International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Deep Learning Based Traffic Classification with Feature Selection Mechanism and Explainable Artificial Intelligence (Xai)

[1]Dr. Aparna Joshi, [2]Dr.P.Namratha, [3]Dr. T Venkata Naga Jayudu, [4]Dr Ashwini Sapkal, [5]Dr.Rupali Amit Bagate, [6]Dr. Gajanan Walunjkar

Submitted: 28/01/2024 **Revised**: 06/03/2024 **Accepted**: 14/03/2024

Abstract: In computer networking and cybersecurity, traffic categorization, which identifies the kind and nature of network traffic flows, is a vital activity. For the purpose of controlling Quality of Service (QoS), optimising network resources, and increasing security measures, accurate traffic classification is essential. Recently, the field of traffic classification has undergone a revolution and explicable artificial intelligence. To improve the comprehension and interpretability of classification findings, this study investigates the application of explainable AI methodologies with deep learning models for efficient traffic categorization and dominant feature selection. Network functions like software-delivered networking structures use traffic cataloguing extensively. Numerous techniques for classifying traffic without looking at the packet payload have been developed, including deep learning models. They have a significant obstacle, though, given that deep learning's method is puzzling. Malfunction yields when training dataset with the improper data hence we have insufficient deep leaning model. This can be fixed with the help of XAI to get better deep learning model. In this work we presented genetic algorithm which works on support of XAI to explore traffic classification by using deep learning model. Model can be applied on traffic classifier after each feature can be evaluated properly. The role of genetic algorithm is to generate mask of feature. In comparative works our model proved with better accuracy and good dominance rate.

Keywords: *Deep learning, feature selection, XAI, traffic classification, QoS.*

1. Introduction

Network traffic flows are categorized into separate classes according to different characteristics like the application, protocol, or service. Traditional approaches that rely on payload- or port-based strategies sometimes struggle to handle the encrypted and obfuscated traffic of today. Due to their capacity to automatically extract complex features from raw data, CNNs, RNNs are better to demonstrated outstanding success in traffic classification.

As mobile devices become more widely used, the network traffic landscape has undergone significant change. Traffic Classification (TC) has emerged as a key player while also facing novel and unheard-of difficulties. Deep Learning (DL) methodologies ensure newly added admiration and need emerged as a viable alternative to machine learning (ML) methods that rely on laborious and time-consuming handmade feature creation. However, because DL models are black boxes, they cannot be used in situations where the accuracy of the results and the legitimacy of the policies are crucial. eXplainable Artificial Intelligence (XAI) approaches have lately piqued the community's interest as a means of overcoming these restrictions. As a result, we explore trustworthiness and interpretability in this work exhausting XAI-based systems to understand, interpret, and enrich the behavior of cutting-edge multimodal DL traffic classifiers.

In contrast to typical XAI results, the proposed methodology makes an effort to deliver global interpretations relatively than sample-based ones. Outcomes from a sweeping dataset permit for the addition of area belongs to the aforementioned conclusions.

A flow-behavior based TC faces a significant hurdle, nevertheless, due to the environment of machine learning paradisims.

A black box adversarial attack can compromise an ow-behavior-based TC due to a critical weakness in the black box problem [4]. A machine-learning model is tricked in a black-box argumentative stabbing situation by an invader who introduces adversarial perturbations, a form of noise, into the input data. A significant attack scenario could happen if the machine-learning model misclassifies the tainted data when it receives them. As a result, resources for applications requiring high priority QoS may run out, and without sufficient QoS

[1]*Asst.Prof, Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune*
[2]*Assoc.Professor, Department of CSE, GATES Institute of Technology, Gooty*
[3]*Assoc.Professor, Department of CSE, Srinivasa Ramanujan Institute of Technology (A),Anantapur*
[4]*Associate Professor, Dept of IT, AIT Pune, Army Institute of Technology, Pune*
[5]*Asst.Prof, Department of Information Technology, Army Institute of Technology, Pune*
[6]*Asst.Prof, Department of Information Technology, Army Institute of Technology, Pune.*
** Corresponding Author Email: changalaravindra@gmail.com*

requirements, mission-critical apps may not be able to continue operating normally.

Network engineers can learn a lot about how to enhance the machine-learning-based traffic classification model by identifying aberrant data in the dataset. A method known as explainable artificial intelligence (XAI) describes how ML models function [6]. Traditional ML algorithms compare the training and test dataset distributions by creating metrics like the distance or score. These metrics develop the classification criteria as hyper-planes that separate data after adequate training. For instance, visualising or creating classification criteria might help explain the workings of traditional machine-learning models like decision trees and support vector machines [7]. As a result, it is more challenging to explain the working of deep learning than it is to explain the working of classical ML. Eventually, the arrival of deep learning for traffic classification brings the black-box dilemma to the forefront for ow-behavior-based techniques as well.

In order to clarify the suggested DL based classifier traffic functions, we suggest a dominating feature selection approach.With the help of a genetic algorithm, we construct a fitting score as the quantification of the value of each feature and generate a better feature selection for mask that achieves the best balance of a high level of accuracy classification and the elimination of superfluous features. An evolutionary method known as a genetic algorithm is capable of solving a variety of NP-hard issues, including the travelling salesman problem (TSP). Finally, we define a dominance rate that indicates how better model. The suggested method has two technical innovations, to sum up.

To describe how the DL based traffic classifier of functions, we suggest a dominant feature selection technique utilising a genetic algorithm. By measuring the significance of each feature, the suggested technique, in particular, can choose which portion of the complete feature the classifier concentrates on.The model can evaluate that classifies the traffic and generates the best accuracy to evaluate the fitting score, we develop the ow-behavior-based traffic classifier. Although the suggested approach is equally effective for classifying traffic at any level of granularity, we use a service oriented traffic classification method to identify the features of services with internet.

The remaining portions of this essay are divided into four pieces.

Section II introduces related works on traffic classification and XAI. In Section III, the development of a deep-learning-based traffic classifier and a dominating feature selection approach are presented. Section IV presents the experimental findings along with a performance assessment. In Sub-section c of Section IV, it is also discussed how traffic can be analysed and divided into each service. In Section V, we offer some last observations.

2. Background Work

But the use of sophisticated AI algorithms inevitably results in the creation of "black-box" models that exhibit questionable actions and an unwanted (and sometimes intolerable) lack of transparency. Modern DL models are severely impacted by this problem since they are complicated AI models with high-dimensional inputs.

Disparities between web apps in traffic and communication

High SNI classification accuracy demonstrates that such protocols are unable to properly protect user privacy from side-channel assaults, which might constitute a substantial danger to ESNI and other methods to circumvent SNI identification.

Examining deep learning's efficacy for HTTPS SNI categorization is the primary objective of this work. The SNI will serve as our ground truth labels, and we will solely depend on encrypted TLS packet contents devoid of the SNI extension. We will investigate if service identification accuracy can be increased with deep learning under the presumption that SNI is neither fabricated or counterfeit. This is the first study that we are aware of that uses deep learning on HTTPS data to categorise SNI.

Previously, great accuracy was attained for various network traffic systems by training supervised Naive Bayes classifiers as header-driven discriminators [10]. These strategies are no longer viable owing to the increase in encrypted traffic.

Recent methods have concentrated on application level identification without requiring IP addresses, port numbers, or payload data that has been encrypted. When categorising Skype and SSH traffic, Decision Trees produce the greatest accuracy, according to Alshammari et al. [11]. By emphasizing the development of statistical characteristics for packet size and packet transfer durations for application categorization (FTP, DNS, HTTP, etc.), Okada et al. further this work [12]. When used with SVM classifiers, these characteristics attain great accuracy. However, because our study issue is more about identifying the underlying service name than it is about the kind of traffic, application level identification is not precise enough to answer it.

One of the first to address this more particular issue of service identification for HTTPS-specific traffic (for

example, maps.google.com vs. drive.google.com) was Shabir et al. [8].

They also mark each connection using the SNI extension and gather HTTPS traces from user sessions as part of their job.

They provide extra statistical aspects pertaining to the encrypted payload in addition to the common packet and inter-arrival time statistics in their suggested statistical framework. Using Decision Tree and Random Forest classifiers, they get the greatest results.

3. Explainable Artificial Intelligence (Xai)

The mechanism of the machine learning model has been examined using explainable artificial intelligence (XAI) approaches. Blackbox predictors, like deep learning, might benefit from some interpretation rules provided by the input-output connection. In [19], a visual attention approach for neural picture caption creation is described.

Convolutional feature extraction was used by the authors to extract the important characteristics from the image. The RNN is trained for image captioning using the retrieved characteristics. Using a convolutional feature extraction throughout this process, the attention mechanism may draw attention to a specific area of the picture.

In several XAI investigations, the machine learning model for image categorization is explained. The traffic classification issue, however, differs from the picture classification problem in a number of ways. Every piece of information in the picture classification issue has the same semantic meaning, such as the RGB colour value. By identifying an object in data made up of pixels with the same meaning, the attention process suggested in [19] chooses a feature subset. The dimension of the data in the traffic classication problem is smaller than the dimension of the picture data. Additionally, a feature selection approach that can take into account all of these qualities is needed because each data element has a unique significance. Based on a genetic algorithm, we created a dominant feature selection technique that is appropriate for low-dimensional behavioural data.

4. Flow Behavior Based Traffic Classification

The categorization of encrypted communication is the most important challenge raised by current research on traffic. The payload itself was a barrier to the categorization of encrypted communications.

Due to their ability to be recovered without requiring the inspection of a scrambled payload, behaviour statistics have become a valuable tool for categorising encrypted data. Encrypted traffic may be categorised using flow-behavior-based techniques by utilising behaviour statistics. The authors of [8] offered three exemplary traffic encryption techniques and decoded statistics from the traffic that was encrypted. Additionally, they assessed the performance of a number of machine-learning methods, including neural networks, support vector machines, random forests, and naive Bayes. They demonstrated the applicability of ow-behavior-based methods by comparing different machine-learning techniques.

Numerous research on traffic categorization have incorporated deep learning's advantages as a result of its considerable advancements. Deep learning technologies have a major benefit over conventional machine learning techniques in that they allow the classifier to automatically extract characteristics from the raw input. Representation learning is a technique for automatically extracting characteristics from unprocessed data, and in deep learning, the CNN is a typical representation learning technique.

According to the authors, it is impossible to use traffic classification algorithms that rely on manually derived feature sets for mobile traffic produced by moving targets. Additionally, they take advantage of the deep learning, that automatically leads to extracting the feature set, to overcome the shortcomings of conventional traffic categorization schemes.

Deep learning exhibits fantastic performance, yet integrating deep learning directly may cause performance to decline. Due to the nature of the features displayed by behaviour data, a novel model has to be revised and put into use. The authors of [12] discussed a problem where many research on deep learning-based traffic categorization often took all the characteristics equally without taking the kind of statistics into consideration. The authors use a multimodal deep learning model to mimic the multimodality of behavioural statistics. The concept of taking anonymity tool (AT) traffic into account was first presented in [13]. A number of ATs, including Tor, have been created as it becomes crucial to protect users' online privacy.

As a result, a number of malicious uses of ATs result in significant problems. The authors suggested categorising AT-specific traffic using a hierarchical categorization that allows for effective fine-grained tailoring. Using a hierarchical categorization, the authors of [14] devised a traffic classification technique. The drawback of flow-behavior-based techniques is that, due to the nature of machine learning, they are unable to categorise unknown traffic classes. Further, improving the traffic class's granularity worsens classification performance. The granularity of the traffic class is used by the authors to construct the sub-classifier hierarchically.

The authors of [15] addressed the issue of an unknown traffic class being unable to categorise using meta-learning. Deep learning requires enough data to do ne-tuning when an unknown traffic class shows up. It is challenging to gather a large enough dataset of an unknown traffic class, though. Deep learning can teach the relationship between each piece of data thanks to few-shot learning, or meta-learning.

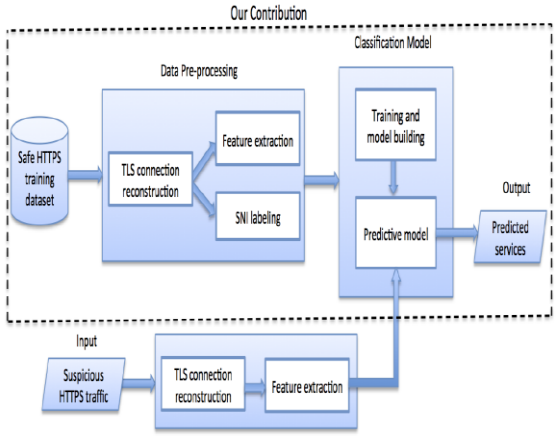


Fig 1. Workflow of the proposed HTTPS Identification model

Technique	Scope	Resources
Feature's role	Classification of images	[13]
Primary feature selection	Classification of traffic	[15]
Working of perturbation	Images segmentation	[16]
Alert mechanism	Caption generation for images	[17]

Table 1. Comparative study.

Algorithm 1 Procedure of Traffic Classification

Require: Training packet trace P collected with short time duration, the traffic flow F composed of packets p_i, the function $\Omega(F)$ returning the 5-tuple of the flow F.

Ensure: Pre-trained traffic classifier.

$D \leftarrow \varnothing$

$\Omega = \{\Omega(F_1),\ \Omega(F_2),\ \ldots,\ \Omega(F_N)\}$

Perform clustering packets in P by 5-tuple set Ω to form a bidirectional flow set $\{F_1, F_2, \ldots, F_N\}$

for $i = 1 \rightarrow N$ **do**

  $\mathbf{t} \leftarrow \bigcup_{j=1}^{n-1} \{\tau(p_{j+1}) - \tau(p_j)\}$

  $\mathbf{s} = \{s_k | s_k$ is packet size of packet $p_k, 1 \leq k \leq n\}$

  Compute total bytes b in the flow

  Compute feature vector by using traffic flow statistical features

$$\psi = [m_t\ M_t\ \mu_t\ \sigma_t\ m_s\ M_s\ \mu_s\ \sigma_s\ n\ b]$$

  Compute reverse directional feature vector $\bar{\psi}$

  $\mathbf{x_i} = [\psi,\ \bar{\psi}]$

  Detect the application layer l_i by the packet gathering step

  $D \leftarrow D \cup \{(\mathbf{x_i}, l_i)\}$

end for

Normalize dataset D

for $i = 0 \rightarrow N$ **do**

  Pick $(\mathbf{x_i}, l_i) \in D$

  for $j = 0 \rightarrow$ number of ResNet layers **do**

    $e := x_i$

    $x_i :=$ batch_normalization (x_i)

    $x_i :=$ ReLU (x_i)

    $x_i :=$ convolution (x_i)

    $x_i := e + x_i$

  end for

  Calculate the loss between the result of ResNet and l_i, and backpropagate the gradient of the loss to the model.

end for

5. Proposed Approach

Figure 2 shows a summary of the suggested prominent feature selection technique. The creation of a traffic classifier and the selection of dominating features make up the two components of the suggested methodology. A residual network (ResNet), a cutting-edge deep learning method, is used to create the traffic classifier [20]. Data pre-processing and training are applied during the traffic categorization process.

Using a masked input dataset, a pre-trained classifier, and a count of the zero elements, the mask selection analyses the masks and determines their correctness. Following the review, a few masks are selected for the following generation's mask development utilising a roulette wheel selection process.

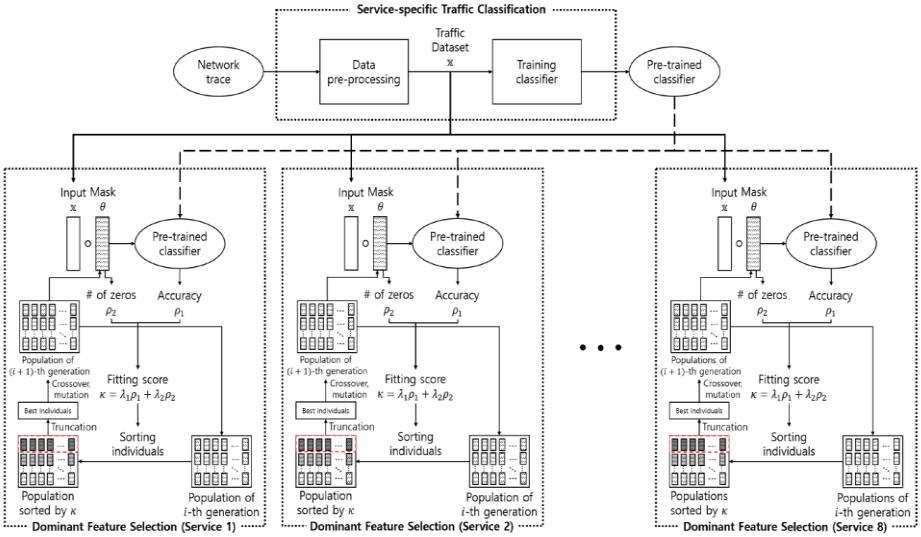


Fig 2. Basic view of the traffic classification and dominant feature selection.

The data pre-processing phase computes after collecting packets from a network flow the statistical characteristics from the collection of packets. The classifier training stage, which is the final step, develops the deep-learning utilizing the dataset as a classifier. Using Algorithm 1, the steps involved in creating a traffic classifier.

6. Data Preprocessing

The bidirectional ow set's statistical characteristics are calculated at the data pre-processing stage and are displayed in lines 4 through 14 of Algorithm 1. The statistical characteristics used to depict the behaviours of the packets in the network include inter arrival time, size of packet, bytes count and total number of packets[21]. Despite the fact that the packets are encrypted, the behaviour of the packets serving the same application layer protocol is distinct, while the behaviour of the protocols offering the same type of service is identical. For instance, instant messaging systems may result in bursty traffic, as evidenced by statistical characteristics like a brief inter arrival time and a small packet size. In order to categorise packets by service independent of encryption, the deep-learning based traffic classifier learns the circulation of statistical attributes that vary for service to service. As demonstrated in Table 2, the traffic classifier extracts 20 different types of characteristics using bidirectional ow features.

Specification	Value	Features
Standard deviation flow in packets	8	Size of packet
Inter packet time with respect to standard deviation	8	Mean arrival time
All packets in flow	2	Packets
Maximum byte size in flow	2	Bytes

Table 2. Statistical flow features.

Feature extraction and service labelling are two aspects of data preparation. The ow F=(p_1,p_2,p_3….p_n) where F has n packets and p_i is the i-th packet, is intended to be processed for its characteristics in the first portion. We need the characteristics of the reverse direction flow F and F` because our ow is bidirectional. As a result, it is possible to extract 10 different statistical feature types in one direction, and 20 different feature types may be found in a single bidirectional ow made up of F and F`. The inter-arrival time and packet size are statistical characteristics that we compute in ow F as follows.

The following statistics are generated from the behaviour vector F=(p_1,p_2,p_3….p_n) ,minimum, maximum, average, and standard deviation.

$$m_f = \min\{f_1, f_2, \ldots, f_k\},\ M_f = \max\{f_1, f_2, \ldots, f_k\}$$
$$\mu_f = \frac{1}{k}\sum_{i=1}^{k} f_i,\ \sigma_f = \frac{1}{k}\sum_{i=1}^{k}(f_i - \mu_f)^2$$

-------Eq (1).

Inter-arrival time: UNIX time (p) is used to calculate each packet's arrival time. The following is how the inter-arrival time characteristics between two packets are calculated:

$$t_i = \tau(p_i) - \tau(p_{i-1})$$

----Eq (2).

Finally, the input vector **x** is composed as follows

$$\psi = [m_t\ M_t\ \mu_t\ \sigma_t\ m_s\ M_s\ \mu_s\ \sigma_s\ n\ b]$$
$$\mathbf{x} = [\psi\ \bar{\psi}]$$

-------Eq (3).

The deep learning-based traffic classifier requires two-dimensional input, hence extra pre-processing processes, such as normalisation and reshaping, must be carried out before training. A greater capacity model should be employed to utilise complicated data in order to prevent the over fitting issue, and the deep-learning-based traffic classifier utilises a ResNet model in order to do so.

7. Dominant Feature Selection

To clarify how the deep learning model categorises traffic, we suggested a dominating feature selection approach. There are crucial data components that form the foundation of classification in situations involving classification. Using data for training that has an excessive number of or unneeded components may make the model more complicated. In actuality, more complex data may result in greater accuracy. As a result, there is a trade-off between the dimensions of the data and the accuracy of the classification; hence, the classification requires a dimension-reduction approach that maximises accuracy.

As a methodology for dimension reduction, we suggest a dominant feature selection approach based on a genetic algorithm. Here, 1 represents the amount of deleted features, and 2 represents the precision of the classification. Additionally, we maximise 1 since maximising the number of features that are dropped is equivalent to minimising the number of features that are picked.

The formulation of this issue is as follows.

$$\begin{aligned} \underset{\rho_1}{\text{maximize}} \quad & \lambda_1\rho_1 + \lambda_2\rho_2 \\ \text{subject to} \quad & \rho_1 = 0, 1, \ldots, l_{mask} \\ & 0 \le \rho_2 \le 1 \\ & \lambda_1 + \lambda_2 = 1 \\ & 0 \le \lambda_1 \le 1, 0 \le \lambda_2 \le 1 \end{aligned} \quad \text{-----Eq (4).}$$

Given that 1 is an integer, it is challenging to maximise the objective function. Furthermore, even if the mask has the same amount of zeros, 2 might be different because the location of the zero components in the mask defines the crucial piece of information for categorising the traffic. In other words, employing optimisation techniques that only modify 1, it might be challenging to maximise the target function. As a result, the suggested technique uses a genetic algorithm to determine the feature selection masks, which can maximise accuracy by taking the position of the zero components into account. A genetic algorithm is a meta-heuristic algorithm that draws inspiration from the idea of passing down the best chromosomes from one generation to the next so that the quickest can survive. The algorithm creates a mask pool to maximise the objective function using the suggested technique, which uses the chromosomes as feature selection masks.

The suggested technique creates an output based on the algorithm that optimum feature selection mask that chooses the fewest characteristics while maintaining a high level of classification accuracy. The best mask selection and progeny mask creation processes make up the several rounds of the feature selection approach. The best mask selection stage assesses the parent masks' fitting scores and chooses a couple of the finest masks using a roulette-wheel method. The offspring are produced by a crossover and mutation after mask selection. The ideal masks are produced by doing the aforementioned processes several times in order to maximise the target function indicated by the fitting score. Keep in mind that the fitting score is a measure that reflects optimality. The algorithm illustrates the full feature selection process.

8. Performance Evaluations

Submitted: 28/01/2024 Revised: 06/03/2024 Accepted: 14/03/2024

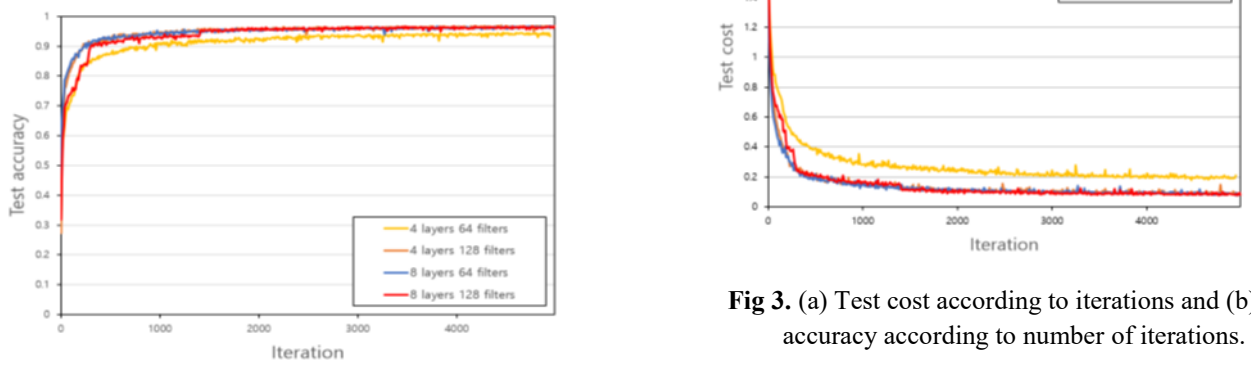
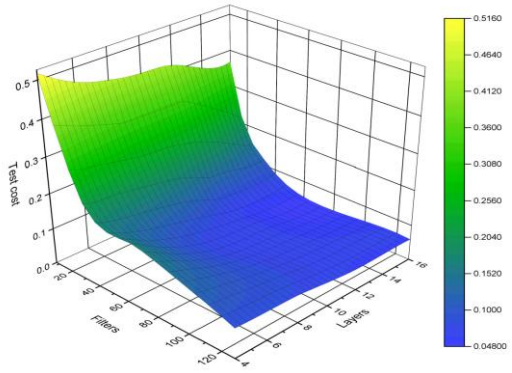


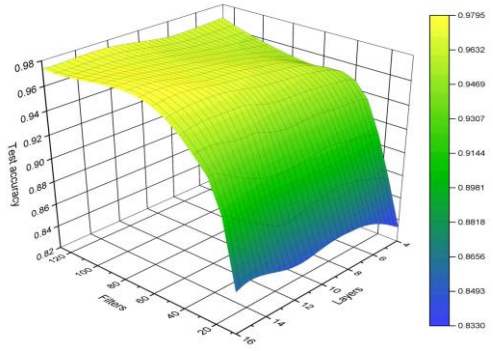
Fig 3. (a) Test cost according to iterations and (b) test accuracy according to number of iterations.

The data classification accuracy is indicated by the confusion matrix, which also aids in computing metrics like true positive, true negative, false positive, and false negative. The confusion matrix according to classes is displayed in Figure 4(a). The confusion matrix reveals that the model's overall classification accuracy is close to 96.54%. Additionally, as shown in Figure 4(b), the precision, recall, and F1-score of each service may also be computed using the confusion matrix. Recall is the ratio of the amount of data anticipated as 'A' to the entire amount of real data 'A,' where precision is the ratio of the quantity of actual data 'A' to the total amount of projected data 'A'. However, judging

performance based on accuracy and recall could be challenging since an unbalanced dataset might show inconsistent patterns in these metrics. Since the F1-score is the harmonic mean of accuracy and memory, it may illustrate how well the model performed despite the fact that precision and recall have undergone various trends.



(a)



(b)

Fig 4. (a) Test cost and (b) test accuracy according to the number of layers and filters.

The deep learning model can categorise "web surfing" with far less features than other services since it belongs to a generic class and hence has more general characteristics than other particular services. Figure 6 displays the final generation of the suggested method's fitting score, accuracy, and number of deleted features. It is evident from each service that the weight values of 1 and 2 have an impact on the average accuracy and quantity of lost features. Because discovering masks with more dropped features needs less research, fitting scores are often greater when the weight λ 1 is higher.

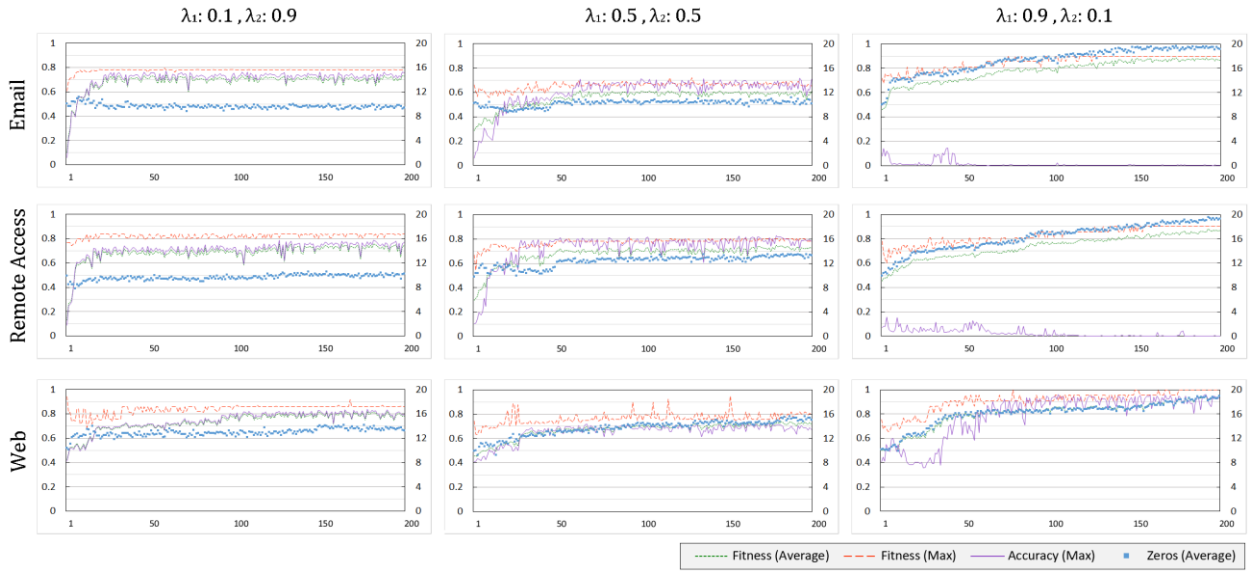


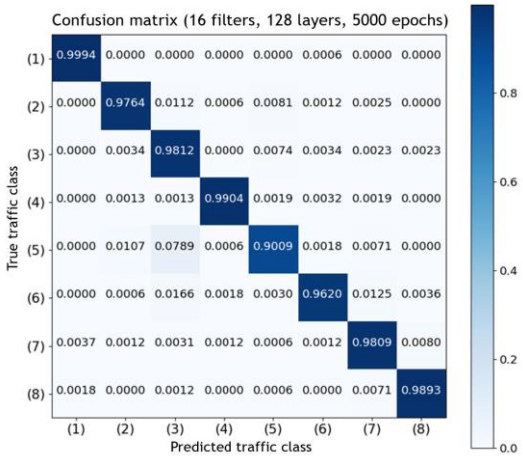
Fig 5. Fitting score, accuracy, and number of zeros per generation for 3 services

In contrast, when 2 is greater, comparatively lower fitting scores are displayed since finding the right masks with a higher degree of precision necessitates a lot more comprehensive study. The CNN model often uses a collection of local characteristics to conduct the classification process. It is crucial to identify the crucial characteristics that serve as the classification criteria for the deep learning model. The XAI is a way for explaining how a deep learning model can categorise an object's properties by separating out important elements like the eyes, nose, and ears [16]. By creating the dominant feature selection approach based on a genetic algorithm as a method of the XAI, we were able to identify the important characteristics of the flow.

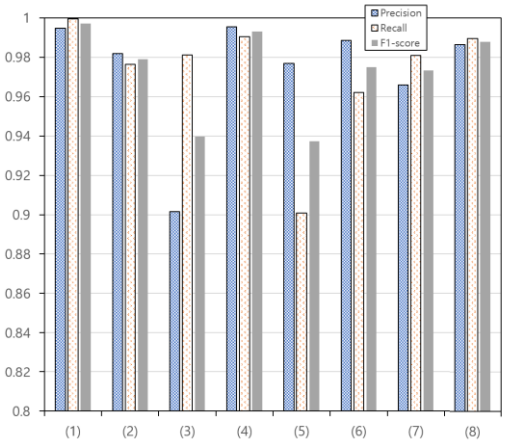
By altering the hyper-parameters, such as the accuracy and weight of the deleted features, we used the suggested dominant feature selection approach. To show how dominant each attribute is in categorising the traffic, we defined the dominance rate. The dominance rate is determined by dividing the total number of features in the experiment by the number of important features that were chosen.

$$I = \frac{K}{N} \times 100, \ 0 \le K \le N,$$

-----Eq (5).



(a)



(b)

Fig 6. (a) Confusion matrix according to traffic classes (the ratio of predicted results to the true traffic class). (b) Precision, recall, and F1-score according to each service.

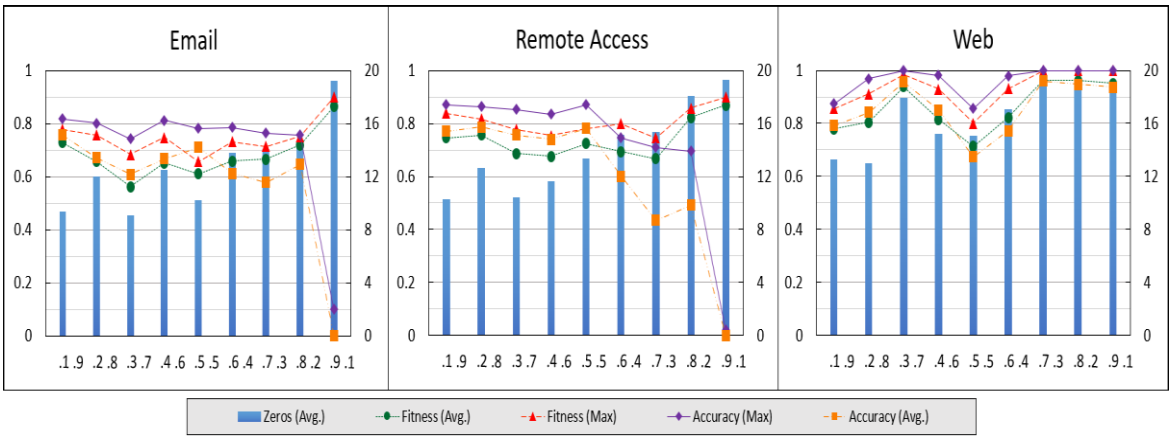


Fig 7. Number of zeros, fitting score, and accuracy depending on different weight settings.

As can be observed, the deep-learning-based traffic classifier categorises traffic into services using a subset of characteristics rather than all features. By altering the λ1 and λ2, we ran nine trials, and the results were averaged. The suggested technique creates 200 feature selection masks for each service for one experiment and selects the top 10 masks with the best accuracy after a sufficient number of rounds.

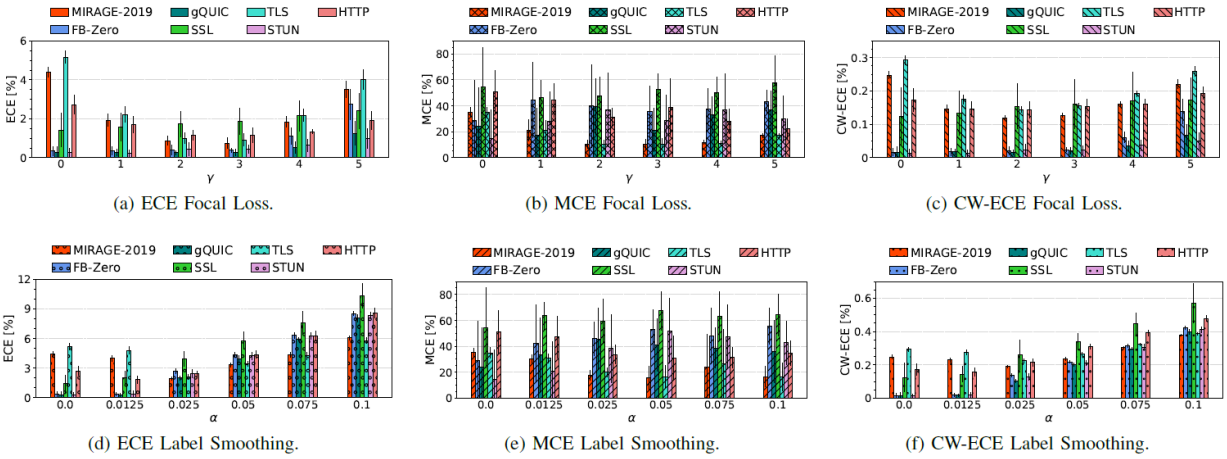
The dominance rate for each statistical variable that influences accuracy is displayed in Figure 7. If the dominance rate is high, it may be a contender for the main feature and aid improve classification accuracy or fitting score. If not, these attributes have less of an impact on how traffic is classified, making them candidates for superfluous features. A dominance rate of 100% indicates that the feature is utilised as the primary feature of the service and that the classifier always utilises it to categorise traffic into the service.

The characteristic is meaningless for classification since the dominance rate of 0% suggests that it has no impact on categorization. A feature is eliminated if its dominance rate falls below the threshold for deletion. Figure 8 displays the accuracy in relation to the threshold for removal. The number of characteristics deleted rises as the threshold for deletion does, decreasing the total classification accuracy. The relationships between characteristics are ignored since features with a low dominance rate are simply eliminated using the elimination threshold. Consequently, a fluctuation that momentarily reduces accuracy may happen when a feature that is connected to other characteristics is removed. There is minimal association for each feature in the "instant messaging" class since there is a repetitive reduction in the link between the number of characteristics deleted and accuracy. Although there may be some variations in accuracy for the "web surfing" class, overall accuracy does not decline noticeably when the number of characteristics deleted rises since most features have a low dominance rate.

Although the "web surfing" class has few contenders, the majority of the services contain a number of essential qualities. This may be explained by the fact that "web surfing" traffic has a tendency to display a common characteristic that typifies the typical Internet service behaviours. Since the majority of Internet services rely on the hyper-text transmission protocol (HTTP), for instance, the network behaviour of HTTP-based "web surfing" services might vary. Contrarily, there are numerous connections between the characteristics, and as a result, the accuracy is constantly subject to change.

Metric	Definition
Accuracy	$\text{acc} = \frac{\sum_{i=1}^{L} tp_i}{tp_i + tn_i + fp_i + fn_i}$
F-measure	$\text{F-meas} = \frac{\sum_{i}^{L} \frac{\text{prec}_i \cdot \text{rec}_i}{\text{prec}_i + \text{rec}_i}}{L}$
G-mean	$\text{G-mean} = \frac{\sum_{i}^{L} \sqrt{\text{spec}_i \cdot \text{rec}_i}}{L}$
Recall	$\text{rec}_i = \frac{tp_i}{tp_i + fn_i}$
Precision	$\text{prec}_i = \frac{tp_i}{tp_i + fp_i}$
Specificity	$\text{spec}_i = \frac{tn_i}{tn_i + fp_i}$

Table 3. F1 score description for the work.



(a) ECE Focal Loss. (b) MCE Focal Loss. (c) CW-ECE Focal Loss.

(d) ECE Label Smoothing. (e) MCE Label Smoothing. (f) CW-ECE Label Smoothing.

Fig 8. ECE and MCE works.

γ	*Accuracy [%]*	*F-measure [%]*	*G-mean [%]*
0	92.14 (± 0.28)	91.47 (± 0.27)	95.36 (± 0.19)
1	92.33 (± 0.25)	91.68 (± 0.34)	95.44 (± 0.21)
🏆 2	92.43 (± 0.31)	91.83 (± 0.35)	95.50 (± 0.19)
⚖ 3	92.21 (± 0.21)	91.61 (± 0.20)	95.39 (± 0.15)
4	92.14 (± 0.20)	91.55 (± 0.30)	95.36 (± 0.16)
5	91.90 (± 0.33)	91.32 (± 0.41)	95.24 (± 0.21)
⚖ − 🏆	− 0.21 (± 0.27)	− 0.21 (± 0.30)	− 0.10 (± 0.14)

Table 4. TC-performance Focal Loss.

α	*Accuracy [%]*	*F-measure [%]*	*G-mean [%]*
0	92.14 (± 0.28)	91.47 (± 0.27)	95.36 (± 0.19)
0.0125	92.22 (± 0.28)	91.52 (± 0.26)	95.37 (± 0.18)
⚖ 0.025	92.39 (± 0.27)	91.77 (± 0.35)	95.41 (± 0.22)
0.05	92.44 (± 0.36)	91.90 (± 0.44)	95.44 (± 0.28)
0.075	92.47 (± 0.24)	91.88 (± 0.30)	95.40 (± 0.19)
🏆 0.1	92.58 (± 0.26)	92.01 (± 0.31)	95.48 (± 0.21)
⚖ − 🏆	− 0.19 (± 0.35)	− 0.24 (± 0.42)	− 0.07 (± 0.26)

Table 5. TC-performance Label Smoothing.

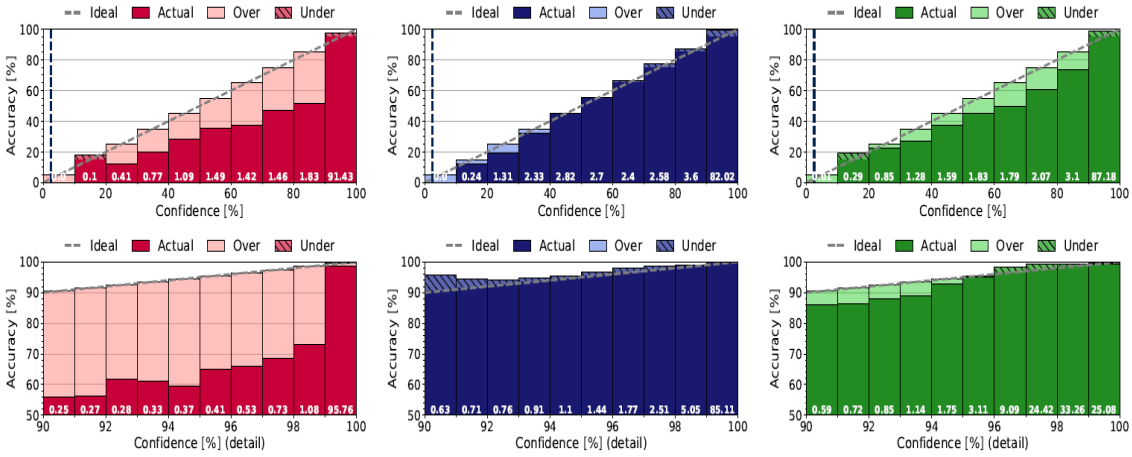


Fig 9. Accuracy and confidence details.

9. Conclusion

We worked out a deep learning based model for traffic classification with the support of genetic algorithm. ResNet model used to explain more by using XAI for traffic classification. A dominant feature selection was entitled here to create an ideal feature selection mask along with GA (Genetic Algorithm). The suggested explanatory technique creates the best feature assortment covers by joining the traffic classifier's DL results onto the chromosomal assessment in a genetic algorithm. By weighing the interchange of the accuracy of the classifier and the quantity of superfluous features, the article assortment masks are used to eliminate the significant feature subclass from the complete feature set. The stochastic nature of a genetic algorithm was reflected in a number of tests, and the importance rate was calculated using feature selection masks. By examining the salient characteristics of each Internet service, we were able to define the operation of the DL centered traffic classifier consuming the prominence rate. We intend to create a key feature selection method in the future for more precise application-specific traffic classifiers. In order to allow real-time important feature selection, we will also speed up the genetic algorithm's convergence.

References

[1] H. Hagras, "Toward human-understandable, explainable AI," IEEE Computer, vol. 51, no. 9, pp. 28–36, 2018.

[2] M. Kull, M. Perello-Nieto, M. Kängsepp, H. Song, P. Flach et al., "Beyond temperature scaling: Obtaining well-calibrated multiclass probabilities with dirichlet calibration," in 32th Conference on Neural Information Processing Systems (NeurIPS), 2019.

[3] Montieri, D. Ciuonzo, G. Bovenzi, V. Persico, and A. Pescape, ``A dive into the dark Web: Hierarchical traf_c classi_cation of anonymity tools," IEEE Trans. Netw. Sci. Eng., vol. 7, no. 3, pp. 1043_1054, Jul. 2020.

[4] L. Grimaudo, M. Mellia, and E. Baralis, ``Hierarchical learning for _ne grained Internet traf_c classi_cation," in Proc. 8th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC), Aug. 2012, pp. 463-468.

[5] W. Zheng, C. Gou, L. Yan, and S. Mo, ``Learning to classify: A flow-based relation network for encrypted traffic classification," in Proc. Web Conf., Apr. 2020, pp. 13-22.

[6] M. I. Thariq Hussan, D. Saidulu, P. T. Anitha, A. Manikandan and P. Naresh (2022), Object Detection and Recognition in Real Time Using Deep Learning for Visually Impaired People. IJEER 10(2), 80-86. DOI: 10.37391/IJEER.100205.

[7] Niculescu-Mizil and R. Caruana, "Predicting good probabilities with supervised learning," in 22nd International Conference on Machine learning (ICML), 2005, pp. 625–632.

[8] D. Widmann, F. Lindsten, and D. Zachariah, "Calibration tests in multiclass classification: A unifying framework," in 33th Conference on Neural Information Processing Systems (NeurIPS), 2019.

[9] J. Mukhoti, V. Kulharia, A. Sanyal, S. Golodetz, P. H. Torr, and P. K. Dokania, "Calibrating deep neural networks using focal loss," in 34th Conference on Neural Information Processing Systems (NeurIPS), 2020.

[10] Naresh, P., & Suguna, R. (2021). IPOC: An efficient approach for dynamic association rule generation using incremental data with updating supports. Indonesian Journal of Electrical Engineering and Computer Science, 24(2), 1084. https://doi.org/10.11591/ijeecs.v24.i2.pp1084-1090.

[11] W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K. R. Müller, "Explaining deep neural networks and beyond: A review of methods and applications," Proceedings of the IEEE, vol. 109, no. 3, pp. 247–278,2021.

[12] W. Wang, M. Zhu, X. Zeng, X. Ye, and Y. Sheng, ``Malware traffic classification using convolutional neural network for representation learning,"in Proc. Int. Conf. Inf. Netw. (ICOIN), 2017, pp. 712-717.

[13] B. Narsimha, Ch V Raghavendran, Pannangi Rajyalakshmi, G Kasi Reddy, M. Bhargavi and P. Naresh (2022), Cyber Defense in the Age of Artificial Intelligence and Machine Learning for Financial Fraud Detection Application. IJEER 10(2), 87-92. DOI: 10.37391/IJEER.100206

[14] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescape, ``Mobile encrypted traffic classification using deep learning: Experimental evaluation, lessons learned, and challenges," IEEE Trans. Netw. Service Manage., vol. 16, no. 2, pp. 445_458, Jun. 2019.

[15] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapè, ``MIMETIC: Mobile encrypted traf_c classi_cation using multimodal deep learning," Com-put. Netw., vol. 165, Dec. 2019, Art. no. 106944.

[16] P. Naresh, S. V. N. Pavan, A. R. Mohammed, N. Chanti and M. Tharun, "Comparative Study of Machine Learning Algorithms for Fake Review Detection with Emphasis on SVM," 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2023, pp. 170-176, doi: 10.1109/ICSCSS57650.2023.10169190.

[17] Nagesh, C., Chaganti, K.R. , Chaganti, S. , Khaleelullah, S., Naresh, P. and Hussan, M. 2023. Leveraging Machine Learning based Ensemble Time Series Prediction Model for Rainfall Using SVM, KNN and Advanced ARIMA+ E-GARCH. International Journal on Recent and Innovation Trends in Computing and Communication. 11, 7s (Jul. 2023), 353–358. DOI:https://doi.org/10.17762/ijritcc.v11i7s.7010.

[18] M. Sadeghzadeh, S. Shiravi, and R. Jalili, "Adversarial network traffic: Towards evaluating the robustness of deep-learning-based network traffic classification," IEEE Transactions on Network and Service Management, vol. 18, no. 2, pp. 1962–1976, 2021.

[19] S. Khaleelullah, P. Marry, P. Naresh, P. Srilatha, G. Sirisha and C. Nagesh, "A Framework for Design and Development of Message sharing using Open-Source Software," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 639-646, doi: 10.1109/ICSCDS56580.2023.10104679.

[20] W. Wang, M. Zhu, J. Wang, X. Zeng, and Z. Yang, "End-to-end encrypted traffic classification with one-dimensional convolution neural networks," in IEEE International Conference on Intelligence and Security Informatics (ISI), 2017, pp. 43–48.

[21] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, ``Show, attend and tell: Neural image caption generation with visual attention," in Proc. Int. Conf. Mach. Learn., 2015, pp. 2048-2057.

[22] B.M.G. Prasad, P. Naresh, V. Veeresh, "Frequent Temporal Patterns Mining With Relative Intervals",

[23] International Refereed Journal of Engineering and Science ,Volume 4, Issue 6 (June 2015), PP.153-156.

[24] T. Aruna, P. Naresh, A. Rajeshwari, M. I. T. Hussan and K. G. Guptha, "Visualization and

Prediction of Rainfall Using Deep Learning and Machine Learning Techniques," 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022, pp. 910-914, doi: 10.1109/ICTACS56270.2022.9988553.

[25] J. Zhang, X. Chen, Y. Xiang,W. Zhou, and J.Wu, ``Robust network traffic classification," IEEE/ACM Trans. Netw., vol. 23, no. 3, pp. 1257-1270, Aug. 2015.