JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Network Traffic Classification Techniques: A Review

[1]Namita, [2]Rahul
[1]Research Scholar, [2]Assistance Professor
[1,2]Department of Computer Science and Engineering,
Ganga Institute of Technology and Management, Kablana,

Abstract : The network traffic classification task is focused on recognizing diverse kinds of applications or traffic data for which the received data packets are analyzed that is essential in communication networks in these days. A network controller must have efficient understanding of applications and protocols in the network traffic to deploy the suitable security solutions. In addition, the name or kind of application is recognized and classified in the network for treating some aspects in advance. The process to classify the network traffic has become popular among research community along with the industrial field. A number of schemes have been put forward and constructed over the last two decades. The network traffic can be classified in several stages, in which pre-processing is done, attributes are extracted and classification is performed. The various machine learning models are reviewed in this paper for the network traffic classification.

Keywords - Network Traffic Classification, Machine learning, Classification, KDD

I. Introduction

The escalating growth of network technology has pushed the expansion of the forms and amounts of traffic data in the virtual world. Identifying and classifying the traffic flow within a network is an important research work in the field of network protection and management. It is the basis for dynamic access control, network resource planning, content-based invoicing, interference and malware finding among others. Service quality guarantee, dynamic access control and anomalous network behaviour recognition are some of the tasks in which classifying traffic efficiently and accurately is of great realistic significance [1]. Network traffic classification (NTC) innovations based on port and deep packet inspection are steadily deprecated with full traffic encryption. Machine learning (ML) technology has grown into the best performing and prevailing approach. A great deal of work on efficient traffic feature mining and the search for optimum classification networks have appeared in the education sector, and have yielded promising outcomes. Nevertheless, the majority of works ignored two big challenges. First, network traffic leads to a usual imbalance distribution. There is broad difference in the amount of traffic produced by the various protocols and applications. Next, the design objective of almost all machine learning algorithms is to maintain the maximum generalized accuracy irrespective of class imbalance, which switches the classifier's training to the majority class. Typically, the class with the greatest and the smallest sample dimension is respectively known as the majority and minority class. As a result, the productivity of existent network traffic classification-based machine learning approaches deteriorates greatly in practical unbalanced traffic classification operations [2]. In some cases, for example network control, intrusion finding, etc. where high-quality traffic occupies only a small portion, the performance deterioration on the minority class is disastrous. Hence, it is required to give sufficient consideration to the imbalance problem in NTC.

1.1 Description of Class imbalance

There are several applications that give rise to natural allocation of skewed data. In these applications, positive class arises with low frequency, such as data originated in disease diagnostics, fraud detection, computer security and image recognition. At one side, internal imbalance occurs due to normally existing data frequencies, for example, clinical diagnosis where most patients are physically fit. On the other side, external imbalance is the result of external factors, for example accumulation or storage processes. The representative of minority and majority classes must be considered when learning from unbalanced data. It is possible to obtain high-quality results without considering class inequality through good representation of both groups coming from non-overlapping distributions. To study the impacts of class imbalance, few investigators have created artificial data sets with different blends of complexity, training set dimension, and imbalance levels [3].

According to results, imbalance sensitivity increases with the increase in issue complexity and that simple, linearly separable problems remain untouched by all degrees of class imbalance. There is really a lack of data in some areas owing to the low frequency with which events happen, such as Oil leak detection. It is highly important to learn from acute class imbalanced data, where the minority class has the proportion up to 0.1% of the training data, as it is usually these rare events that are of utmost interest. The overall existing minority samples is of more interest than the proportion or ratio of the minority. Assume a minority group that accounts for just 1% of a data set covering 10 lakh samples. Several positive samples (10,000) still occur for a model training despite the imbalance of high degree [4]. In contrast, an imbalanced data set where the minority class exhibits infrequency or under-representations has higher degree of probability to undermine the classifier's performance.

$$\rho = \frac{max_i\{|C_i|\}}{mini_i\{|C_i|\}}$$

In the above example, a ratio ρ depicts the greatest between-class imbalance level. C_i specifies an example set in class i, and $max_i\{|C_i|\}$ and $min_i\{|C_i|\}$, respectively return the maximal and minimal class size over all i classes.

1.2 Existing Solutions to Class Imbalance

The solutions to class imbalance in network traffic classification are divided into three major levels: data level, algorithm level and cost-sensitive level [5].

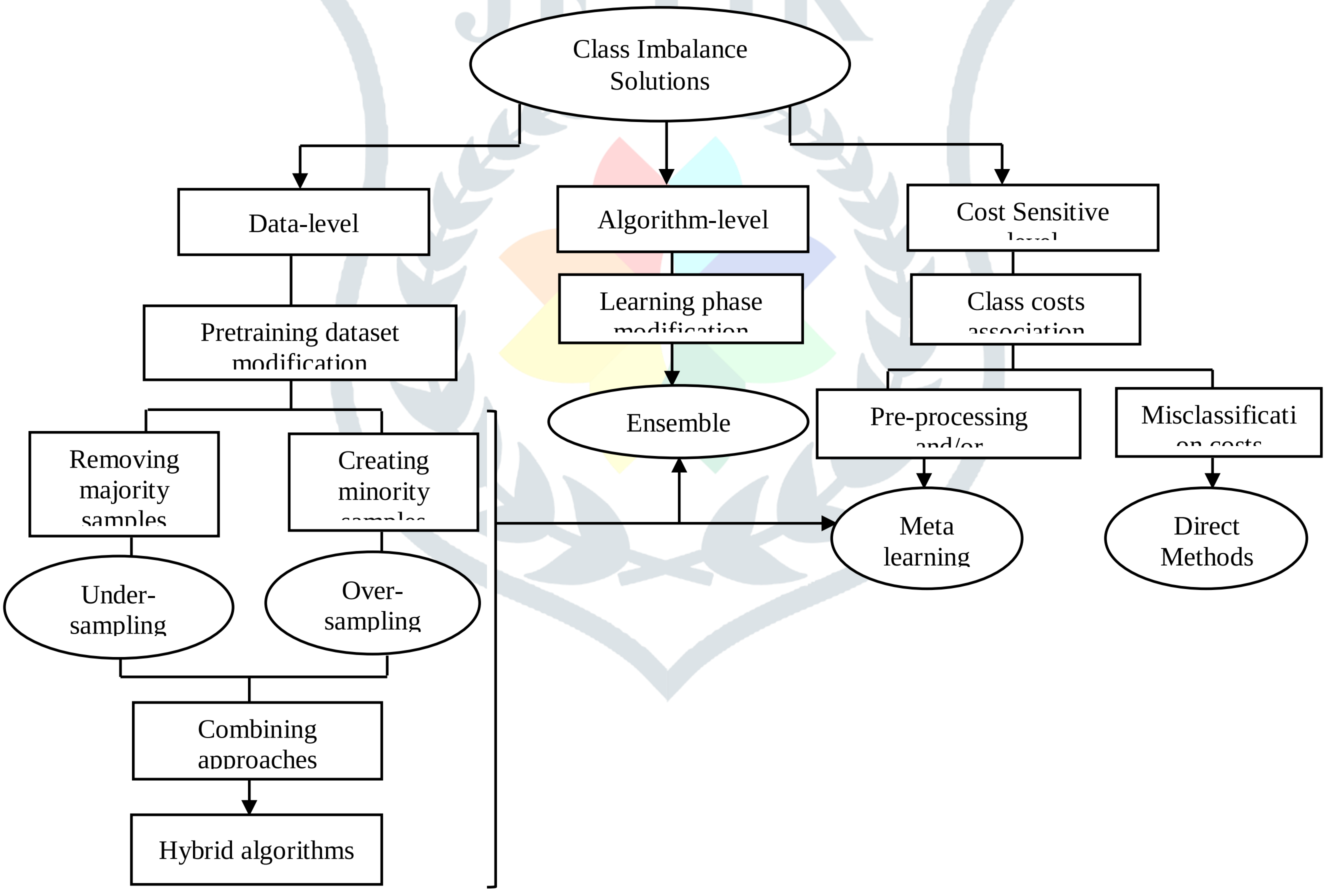

Figure 1: Existing Solutions to Class Imbalance

A. Data level methods

Data level techniques deal with class imbalances by oversampling and under-sampling, respectively, by enlarging the sample size of minority classes or reducing the sample number of majority classes. These techniques reduce the imbalance degree or noise by modifying the training distributions for example, mislabeled samples or irregularities. In its most non-complex types, random under-sampling (RUS) rejects random samples from the majority class, whereas random over-sampling (ROS) replicates random samples from the minority class.

a. Under sampling methods: Under-sampling willingly rejects the data in order to decrease the overall information the model needs to learn. Various smart sampling approaches have been devised as a means to balance these shifts. The purpose of smart under-sampling schemes is to maintain important information for learning. Some researchers have obtained under-sampling using the K-Nearest Neighbors (KNN) classifier [6]. The features of the provided data distribution have contributed in the development of four KNN under-sampling techniques called near miss-1, near miss-2, near miss-3 and "most distant" technique. Rather than using the whole set of over-represented majority training samples, a small subset of these samples is chosen so that the resultant training data is less heterogeneous.

- The Near Miss-1 approach chooses the majority of instances that have the shortest average distance from the three nearest minority instances.
- The near miss-2 approach chooses the majority class instances whose mean distance is the shortest from the three most distant minority class instances.
- NearMiss-3 chooses a provided number of nearest majority instances to each minority instance to guarantee that each minority instance is encircled by some majority instance.
- At last, the ‘most distant’ approach selects the majority class instances that have the greatest average distance of the three nearest minority class instances [7].

b. Oversampling methods: Random oversampling attempts to balance the class distribution by randomly repeating the minority class examples. However, many authors come to a conclusion that this approach may improve the probability of overfitting, as it creates same replicas of existent instances Several reported over-sampling techniques have also been developed to toughen class boundaries, decrease over-fitting, and increase differentiation.

- Synthetic Minority Over-sampling Technique (SMOTE): It is well-known over-sampling technique. The SMOTE (Synthetic Minority Over-sampling Technique) is implemented for generating the novel minority class examples for which various minority class instances which are put together are interrupted. This technique assists in avoiding the over fitting problem. But, its procedure is naturally dangerous as the minority class is often generalized blindly by it without considering any majority class and this strategy becomes problematic during highly skewed class distributions due to the sparse nature of the minority class in terms of the majority class in some scenarios. Consequently, a greater chance of class mixture is obtained. Various enhanced oversampling algorithms are implemented for retaining the benefits of SMOTE and lessening its drawbacks [8].
- MSMOTE (Modified SMOTE): This algorithm is an enhancement of SMOTE (Synthetic Minority Over-sampling Technique). This algorithm is adopted for partitioning the instances of the minority class into three groups: safe, border and latent noise instances after computing the distances among all examples. The new instances are created by (Modified Synthetic Minority Over-sampling Technique). The strategy of selecting the nearest neighbors is changed in accordance with earlier technique and the group previously assigned to the instance. This algorithm emphasizes on selecting a data point at random from the KNNs for safe instances, choosing the nearest neighbor for border instances and selecting nothing for latent noise instances. This technique efficiently decreases the risk of introducing artificially mislabelled instances. Thus, this technique is useful for performing more accurate classification in contrast to earlier technique.

B. Algorithm-level methods

Different from other techniques, these techniques have not changed the training data distribution to handle the class imbalance. Its learning or decision process is adjusted to maximize the efficacy of the positive class. The algorithms which assist in awarding the minority classes and punishing the majority under the training phase are deployed in these techniques. A class penalty or weight is considered by modifying the algorithms or the shifting of the decision threshold for mitigating the bias towards the negative class [9]. The major intend of the algorithm-level category is to modify the existing learner for eliminating its bias towards the majority classes. The ensemble methods are well-known algorithm-level techniques in which a re-sampling stage is included during the development of ensembles. Ensemble classification algorithms are utilized for enhancing the accuracy of single classifiers with the integration of various models and can be implemented on the imbalanced data-sets.

C. Cost-sensitive methods

In these techniques, the data level transformations are integrated with the algorithm level modifications. The costs related to the misclassifying samples are taken in account using the cost matrix. The learner is pressurized for classifying the minority class samples in exact manner for which a high cost is set to the misclassified samples of minority class. There is not any penalty for correct classification samples. However, the cost of misclassification is greater as compare to the majority samples. The cost-sensitive techniques concentrate on diminishing the total cost of the training dataset. But, it is challenging task to determine the cost values due to their dependency on multi factors having trade-off relationships. These techniques have two categories namely direct methods and meta-learning methods.

a. Direct methods: These techniques have cost-sensitive potentials which are attained by enhancing the learner's underlying algorithm such as the costs are considered under the learning process. The optimization process is changed amid one of minimizing total error and cost [10].

b. Meta-learning: A wrapper is utilized for converting the cost-insensitive learners into cost-sensitive systems in this technique. In case of generation of a cost-insensitive classification algorithm, a new threshold p^* is defined using the cost matrix as:

$$p^* = \frac{c_{10}}{c_{10} + c_{01}}$$

In general, p^* is utilized in the thresholding techniques for redefining the output decision threshold when the samples are classified. The above-mentioned equation is utilized to perform the threshold moving or post-process the output class probabilities and it is a meta-learning approach using which the cost-insensitive learner is transformed into a cost-sensitive system.

D. Hybrid methods

The integration of data-level is done with the algorithm-level techniques in several ways and deployed for dealing with the class imbalance issues. This strategy has sampled the data for mitigating the class noise and imbalance. Thereafter, the cost-sensitive learning or thresholding is utilized for further alleviating the bias towards the majority group. A number of methods in which the ensemble techniques are put together with the sampling and cost-sensitive learning have introduced. Two well-known hybrid algorithms are Easy Ensemble and Balance Cascade utilized for learning multiple classifiers. For this purpose, the subsets of the majority group are combined with the minority group and pseudo-balanced training sets are generated for every individual classification algorithm.

a. Easy Ensemble: The chief objective of this technique is that a high efficiency of under-sampling has to be maintained and the risk of avoidance of potentially useful information contained in majority class examples must be diminished [11]. A simple strategy is utilized in this technique. Initially, this technique emphasizes on creating the multiple subsamples $S_{maj\,1}, S_{maj\,2}, \ldots\ldots, S_{maj\,n}$from the majority class sample. The size of each subsample is similar to the minority class sample S_{min}, that is,$|S_{maji}| = |S_{min}|, 1 \le i \le n$. Subsequently, an adaboost ensemble is trained using the union of each possible pair (S_{maji}, S_{min}). All the base learners are integrated in all the adaboost ensembles to develop the final ensemble. The technique provides optimal outcomes in comparison with adaboost, bagging, RF (random forest), SMOTE Boost and BRF while tackling the issues of binary imbalance.

b. Balance Cascade: This technique focuses on deploying the guided instead of random deletion of majority class examples. Unlike the Easy Ensemble, this method performs in a supervised way. The ithround is executed to generate a subsample $S_{maj\,i}$ at random from the current majority class data set $S_{maj}$with sample size $|S_{maj\,i}| = |S_{min}|$.Therefore, adaboost is adopted to train an ensemble Hi from the union of $S_{maj\,i}$and S_{min}. Later on, the majority class data examples that the Hi algorithm has classified are eliminated from theS_{maj}. The Balance Cascade assists in eliminating the correctly classified majority class examples in every iteration. Thus, its efficacy must be enhanced for highly imbalanced data sets.

2. LITERATURE REVIEW

2.1 Class Imbalance for Network Traffic Classification using Machine Learning

Santiago Egea Gómez, et.al (2019) emphasized on tackling the issue related to class imbalance while classifying the network traffic [12]. The presence of this phenomenon was analyzed and various solutions were examined in 2diverse Internet environments. Twenty one data-level algorithms, 6 ensemble techniques and a cost-level method were employed in the experimentation. The issue related to imbalance was resolved considering the methodological aspects such as DOB-SCV validation technique. Moreover, for this, the binary techniques, in which 2ensemble techniques included, were also adopted in ML (Machine Learning). The experimental outcomes depicted that some methods led to diminish the class imbalance and boost the accuracy by 8% under diverse scenarios. A FLAGB (focal loss based adaptive gradient boosting) model was suggested by Yu Guo, et.al (2020) to classify the imbalanced traffic [13]. This model was adaptable in classifying the network traffic at diverse imbalance levels and tackling the imbalance without any prior knowledge regarding process to distribute the data. BOT and KDD99' datasets were applied to conduct the experiments in which binary and multiple classes were covered. The experimental outcomes demonstrated the supremacy of the suggested model over the traditional methods. This model consumed least time in the training phase, thus, it becomes an effective tool to classify the highly imbalanced traffic. A modified SVM algorithm known as CMSVM(cost-sensitive support vector machine) was introduced by Shi Dong, et.al (2021) for addressing the imbalance issue while classifying the network traffic [14]. A multi-class SVM algorithm was implemented with active learning due to its proficiency of assigning a weight for applications in dynamic way. The MOORE_SET and NOC_SET datasets were utilized for the quantification of introduced algorithm with regard to accuracy and efficiency. The experimental outcomes confirmed that the introduced algorithm was applicable for lessening the computation cost, increasing the accuracy and tackling the imbalance problem in comparison with other ML (machine learning) methods. An IDGC (Imbalanced Data Gravitation-based Classification) system was constructed by Lizhi Peng, et.al (2017) for dealing with the issue related to classify the Internet traffic [15]. Initially, the generation of 6 imbalanced traffic data sets was done from 3 original traffic data sets. Subsequently, their packet sizes were considered to extract the attributes. A comparative analysis was conducted on the constructed system against various algorithms in the experimentation. The results obtained in experiments revealed the effectiveness and stability of the constructed system with regard to diverse parameters while classifying the imbalanced traffic.

2.1 Comparison Table

Author	Year	Technique Used	Dataset	Results
Santiago Egea Gómez, et.al	2019	DOB-SCV validation approach	HOST datasets	Some methods were assisted in diminishing the class imbalance and boosting the accuracy by 8% under diverse scenarios.
Yu Guo, et.al	2020	Focal loss based adaptive gradient boosting framework (FLAGB)	BOT and KDD99' dataset,	This model consumed least time in the training phase, thus, it becomes an effective tool to classify the highly imbalanced traffic.
Shi Dong, et.al	2021	A cost-sensitive SVM (CMSVM)	MOORE_SET and NOC_SET datasets	The introduced algorithm was applicable for lessening the computation cost, increasing the accuracy and tackling the imbalance problem
Lizhi Peng, et.al	2017	IDGC (Imbalanced Data Gravitation-based Classification)-based model	UNIBS-SKYPE and UJN-CD data sets	The results obtained in experiments revealed the effectiveness and stability of the constructed system with regard to diverse parameters while classifying the imbalanced traffic.

2.2 Class Imbalance for Network Traffic Classification using Deep Learning

Pan Wang, et.al (2020) developed a new technique known as Packet CGAN using CGAN with the objective of controlling the modes of data before their construction [16]. The efficiency of CGAN (Conditional Generative Adversarial Network) was adopted in this technique for creating the specified samples with the input of types of applications as conditional. Hence, the data was balanced. Four kinds of encrypted traffic datasets were classified using three traditional DL (Deep Learning) models. The ROS (Random over Sampling), SMOTE (Synthetic Minority Over-sampling Technique), Vanilla GAN and Packet CGAN were deployed to augment these datasets. The results of experiments indicated that the developed technique performed more successfully against other algorithms for classifying the encrypted traffic. A technique planned on the basis of Text CNN (convolution neural network) algorithm was established by Mingze Song, et.al (2019) in which the traffic data was represented as vectors [17]. The key attributes were extracted using this algorithm to classify the traffic. ISCX VPN-non VPN dataset was applied to evaluate the established algorithm. The results proved that the established algorithm outperformed the earlier technique concerning F1 score. Moreover, the issue regarding class imbalance was resolved with the implementation of a novel loss function and a suitable technique of allocating the class weight to perform multi-class classification. The adaptability of these techniques was proved. An innovative GAN (Generative Adversarial Network) algorithm was formulated by Yu Guo, et.al (2021) for generating the traffic samples [18]. Furthermore, the stability and efficiency was offered to this procedure using the classification algorithm and the pre-training module. An E2E (end-to-end) model recognized as ITCGAN was put forward for generating traffic samples for minority classes so that the original traffic was rebalanced in adaptive way. A publicly available dataset namedISCXVPN2016 was utilized to validate the formulated algorithm considering the global and individual parameters. The experimental outcomes exhibited that the formulated algorithm was effective to classify the imbalanced network traffic subsequent to decrease the performance degradation. A GAN (Generative Adversarial Network) algorithm named Flow GAN was designed by ZiXuan Wang, et.al (2019) for addressing the class imbalance issue during classifying the traffic [19]. The designed algorithm was computed by training the MLP (Multilayer Perceptron) based classification algorithm. The experiments were conducted on ISCX dataset. The results of experiments validated the supremacy of the designed algorithm over the existing techniques and assisted in enhancing the precision up to 13.2%, recall around 17.0% and F1-score up to15.6%

2.2 Comparison Table

Author	Year	Technique Used	Dataset	Results
Pan Wang, et.al	2020	Packet CGAN using Conditional GAN	ISCX2012 and USTC-TFC2016	The developed technique performed more successfully against other algorithms for classifying the encrypted traffic.
Mingze Song, et.al	2019	Text convolution neural network	ISCX VPN-non VPN dataset	The issue regarding class imbalance was resolved with the implementation of a novel loss function and a suitable technique of allocating the class weight to perform multi-class classification.
Yu Guo, et.al	2021	Generative Adversarial Network (GAN)	ISCXVPN2016 dataset	The formulated algorithm was effective to classify the imbalanced network traffic subsequent to decrease the performance degradation.
ZiXuan Wang, et.al	2019	Flow GAN	ISCX dataset	The designed algorithm was superior to the existing techniques and assisted in enhancing the precision up to 13.2%, recall around 17.0% and F1-score up to15.6%.

2.3 Class Imbalance for Network Traffic Classification using Ensemble Technique

Phuylai Oeung, et.al (2019) projected a mechanism for developing an effectual classification algorithm from NetFlow to classify the traffic [20]. First of all, the C4.5 DT (decision tree) algorithm was utilized with attributes of Net Flow records for analyzing the per-application. In addition, an ensemble FS (feature selection) technique was suggested with the objective of boosting the accuracy and mitigating the computational complexity. In the end, an integration of clustering-based under-sampling with SMOTE (synthetic minority over-sampling technique) was implemented so the issue related to data imbalance was resolved. The experimental outcomes indicated that the projected mechanism offered higher F-measure and least computational complexity. A hybrid algorithm was presented known as FWFS (Filter-Wrapper Feature Selection) by Fatimah Audah Md. Zaki, et.al (2019) in order to classify the network traffic [21]. The robust attributes were selected using this algorithm for providing the resistance against concept drift. The wrapper function was utilized to discard the redundant attributes. The results confirmed the reliability and stability of the presented algorithm to classify the new data and the acquired accuracy was calculated 98.7% while classifying the new data and the F-measure was found 0.8 above in every class. An innovative algorithm was suggested by Luyang Xu, et.al (2019) to classify the traffic on the basis of packet transport layer payload for which an EL (ensemble learning) was implemented [22]. Three types of NNs (neural networks) were deployed for generating an effective classification algorithm. The training of every model was done at individual level. The weight voting was implemented to decide the predictive outcome. The experimental outcomes demonstrated that the suggested algorithm yielded the accuracy around 96.38% and proved superior over the traditional techniques. An E2E (end-to-end) technique of classifying the encrypted traffic was intended by Wei Wang, et.al (2017) with 1D-CNN (one-dimensional convolution neural networks) [23]. The techniques of extracting the attributes, selecting the attributes and a classification algorithm were integrated into unified E2E model to learn then on linear relationship amid raw input and expected output. ISCX VPN-non VPN traffic dataset was executed to authenticate the intended algorithm. The experimental outcomes revealed that the intended technique was more adaptable in comparison with the traditional technique.

2.3 Comparison Table

Author	Year	Technique Used	Dataset	Results
Phuylai Oeung, et.al	2019	Ensemble feature selection (FS) method	UNIBS and Auckland dataset	The experimental outcomes indicated that the projected mechanism offered higher F-measure and least computational complexity.
Fatimah Audah Md. Zaki, et.al	2019	Filter and Wrapper Feature Selection (FWFS)	ISCX dataset	The presented algorithm offered the accuracy of 98.7% to classify the new data and the F-measure was found 0.8 above in every class.
Luyang Xu, et.al	2019	A novel traffic classification approach	UNB ISCX VPN-nonVPN dataset	The suggested algorithm yielded the accuracy around 96.38% and proved superior over the traditional techniques.
Wei Wang, et.al	2017	An end-to-end encrypted traffic classification method with one-dimensional convolution neural networks	ISCX VPN-nonVPN traffic dataset	The experimental outcomes revealed that the intended technique was more adaptable in comparison with the traditional technique.

CONCLUSION

The network traffic classification techniques have 3 categories namely port-based, payload-based and flow statistics-based. The process to classify the network traffic is deal with recognizing distinct kinds of applications or traffic data for which the received data packets are investigated that is essential in the communication networks of real world. The new network management functions such as to ensure the network QoS (Quality of Service) and for detecting the network anomaly on the basis of accuracy obtained so as the network traffic can be classified. The network is capable of various relevant applications and services. Thus, the differentiation of the network packets or flows is done on the basis of the applications or services offered through the network. It is analyzed that machine learning algorithms performance best for the network traffic classification in terms of accuracy, precision and recall.

REFERENCES

[1] RaminHasibi, Matin Shokri, and Mehdi Dehghan Takht Fooladi, “Augmentation Scheme for Dealing with Imbalanced Network Traffic Classification Using Deep Learning”, 2019, arXiv

[2] Muhammad Shafiq, Xiangzhan Yu, Ali Kashif Bashir, Hassan Nazeer Chaudhry, and Dawei Wang, “A machine learning approach for feature selection traffic classification using security analysis”, 2018, The Journal of Supercomputing, 74(10):4867–4892

[3] Lizhi Peng, Haibo Zhang, Yuehui Chen, and Bo Yang, “Imbalanced traffic identification using an imbalanced data gravitation-based classification model”, 2017, Computer Communications, 102:177 – 189

[4] Ly Vu, Cong Thanh Bui, Quang Uy Nguyen, “A Deep Learning Based Method for Handling Imbalanced Problem in Network Traffic Classification”, 2017, the Eighth International Symposium

[5] Jafar Tanha, Yousef Abdi, NeginSamadi, NazilaRazzaghi and Mohammad Asadpour, “Boosting methods for multi class imbalanced data classification: an experimental review”, 2020, Journal of Big data

[6] Qiong Liu & Zhen Liu, “A comparison of improving multi-class imbalance for internet traffic classification”, 2014, Information Systems Frontiers

[7] Liu Zhena, Liu Qiong, “A New Feature Selection Method for Internet Traffic Classification Using ML”, 2012, International Conference on Medical Physics and Biomedical Engineering

[8] J. Yang, Y. Wang, C. Dong and G. Cheng, "The evaluation measure study in network traffic multi-class classification based on AUC," 2012 International Conference on ICT Convergence (ICTC), 2012, pp. 362-367

[9] Y. Dhote, S. Agrawal and A. J. Deen, "A Survey on Feature Selection Techniques for Internet Traffic Classification," 2015 International Conference on Computational Intelligence and Communication Networks (CICN), 2015, pp. 1375-1380

[10] Z. Wang, P. Wang, X. Zhou, S. Li and M. Zhang, "FLOWGAN:Unbalanced Network Encrypted Traffic Identification Method Based on GAN," 2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), 2019, pp. 975-983

[11] M. S. Sharif and M. Moein, "An Effective Cost-Sensitive Convolutional Neural Network for Network Traffic Classification," 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), 2021, pp. 40-45

[12] Santiago Egea Gómez, Luis Hernández-Callejo, Antonio J. Sánchez-Esguevillas, “Exploratory study on Class Imbalance and solutions for Network Traffic Classification”, 2019, Neurocomputing

[13] Yu Guo, Zhenzhen Li, Zhen Li, Gang Xiong, Minghao Jiang, Gaopeng Gou, “FLAGB: Focal Loss based Adaptive Gradient Boosting for Imbalanced Traffic Classification”, 2020, International Joint Conference on Neural Networks (IJCNN)

[14] Shi Dong, “Multi class SVM algorithm with active learning for network traffic classification”, 2021, Expert Systems with Applications

[15] Lizhi Peng, Haibo Zhang, Bo Yang, "Imbalanced traffic identification using an imbalanced data gravitation-based classification model", 2017, Computer Communications

[16] Pan Wang, Shuhang Li, Feng Ye, Zixuan Wang, Moxuan Zhang, "PacketCGAN: Exploratory Study of Class Imbalance for Encrypted Traffic Classification Using CGAN", 2020, IEEE International Conference on Communications (ICC)

[17] Mingze Song, Jing Ran, Shulan Li, "Encrypted Traffic Classification Based on Text Convolution Neural Networks", 2019, IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)

[18] Yu Guo, Gang Xiong, Zhen Li, Junzheng Shi, Mingxin Cui, Gaopeng Gou, "Combating Imbalance in Network Traffic Classification Using GAN Based Oversampling", 2021, IFIP Networking Conference (IFIP Networking)

[19] ZiXuan Wang, Pan Wang, Xiaokang Zhou, ShuHang Li, MoXuan Zhang, "FLOWGAN:Unbalanced Network Encrypted Traffic Identification Method Based on GAN", 2019, IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)

[20] PhuylaiOeung, Fuke Shen, "Imbalanced Internet Traffic Classification Using Ensemble Framework", 2019, International Conference on Information Networking (ICOIN)

[21] Fatimah Audah Md. Zaki, Tan Saw Chin, "FWFS: Selecting Robust Features Towards Reliable and Stable Traffic Classifier in SDN", 2019, IEEE Access

[22] Luyang Xu, Xu Zhou, Yongmao Ren, Yifang Qin, "A Traffic Classification Method Based on Packet Transport Layer Payload by Ensemble Learning", 2019, IEEE Symposium on Computers and Communications (ISCC)

[23] Wei Wang, Ming Zhu, Xuewen Zeng, Xiaozhou Ye, Yiqiang Sheng, "Malware traffic classification using convolutional neural network for representation learning", 2017, International Conference on Information Networking (ICOIN)