are shown in Fig. 4.

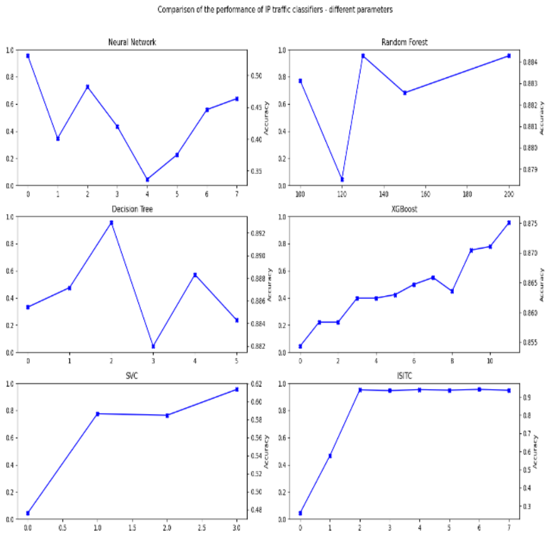

Fig. 4. Comparison of the performance of IP traffic classifiers—different parameters.

I found that the IP traffic classifier using the neural network achieved an accuracy of 0.5195%, with the best configuration comprising hidden layers of size 150 and a learning rate of 0.01. The IP traffic classifier using the random forest model achieved an accuracy of 0.8897% with 100 estimators. The IP traffic classifier using the decision tree model showed an accuracy of 0.9103% at a maximum depth of 30.

The IP traffic classifier using the XGBoost model achieved an accuracy of 0.9609% with 120 estimators and a learning rate of 0.1. The IP traffic classifier using the SVC model showed an accuracy of 0.6483% with an "rbf" kernel. The IP traffic classifier using the ISITC, the proposed solution, achieved an accuracy of 0.9678% with an "rbf" kernel.

To further analyse the performance of the classifiers, a confusion matrix (CM) was created for each IP traffic classifier, providing information on the number of correct and incorrect classifications for each class, as shown in Fig. 5, below.

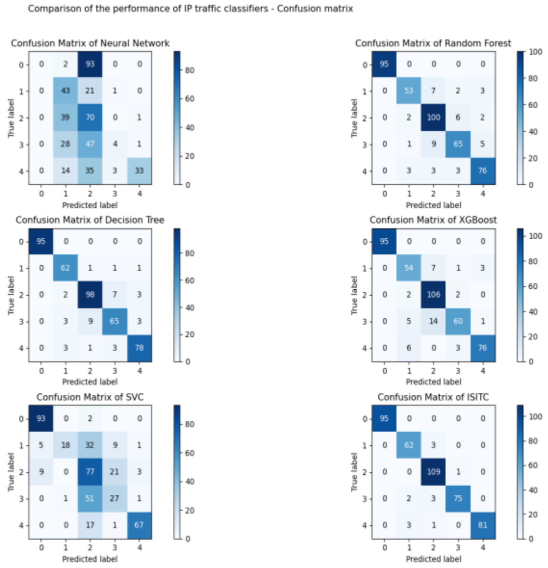

Fig. 5. Confusion matrix for the IP traffic classifiers.

C. Statistical Analysis

To assess the statistical significance of the observed differences in the performance of the IP traffic classifiers, t-tests were performed between the results of the cross-validation of the IP traffic classifiers. The t-test results are shown in Table II.

TABLE II. T-TEST RESULTS BETWEEN IP TRAFFIC CLASSIFIERS

Classifier 1	Classifier 2	t-statistic	p-value
ISITC	Neural Network	7.675364069946557	0.001549315592527276
ISITC	Random Forest	4.29261462099269	0.012719774710311835
ISITC	Decision Tree	5.2256182997013205	0.006402824672843334
ISITC	XGBoost	5.163652433269795	0.006681160371121216
ISITC	SVC	8.813500048755037	0.0009144801911715298
ISITC	Neural Network	7.675364069946557	0.001549315592527276
ISITC	Random Forest	4.29261462099269	0.012719774710311835
ISITC	Decision Tree	5.2256182997013205	0.006402824672843334

V. DISCUSSION ON THE PERFORMANCE OF THE IP TRAFFIC CLASSIFIERS

In this section, the performance evaluation of IP traffic classifiers is discussed. From the accuracy results mentioned above (Fig. 3), the evaluation of the different IP traffic classifiers shows that the ISITC (a stacked model of RF and XGB, with SVC as the final estimator) provides promising results and significantly outperforms all the individual classifiers, including the standalone XGB model, which has the higher accuracy among the individual classifiers.

When evaluating the different IP traffic classifiers with the different hyperparameters (Fig. 4), I found that the performances of the NN, RF, DT and XGBoost classifiers were very sensitive to the learning rate, with the NN accuracy decreasing significantly at higher learning rates. The SVC and ISITC classifiers showed consistent performance at different learning rates and different numbers of estimators, with the highest accuracy achieved with the ISITC, highlighting the advantage of hybrid classifiers with stacking.

The ISITC classifier outperformed all single classifiers and the hybrid model, confirming that the combination of different models can further improve performance in IP traffic classification. The IP traffic classifier with the ISITC, the proposed solution, achieved an accuracy of 0.9678% with an "rbf" kernel, emphasising the advantage of hybrid classifiers with stacking.

The higher performance of the ISITC can be attributed to its ability to capture various patterns in data that individual classifiers may miss. The random forest classifier captures complex relationships, while XGBoost recognises interactions between features perfectly.

Confusion matrices (Fig. 5) enable further analysis of the performance of each classifier by providing insight into the number of correct and incorrect classifications for each class. The NN classifier only performed well in class 2 classification (contributing to its lower overall accuracy), while class 0 is recognised with high accuracy by all other classifiers (RF, DT, XGBoost, and SVC). The ISITC's confusion matrix shows strong performance with a balanced classification, similar to the standalone XGB model, with the exception of class 1 and class 3 classification (but also a slight increase in false positives).

To ensure the scalability of the ISITC, cross-validation (cv=3) is used to evaluate the ISITC on three different datasets derived from the entire dataset during the training and testing phase, and the performance of the ISITC with different parameters and different datasets is shown in Fig. 4.

The results of this study are consistent with those of previous studies that have emphasised the effectiveness of hybrid methods in classification in general. However, the specific combination of random forest and XGBoost classifiers with SVC as the final estimator has not been extensively studied, making this work a novel contribution.

The t-tests performed to compare the ISITC and other IP traffic classifiers showed that the performance improvements observed were always statistically significant. For example, the t-test between the ISITC and the neural network showed a statistically significant difference in performance with a p-value of 0.15. The t-tests between the ISITC and the other classifiers (the random forest, decision tree, XGBoost, and SVC) also showed statistically significant differences in performance with p-values of 1.2, 0.6, 0.6, and 0.09, respectively. This means that the ISITC is different from the other IP traffic classifiers.

VI. Conclusion

In this paper, I propose a hybrid system called the ISITC, which simply and efficiently combines random forest (RF) and XGBoost (XGB) classifier techniques with an SVC as the final estimator to classify IP traffic with a small dataset. The ISITC classifier presented here can efficiently classify IP traffic with a high accuracy of 96.7%, which holds promise for improving network management, security measures, and quality of service (QoS). The ISITC outperforms IP traffic classifiers with NN, RF, DT, or XGB classifiers or SVCs. The t-test values show that there is a statistically significant difference in the accuracy of the ISITC and the other IP traffic classifiers. These results emphasise the potential of advanced hybrid classifiers to significantly improve the accuracy and reliability of IP traffic classification. Furthermore, the ISITC results confirm that the combination of different models can further improve performance in IP traffic classification. In the future, further research should be conducted on combinations of classifiers and their performance should be tested on different datasets. In addition, future work could explore advanced ensemble techniques and further tuning of hyperparameters to improve the performance of hybrid models.

References

[1] Rahul, A. Gupta, A. Raj and M. Arora, "IP Traffic Classification of 4G Network using Machine Learning Techniques," in 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 127-132, doi: 10.1109/ICCMC51019.2021.9418397.

[2] Jun, Li & Shunyi, Zhang & Yanqing, Lu & Zailong, Zhang, "Internet Traffic Classification Using Machine Learning," in CHINACOM 239 - 243. 10.1109/2007.4469372.

[3] T. T. T. Nguyen and G. Armitage, "A survey of techniques for internet traffic classification using machine learning," In IEEE Communications Surveys & Tutorials, vol. 10, no. 4, pp. 56-76, Fourth Quarter 2008.

[4] J. Gómez, V. H. Riaño and G. Ramirez-Gonzalez, "Traffic Classification in IP Networks Through Machine Learning Techniques in Final Systems," In IEEE Access, vol. 11, pp. 44932-44940, 2023, doi: 10.1109/ACCESS.2023.3272894.

[5] F. Pacheco, E. Exposito, M. Gineste, C. Baudoin and J. Aguilar, "Towards the Deployment of Machine Learning Solutions in Network Traffic Classification: A Systematic Survey," In IEEE Communications Surveys & Tutorials, vol. 21, no. 2, pp. 1988-2014, Secondquarter 2019

[6] Donghong Qin, Jiahai Yang, Jiamian Wang and Bin Zhang, "IP traffic classification based on machine learning," in 2011 IEEE 13th International Conference on Communication Technology, Jinan, China, 2011, pp. 882-886, doi: 10.1109/ICCT.2011.6158005.

[7] Kuldeep Singh, S. Agrawal, B.S. Sohi, "A Near Real-time IP Traffic Classification Using Machine Learning," in International Journal of Intelligent Systems and Applications (IJISA), vol.5, no.3, pp.83-93, 2013. DOI:10.5815/ijisa.2013.03.09

[8] Shaoxuan Zhou, "An Analysis of The Small Sample Datasets Based on Machine Learning," in 2022 6th International Conference on Electronic Information Technology and Computer Engineering (EITCE 2022), October 21-23, 2022, Xiamen, China. ACM, New York, NY, USA,

[9] Bhowmik, Pritom and Arabinda Saha Partha, "A Data-Centric Approach to Improve Machine Learning Model's Performance in Production," International Journal of Engineering and Advanced Technology (2021): n. pag.

[10] Singh, Kuldeep and Sunil Agrawal, "Comparative analysis of five machine learning algorithms for IP traffic classification," in 2011 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC) (2011): 33-38.

[11] V. K. BP, K. SM and P. LV, "Deep machine learning based Usage Pattern and Application classifier in Network Traffic for Anomaly Detection," in 2023 International Conference on Advances in Electronics, Communication, Computing and Intelligent Information Systems (ICAECIS), Bangalore, India, 2023, pp. 50-54, doi: 10.1109/ICAECIS58353.2023.10169914.

[12] P. Amaral, J. Dinis, P. Pinto, L. Bernardo, J. Tavares, and H. S. Mamede, "Machine learning in software defined networks: Data collection and traffic classification," in Proc. - Int. Conf. Netw. Protoc. ICNP, vol. 2016-Decem, no. NetworkML, pp. 1-5, 2016, doi: 10.1109/ICNP.2016.7785327.

[13] W. A. Aziz, H. K. Qureshi, A. Iqbal, A. Al-Dulaimi and S. Al-Rubaye, "Towards Accurate Categorization of Network IP Traffic Using Deep Packet Inspection and Machine Learning," in GLOBECOM 2023 - 2023 IEEE Global Communications Conference, Kuala Lumpur, Malaysia, 2023, pp. 01-06, doi: 10.1109/GLOBECOM54140.2023.10437078.

[14] P. Xiao, W. Qu, H. Qi, Y. Xu, and Z. Li, "An efficient elephant flow detection with cost-sensitive in SDN," in Proc. 2015 1st Int. Conf. Ind. Networks Intell. Syst. INISCom 2015, pp. 24-28, 2015, doi: 10.4108/icst.iniscom.2015.258274.

[15] Zafar Ayyub Qazi, Jeongkeun Lee, Tao Jin, Gowtham Bellala, Manfred Arndt, and Guevara Noubir, "Application-awareness in SDN," in SIGCOMM Comput. Commun. Rev. 43, 4 (October 2013), 487-488. https://doi.org/10.1145/2534169.2491700.

[16] M. Uddin and T. Nadeem. Traffic Vision: A Case for Pushing Software Defined Networks to Wireless Edges. Proc. - 2016 IEEE 13th Int. Conf. Mob. Ad Hoc Sens. Syst. MASS 2016, pp. 37-46, 2017, doi: 10.1109/MASS.2016.016.

[17] D. Rossi and S. Valenti, "Fine-grained traffic classification with Netflow data," in IWCMC 2010 - Proc. 6th Int. Wirel. Commun. Mob. Comput. Conf., pp. 479-483, 2010, doi: 10.1145/1815396.1815507.

[18] I. Paper, "INVITED PAPER Special Section on Communication Quality in Wireless Networks Toward In-Network Deep Machine Learning for

Identifying Mobile Applications and Enabling Application Specific Network Slicing," in transcom, No. 7, pp. 1536-1543, 2018, doi: 10.1587/transcom.2017CQI0002.

[19] Manju, N., Harish, B.S. (2020), "Classification of Internet Traffic Data Using Ensemble Method. In: Das, H., Pattnaik, P., Rautaray, S., Li, KC. (eds) Progress in Computing, Analytics and Networking. Advances in Intelligent Systems and Computing," vol 1119. Springer, Singapore. https://doi.org/10.1007/978-981-15-2414-1_39.

[20] Haitao He, Chunhui Che, Feiteng Ma, Jun Zhang, Xiaonan Luo, "Traffic Classification Using En-semble Learning and Co-training," in 8th WSEAS International Conference on Applied Informatics and Communications (AIC'08), Rhodes, Greece, August 20-22, 2008.

[21] L. Xu, X. Zhou, Y. Ren and Y. Qin, "A Traffic Classification Method Based on Packet Transport Layer Payload by Ensemble Learning," in 2019 IEEE Symposium on Computers and Communications (ISCC), Barcelona, Spain, 2019, pp. 1-6].

[22] Santiago Egea Gómez, Belén Carro Martínez, Antonio J. Sánchez-Esguevillas, Luis Hernández Callejo, "Ensemble network traffic classification: Algorithm comparison and novel ensemble scheme proposal. Computer Networks," Volume 127, 2017, Pages 68-80, ISSN 1389-1286, https://doi.org/10.1016/j.comnet.2017.07.018.

[23] IP Network Traffic Flows Labeled With 75 Apps, Apr. 2019, [online] Available: https://kaggle.com/jsrojas/ip-network-traffic-flows-labeled-with-87-apps. (accessed on 10 July 2024).