International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

# A Dominant Feature Selection Method for Deep Learning Based Traffic Classification Using a Genetic Algorithm

**Uma Maheswari Gali[1], Yasmeen[2], Mudimela Madhusudhan[3], Ravindra Changala[4], Dr. Mahesh Kotha[5]**

[1]Assistant Professor, CSE Department, Sri Indu college of Engineering & Technology, Hyderabad, India
[2]Assistant Professor, CSE Department, CMR Technical Campus Hyderabad, India
[3]Assistant Professor, Guru Nanak Institutions Technical Campus, Hyderabad, India
[4]Assistant professor, CSIT Department, CVR College of engineering, Hyderabad, India
[5]Assistant Professor, Department of CSE (AI&ML), CMR Technical Campus, Hyderabad, India

**Article Info**

**Publication Issue :**
Volume 8, Issue 6
November-December-2022

**Page Number :** 173-181

**Article History**
Accepted: 05 Nov 2022
Published: 18 Nov 2022

## ABSTRACT

Internet data handling is becoming a challenging issue to networking organizations now a days. To have an efficient throughput of network functions various traffic classification techniques were propose so far. These types describe encrypted traffic classification which uses the support of deep learning approaches. While packet inspecting payload is another stuff of classification. A malfunction of the deep learning model may occur if the training dataset includes malicious or erroneous data. Security and confidentiality of users data while in networking is another open issue which can be solved by using Explainable artificial intelligence (XAI) somehow. In this paper, we propose a strategy for making sense of the functioning system of deep learning-based traffic grouping as a technique for XAI in view of a hereditary calculation. We depict the component of the deep learning-based traffic classification by measuring the significance of each element by using genetic algorithms. Moreover, we influence the hereditary calculation to produce an element determination veil that chooses significant highlights in the whole list of capabilities. To exhibit the proposed clarification technique, we carried out a deep-learning based traffic classifier with an exactness of roughly 96.55%. Likewise, we present the significance of each component got from the proposed clarification technique by characterizing the predominance rate.

**Keywords :** XAI, classification, deep learning, genetic algorithms, security, network traffic.

## I. INTRODUCTION

With the expansion of Web associated gadgets and different Internet providers, it is vital to control the huge traffic volume in an efficient way. Traffic classification can be utilized to control different kinds of traffic in software designed organizing (SDN) or to recognize noxious traffic in network interruption discovery framework (NIDS) [1]. On account of SDN, QoS the executives is critical to alleviate the weight of the whole organization and to fulfill the prerequisites of each kind of administration [2]. As the Internet providers are more different, It is vital to give every Network access the differential QoS. Dynamic QoS can give the differential QoS by partitioning the QoS class to help a more intricate QoS.

Moreover, in light of the fact that various gadgets are associated to the Web, the significance of advances for distinguishing furthermore, safeguarding against different assaults that might happen on the network has been accentuated. NIDS fills in as a center capability in network security by identifying assaults, for example, the refusal of administration (DoS) assault in light of traffic classification. Customary traffic classifications (TCs) are generally based on a payload-investigation, which is known as a payload-based TC. A payload-based TC straightforwardly reviews the payload of parcels furthermore, matches the pre-defined designs. Albeit a payload based TC shows a superior exhibition, there are two basic issues. One issue is that payload-based TC can't investigate the scrambled payload. Since secure correspondence plans, for example, SSH and TLS encode the payload, the payload-based approaches can't review the payload mixed by the encryption conspire. Another issue is that reviewing the payload of bundles requires tremendous computational assets.

## II. RELATED WORK

We propose a dominant feature selection method to explain how the proposed deep learning-based traffic classier operates. We define a fitting score as a feature importance quantification and create a feature selection mask that finds the optimal trade-off between the high classification accuracy and a reduction of the unnecessary features based on a genetic algorithm. A genetic algorithm is an evolutionary algorithm that can solve the various NP-hard problems such as the traveling salesman problem (TSP) or the design of very large scale integration (VLSI). Finally, we describe the deep learning based traffic classifier by defining a dominance rate indicating the extent to which each deep learning model refers to each feature. In conclusion, the proposed method has two technical contributions.

We propose a dominant feature selection method using a genetic algorithm to explain how the deep learning based traffic classifier operates. In particular, the proposed method can determine which part of the entire feature the classifier focuses on by quantifying the importance of each feature.

We implement the flow-behavior-based traffic classifier as the evaluation method that classifies the traffic and produces the accuracy to compute the fitting score. Although the proposed method also works well in any granularity of the types of traffic, we implement a service specific traffic classification model to figure out the characteristics of internet services.

Customary traffic classification (TCs) are typically founded on a payload-review, which is known as a payload-based TC. A payload-based TC straightforwardly investigates the payload of bundles and matches the precharacterized designs. Deep learning model a payload-based TC shows an elite presentation, there are two basic issues. One issue is

that payload-based TC can't review the encoded payload. Since secure communication plans, for example, SSH and TLS encode the payload, the payload-based approaches can't review the payload mixed by the encryption conspire. Another issue is that investigating the payload of parcels requires huge computational assets.

Behavior statistics became a clue for classifying encrypted traffic because the statistics can be extracted without inspecting a scrambled payload. Flow-behavior-based approaches enable encrypted traffic to be classified by leveraging the behavior statistics. The authors of [8] introduced three representative encryption mechanisms of traffic and extracted the statistics from the encrypted traffic. Moreover, they evaluated the performance of several machine-learning algorithms such as a support vector machine, random forest, naive Bayes, logistic regression, and neural networks. They presented the practicality of flow-behavior-based approaches by evaluating various machine-learning technologies.

The core advantage of deep learning over traditional machine learning technologies is to enable the classifier to automatically extract features from the raw data. The authors of [9] adopted a convolutional neural network (CNN) for traffic classification. Representation learning is a method used to automatically extract features from raw data and the CNN is a typical method of representation learning in deep learning. The convolution layer enables a CNN to extract the local features from the raw data. The authors integrate feature extraction and training by leveraging the advantages of the CNN. In [1], the authors evaluated the two different types of typical deep learning models, a CNN and a recurrent neural network (RNN).

The authors suggested that traffic classification schemes using a manually extracted feature set for mobile traffic generated by a moving target are impractical. In addition, they address the limitations of traditional traffic classification schemes by leveraging the advantages of deep learning, which can automatically extract the feature set.

## III. WORKING OF NOVEL MODEL

Revising and applying a novel model is necessary because of the nature of features shown by behavior statistics. In [12], the authors described an issue in which many studies on deep learning-based traffic classification have usually adopted all the features equally without considering the type of statistics. The authors reflect the multimodality of behavioral statistics using a multimodal deep learning model. Consideration of traffic generated by anonymity tools (ATs) was introduced in [13]. Because it becomes important to preserve the privacy of users on the Internet, several ATs have been developed, including Tor.

Consequently, several malicious usages of ATs produce crucial issues. The authors proposed an AT-specific traffic classification by leveraging a hierarchical classification that enables an efficient fine-grained tuning. The authors of [4] proposed a traffic classification scheme using a hierarchical classification. Flow-behavior-based approaches have a disadvantage in that they cannot classify unknown traffic classes because of the nature of machine-learning. Moreover, increasing the granularity of the traffic class exacerbates the classification performance. The authors compose the sub-classifier hierarchically based on the granularity of the traffic class.

## IV. EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI)

Explainable artificial intelligence (XAI) techniques have been studied to demonstrate the mechanism of the machine learning model. In [6], the authors proposed the concept of explainable artificial

intelligence (XAI), and devised a novel explainable model that allows the machine learning model to derive such classification results based on the feature subset of the input data. In [16], the authors visualized the important features that are used for classifying certain input data
and explained why the deep learning model can recognize such data.

To explain this, the authors proposed a sensitivity analysis (SA) that explains the impact of the transition of each pixel. Moreover, the authors also proposed layer-wise relevance propagation (LRP), which explains the importance of each pixel. In [7], the authors proposed EXPLAIN-IT, a framework that explains how to cluster an unlabeled YouTube traffic dataset acquired from the network using an unsupervised learning technique. EXPLAIN-IT explains the clustering method using LIME, which selects the feature most relevant to a specific decision from the input data. Hence, the key feature selection can explain why the deep learning model classifies the data. In [18], the authors describe the relationship between the input and output by inserting artificial perturbations in certain features. The input-output relationship can provide some interpretation rules for black box predictors such as deep learning. Neural image caption generation with a visual attention scheme is proposed in [9].
The authors extracted the key features in the image using convolutional feature extraction. The extracted features are used to train the RNN for image captioning. During this procedure, the attention mechanism implemented through a convolutional feature extraction can highlight an important part of the image.

## V. THE PROPOSED METHOD

The overview of the proposed dominant feature selection method is illustrated in Figure 1. The proposed method consists of two parts: (1) the construction of a traffic classifier, and (2) dominant feature selection. The traffic classifier is designed by a residual network (ResNet), which is known
as a state-of-the-art deep learning technique [20]. The traffic classification applies data pre-processing and training step. The data pre-processing step collects packets from traffic flows and extracts the statistical features of each flow. After the pre-processing step, in which the traffic dataset is created, the traffic classifier is trained using the dataset composed of statistical features.
After the classifier is trained, the proposed dominant feature selection method generates a feature selection mask based on a genetic algorithm.

The dominant feature selection conducts a mask selection and an offspring mask generation. The mask selection evaluates the masks by counting the zero elements and calculating the accuracy using a masked input dataset and a pre-trained classifier. After the evaluation, with the mask selection picks a few masks are chosen using a roulette wheel selection method for the mask creation of the next generation. With the roulette wheel selection method, the probability of selecting the masks with a higher fitting score is higher than the others. The offspring mask generation creates a mask pool using the selected masks and gives variety to the mask pool through a crossover and mutation. The mask pool generated by the offspring mask generation is inherited by the next generation. After the iteration of two steps, the feature selection masks for each service are made and the masks are used to pick the features necessary to classify each service from all statistical features. Finally, we analyze the mechanism of the traffic classifier by computing the importance of each feature using the feature selection masks.

## VI. THE CONSTRUCTION OF A TRAFFIC CLASSIFIER

The construction of a traffic classifier consists of three steps: packet gathering, data pre-processing, and classifier training. The packet gathering step collects packets and groups them by the traffic to construct the training dataset. Because most packets are encrypted, a packet itself cannot be used as a training dataset, although the grouped packet dataset that shares the same end-to-end network address such as the IP address or TCP port number is needed. The packets in a grouped dataset may serve the same application service because they have the same application source, and such packets form a network flow.

The packets in a network flow have a similar behavior, which is represented by the statistical features such as the packet size and inter-arrival time. After gathering packets from a network flow, the data pre-processing step computes the statistical features from the group of gathered packets. Finally, the classifier training step trains the deep-learning based classifier using the dataset. Algorithm 1 shows the procedure of the construction of a traffic classifier.

### DATA PREPROCESSING

The data pre-processing step computes the statistical features of the bidirectional flow set shown in lines 4-14 of Algorithm 2. The behaviors of the packets in the network are represented as statistical features, which are mainly revealed by the inter-arrival time, packet size, number of packets, and number of bytes [2]. Although the packets are encrypted, the packets serving the same application layer protocol have unique behaviors, and the protocols that serve a similar type of service show similar behaviors. For example, instant messaging services can cause bursty traffic, which can be shown in the statistical features such as a relatively short inter-arrival time and packet size. Therefore, the deep-learning-based traffic classifier can classify packets by service regardless of encryption by learning the distribution of statistical features that are different for each service. The traffic classifier uses bidirectional flow features and extracts 20 types of features, as shown in Table 1.

| Features | Description | Value |
|---|---|---|
| Packet Size | The maximum, minimum, average, standard deviation of packet sizes in a flow. | 8 |
| Inter Arrival Time | The maximum, minimum, average, standard deviation of inter packet time in a flow. | 8 |
| Packets | The total number of packets in a flow | 2 |
| Bytes | The amount of bytes in a flow | 2 |

Table 1. Statistical flow features

### DOMINANT FEATURE SELECTION

We proposed a dominant feature selection method to explain how the deep learning model classifies traffic. In classification problems, there are key elements within the data that are the basis for classification. For example, in natural language processing (NLP), the subject and verb are the key elements, and the others are qualifiers used to explain them in the word tokens. Utilizing data with too many or unnecessary components for training may cause a higher complexity of the model. In fact, data with many components may lead to a higher accuracy. In other words, the classification accuracy also decreases because low-dimensional data have less information for the decision. Therefore, there is a trade-off between the classification accuracy and dimensions of the data, and the classifier needs a dimension-reduction technique that maximizes the accuracy.

We propose a dominant feature selection method based on a genetic algorithm as a dimension reduction technique. The aim of the proposed method is to find the optimal feature selection masks, minimizing the number of selected features and maximizing the classification accuracy. Hence, we formulated the objective function as a linear combination of two factors, namely, the number of masked elements and the classification accuracy.

**Algorithms:** Procedure of Dominant Feature Selection

**Require:** Pre-trained traffic classifier, flow statistical feature dataset $\mathbf{x}$, the weight of dropped feature numbers $\lambda_1$, the weight of classification accuracy $\lambda_2$, the number of whole generation $N$, the number of individuals in a population $M$.

**Ensure:** Optimal feature selection mask $\theta^\star$ of a service.

Randomly generates the initial population $\theta_0$.
**for** $i = 0 \rightarrow N-1$ **do**
  $\kappa_i = \varnothing$
  **for** $j = 1 \rightarrow M$ **do**
    Pick a individual $\theta_i^j$ from the $i$-th generation population $\theta_i$.
    Compute $\rho_1^j$, which is the number of zeros in $\theta_i^j$.
    $\mathbf{x}' = \theta_i^j \circ \mathbf{x}$
    Compute the accuracy $\rho_2^j$ from the pre-trained traffic classifier using $\mathbf{x}'$.
    Compute $\kappa_i^j = \lambda_1 \rho_1^j + \lambda_2 \rho_2^j$
    $\kappa_i = \kappa_i \cup \{\kappa_i^j\}$
  **end for**
  Compute best individuals $\hat{\theta}_i$ by truncating the population based on $\kappa_i$
  $\theta_{i+1} = \varnothing$
  **for** $j = 1 \rightarrow M$ **do**
    Decide to perform elitism, crossover and mutation operation in Monte-Carlo manners.
    **if** Perform elitism operation **then**
      Randomly pick a individual $\hat{\theta}_i^1$ from $\hat{\theta}_i$
      $\theta_{i+1}^j = \hat{\theta}_i^1$
    **end if**
    **if** Perform crossover operation **then**
      Randomly pick two individuals $\hat{\theta}_i^1, \hat{\theta}_i^2$ from $\hat{\theta}_i$
      Compute $\theta_{i+1}^j$ by crossover operation.
    **end if**
    **if** Perform mutation operation **then**
      Randomly pick two individuals $\hat{\theta}_i^1, \hat{\theta}_i^2$ from $\hat{\theta}_i$
      Compute $\theta_{i+1}^j$ by mutation operation.
    **end if**
    $\theta_{i+1} = \theta_{i+1} \cup \{\theta_{i+1}^j\}$
  **end for**
**end for**

The population of selected chromosomes consists mostly of masks evaluated with a high fitting score and is inherited by the next generation. The masks converged to the optimal masks that have a high fitting score from sufficient iterations of the survival of the fittest.

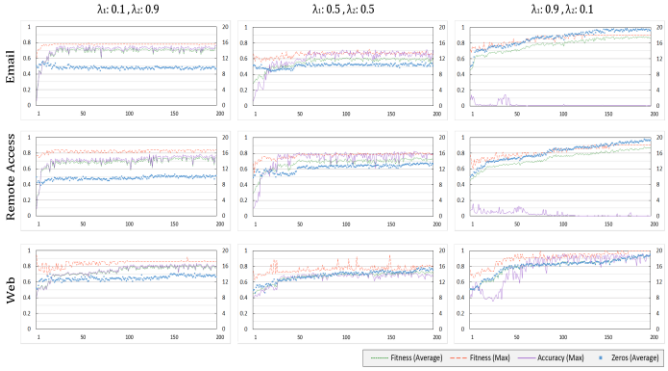


Fig 1. Fitting score, accuracy, and number of zeros per generation for 3 services

## VII. PERFORMANCE EVALUATION AND EMPIRICAL ANALYSIS

In this section, we describe the performance of the deep learning-based traffic classifier used to evaluate the accuracy, learning cost. To evaluate the performance of the proposed method, we carried out numerous experiments using real world data. For fair evaluations, the public pcap datasets are used to build the training dataset. We adopted public pcap datasets from ISCX VPN-nonVPN, MACCDC, and WRCCDC, which have also been frequently used in other studies in traffic classification and include both encrypted and non-encrypted packets.

Although the public pcap dataset has many packets that operate various protocols, the number of flows grouped by packets that share the same 5-tuple is insufficient to train deep-learning-based traffic classification model. To supply more training data, we gather the additional pcap data utilizing the server which generates packets of various protocols from a campus network. As a result, the entire dataset is composed of 49 applications, as shown in Table 1. In Table 1, a number column represents the number of

flows. We set the number of data by each service similar to avoid biased training. Moreover, for practical use, packets of one flow are gathered for 900 seconds without considering a TCP session timeout.

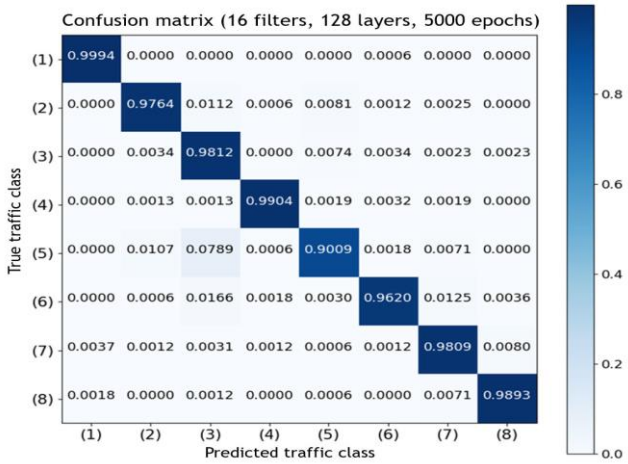


Fig 2. Confusion matrix according to traffic classes

For training the deep learning model, we divided the 70% of the dataset into training dataset and 30% into test dataset, and all evaluations are based on the test dataset. The parameters are initialized at random, and a batch normalization layer is used to mitigate the effort required to regularize the parameters by forming a similar distribution in each layer. There are some hyper-parameters to be tuned before the training, such as batch size and number of epochs. We found the

two hyper-parameters above through an adequate number of experiments with a batch size of 300 and 5,000 epochs. Moreover, we conducted experiments by adjusting other hyperparameters such as the number of filters in the convolution layer and the number of layers in the residual block. Note that one residual block consists of several convolution layers and batch normalization layers, and the entire model is constructed by stacking several residual blocks. Figures 2 show the test cost and test accuracy according to the iterations. We conducted experiments by changing the number of layers from 4 to 8 and using 64 and 128 filters. It can be seen that the greater the number of filters and the number of layers, the higher the classification accuracy, and the faster the cost convergence.

Therefore, the result shows that the number of dropped features is much higher than that of the other two results with different weights. However, it shows the lowest accuracy result in general. In addition, it can be observed that the result of a "web surfing" service shows a relatively higher

accuracy even when the weight _1 is higher and _2 is lower. That is because a web surfing service is a generic class and therefore has more general characteristics than other specific services, it allows the deep learning model to require much fewer features to classify the "web surfing" than the other services.

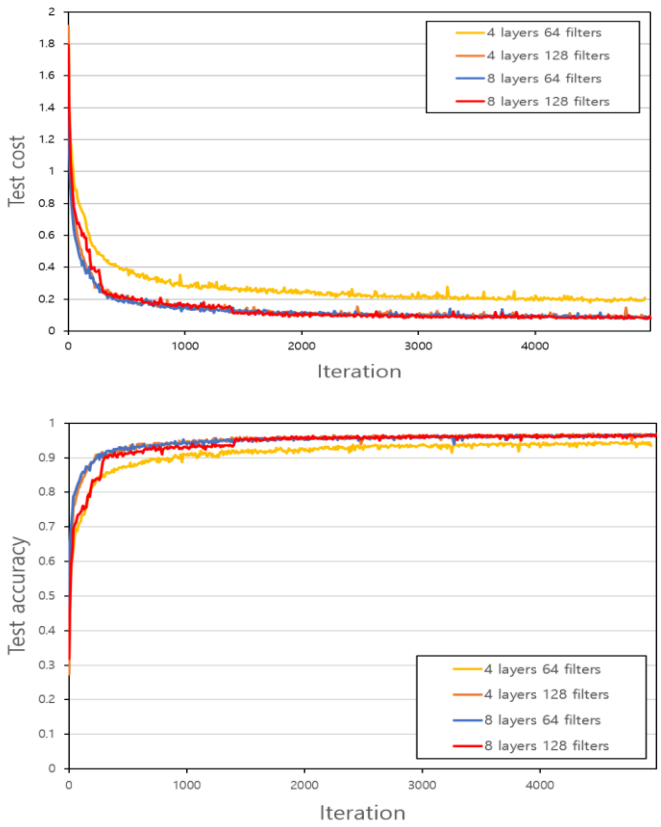


Fig 3. Test cost according to iterations and test accuracy according to number of iterations.

In general, the CNN model operates the classification process by collecting several local features. It is important to discover key features that are used as

criteria used for the deep learning model to classify. The XAI is an explanatory method that describes how the deep learning model can classify the characteristics of a certain object by extracting the key features such as eyes, nose, and ear [16]. We found the key features of the flow by designing the dominant feature selection method based on a genetic algorithm as a method of the XAI. We applied the proposed dominant feature selection

It can be seen that the deep-learning-based traffic classifier does not classify traffic into a service using all features, but classifies it using only specific features. We conducted nine experiments by changing the 1 and 2 and averaged the results of the experiments. When the proposed method completes a sufficient number of iterations, it generates 200 feature selection masks of each service for one experiment and picks the top-10 masks, which achieve the highest accuracy. Figure 3 shows the dominance rate for each statistical feature that affects the accuracy. If the dominance rate is high, it can help increase the classification accuracy or fitting score, namely, it can be a candidate for the key feature.

Otherwise, these features have less influence on the traffic classification, and thus they can be candidates of unnecessary features. A dominance rate of 100% implies that the classifier always uses the features to classify the traffic into the service, namely, the feature is used as the core feature of the service. Because the 0% dominance rate implies that the feature does not affect the classification, it indicates that the feature is irrelevant for classification. To measure how much each feature contributes to the classification accuracy, we define the elimination threshold and measure the accuracy of each elimination threshold. If the dominance rate of a feature is lower than the elimination threshold, the feature is removed.

Figure 2 shows the accuracy according to the elimination threshold. As the elimination threshold increases, the number of features removed increases, and thus the overall classification accuracy tends to decrease. Because features with a low dominance rate are simply removed through the elimination threshold, the correlations between features are not considered. Consequently, when removing a feature that is related to other features, fluctuation that temporarily decreases the accuracy may occur. In the case of the "instant messaging" class, the relationship between the number of removed features and accuracy has a monotonous decrease, which means that there is little correlation for each feature. In the case of the "web surfing" class, there may be slight fluctuations in accuracy, although the overall accuracy does not decrease significantly as the number of removed features increases because most of the features have a low dominance rate. Most of the services have several key features, although the "web surfing" class has few candidates.

## VIII. CONCLUSION

In this study, we proposed a explanatory method of the deep learning-based traffic classifier based on a genetic algorithm. Further, we implemented the deep-learning-based traffic classifier based on the ResNet model for demonstrating the proposed explanatory method. We designed the dominant feature selection method as a explanatory method based on a genetic algorithm to generate an optimal feature selection mask. The proposed explanatory method generates the optimal feature selection masks by grafting the deep-learning based traffic classifier's result onto the evaluation of the chromosome in a genetic algorithm. The feature selection masks are used to extract the key feature subset from the entire feature set by considering the trade-off between the classifier's accuracy and the number of unnecessary features. We conducted several experiments for reflecting the stochastic property of a genetic

algorithm and computed the importance rate through the feature selection masks. Through the importance rate, we explained the mechanism of the deep-learning-based traffic classifier by investigating the key features of each Internet service. In the future, we plan to design a key feature selection algorithm for finger-grained application-specific traffic classifiers. In addition, we will improve the convergence speed of the genetic algorithm to enable real-time key feature selection.

## IX. REFERENCES


1. 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.
2. A. Adadi and M. Berrada, ``Peeking inside the black-box: A survey on explainable artificial intelligence (XAI),'' IEEE Access, vol. 6, pp. 52138_52160, 2018.
3. Ravindra Changala, "A Survey1 on Clustering Techniques to Improve Energy Efficient Routing in Wireless Sensor Networks" in International Journal of Applied Engineering Research, 10(58), pp.-1-5,2015.
4. M. Usama, A. Qayyum, J. Qadir, and A. Al-Fuqaha, ``Black-box adversarial machine learning attack on network traffic classification,'' in Proc.15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC), Jun. 2019, pp. 84_89.
5. D. Gunning, ``Explainable artificial intelligence (XAI),'' in Defense Advanced Research Projects Agency (DARPA), nd Web. Arlington, VA, USA: Defense Advanced Research Projects Agency (DARPA), 2017.
6. Ravindra Changala, "Diminution of Deployment Issues in Secure Multicast System with Group Key Management" published in International Journal of Computer Application (IJCA), Impact Factor 2.52, ISSN No. : 2250-1797, Volume 2, Issue 3, June 2012.
7. K. Zhou, W. Wang, C. Wu, and T. Hu, ``Practical evaluation of encrypted traffic classification based on a combined method of entropy estimation and neural networks,'' ETRI J., vol. 42, no. 3, pp. 311_323, Jun. 2020.
8. Ravindra Changala, "Intrusion Detection System Using Genetic Algorithm" published in International Journal of Emerging Trends in Engineering and Development [IJETED], Impact Factor 2.87, ISSN NO:2249-6149, Issue 2,Vol. 4 May 2012.
9. 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.
10. Ravindra Changala, "Secured Activity Based Authentication System", in Journal of innovations in computer science and engineering (JICSE), Volume 6, Issue 1,Pages 1-4, September 2016.ISSN: 2455-3506.



**Cite this article as :**

Uma Maheswari Gali, Yasmeen, Mudimela Madhusudhan, Ravindra Changala, Dr. Mahesh Kotha, "A Dominant Feature Selection Method for Deep Learning Based Traffic Classification Using A Genetic Algorithm", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 6, pp. 173-181, November-December 2022.
Journal URL : https://ijsrcseit.com/CSEIT228624