IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 22 January 2023, accepted 26 February 2023, date of publication 2 March 2023, date of current version 8 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3251745

RESEARCH ARTICLE

RFSE-GRU: Data Balanced Classification Model for Mobile Encrypted Traffic in Big Data Environment

MURAT DENER, SAMED AL, AND GÖKÇE OK
Department of Information Security Engineering, Graduate School of Natural and Applied Sciences, Gazi University, 06560 Ankara, Turkey
Corresponding author: Murat Dener (muratdener@gazi.edu.tr)

ABSTRACT With the widespread use of mobile technologies and the Internet, traffic in mobile networks is increasing. This situation has made the classification of traffic an important element for data security and network management. However, encryption of traffic in modern networks makes it difficult to classify traffic with traditional methods. In this study, a unique deep learning-based classification model is proposed for the classification of encrypted mobile traffic data. The proposed model is a classification model called RFSE-GRU, which combines the Gated Recurrent Units (GRU) algorithm, feature selection and data balancing. The features that are more meaningful in the classification process are determined by selecting the features with the Random Forest algorithm. In addition, Synthetic Minority Oversampling Technique (SMOTE) oversampling algorithm and Edited Nearest Neighbor (ENN) undersampling algorithm were used together to reduce the negative impact of data imbalance on classification performance. The study was carried out on Apache Spark's big data platform in the Google Colab environment. In the study, ISCX VPN-Non VPN and UTMobileNet2021 datasets were used. Binary and multiclass classifications were made for the ISCX VPN-Non VPN dataset, and multiclass classifications were made for the UTMobileNet2021 dataset by using various algorithms on the datasets. The proposed model has been compared with eleven different algorithms and hybrid methods. At the same time, the effect of data balancing and feature selection on classification performance is examined. As a result, the proposed model achieved 93.91%, 82.68% and 96.83% accuracy rates in ISCX VPN-Non VPN binary and multiclass, UTMobileNet2021 multiclass classifications, respectively.

INDEX TERMS Mobile encrypted traffic, VPN, big data, machine learning, deep learning, Apache Spark, classification.

I. INTRODUCTION

With the development and spread of the internet and mobile technologies, traffic continues to increase day by day. This makes traffic classification critical in modern networks [1]. Since mobile technologies are a dynamic platform where new applications are quickly introduced, the classification of emerging traffic has become an important research topic in terms of the management of network resources and evaluation of security [1], [2]. Classification of mobile traffic provides various opportunities for applications such as increasing service quality, content billing, providing access control, anomaly detection and security, and implementation of necessary policies [2], [3], [4]. Traffic classification can be at different levels of detail, depending on the intended purpose. Binary classification can be made for anomaly detection, malware detection or VPN detection, while classification according to protocols can be made for the management of network resources. In addition, traffic classifications are made according to a service group, applications, user behaviors, and user credentials in applications or smart devices [2].

With the rapid transformation and spread of the digital world, ensuring secure communication and data security has made it necessary to encrypt traffic. For this reason, security protocols such as SSH, PKI, SET, SSL, HTTPS have been created and the use of encrypted tunnels called Virtual Private Networks (VPN) has become widespread [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Aasia Khanum.

In encrypted traffic, camouflaging the data by converting it into a certain format with various encryption methods makes traffic classification difficult [5].

Mobile encrypted traffic classification methods are generally divided into three as port-based, payload-based and flow-based statistical methods. The port-based approach is a traffic classification method using the relationship of port numbers in the TCP/UDP header. It is a simple and fast method. However, developments such as the widespread use of VPN technology, the use of hidden or random port numbers, and the inability to classify all protocol types due to private ports reduce the prevalence of this method [2], [6]. The payload-based approach is also called DPI (Deep Packet Inspection). In this method, the payload must be examined and matched to the stored signature or patterns. It cannot classify encrypted traffic because it cannot be matched in encrypted or obfuscated traffic. In addition, the DPI method is not an efficient method due to high computing resource consumption and costly updating of the package signature library [1], [7]. Flow-based statistical methods, on the other hand, classify network traffic with machine learning methods by extracting flow properties such as packet length and flow time. Machine learning removes the limitations of port-based and payload-based methods on encrypted traffic classification. However, machine learning methods also have their own limitations. Some of these constraints are that manual feature extraction is cumbersome and flow-based features require constant updating. This situation limits the generalization and representation ability of the models [2], [8]. It has been seen that deep learning methods have some advantages against the limitations of machine learning methods.

Deep learning is a popular approach to the classification of encrypted network traffic. Automatic feature extraction is the biggest advantage of deep learning. It also performs better by designing end-to-end when given enough labeled traffic data [9]. However, deep learning methods require large amounts of data to achieve the desired success. Insufficient data or a small number of labeled data are factors that negatively affect the classification result of the models [10]. The often imbalanced distribution of encrypted network traffic data is another important factor affecting the accuracy of classification.

In this study, a classification model called RFSE-GRU, which is formed by combining the GRU algorithm, feature selection and data balancing, is presented. In the study, ISCX VPN-Non VPN and UTMobileNet2021 datasets containing encrypted mobile network traffic data are used. In the ISCX VPN-Non VPN dataset, binary and multiclass classification covering VPN-Non VPN classes and service groups has been performed. The multiclass classification was made according to mobile application types in the UTMobileNet2021 dataset. A hybrid data balancing algorithm, which combines SMOTE and ENN methods, is used to eliminate the imbalance of classes in the datasets. Feature selection has been made based on the Random Forest algorithm to increase the classification performance and reduce the processing load. The big data approach was used because of the large volume of encrypted mobile network traffic. The study was carried out on Google Colab using Apache Spark's big data platform. Classifications were performed using twelve different machine learning, deep learning and hybrid methods. These are RF (Random Forest), DT (Decision Tree), NB (Naive Bayes), LR (Logistic Regression), MLP (Multilayer Perceptron), LSTM (Long Short-Term Memory), CNN (Convolutional Neural Network), CNN-LSTM, LSTM-CNN, CNN-GRU, GRU-CNN and proposed model (RFSE-GRU). The results were compared using Accuracy, F-score, Precision and Recall performance metrics.

The main contributions of this study can be summarized as follows:

1) In the study, studies were carried out to classify mobile encrypted traffic data. A new classification model has been developed as encrypted traffic data is increasing day by day in terms of both dimensional and overall traffic. Considering the significant increase in mobile traffic data, the study was carried out in a big data environment and the results were verified.

2) Feature selection was performed on the ISCX VPN-Non VPN and UTMobileNet2021 datasets to both reduce computational complexity and increase classification accuracy. Random Forest algorithm was used for feature selection in the study. As a result of this process, more meaningful features were used for classification. In addition, a faster model was developed because fewer data would be processed.

3) Another contribution of the study is to compare the effects of different feature selection algorithms on classification performance. It has been shown that the Random Forest algorithm can be used as a feature selection algorithm as well as a classification algorithm and it improves classification performance.

4) In addition, it has been emphasized that imbalanced datasets affect classification performance and that these datasets need to be balanced. SMOTE oversampling and ENN undersampling algorithms are combined for data balancing. Thanks to this combination, the disadvantages of both oversampling and undersampling techniques are eliminated. As a result, the classification performance of the intrusion detection system has been improved and the obtained results has been confirmed the improvement.

5) The proposed model is compared with eleven different machine learning and deep learning classification techniques. The results showed that the proposed model outperforms other machine learning and deep learning methods.

The rest of the study is organized as follows. Related studies are given in Chapter 2. In Chapter 3, the materials and methods used in the study are mentioned. In Chapter 4, the proposed method is explained, and in Chapter 5, model parameters and experimental results are given. Finally, Chapter 6 contains the conclusion of the study.

II. RELATED WORKS

Encrypted mobile network traffic classification is an important and studied subject with the widespread use of the internet and mobile applications. Various models have been developed in the literature to provide a classification of mobile network traffic. Open or private datasets are produced to test these models. One of these datasets is the ISCX VPN-Non VPN dataset, created by Draper-Gil et al. [1], considering the difficulty in detecting virtual private network (VPN) traffic. Whether the dataset belongs to the VPN and the traffic characterization are tested by extracting the time-related features based on the flow. Experiments were carried out on various scenarios using machine learning methods. Wang et al. [11] performed a CNN-based study for malicious traffic classification by converting traffic data to gray images. In the study, two different scenarios were carried out with three types of classifiers. As a result of this study, the USTC-TRC2016 dataset and the USTCTK2016 preprocessing tool were developed.

Many studies are using various datasets. Pektas and Acarman [12] performed classification studies on flow-based static properties in order to identify applications from network traffic. They compared RF, SVM and LR algorithms by selecting features on two different datasets. According to the results, the RF algorithm provided higher accuracy rates for both datasets. Al-Obeidat and El-Alfy [13] proposed a hybrid method based on multi-criteria fuzzy decision trees to classify network traffic. The proposed method has been tested on 13 different datasets, and compared with the performances of SVM, KNN, and NB machine learning algorithms. The authors noted that the method achieves good results both in imbalanced datasets and in detecting unknown traffic patterns, including Zero-day attacks. In another study, Wang et al. [14] focused on balancing data in traffic classification. A deep learning approach called FlowGAN, which balances by generating synthetic data with GAN, is presented. Evaluations were made on the ISCX VPN-Non VPN dataset using MLP. Accordingly, the imbalanced dataset and oversampled dataset provided 89.95% accuracy and 97.94% accuracy in the classification of applications, while the recommended approach was 99.10% accuracy.

Song et al. [15] treat bytes in traffic data and words in the text in the same way. Based on this, it has been suggested that it can be classified with Text-CNN by creating vectors from traffic data. The proposed approach has been evaluated on the ISCX VPN-non VPN dataset and it has been shown that it performs better in CNN and C4.5 algorithms. In the study of Nazari et al. [16] a DPI-based classification method called DSCA was proposed to identify applications from real-time encrypted network traffic. DSCA Feature Extractor consists of four modules: Inline Deep Packet Inspector, Stream Processor and Stream Classifier. In the study, the proposed method was tested on ISCX VPN-Non VPN and ISCX Tor-Non Tor datasets by integrating DSCA with 10 different algorithms. The authors mentioned that the best performance rates for the datasets were 96.75% and 86.92%, respectively. In another study, Chiu et al. [17] investigated faster processing of traffic packets. For this, they presented the package-based Convolutional Autoencoder Packet Classifier (CAPC) approach, which consists of 1D-CNN and AE models. This approach, it is aimed to classify dynamic port numbers and applications as well as encrypted traffic. The proposed method has been tested on the ISCX VPN-Non VPN dataset and a special dataset, and it has been shown that it achieves higher accuracy rates than DNN, DAE and 1D-CNN.

Baldini [18] classified the traffic characterizations by transforming the time-based features in network traffic into time series with Constant-Q Transform (CQT), Continuous Wavelet Transform and Wigner-Ville Distribution methods. In the study, evaluations were made with SVM, KNN and DT algorithms on 15, 30, 60 and 120 second processes. The highest accuracy rate was achieved by DT for the 15 second process using Wigner-Ville Distribution. In another study by Baldini et al. [19], during the same period, it was mentioned that encrypted communication would not be sufficient to ensure confidentiality. In this direction, three components are presented and frameworks are explained to ensure confidentiality. Integration of components was evaluated in the classification process of traffic characterization in the ISCX VPN-non VPN dataset. Chen et al. [20] proposed a classification method based on a capsule neural network using PDU lengths for encrypted traffic classification. The method is a three-layer neural network model with filter, N-gram and classification layer. ISCX VPN-Non VPN dataset was used to test the proposed method. As a result of the study, an F1 score of 0.9855 was obtained.

Iliyasu and Deng [3] studied the difficulty of collecting and labeling large encrypted network traffic data. Using the semi-supervised Deep Convolutional Generative Adversarial Network (DCGAN) approach, it is aimed to perform high-performance classification with a small number of labeled training samples. QUIC and ISCX VPN-Non VPN datasets were used to evaluate the proposed method. In experiments performed by labeling 10% of the datasets, 89% accuracy was obtained in the QUIC dataset and 78% in the ISCX VPN-Non VPN dataset. At the same time, it is emphasized that the proposed method outperforms MLP and CNN algorithms on streams containing a large number of various packets such as QUIC. Lotfollahi et al. [4] presented the Deep Packet approach, which combines the stages of classification and attribute extraction of encrypted traffic. The Deep Packet framework can identify traffic, classifying VPN-non-VPN networks, traffic characterization, and application classification. The authors tested the proposed approach on the ISCX VPN-Non VPN dataset using Stacked Autoencoder (SAE) and Convolution Neural Network (CNN) methods. As a result of the study, the CNN network showed the best classification performance by obtaining 94% Recall values in the classification of traffic characterization and 98% in the classification of applications.

Bu et al. [8] carried out a study to classify encrypted network traffic by constructing deep and parallel network-in-network (NIN) models. The designed models have MLP convolutional layers. A decision mechanism is executed that processes the header and body of the packet over two subnets. F1 scores of 0.983 and 0.985 were obtained, respectively, for traffic characterization and application classification in the ISCX VPN-Non VPN dataset. Comparing the performance of the proposed method to the CNN algorithm, it was revealed that the proposed method was more successful. Zhang et al. [21] addressed the problem that deep learning approaches could not properly classify unknown classes in encrypted network traffic. As a solution to this problem, an autonomous learning framework that filters and tags packets from unknown classes is presented. The proposed framework was evaluated using the ISCX VPN-Non VPN dataset. Three different scenarios with different classes are created and the ability of the proposed framework to update deep learning-based classifiers is demonstrated by using MLP and CNN models of different sizes. In another study, Wang et al. [9] proposed App-Net to identify applications from TLS traffic. App-Net is a hybrid neural network approach consisting of bi-LSTM and 1D-CNN. The approach has been tested using a real-world dataset containing 80 different classes of applications. The accuracy rate of the classification was stated as 93.2%.

Aceto et al. [22] emphasized the importance of deep learning methods in encrypted traffic classification. A framework for encrypted traffic classification with deep learning is proposed. The proposed framework was evaluated on three datasets of user activities: Android, IOS and Facebook Messenger. In another study, Ugurlu et al. [23] made binary and multiclass classifications with various scenarios on the ISCX VPN-Non VPN dataset to analyze encrypted network traffic. Classifications were tested using XGBoost, Decision Tree and Random Forest algorithms. The algorithm with the highest accuracy rate in the results was XGBoost with 94.53%. Heng et al. [24] presented the labeled UTMobileNet-Traffic2021 dataset containing popular application and activity levels for the evaluation and development of algorithms in network traffic classification studies. Montieri et al. [25] focused on estimating network traffic characteristics with packet detail information rather than analysis based on traffic volume. In their study, they evaluated the packet-level prediction performance of DSANET and SERIESNET architectures from CNN, GRU, LSTM and composite neural networks by performing experiments on MIRAGE-2019 and MIRAGE-VIDEO datasets.

Aceto et al. [26] presented an approach consisting of Markov chains and Hidden Markov models. In this approach, the characterization and applications of mobile network traffic at packet and message levels are classified. The performance efficiency is demonstrated by comparing the proposed approach with RF, LR, and KNN algorithms. Dong et al. [27] presented a classification model based on Generating Adversarial Deep Convolutional Networks (GADCN) by addressing the imbalance of traffic data. The proposed model has been tested on the USTC-TFC2016 dataset for the classification of benign and malignant traffic. The proposed method gave better results than other models. At the same time, when compared with other balancing methods, it was observed that it achieved higher accuracy than other methods except for DCGAN. In another study, Lu et al. [28] proposed a new method called Inception-LSTM (ICLSTM) for the classification of encrypted traffic. After the network traffic is converted to gray images, the spatial information of the packets and the temporal information between the packets are processed simultaneously with the ICLSTM consisting of LSTM and the Inception layer. At the same time, a weight parameter is assigned to each traffic category to avoid data imbalance. Five different scenarios were applied on ISCX VPN-non VPN to measure the success of the proposed method. While 99% accuracy rates were obtained for VPN classes and 98.7% for non-VPN classes, 98.1% accuracy was achieved in the classification made for all categories.

Liu and Lang [29] performed classification studies on the detection of unknown protocols. In their proposed method, they aimed to classify known and unknown network traffic at the same time by combining CNN and density-based DBSCAN clustering algorithms. In the analyzes made, they obtained an accuracy rate of 97.03% on the ISCX VPN-Non VPN dataset and 98.50% on the DARPA 1998 dataset. In addition, the protocol profiling method and automatic rule extraction method are suggested in the study. Later on, Lu et al. [30] proposed a network-in-network (NIN) model to reduce the need for computing and storage resources. Gradual pruning and knowledge distillation (KD) compression were used to train the model. An F1 Score of 0.9805 was obtained in the experiments performed on ISCX VPN-Non VPN. A better performance than the CNN model has been demonstrated. At the same time, significant savings and improvements were achieved in parameters and processing time. Hu et al. [31] developed the CLD-Net method for the classification of encrypted network traffic. In the proposed method, flow segmentation of raw traffic packets and then recombination is used to use important payload information in the packets. Thanks to CNN and LSTM, it is aimed to benefit from both image classification and time series. The proposed method has been tested with binary and application classification on ISCX VPN-non VPN and its performance has been demonstrated.

Soleymanpour et al. [6] carried out a study to solve the class imbalance of datasets used in the classification of encrypted network traffic. Cost-Sensitive Convolutional Neural Network (CSCNN) classification method is proposed to solve the unbalanced data problem in the training phase. In this method, a cost matrix based on class distribution is used. In this way, false classifications are costly and classification accuracy is increased during the training phase. The proposed method achieved 97.9% accuracy for application

identification and 97.7% accuracy for traffic description. In the study of Zheng et al. [32] the multi-task transformer (MTT) method, which describes the characterization of encrypted network traffic and the application at the same time using a single packet, is proposed. To test the MTT model, experiments were conducted on the ISCX VPN-Non VPN dataset and compared with one-dimensional and two-dimensional CNN models. The proposed model achieved accuracy rates of 98.75% for application classification and 99.34% for traffic characterization in multi-tasking. An improvement was achieved in the calculation time as well as the classification performance with this study.

Yao et al. [33] treated the classification problem of encrypted network traffic as time series. Two different models were proposed in the study. The first of the proposed models is the attention-based LSTM method. The other method is hierarchical attention network (HAN). Both proposed models have been tested on ISCX VPN-Non VPN. As a result, the attention-based LSTM method achieved 91.2% accuracy and the HAN method 89.5% accuracy. In the study of Izadi et al. [10], the popular deep learning approach in network traffic analysis requires a lot of training data and otherwise the negative effect on the accuracy of classification is discussed. The Data Fusion method has been proposed to solve this problem. In the study, classifications were made using Deep Belief Networks (DBN), one-dimensional Convolution Neural Networks (CNN), and Multi-layer Perceptron (MLP) networks and the results obtained by Bayesian decision fusion were combined. The proposed method is tested on the ISCX VPN-Non VPN dataset. The approach outperformed single classifiers with an accuracy rate of 97%. However, the method has higher processing time and memory requirement. This is a disadvantage of the proposed classification approach. Telikani et al. [34] proposed a cost-sensitive deep learning-based approach as a solution to the problem of balancing encrypted traffic data. In this approach, cost matrices are created for each section according to their distribution by dividing the data set into parts. Classification errors are penalized using the cost function. The cost sensitive learning method was applied to the ISCX VPN-Non VPN dataset with SAE and CNN algorithms. As a result of the study, the proposed method achieved a 98.6% accuracy rate and outperformed other network traffic classification methods.

In this study, experiments have been carried out using twelve machine learning and deep learning algorithms in the classification of encrypted mobile network traffic. By using different attribute types, different traffic features are classified. With ISCX VPN-Non VPN, VPN-Non VPN and service groups are classified on flow-based attributes. In the UTMobileNet2021 dataset, mobile application classification was made by analyzing the packet header information attributes. In addition, the effects of data balancing and feature selection on the classification performance of encrypted mobile network traffic are evaluated. Some studies on mobile traffic classification in the literature are summarized in Table 1.

TABLE 1. Comparison of other works on encrypted mobile network traffic.

Year	Authors	Model	Dataset	Type	Accuracy (%)	F-Score (%)
2017	Wang, et al. [11]	CNN	USTC-TFC2016	Malware	99,41 (avg)	
2018	Pektas and Acarman [12]	RF	NISM Private Dataset	Application	99,00 96,00	
2019	Nazari, et al. [16]	DSCA-DPI technique	ISCX VPN-Non VPN ISCX Tor-Non Tor	Application	96,75 86,92	
2019	Iliyasu and Deng [3]	DCGAN	ISCX VPN- Non VPN QUIC	Service Group	78,00 89,00	
2019	Al-Obeidat and El-Alfy [13]	Multicriteria Fuzzy DT	KDD'99 WEP\WPA WPA2	Service Group (Attack)	99,95 87,13 93,16	
2019	Song, et al. [15]	TextCNN	ISCX VPN- Non VPN	Binary		91,80 (avg)
2019	Wang, et al. [14]	GAN + MLP	ISCX VPN- Non VPN	Application	99,10	
2020	Chiu, et al. [17]	1D-CNN + AE	Private Dataset ISCX VPN- Non VPN	Application	99,98 97,42	
2020	Lotfollahi, et al. [4]	CNN SAE	ISCX VPN- Non VPN	Application Service Group		98,00 95,00 93,00 92,00
2020	Chen, et al. [20]	Capsule NN	ISCX VPN- Non VPN	Service Group		98,60
2020	Bu, et al. [8]	NIN models with multiple MLP convolutional layers	ISCX VPN-Non VPN	Application Service Group		98,50 98,30
2020	Wang, et al. [9]	bi-LSTM + 1D-CNN	Private Dataset	Application	93,20	
2020	Baldini [18]	DT	ISCX VPN- Non VPN	Service Group	81,80	
2021	Heng, et al. [24]	XGBoost	UTMobileNetTraffic2021	Application	79,10	
2021	Hu, et al. [31]	CNN + LSTM	ISCX VPN-Non VPN	Binary Application	98,00 92,89	
2021	Lu, et al. [28]	Inception-LSTM	ISCX VPN- Non VPN	Service Group	98,10	
2021	Dong, et al. [27]	GADCN	USTC-TFC2016	Malware	99,22	
2021	Liu and Lang [29]	CNN + DBSCAN	ISCX VPN- Non VPN DARPA 1998	Protocol	97,03 98,50	
2021	Lu, et al. [30]	NIN + stepwise pruning and knowledge distillation (KD)	ISCX VPN- Non VPN	Application	98,86	
2021	Soleymanpour, et al. [6]	CSCNN	ISCX VPN- Non VPN	Application Service Group	97,90 97,70	
2022	Zheng, et al. [32]	MTT(multi-task transformer)	ISCX VPN- Non VPN	Application Service Group	98,75 99,34	
2022	Yao, et al. [33]	LSTM HAN	ISCX VPN- Non VPN	Service Group	91,20 89,50	
2022	Izadi, et al. [10]	Fusion (CNN, DBN, MLP)	ISCX VPN- Non VPN	Service Group	97,00	
2022	Telikani, et al. [34]	Cost + CNN SMOTE+ CNN Cost +SAE SMOTE +SAE	ISCX VPN- Non VPN	Application		95,50 92,40 94,00 90,30
2022	Proposed Model RFSE-GRU	GRU + RF feature selection + SMOTE-ENN	ISCX VPN- Non VPN ISCX VPN- Non VPN UTMobileNetTraffic2021	Binary Service Group Application	93,91 82,69 96,83	93,91 81,69 96,92

As can be seen from Table 1, many studies have been carried out in the literature for the classification of encrypted network traffic. For this purpose, especially the ISCX VPN-Non VPN dataset is frequently used in the literature. However, when the studies in the literature are examined in detail, the ISCX VPN-Non VPN datasets used differ. For instance, Wang et al. [14] made evaluations on the

ISCX VPN-Non VPN dataset in their study, but although the dataset used includes 206688 samples, there is also a difference in the number of classified applications. Similarly, Chiu et al. [17] made evaluations on the ISCX VPN-Non VPN dataset in their study, but the dataset used includes 222797 samples, but there is also a difference in the number of classified applications. Similar to these examples, when other studies such as Lotfollahi et al. [4], Bu et al. [8], Chen et al. [20], Izadi et al. [10] and Hu et al. [31] mentioned in Table 1 are examined, it can be seen that they have different dataset classes and sizes. The dataset in [35] was used in this study. It is considered that the data set may have been changed over the years as the reason for the difference in the studies.

III. MATERIAL AND METHODS

In this study, a new classification model named RFSE-GRU is proposed. The proposed model aims to classify rapidly growing encrypted mobile network traffic data according to various characteristics with high accuracy performance. In the proposed model, GRU deep learning algorithm, feature selection based on the RF algorithm and SMOTE-ENN data balancing technique are used. In this section, first, the datasets used in the study are introduced. Then, the pre-processing steps performed to prepare the datasets for classification are mentioned. The feature selection method used to improve the classification performance has been introduced and the feature size has been determined. Then, oversampling and undersampling algorithms used for data balancing are explained. Finally, information is given about the GRU algorithm used in the developed model.

A. DATASETS

1) ISCX VPN-Non VPN

ISCX VPN-Non VPN [35] is a publicly published dataset by the Canadian Institute for Cybersecurity. It contains traffic data collected by the Information Security Center of Excellence. Traffic was captured using Wireshark and tcpdump. The original dataset is shared in pcap format. The dataset contains two types of traffic, virtual private networks (VPN) and regular traffic. Traffic types cover various service groups such as Chat, VoIP, and Streaming. Service groups and sample distribution are given in Table 2. CICFlowMeter [36] available by the Canadian Institute for Cybersecurity is used to extract flow characteristics from pcap files containing raw data. CICFlowMeter is a Java-based tool that generates 84 attributes from pcap files. Flow features were extracted from the original pcap files with CICFlowMeter and saved in csv format. Data files in csv format belonging to 10 service groups used in the study were labeled and combined using Python. There are 45,537 records in the created data set. The attributes and definitions created for the ISCX VPN-Non VPN dataset are given in Table 3.

TABLE 2. The number of samples in each class of ISCX VPN-Non VPN dataset.

Service Group	Content	Sample
Chat	AimChat, Facebook, Hangouts, ICQ, Skype	6596
Email	Email	5071
File Transfer	FTPS, SFTP, Skype	1598
Streaming	Vimeo, Youtube, Spotify, Netflix	3335
VoIP	Facebook, Hangouts, Skype, VOIP Buster	7101
VPN-Chat	AimChat, Facebook, Hangouts, ICQ, Skype	4540
VPN-Email	Email	569
VPN-File Transfer	FTPS, SFTP, Skype	1794
VPN-Streaming	Vimeo, Youtube, Spotify, Netflix	1137
VPN-VoIP	Facebook, Hangouts, Skype, VOIP Buster	13796
	Total	45537

2) UTMobileNet2021

UTMobileNet2021 [37] is a publicly available dataset for mobile traffic research, released in 2021. The dataset was created by capturing more than 21 million mobile traffic packets with tcpdump in a controlled environment and presented in csv format. The dataset includes 16 popular mobile traffic applications such as Facebook, Twitter, Instagram and the activities performed on them. Attributes consist of packet header information. A total of 42 selected attributes of Frame, Link, IP, TCP and UDP header information are used. The dataset creates four different scenarios: Wild Test Data, Action-Specific Wild Test Data, Deterministic Automated Data and Randomized Automated Data.

In the study, analyses were carried out on the Wild Test Data scenario. The Wild Test Data scenario consists of csv files containing 14 different application traffic. The data types saved in different formats in the data set files are organized. Similarly, the Location column, which only exists in some data files, has been removed. Each data file was labeled according to the 14 applications used in the study and assembled using Python. There are 634,958 records in the created dataset. The applications and sample numbers used in the study are given in Table 4. The attributes and definitions created for the UTMobileNet2021 data set are given in Table 5.

B. DATA PREPROCESSING

Generally, datasets need to go through some preprocessing in order to adapt them to the format that machine learning and deep learning algorithms can process and to increase the efficiency obtained. ISCX VPN-Non VPN and UTMobileNet2021 datasets used in this study have gone through some preprocessing steps.

TABLE 3. Detailed description of the attributes of the ISCX VPN-Non VPN dataset.

ID	Attribute Name	Attribute Description
0	Flow ID	Id number of the flow
1	Src IP	Source IP address
2	Src Port	Source Port
3	Dst IP	Destination IP address
4	Dst Port	Destination Port
5	Protocol	Transport Protocol
6	Timestamp	Start time flow first seen
7	Flow Duration	Duration of the flow in Microsecond
8	Tot Fwd Pkts	Total packets in the forward direction
9	Tot Bwd Pkts	Total packets in the backward direction
10	TotLen Fwd Pkts	Total size of packet in forward direction
11	TotLen Bwd Pkts	Total size of packet in backward direction
12	Fwd Pkt Len Max	Maximum size of packet in forward direction
13	Fwd Pkt Len Min	Minimum size of packet in forward direction
14	Fwd Pkt Len Mean	Mean size of packet in forward direction
15	Fwd Pkt Len Std	Standard size of packet in forward direction
16	Bwd Pkt Len Max	Maximum size of packet in backward direction
17	Bwd Pkt Len Min	Minimum size of packet in backward direction
18	Bwd Pkt Len Mean	Mean size of packet in backward direction
19	Bwd Pkt Len Std	Standard size of packet in backward direction
20	Flow Byts/s	Number of flow bytes per second
21	Flow Pkts/s	Number of flow packets per second
22	Flow IAT Mean	Mean time between two packets sent in the flow
23	Flow IAT Std	Standard deviation time between two packets sent in the flow
24	Flow IAT Max	Maximum time between two packets sent in the flow
25	Flow IAT Min	Minimum time between two packets sent in the flow
26	Fwd IAT Tot	Total time between two packets sent in the forward direction
27	Fwd IAT Mean	Mean time between two packets sent in the forward direction
28	Fwd IAT Std	Standard deviation time between two packets sent in the forward direction
29	Fwd IAT Max	Maximum time between two packets sent in the forward direction
30	Fwd IAT Min	Minimum time between two packets sent in the forward direction
31	Bwd IAT Tot	Total time between two packets sent in the backward direction
32	Bwd IAT Mean	Mean time between two packets sent in the backward direction
33	Bwd IAT Std	Standard deviation time between two packets sent in the backward direction
34	Bwd IAT Max	Maximum time between two packets sent in the backward direction
35	Bwd IAT Min	Minimum time between two packets sent in the backward direction
36	Fwd PSH Flags	Number of times the PSH flag was set in packets travelling in the forward direction
37	Bwd PSH Flags	Number of times the PSH flag was set in packets travelling in the backward direction
38	Fwd URG Flags	Number of times the URG flag was set in packets travelling in the forward direction
39	Bwd URG Flags	Number of times the URG flag was set in packets travelling in the backward direction
40	Fwd Header Len	Total bytes used for headers in the forward direction
41	Bwd Header Len	Total bytes used for headers in the backward direction
42	Fwd Pkts/s	Number of forward packets per second
43	Bwd Pkts/s	Number of backward packets per second
44	Pkt Len Min	Minimum length of a packet
45	Pkt Len Max	Maximum length of a packet
46	Pkt Len Mean	Mean length of a packet
47	Pkt Len Std	Standard deviation length of a packet
48	Pkt Len Var	Variance length of a packet
49	FIN Flag Cnt	Number of packets with FIN
50	SYN Flag Cnt	Number of packets with SYN
51	RST Flag Cnt	Number of packets with RST
52	PSH Flag Cnt	Number of packets with PUSH
53	ACK Flag Cnt	Number of packets with ACK
54	URG Flag Cnt	Number of packets with URG
55	CWE Flag Count	Number of packets with CWR
56	ECE Flag Cnt	Number of packets with ECE
57	Down/Up Ratio	Download and upload ratio
58	Pkt Size Avg	Average size of packet
59	Fwd Seg Size Avg	Average size observed in the forward direction
60	Bwd Seg Size Avg	Average size observed in the backward direction
61	Fwd Byts/b Avg	Average number of bytes bulk rate in the forward direction
62	Fwd Pkts/b Avg	Average number of packets bulk rate in the forward direction
63	Fwd Blk Rate Avg	Average number of bulk rate in the forward direction

TABLE 4. The number of samples in each class of UTMobileNet2021 dataset.

Application	Samples
Facebook	21004
Gmail	4338
Google Drive	218098
Google Maps	41919
Hangouts	1927
Hulu	85044
Instagram	46207
Messenger	1288
Netflix	58684
Pinterest	59167
Reddit	39493
Spotify	33726
Twitter	23672
Youtube	391
Total	634958

First of all, service groups were labeled in the ISCX VPN-Non VPN dataset. Then, ''Flow ID'', ''Src IP'', ''Dst IP'', ''Timestamp'' attributes that might cause bias and 14 features with a standard deviation of 0, i.e., affecting each sample equally, were extracted. 66 features in total remained after these processes. Categorical Label values from Scikitlearn library have been converted to numeric values by Label Encoder. The numeric labels of the classes can be seen in Table 6. For the remaining features, normalization was applied using MinMaxScaler.

In the binary classification of traffic types, VPN and Non-VPN, of the ISCX VPN-Non VPN dataset, traffic types were converted to numerical values with Label Encoder. The numeric labels of the classes can be seen in Table 7.

In the UTMobileNet2021 dataset, labeling the study, the ''udp.srcport'', ''udp.dstport'', ''udp.length'', ''udp.checksum'', ''gquic.puflags.rsv'', ''gquic.packet_number'' attributes containing UDP header information were removed. Also, the ''frame.time'' and ''frame.number'' attributes have been removed. Missing values are filled with ''most_frequent'' and ''median'' using SimpleImputer from Scikitlearn library. The labeled ''Application'' column and other string attributes were converted to numeric with Label Encoder. Then, 8 features with a standard deviation of 0, i.e., affecting each sample equally were extracted. After these operations, a total of 26 features remained. The numeric labels of the Application classes are seen in Table 8. Normalization was applied using MinMaxScaler for the remaining features.

TABLE 5. detailed description of the attributes of the UTMobileNet2021 dataset.

ID	Attribute Name	Attribute Description
0	frame.number	Frame Number
1	frame.time	Arrival Time
2	frame.len	Frame length on the wire
3	frame.cap_len	Frame length stored into the capture file
4	sll.pkttype	Packet type
5	sll.hatype	Link-layer address type
6	sll.halen	Link-layer address length
7	sll.src.eth	Source of Ethernet or other MAC address
8	sll.unused	Unused
9	sll.etype	Protocol
10	ip.hdr_len	Header Length
11	ip.dsfield.ecn	Explicit Congestion Notification
12	ip.len	Total Length
13	ip.id	Identification
14	ip.frag_offset	Fragment Offset
15	ip.ttl	Time to Live
16	ip.proto	Protocol
17	ip.checksum	Header Checksum
18	ip.src	Source Address
19	ip.dst	Destination Address
20	tcp.hdr_len	Header Length
21	tcp.len	TCP Segment Len
22	tcp.srcport	Source Port
23	tcp.dstport	Destination Port
24	tcp.seq	Sequence Number
25	tcp.ack	Acknowledgment Number
26	tcp.flags.ns	Nonce
27	tcp.flags.fin	Fin
28	tcp.window_size_value	Window
29	tcp.checksum	Checksum
30	tcp.urgent_pointer	Urgent Pointer
31	tcp.option_kind	Kind
32	tcp.option_len	Length
33	tcp.options.timestamp.tsval	Timestamp value
34	tcp.options.timestamp.tsecr	Timestamp echo reply
35	udp.srcport	Source Port
36	udp.dstport	Destination Port
37	udp.length	Length
38	udp.checksum	Checksum
39	gquic.puflags.rsv	Reserved
40	gquic.packet_number	Packet Number
41	application	Application class

In the study, before the feature selection stage, the datasets used at the end of the preprocessing steps, the number of samples and features and the class numbers in each dataset are summarized in Table 9.

TABLE 6. Class name and numeric labels of the ISCX VPN-Non VPN dataset.

Label	Class
0	Chat
1	Email
2	File Transfer
3	Streaming
4	VoIP
5	VPN-Chat
6	VPN-Email
7	VPN-File Transfer
8	VPN-Streaming
9	VPN-VoIP

TABLE 7. Class name and numeric labels of the ISCX VPN-Non VPN dataset for binary classification.

Label	Class
0	Non-VPN
1	VPN

TABLE 8. Class name and numeric labels of the UTMobileNet2021 dataset.

Label	Class
0	Facebook
1	Gmail
2	Google Drive
3	Google Maps
4	Hangouts
5	Hulu
6	Instagram
7	Messenger
8	Netflix
9	Pinterest
10	Reddit
11	Spotify
12	Twitter
13	Youtube

C. RANDOM FOREST BASED FEATURE SELECTION

Feature selection is the process of determining the most relevant and appropriate feature set for classification by removing unnecessary features. Feature selection is an important process in machine learning and deep learning algorithms to reduce the processing load and improve the performance of models. Therefore, feature selection is a critical factor in classifying Network traffic data with many features. In this study, the Random Forest algorithm was used for feature selection.

Random Forest [38] is a popular supervised ensemble algorithm used for classification and regression. The Random Forest algorithm is based on the conclusion of a large number of decision trees consisting of subsets of the dataset by majority voting. Decision trees start branching with the attribute with the highest information gain in the nodes. The selection of the feature to be branched depends on the importance of the feature. Calculating the importance of the features in the dataset is effective in feature selection and reduction of features with relatively low impact levels. The ability to sort the features according to their importance makes it possible to use the Random Forest algorithm for feature selection. The importance of an attribute in the decision tree is determined by how much the nodes in the tree increase the purity of other trees in the forest. The Gini Index is a measure of the impurity ratio in feature selection. Accordingly, the attribute with the smallest Gini Index is considered the attribute with the purest and highest importance. To determine the purity of a v-node, the Gini

Index is calculated as follows [39]:

$$Gini\,(v) = \sum_{i=1}^{l} f_i\,(1 - f_i) \tag{1}$$

where l represents the total number of classes, and *fi* represents the frequency of class i. To branch the *v* -node, the gain of the *Xi*attribute is calculated as follows:

$$gain\,(X_i, v) = Gini\,(X_i, v) - W_L Gini\left(X_i, v^L\right) - W_R Gini\left(X_i, v^R\right) \tag{2}$$

where, *vL* and *vR* represent the left and sal child nodes of *v*, and *WL* and *WR* represent the number of samples reaching the child nodes. The severity for the *Xi* attribute is calculated as follows:

$$Importance_i = \frac{1}{n_{tree}} \sum_{k \in SX_i} gain\,(X_i, v) \tag{3}$$

where *ntree* is the number of trees, and $k \in SXi$ is the set of branched nodes. The obtained significance/score is normalized and reduced to the [0,1] range. As a result of the Random Forest fetaure selection process, the importance degrees of the features in the ISCX VPN-Non VPN and UTMobileNet2021 datasets are shown in Table 10.

TABLE 9. Comparison of datasets.

Datasets	# of records	# of features used	Classes
ISCX2016 Vpn- Non Vpn (Binary)	45537	65	2
ISCX2016 Vpn- Non Vpn	45537	65	10
UTMobileNet2021	634958	25	14

The threshold value of the significance levels obtained by the Random Forest feature selection method was determined as 0.02. Accordingly, the features above the threshold value were selected for classification studies. The graphs in Fig. 1a and Fig. 1b show comparative importance of the attributes of the ISCX VPN-Non VPN and UTMobileNet2021 datasets.

As a result, feature selection was made to improve the performance of the proposed classification model in this study. As a result of the feature selection process, the number of features was reduced, and 15 features were used in the ISCX VPN-Non VPN dataset and 13 features in the UTMobileNet2021 dataset. While the attributes used in the ISCX VPN-Non VPN dataset are 0, 1, 3, 16, 17, 18, 20, 21, 27, 28, 30, 31, 34, 36, 55; the attributes used in the UTMobileNet2021 dataset are 0, 2, 5, 8, 9, 11, 12, 13, 16, 17, 18, 20, 24.

D. DATA BALANCING

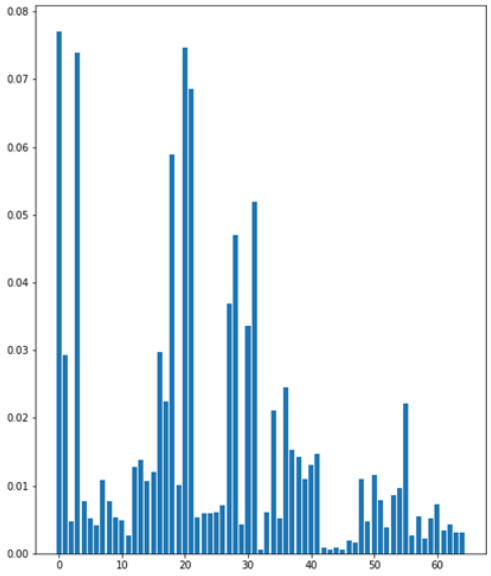
ISCX VPN-Non VPN and UTMobileNet2021 datasets consist of imbalanced classes. imbalanced class distributions negatively affect the performance of classification models. The use of only oversampling techniques in data balancing processes can cause an increase in unnecessary data. On the other hand, using only undersampling techniques can result in serious reduction of training data. In this study, SMOTE oversampling technique and ENN undersampling technique are used together to solve the data balancing problem.

TABLE 10. Importance score of features of datasets according to random forest Feature selection method.

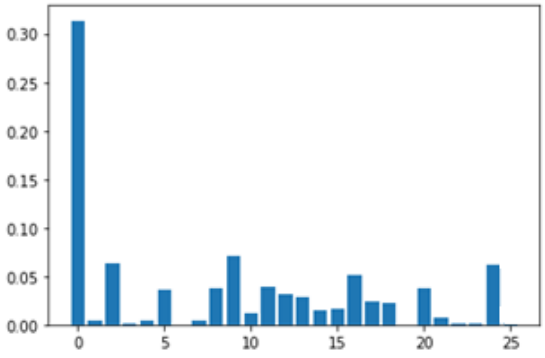
	UTMobileNet2021	ISCX VPN-Non VPN
Feature Number	Score	Score
0	0.31353	0.07705
1	0.00383	0.02925
2	0.06279	0.00467
3	0.00191	0.07397
4	0.00399	0.00765
5	0.03533	0.00521
6	0.00030	0.00410
7	0.00367	0.01083
8	0.03746	0.00774
9	0.07101	0.00538
10	0.01141	0.00484
11	0.03964	0.00261
12	0.03124	0.01274
13	0.02854	0.01383
14	0.01462	0.01070
15	0.01664	0.01202
16	0.05059	0.02980
17	0.02352	0.02250
18	0.02286	0.05893
19	0.00001	0.01005
20	0.03810	0.07474
21	0.00682	0.06863
22	0.00122	0.00529
23	0.00074	0.00588
24	0.06188	0.00598
25	-	0.00609
26	-	0.00718
27	-	0.03693
28	-	0.04697
29	-	0.00436
30	-	0.03368
31	-	0.05190
32	-	0.00059
33	-	0.00613
34	-	0.02103
35	-	0.00513
36	-	0.02445
37	-	0.01527
38	-	0.01430
39	-	0.01097
40	-	0.01313
41	-	0.01474
42	-	0.00087
43	-	0.00060
44	-	0.00085
45	-	0.00062
46	-	0.00194
47	-	0.00161
48	-	0.01098
49	-	0.00480

1) SMOTE

Synthetic Minority Oversampling Technique [40] is a widely used oversampling method. It is aimed to balance the number of samples between classes by generating synthetic data from minority class samples. SMOTE works by using the



(a)



(b)

FIGURE 1. Random forest feature scores as a result of random forest feature selection for: a) ISCX VPN-Non VPN dataset; b) UTMobileNet2021 dataset.

similarity of minority samples in feature space instead of data space. In the SMOTE algorithm, the k-nearest neighbors are determined by calculating the Euclidean distance for each sample in the minority class. One of the determined neighbors is selected and a vector is created between it and the sample. The vector is added to the sample by multiplying it with a value between 0 and 1. In this way, a new synthetic sample is created to be included in the minority class. The desired number of random nearest neighbors is selected and the synthetic sampling process continued depending on the desired oversampling rate. The problem of excessive learning due to the imbalance between classes can be solved [40] by producing synthetic examples with SMOTE. Oversampling with SMOTE is expressed as:

$$n = s + d \cdot \left(s^R - s\right), 0 \leq d \leq 1 \tag{4}$$

where, s represents the sample belonging to the minority class, sRrepresents the neighbor determined by the k-nearest neighbors, d represents the randomly selected value in the range [0,1], and the n generated synthetic samples.

2) ENN

Edited Nearest Neighbor [41] is a subsampling method based on the k-nearest neighbor method. The ENN method is based on removing misclassified samples. By calculating the Euclidean distance of each sample, the k-nearest neighbors are determined. The number of nearest neighbors is generally accepted as $k = 3$. If the class of a determined instance is different from the class of the majority of its k-nearest neighbors, both the determined instance and its k-nearest neighbors are removed. The ENN algorithm will continue the sample reduction process until the desired equilibrium ratio between the classes is achieved. Samples belonging to both majority and minority classes can be removed from the training dataset with the ENN method.

3) SMOTE-ENN

It is a hybrid balancing approach developed by Batista et al. [42]. In this approach, SMOTE and ENN algorithms are implemented with a pipeline. First, synthetic minority samples are created with the SMOTE algorithm and the effect of the minority class on the data set is increased. Oversampling continues until the prejudice created by the minority class in the data set is broken. Then, the data reduction process is applied by checking the similarity between the class of the sample with the ENN algorithm and the class determined by k-NN. The subsampling process continues until the desired optimal learning success is achieved [43]. Fig. 2 and Fig.3 show the final state of the balanced datasets resulting from the SMOTE+ENN process applied to the ISCX VPN-Non VPN and UTMobileNet2021 datasets, respectively.

E. GATED RECURRENT UNITS(GRU)

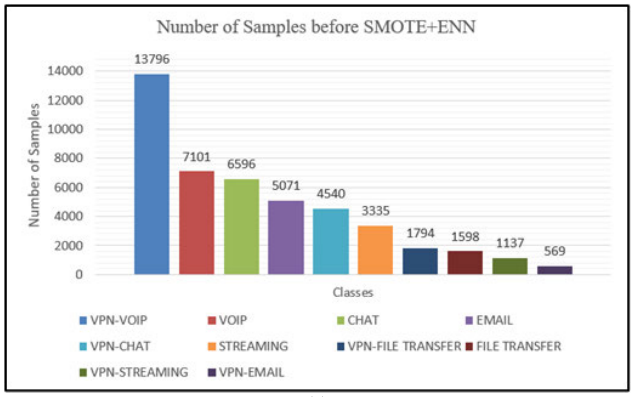
It is a type of feedback neural network designed to solve the lost gradient in recurrent neural networks (RNNs). It has similar structures with to the LSTM algorithm. As seen in Fig. 4, the cells have two gates as an upgraded gate and a reset gate. The upgrade gate is the gate that decides how much of the historical information obtained up to a certain point will be transferred forward. The reset gate is the gate that decides how much of the information should be discarded from memory at a given moment. The mathematical formulas of the gates are as shown below:

$$Upgrade\ Gate\ (z_t) = \sigma_g(W_z x_t + U_z h_{t-1} + b_z) \tag{5}$$

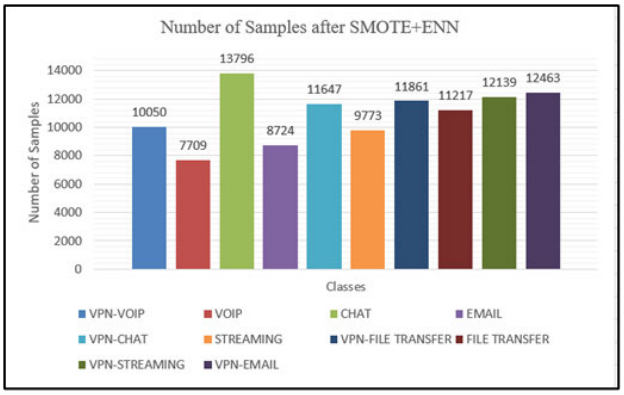
$$Reset\ Gate\ (r_t) = \sigma_g(W_r x_t + U_r h_{t-1} + b_r) \tag{6}$$

IV. PROPOSED METHOD

In this study, a unique classification model based on deep learning is proposed for the classification of encrypted mobile traffic data called RFSE-GRU. With the widespread use

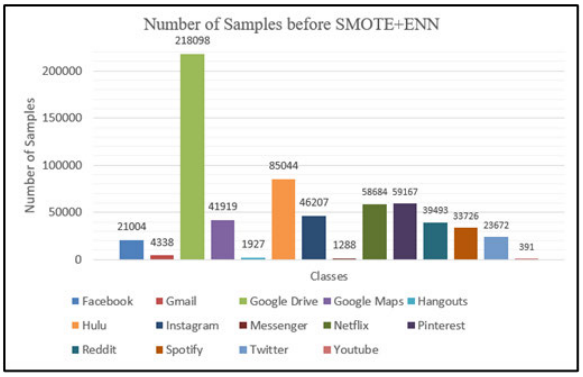


(a)

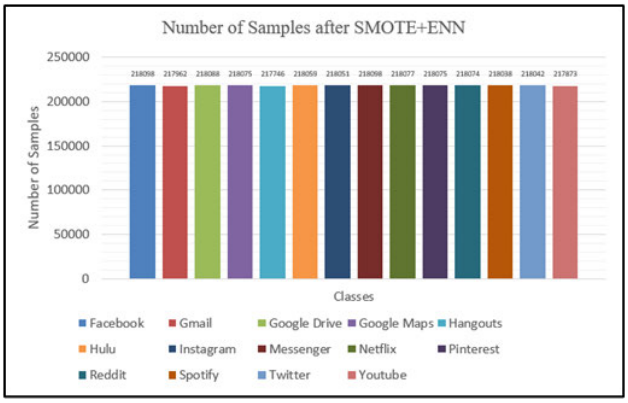


(b)

FIGURE 2. ISCX VPN Non-VPN dataset a) before data balancing b) after data balancing.



a)



b)

FIGURE 3. UTMobileNet2021 dataset a) before data balancing b) after data balancing.

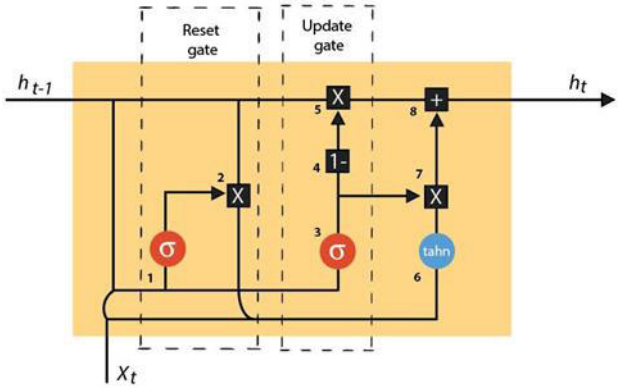


FIGURE 4. Gated Recurrent Units (GRU): Cell Architecture.

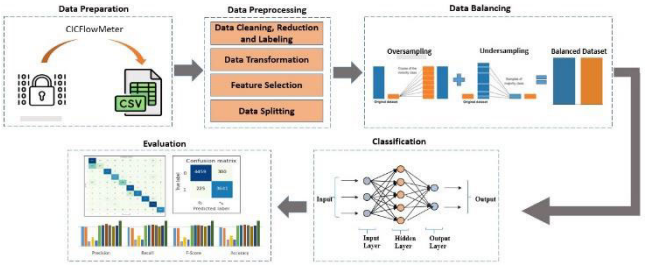


FIGURE 5. A schematic diagram of the proposed method.

of mobile technologies and the Internet, traffic in mobile networks is increasing. This has made traffic classification an important element for data security and network management. However, encryption of traffic in modern networks makes it difficult to classify traffic with traditional methods. The main purpose of the proposed system is to provide a successful classification of increasingly encrypted mobile network traffic. For this purpose, GRU deep learning algorithm is combined with the Random Forest feature selection algorithm and SMOTE+ENN data imbalance processing in the proposed system. The proposed system consists of data preparation, data preprocessing, data balancing, classification and evaluation sections as shown in Fig. 5.

In the data preparation section, the flow properties in the pcap files containing raw data were converted to csv format and the data sets were made processable. In the data preprocessing stage, raw data sets are processed and made ready for classification algorithms. In addition, with the feature selection in the data preprocessing stage, the feature size is adjusted to maximize the algorithm performance. In the dataset splitting phase, the dataset is divided into two a train dataset and a test dataset in accordance with training and testing purposes. In the data balancing phase, oversampling is done by resampling the data with SMOTE according to the minority class. Then, the Edited Nearest Neighbor (ENN) undersampling approach was used in order to avoid overfitting problems and to reduce the noise in the newly

produced data. At this stage, with the combination of SMOTE and Edited Nearest Neighbor methods, the dataset becomes balanced, and the classification performance is increased. In the classification phase, mobile network traffic is classified using the GRU deep learning algorithm. Finally, the results obtained in the evaluation phase are evaluated according to the evaluation parameters and the performance of the method is determined.

V. EXPERIMENTS AND EVALUATIONS

The model proposed in this section is implemented on the ISCX VPN-Non VPN and UTMobileNet2021 datasets. The studies were carried out on Google Colab with Pyspark supported by Apache Spark's big data platform. Scikitlearn, Keras and Spark MLlib libraries were used in the setup of machine learning and deep learning models. The proposed model was compared with eleven different machine learning, deep learning and hybrid methods and their performance was evaluated.

A. MODEL PARAMETERS

In the study, datasets were divided into two an 80% training set and a 20% test set during the data splitting phase. In the model established with the Random Forest algorithm, maxDept= 20 and maxBins=50. Maximum depth expresses the frequency of the tree, and as the depth increases, the chance of capturing more information about the data increases. Similarly, in the model established with the Decision Tree algorithm, it is specified as maxDept= 20 and maxBins=50. For the Naive Bayes algorithm, the smoothing value is 1.0 and the model type is Multinomial Naive Bayes parameters. All parameters used in the Logistic Regression model are used by default. The model established with MLP has four layers. Several neurons equal to the number of features used in the data sets were used in the input layer. There are 5 and 4 nerve cells in the hidden layers. In the study, the output layer took the values of 14 in the UTMobileNet2021 dataset, 10 for multiclass classification and 2 for binary classification in the ISCX VPN-Non VPN dataset. In addition, the maximum number of iterations for the MLP model is 100 and the block size is 128.

Models installed with Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) have three layers. For both algorithms, 128 neurons were used in the input and hidden layers and 100 neurons were used in the output layer. In the first and second convolution layers of the model built with the Convolutional Neural Network (CNN-1D), filter 32 and kernel size 3 are taken. In the output layer, the filter is configured as 100. The CNN-LSTM model has a convolution layer and an LSTM layer. In the first convolution layer, filter 32 and kernel size 3 are accepted. In pooling layers, the stride parameter is 2. There are 128 neurons in the LSTM layer. A Dense layer of 100 neurons was then used. The LSTM-CNN model, on the other hand, has an LSTM and a convolution layer, like the CNN-LSTM model. The same parameters used in the CNN-LSTM model were used in the LSTM-CNN model. The CNN-GRU model has a convolution layer and a GRU layer. In the first convolution layer, filter 32 and kernel size 3 are accepted. In pooling layers, the stride parameter is 2. There are 128 neurons in the GRU layer. A Dense layer of 100 neurons was then used. The GRU-CNN model, on the other hand, has a GRU and a convolution layer, like the CNN-GRU model. The same parameters used in the CNN-GRU model were used in the GRU-CNN model.

In LSTM, GRU, CNN-1D and CNN-LSTM models built using the Keras library, the ReLu activation function is used in input and hidden layers. The output layer has a value of 14 in the UTMobileNet2021 dataset, 10 for multiclass classification and 1 for binary classification in the ISCX VPN-Non VPN dataset. In the output layer, the activation function was determined as Softmax in multiclass classification and Sigmoid in binary classification. The loss function used in the model is categorical cross-entropy in multiclass classification and binary cross-entropy in binary classification. The optimization algorithm is Adam. Models were run for 100 epochs.

B. RESULTS AND COMPARISON

The results obtained in the study were discussed from three different perspectives. First, the performance of the proposed model in classifying various categories in different datasets was measured. The obtained results were compared with various machine learning and deep learning algorithms. At the same time, the success of the proposed RFSE-GRU model against hybrid deep learning methods was also evaluated. Secondly, the effects of data balancing on classification performance are examined. Finally, the feature selection method used in the proposed model is superior to other feature selection methods.

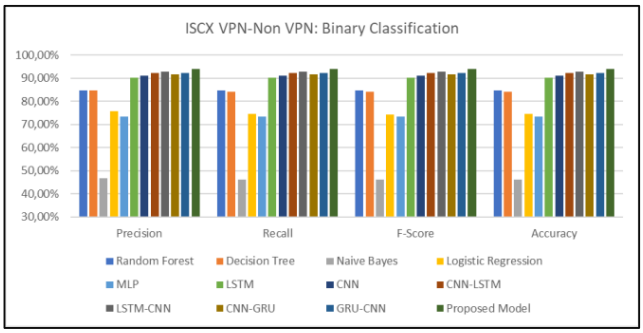


FIGURE 6. Performance comparison of models for binary classification of ISCX VPN-Non VPN dataset.

Fig. 6 presents the comparison of the proposed model and the Accuracy, Precision, Recall and F-Score parameters of various machine learning and deep learning algorithms obtained according to binary classification performed on the ISCX VPN-Non VPN dataset. When Fig. 6 is examined, it is seen that the proposed model achieves the best results for each parameter.

According to Table 11, which shows the detailed results of the algorithms according to the evaluation parameters, the

TABLE 11. Accuracy, precision, recall and F-Score results for ISCX VPN-Non VPN dataset for binary classification.

Classifier	Precision	Recall	F-Score	Accuracy
Random Forest	84,78%	84,72%	84,73%	84,72%
Decision Tree	84,57%	84,10%	84,10%	84,10%
Naive Bayes	46,62%	46,10%	46,16%	46,10%
Logistic Regression	75,73%	74,52%	74,38%	74,52%
MLP	73,34%	73,29%	73,30%	73,29%
LSTM	90,22%	90,20%	90,18%	90,20%
CNN	91,19%	91,18%	91,17%	91,18%
CNN-LSTM	92,17%	92,17%	92,19%	92,17%
LSTM-CNN	92,79%	92,78%	92,77%	92,78%
CNN-GRU	91,78%	91,70%	91,65%	91,70%
GRU-CNN	92,14%	92,13%	92,13%	92,13%
Proposed Model	**93,91%**	**93,91%**	**93,91%**	**93,91%**

proposed model achieved higher classification performance than other models with an accuracy rate of 93.91%. In addition, according to the results, it has been observed that individual and hybrid deep learning algorithms provide higher classification performance than traditional machine learning algorithms. On the other hand, it is noteworthy that Naive Bayes is the algorithm with the lowest performance in binary classification.

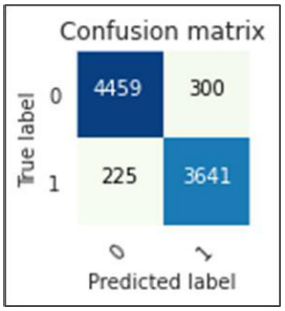


FIGURE 7. Confusion matrix of proposed model for binary classification of ISCX VPN-Non VPN dataset.

The classification results of the proposed method in the binary classification of the ISCX VPN-Non VPN dataset are shown on the confusion matrix in Fig. 7. It can be seen from the confusion matrix that the proposed model is successful in VPN-Non VPN classification. Class labels shown numerically in the confusion matrix are given in Table 7.

TABLE 12. Accuracy, precision, recall, and F-Score results according to classes in THE binary classification of ISCX VPN-Non VPN dataset.

Class	Precision	Recall	F-Score	Accuracy
VPN	92,39%	94,18%	93,28%	94,18%
Non-VPN	95,20%	93,67%	94,44%	93,69%

In Table 12, detailed results of the binary classification of the ISCX VPN-Non VPN dataset are presented according to the evaluation parameters of the classes. According to the results obtained, it has been observed that VPN traffic is classified with a higher accuracy rate.

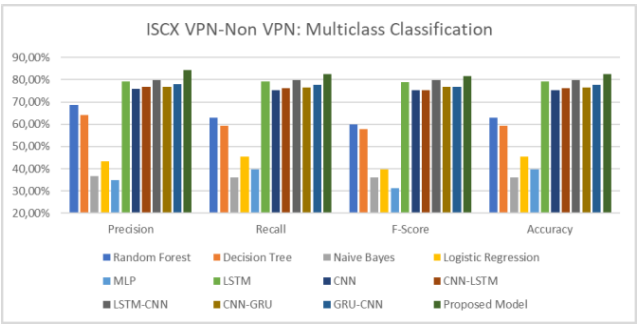


FIGURE 8. Performance comparison of models for multiclass classification of ISCX VPN-Non VPN dataset.

Fig. 8 presents the comparison of the Accuracy, Precision, Recall and F-Score parameters of the machine learning and deep learning algorithms used in the study, obtained according to the multiclass classification performed on the ISCX VPN-Non VPN dataset. When Fig. 8 is examined, it is seen that the proposed method achieves the best results for each parameter.

TABLE 13. Accuracy, precision, recall and F-Score results for ISCX VPN-Non VPN dataset for binary classification.

Classifier	Precision	Recall	F-Score	Accuracy
Random Forest	68,66%	63,07%	59,96%	63,07%
Decision Tree	64,24%	59,45%	57,87%	59,45%
Naive Bayes	36,62%	36,10%	36,16%	36,10%
Logistic Regression	43,47%	45,41%	39,80%	45,41%
MLP	34,76%	39,80%	31,16%	39,80%
LSTM	79,21%	79,22%	78,86%	79,22%
CNN	75,87%	75,37%	75,32%	75,37%
CNN-LSTM	76,83%	76,28%	75,29%	76,28%
LSTM-CNN	79,83%	79,86%	79,90%	79,86%
CNN-GRU	76,67%	76,65%	76,77%	76,65%
GRU-CNN	77,89%	77,86%	76,92%	77,86%
Proposed Model	**84,29%**	**82,68%**	**81,69%**	**82,68%**

According to Table 13, which shows the detailed results of the algorithms according to the evaluation parameters, the proposed model achieved higher classification performance than other models with an accuracy rate of 82.68%. In addition, it is seen that traditional machine learning algorithms perform quite poorly when compared to individual and hybrid deep learning algorithms. Naive Bayes and MLP algorithms are the algorithms that show the lowest performance in multiclass classification.

According to Table 13, which shows the detailed results of the algorithms according to the evaluation parameters, the proposed model achieved higher classification performance than other models with an accuracy rate of 82.68%. In addition, it is seen that traditional machine learning algorithms perform quite poorly when compared to individual and hybrid

TABLE 14. Accuracy, precision, recall, and F-Score results according to classes in THE multiclass classification of ISCX VPN-Non VPN dataset.

Class	Precision	Recall	F-Score	Accuracy
Chat	63,30%	86,17%	72,98%	86,17%
Email	86,48%	26,63%	40,72%	26,63%
File Transfer	92,29%	91,30%	91,79%	91,30%
Streaming	78,45%	83,48%	80,89%	83,48%
VoIP	87,19%	74,92%	80,59%	74,92%
VPN-Chat	77,61%	94,94%	85,40%	94,94%
VPN-Email	94,60%	79,67%	86,49%	79,67%
VPN-File Transfer	91,64%	87,64%	89,59%	87,64%
VPN-Streaming	81,82%	94,94%	87,89%	94,94%

deep learning algorithms. Naive Bayes and MLP algorithms are the algorithms that show the lowest performance in multiclass classification.

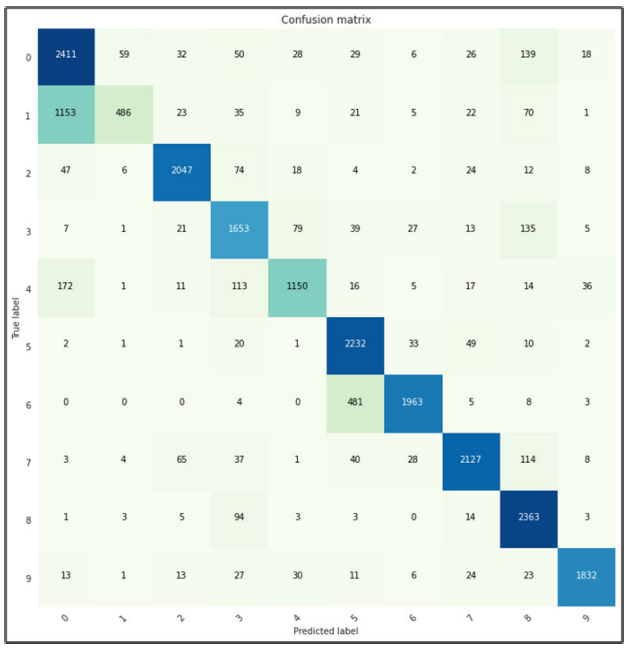


FIGURE 9. Confusion matrix of proposed model for multiclass classification of ISCX VPN-Non VPN dataset.

The classification results of the proposed method in multiclass classification of the ISCX VPN-Non VPN dataset according to service groups are given in Fig. 9 on the confusion matrix. Class labels shown numerically in the confusion matrix are given in Table 6. Accordingly, it is noteworthy that the proposed method is generally successful, but fails to classify the Email class.

In Table 14, detailed results according to the evaluation parameters of the classes in the multiclass classification of the ISCX VPN-Non VPN dataset are presented. According to the results, it has been observed that VPN service groups have higher accuracy rates than Non-VPN service groups. Examining the accuracy rates of the classes, it has been

TABLE 15. Accuracy, precision, recall and F-score results for the UTMobileNet2021 dataset.

Classifier	Precision	Recall	F-Score	Accuracy
Random Forest	81,82%	79,68%	78,34%	79,68%
Decision Tree	81,38%	77,77%	78,38%	77,77%
Naive Bayes	44,21%	47,49%	44,21%	47,49%
Logistic Regression	54,21%	57,49%	54,21%	57,49%
MLP	45,60%	48,13%	47,52%	48,13%
LSTM	83,34%	83,40%	83,33%	83,40%
CNN	85,01%	84,98%	85,02%	84,98%
CNN-LSTM	87,51%	86,71%	85,69%	86,71%
LSTM-CNN	85,41%	84,38%	83,54%	84,38%
CNN-GRU	82,29%	82,36%	81,69%	82,36%
GRU-CNN	83,75%	83,67%	83,65%	83,67%
Proposed Model	**97,03%**	**96,85%**	**96,92%**	**96,83%**

observed that the VPN-Chat and VPN-Streaming classes are classified with higher performance. In contrast, the VPN-Chat class has a lower Precision value than the VPN-Streaming class. On the other hand, although the data in the VPN-Email class is the class with the highest Precision value, it is observed that the Accuracy value has decreased. The fact that the Email class has the lowest accuracy rate in classification with 26.63% is an effective factor in the decrease in the overall classification performance of the model.

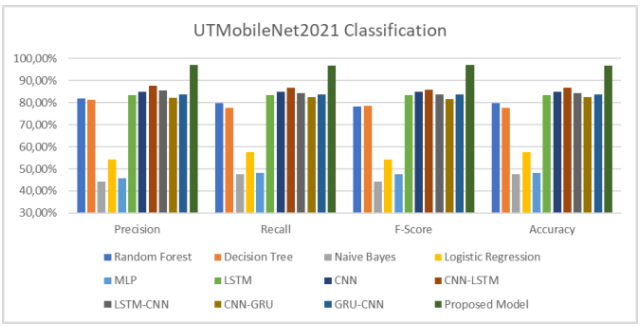


FIGURE 10. Performance comparison of models for UTMobileNet2021 dataset for multiclass classification.

Fig. 10 presents the comparison of the proposed model and the Accuracy, Precision, Recall and F-Score parameters of various machine learning and deep learning algorithms obtained according to the multiclass classification performed on the UTMobileNet2021 dataset. Fig. 10 shows that the proposed model achieves much higher performance for each parameter than other algorithms.

According to Table 15, which shows the detailed results of the algorithms according to the evaluation parameters, it is observed that the proposed model has a much higher classification performance than the other models with an accuracy rate of 96.83%. Also, traditional machine learning algorithms seem to perform worse than individual and hybrid deep learning algorithms. It is noteworthy that Naive Bayes

TABLE 16. Accuracy, precision, recall, and F-Score results according to classes in the multiclass classification of the UTMobileNet2021 dataset.

Class	Precision	Recall	F-Score	Accuracy
Facebook	97,81%	97,09%	97,45%	97,09%
Gmail	92,87%	96,98%	94,88%	96,98%
Google Drive	99,93%	97,60%	98,76%	97,60%
Google Maps	99,99%	97,04%	98,49%	97,04%
Hangouts	93,89%	97,09%	95,47%	97,09%
Hulu	97,88%	92,74%	95,24%	92,74%
Instagram	98,73%	98,27%	98,50%	98,27%
Messenger	92,54%	96,63%	94,54%	96,63%
Netflix	99,11%	97,13%	98,11%	97,13%

and MLP algorithms are the algorithms that show the lowest performance in multiclass classification.

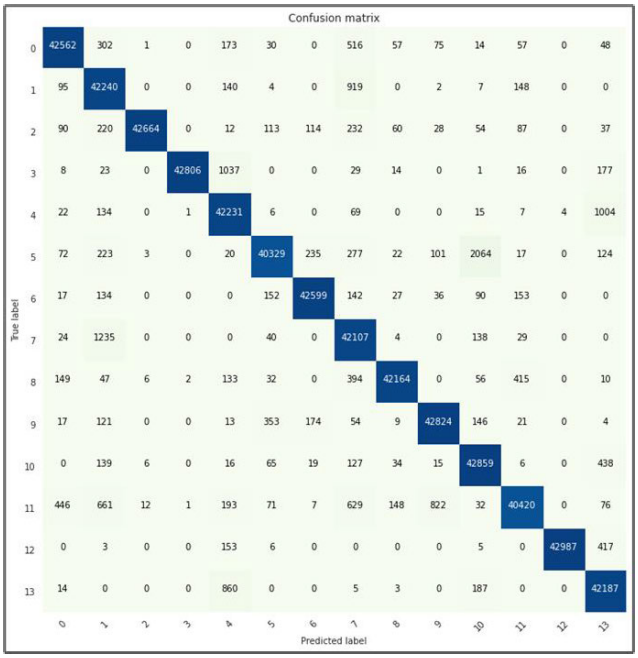


FIGURE 11. Confusion matrix of proposed model for multiclass classification of UTMobileNet2021 dataset.

Classification results of the proposed method in multiclass classification of the UTMobileNet2021 data set according to application types are given in Fig. 11 on the confusion matrix. Class labels shown numerically in the confusion matrix are given in Table 8. Examining the confusion matrix, it is seen that the proposed model is successful in classifying encrypted mobile network traffic according to application types.

In Table 16, detailed results according to the evaluation parameters of the classes in the multiclass classification of the UTMobileNet2021 dataset are presented. According to the results obtained, it is seen that all application types are classified with high accuracy performance. Twitter application is the application with the highest accuracy with 98.66%. The application classified with the lowest accuracy rate is Hulu with 92.74%.

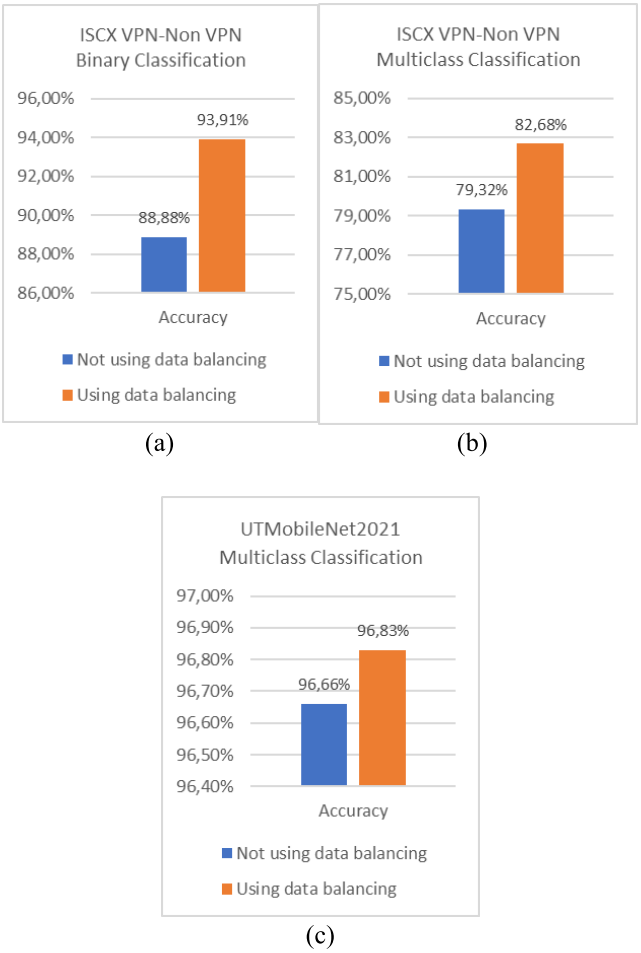


FIGURE 12. The effect of data balancing on the classification performance of the algorithm: (a) ISCX VPN-Non VPN Binary classification; (b) ISCX VPN-Non VPN multiclass classification; (c) UTMobileNet2021 multiclass classification.

Another evaluation method in the study is the data balancing phase. The effect of the SMOTE-ENN algorithm used in the data balancing process on the classification success was evaluated. Fig. 12 shows the effect of data balancing on the classification performance of encrypted mobile network traffic. As seen in Fig. 12a, the accuracy rate increased from 88.88% to 93.91% in binary classification made with the ISCX VPN-Non VPN dataset. Looking at Fig. 12b, it is seen that the accuracy rate increased from 79.32% to 82.68% in the multiclass classification made with the ISCX VPN-Non VPN dataset. Finally, in Fig. 12c, it is seen that the accuracy rate increased from 96.66% to 96.83% in the classification process performed on the UTMobileNet2021 VPN dataset. Accordingly, with the SMOTE-ENN data balancing approach, performance was improved by +5.03, +3.36 and +0.17, respectively, in three classification processes.

Finally, the effect of the algorithm used in feature selection on the classification success of encrypted mobile network traffic was evaluated. The effects of four different feature selection methods on the accuracy rates of the features

TABLE 17. Accuracy rates according to selected features with different feature selection algorithms for ISCX VPN-Non VPN dataset.

Feature Selection Algorithm	Feature No.	Accuracy
Sequential Forward	0,4,5,9,10,11,13,14,16,18	87,78%
Sequential Backward	0,3,4,5,7,11,12,17,18,19	88,93%
XGBoost	2,13,16,20,27,28,31,34,36,43	89,45%
Random Forest	**0,1,3,16,17,18,20,21,27,28,30,31, 34,36,55**	**93,91%**

were compared. In Table 17, the features selected by the feature selection methods in the ISCX VPN-Non VPN dataset and the accuracy rates obtained in binary classification are presented. Accordingly, it has been observed that the features selected with the Random Forest algorithm provide a higher accuracy rate compared to the features determined by other algorithms. With Random Forest, 15 features were selected and an accuracy rate of 93.91% was obtained.

The features selected by the feature selection methods in the UTMobileNet2021 dataset and the obtained accuracy rates are given in Table 18. Accordingly, 13 features were selected with Random Forest. In addition, Random Forest achieved a 96.83% accuracy rate, showing that it makes more efficient feature selection than other algorithms.

TABLE 18. Accuracy rates according to selected features with different feature selection algorithms for UTMobileNet2021 dataset.

Feature Selection Algorithm	Feature No.	Accuracy
Sequential Forward	3,4,5,7,8,10,11,12,14,15,18	88,62%
Sequential Backward	0,2,3,5,10,11,12,13,14,15,19	89,86%
XGBoost	1,3,7,8,11,15,16,19,21,23,24	90,12%
Random Forest	**0,2,5,8,9,11,12,13,16,17,18,20,24**	**96,83%**

VI. CONCLUSION AND FUTURE WORKS

As a result of the digitalizing world, the use of the internet and mobile technologies is increasing. However, with the increasing traffic flow on internet networks, ensuring security in networks has become more important. For data security and management of network resources, traffic needs to be classified. However, the classification process becomes difficult nowadays, with the encryption of the traffic or the use of private networks such as VPN. Conventional classification methods are insufficient for classifying mobile encrypted traffic. This situation gives importance to the studies on the classification of mobile encrypted traffic.

In this study, a new classification model is proposed for the classification of encrypted mobile network traffic. The proposed RFSE-GRU model is an approach that combines GRU deep learning algorithm, data balancing and feature selection methods. In the study, the SMOTE-ENN algorithm was used for data balancing. Random Forest algorithm is used in the feature selection process. The proposed model is compared with eleven different machine learning, deep learning and hybrid methods. The experiments in the study were carried out on the Apache Spark big data platform in the Google Colab environment. In the experiments, VPN-Non VPN classification, classification of service groups and classification of mobile applications were carried out using ISCX VPN-Non VPN and UTMobileNet2021 datasets. The performance of the proposed method was evaluated using Accuracy, Precision, Recall, and F-Score parameters. According to the experimental results, the proposed method obtained 93.91%, 82.68% and 96.83% accuracy values in the classification of VPN-Non VPN, service group and applications, respectively. According to the results, the proposed model showed higher classification performance than other algorithms. At the same time, the effect of data balancing on the classification of encrypted mobile network traffic is investigated. It has been observed that data balancing improves algorithm performance. On the other hand, the superiority of the Random Forest feature selection method proposed in the study over other feature selection methods has been demonstrated.

This study offers a new perspective on mobile encrypted traffic classification using a big data approach. In future studies, different data balancing methods will be evaluated for performance improvement. In addition, research will be carried out on the effectiveness and success level of the proposed method in different datasets.

REFERENCES

[1] G. Draper-Gil, A. H. Lashkari, M. S. I. Mamun, and A. A. Ghorbani, "Characterization of encrypted and VPN traffic using time-related features," in *Proc. 2nd Int. Conf. Inf. Syst. Secur. Privacy*, 2016, pp. 407–414.

[2] P. Wang, X. Chen, F. Ye, and Z. Sun, "A survey of techniques for mobile service encrypted traffic classification using deep learning," *IEEE Access*, vol. 7, pp. 54024–54033, 2019.

[3] A. S. Iliyasu and H. Deng, "Semi-supervised encrypted traffic classification with deep convolutional generative adversarial networks," *IEEE Access*, vol. 8, pp. 118–126, 2019.

[4] M. Lotfollahi, M. J. Siavoshani, R. S. H. Zade, and M. Saberian, "Deep packet: A novel approach for encrypted traffic classification using deep learning," *Soft Comput.*, vol. 24, no. 3, pp. 1999–2012, 2020.

[5] P. Velan, M. Cermák, P. Čeleda, and M. Drašar, "A survey of methods for encrypted traffic classification and analysis," *Int. J. Netw. Manage.*, vol. 25, no. 5, pp. 355–374, Sep./Oct. 2015.

[6] S. Soleymanpour, H. Sadr, and M. N. Soleimandarabi, "CSCNN: Cost-sensitive convolutional neural network for encrypted traffic classification," *Neural Process. Lett.*, vol. 53, no. 5, pp. 3497–3523, Oct. 2021.

[7] M. Finsterbusch, C. Richter, E. Rocha, J.-A. Müller, and K. Hanssgen, "A survey of payload-based traffic classification approaches," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 2, pp. 1135–1156, 2nd Quart., 2014.

[8] Z. Bu, B. Zhou, P. Cheng, K. Zhang, and Z.-H. Ling, "Encrypted network traffic classification using deep and parallel network-in-network models," *IEEE Access*, vol. 8, pp. 132950–132959, 2020.

[9] X. Wang, S. Chen, and J. Su, "Automatic mobile app identification from encrypted traffic with hybrid neural networks," *IEEE Access*, vol. 8, pp. 182065–182077, 2020.

[10] S. Izadi, M. Ahmadi, and A. Rajabzadeh, "Network traffic classification using deep learning networks and Bayesian data fusion," *J. Netw. Syst. Manage.*, vol. 30, no. 2, p. 25, Apr. 2022, doi: 10.1007/S10922-021-09639-Z.

[11] 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, doi: 10.1109/ICOIN.2017.7899588.

[12] T. Acarman, "Identification of application in encrypted traffic by using machine learning," in *Proc. Int. Conf. Man-Mach. Interact.*, vol. 659, 2018, pp. 545–554.

[13] F. Al-Obeidat and E.-S.-M. El-Alfy, "Hybrid multicriteria fuzzy classification of network traffic patterns, anomalies, and protocols," *Pers. Ubiquitous Comput.*, vol. 23, nos. 5–6, pp. 777–791, Nov. 2019.

[14] Z. Wang, P. Wang, X. Zhou, S. Li, and M. Zhang, "FLOWGAN: Unbalanced network encrypted traffic identification method based on GAN," in *Proc. IEEE Int. Conf. Parallel Distrib. Process. Appl., Big Data Cloud Comput., Sustain. Comput. Commun., Social Comput. Netw.*, Dec. 2019, pp. 975–983, doi: 10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00141.

[15] M. Song, J. Ran, and S. Li, "Encrypted traffic classification based on text convolution neural networks," in *Proc. IEEE 7th Int. Conf. Comput. Sci. Netw. Technol. (ICCSNT)*, Oct. 2019, pp. 432–436.

[16] Z. Nazari, M. Noferesti, and R. Jalili, "DSCA: An inline and adaptive application identification approach in encrypted network traffic," in *Proc. 3rd Int. Conf. Cryptogr., Secur. Privacy*, Jan. 2019, pp. 39–43, doi: 10.1145/3309074.3309102.

[17] K.-C. Chiu, C.-C. Liu, and L.-D. Chou, "CAPC: Packet-based network service classifier with convolutional autoencoder," *IEEE Access*, vol. 8, pp. 218081–218094, 2020.

[18] G. Baldini, "Analysis of encrypted traffic with time-based features and time frequency analysis," in *Proc. Global Internet Things Summit (GIoTS)*, 2020, pp. 1–5, doi: 10.1109/GIOTS49054.2020.9119528.

[19] G. Baldini, J. L. Hernandez-Ramos, S. Nowak, R. Neisse, and M. Nowak, "Mitigation of privacy threats due to encrypted traffic analysis through a policy-based framework and MUD profiles," *Symmetry*, vol. 12, no. 9, p. 1576, Sep. 2020, doi: 10.3390/sym12091576.

[20] Z. Chen, G. Cheng, B. Jiang, S. Tang, S. Guo, and Y. Zhou, "Length matters: Fast internet encrypted traffic service classification based on multi-PDU lengths," in *Proc. 16th Int. Conf. Mobility, Sens. Netw. (MSN)*, Dec. 2020, pp. 531–538, doi: 10.1109/MSN50589.2020.00089.

[21] J. Zhang, F. Li, F. Ye, and H. Wu, "Autonomous unknown-application filtering and labeling for DL-based traffic classifier update," in *Proc. IEEE Conf. Comput. Commun.*, Jul. 2020, pp. 397–405, doi: 10.1109/INFOCOM41043.2020.9155292.

[22] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapé, "Toward effective mobile encrypted traffic classification through deep learning," *Neurocomputing*, vol. 409, pp. 306–315, Oct. 2020.

[23] M. Ugurlu, I. A. Dogru, and R. S. Arslan, "A new classification method for encrypted internet traffic using machine learning," *Turkish J. Elect. Eng. Comput. Sci.*, vol. 29, no. 5, pp. 2450–2468, 2021.

[24] Y. Heng, V. Chandrasekhar, and J. G. Andrews, "UTMobileNetTraffic2021: A labeled public network traffic dataset," *IEEE Netw. Lett.*, vol. 3, no. 3, pp. 156–160, Sep. 2021.

[25] A. Montieri, G. Bovenzi, G. Aceto, D. Ciuonzo, V. Persico, and A. Pescapè, "Packet-level prediction of mobile-app traffic using multitask deep learning," *Comput. Netw.*, vol. 200, Dec. 2021, Art. no. 108529, doi: 10.1016/j.comnet.2021.108529.

[26] G. Aceto, G. Bovenzi, D. Ciuonzo, A. Montieri, V. Persico, and A. Pescapè, "Characterization and prediction of mobile-app traffic using Markov modeling," *IEEE Trans. Netw. Service Manage.*, vol. 18, no. 1, pp. 907–925, Mar. 2021.

[27] S. Dong, Y. Xia, and T. Peng, "Traffic identification model based on generative adversarial deep convolutional network," *Ann. Telecommun.*, vol. 77, nos. 9–10, pp. 573–587, Oct. 2022.

[28] B. Lu, N. Luktarhan, C. Ding, and W. Zhang, "ICLSTM: Encrypted traffic service identification based on inception-LSTM neural network," *Symmetry*, vol. 13, no. 6, p. 1080, Jun. 2021, doi: 10.3390/sym13061080.

[29] H. Liu and B. Lang, "Network traffic classification method supporting unknown protocol detection," in *Proc. IEEE 46th Conf. Local Comput. Netw. (LCN)*, Oct. 2021, pp. 311–314, doi: 10.1109/LCN52139.2021.9525009.

[30] M. Lu, B. Zhou, Z. Bu, K. Zhang, and Z. Ling, "Compressed network in network models for traffic classification," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2021, pp. 1–6, doi: 10.1109/WCNC49053.2021.9417340.

[31] X. Hu, C. Gu, and F. Wei, "CLD-Net: A network combining CNN and LSTM for internet encrypted traffic classification," *Secur. Commun. Netw.*, vol. 2021, pp. 1–15, Jun. 2021, doi: 10.1155/2021/5518460.

[32] W. Zheng, J. Zhong, Q. Zhang, and G. Zhao, "MTT: An efficient model for encrypted network traffic classification using multi-task transformer," *Int. J. Speech Technol.*, vol. 52, no. 9, pp. 10741–10756, Jul. 2022.

[33] H. Yao, C. Liu, P. Zhang, S. Wu, C. Jiang, and S. Yu, "Identification of encrypted traffic through attention mechanism based long short term memory," *IEEE Trans. Big Data*, vol. 8, no. 1, pp. 241–252, Feb. 2022.

[34] A. Telikani, A. H. Gandomi, K.-K.-R. Choo, and J. Shen, "A cost-sensitive deep learning-based approach for network traffic classification," *IEEE Trans. Netw. Service Manage.*, vol. 19, no. 1, pp. 661–670, Mar. 2022.

[35] *ISCX VPN-Non VPN*. Accessed: Jul. 12, 2022. [Online]. Available: https://www.unb.ca/cic/datasets/vpn.html

[36] *CICFlowMeter*. Accessed: Jul. 12, 2022. [Online]. Available: https://github.com/ahlashkari/CICFlowMeter

[37] *UTMobileNet2021*. Accessed: Jul. 12, 2022. [Online]. Available: https://github.com/YuqiangHeng/UTMobileNetTraffic2021

[38] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[39] M. T. Uddin and M. A. Uddiny, "A guided random forest based feature selection approach for activity recognition," in *Proc. Int. Conf. Electr. Eng. Inf. Commun. Technol. (ICEEICT)*, May 2015, pp. 1–6, doi: 10.1109/ICEEICT.2015.7307376.

[40] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jul. 2002.

[41] D. L. Wilson, "Asymptotic properties of nearest neighbor rules using edited data," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-2, no. 3, pp. 408–421, Jul. 1972.

[42] G. E. Batista, R. C. Prati, and M. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explor. Newslett.*, vol. 6, no. 1, pp. 20–29, 2004.

[43] H. He and Y. Ma, *Imbalanced Learning: Foundations, Algorithms, and Applications*, 1st ed. Hoboken, NJ, USA: Wiley, 2013.

MURAT DENER was an associate professor in computer science and engineering. He is currently a Faculty Member of Gazi University. He is also the Head of the Information Security Engineering Department. He has been working in the field of the Internet of Things, information security, and smart cities for nearly 15 years. He has published more than 120 international and national studies.

SAMED AL is currently pursuing the Ph.D. degree with the Information Security Engineering Department. He has been working in the field of big data analytics, explainable artificial intelligence, and information security.

GÖKÇE OK is currently pursuing the M.Sc. degree with the Information Security Engineering Department. She has been working in the field of big data analytics and information security.