RESEARCH ARTICLE

ChurnNet: Deep Learning Enhanced Customer Churn Prediction in Telecommunication Industry

SOMAK SAHA[1], **CHAMAK SAHA**[1], **MD. MAHIDUL HAQUE**[1],
MD. GOLAM RABIUL ALAM[1], (Member, IEEE), AND
ASHIS TALUKDER[2], (Member, IEEE)

[1]Department of Computer Science and Engineering, BRAC University, Dhaka 1212, Bangladesh
[2]Department of Management and Information System, University of Dhaka, Dhaka 1000, Bangladesh

Corresponding author: Ashis Talukder (ashis@du.ac.bd)

ABSTRACT In the Telecommunication Industry (TCI) customer churn is a significant issue because the revenue of the service provider is highly dependent on the retention of existing customers. In this competitive market, it is essential for the service providers to figure out the concerns of their existing customers regarding their services as the cancellation of the services by the customers and switching to new service providers will not bring any good to the service provider. In the context of TCI, numerous research have been made to predict customer churn though, after the performance evaluation of these studies, it shows that there is enough room for progress. Therefore, in this study, we proposed a novel customer churn prediction architecture namely ChurnNet to predict customer churn in TCI. In our proposed ChurnNet, the 1D convolution layer is integrated with residual block, squeeze and excitation block, and spatial attention module to improve the performance. Residual block aids in solving the vanishing gradient problem. Squeeze and excitation block and spatial attention module enable the ChurnNet to understand the interdependency between and within the channels respectively. To evaluate the performance, the experiment is performed on 3 publicly available datasets. As the datasets have significant class imbalance issues, three data balancing techniques such as SMOTE, SMOTEEN, and SMOTETomek are performed. Along with 10-fold cross-validation and after going through the rigorous experiment it was found that ChurnNet performed better than the state-of-the-art and obtained 95.59%, 96.94%, and 97.52% accuracy on 3 benchmark datasets respectively.

INDEX TERMS Squeeze and excitation, spatial attention, residual block, churn prediction, ChurnNet.

I. INTRODUCTION

The TCI has had remarkable expansion and progress in development over the past few decades on the basis of the level of competition, the number of operators, services, and so on. But TCI has significant customer turnover difficulties, which are considered a severe problem in this context due to high strong competition, congested markets, changing environment, and alluring & profitable offers [1]. It is very convenient for customers to switch both services and service providers because the market is saturated. Also, there is so much variation and competitiveness in the services of the

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

service providers. The customers who intend to stop taking the service of the company or intend to take services from other companies are considered churned customers from the perspective of first service providers.

In TCI customer turnover decreases the profit and the company's income severely [2]. Besides, obtaining new customers is way more costly than retaining existing customers [3]. Therefore, customer retention for a longer period is very important because it helps telecommunication service provider companies generate more revenues.

Customer churn is an unavoidable consequence when a customer is unsatisfied with the company's services for long periods of time. Unsubscription of service by the customer does not occur overnight; rather it comes from the customer

as a vigorous act due to the accumulation of long-term disappointment. Therefore, it is vital for service providers to identify and overcome their limitations regarding customer service and satisfaction to retain irate customers. Besides, it is necessary for the service providers to point out which customers are likely to churn and which customers are willingly wanting to stay with provided services (non-churn) because it helps the telephone company (TELCO) to revise their strategy to retain potential churn by reducing dissatisfaction. Therefore, customer churn prediction (CCP) is crucial for TELCO because it is very much connected with brand value and revenue. For this purpose, telephone companies usually maintain standing reports regarding clients and services to understand the continuation of their receiving services. For TELCO, the objective is to retain the clients for the long term which makes the CCP so important in TCI [4].

The CCP problem has been studied in several research works and many machine learning-based algorithms have been proposed for customer churn prediction [5], [6] [7], [8]. However, previous studies have struggled to achieve satisfactory outcomes in churn prediction due to the complex and diversified nature of customer churn behaviour. The traditional ML algorithm-based churn prediction models struggle to deal with complex non-linear relationships within the data [9]. Further, dealing with high dimensional data requires manual feature engineering which is labor-intensive and time-consuming [10]. Sometimes it is prone to human biases. Furthermore, traditional ML methods such as SVM, Random Forest, Decision Tree, and so on do not have any built-in mechanism for assigning attention to specific features for extracting key information from churn datasets [11]. Also, in high-dimensional datasets, it is essential to extract local information for better performance which traditional ML algorithms lack. Along with machine learning approaches in previous studies, deep learning methods have also been utilized in customer churn prediction [12]. However, the proposed deep learning-based models in prior works are unable to provide explicit feature importance in churn prediction due to a lack of attention mechanisms. To improve feature extractions from the data, the attention mechanism helps to capture the inter-spatial and inter-channel relationship of the features that vanilla CNN and ANN lack. Considering this fact we design a unique deep learning architecture (ChurnNet) that can assign attention based on specific feature importance. Besides, it can do feature extraction automatically and is capable of finding complex non-linear relationships within features. For example, ANN [13] struggled to perform satisfactorily on the IBM Telco dataset whereas our proposed model performed well because of identifying the important features of the data. The main contributions of the study are:

- We introduce a novel attention mechanism-based customer churn prediction model, namely ChurnNet to predict customer churn in the Telecommunication industry. In the proposed ChurnNet, One-Dimensional (1D) convolution layer is integrated with Residual blocks, Squeeze and Excitation blocks, and Spatial attention

modules to enhance the performance to predict customer churn.

- The datasets that are used in the research are highly imbalanced, so, 3 data balancing techniques such as Synthetic Minority Over-sampling Technique (SMOTE), SMOTE with Edited Nearest Neighbors (SMOTEEN), and SMOTE-Tomek Links (SMOTETomek) are used. After empirical analysis, it was found that SMOTEEN provided better performance. Also, a 10-fold cross-validation technique is used to make the ChurnNet more generalized.

- Along with hyperparameter tuning, extensive experiment is performed on 3 publicly available datasets such as IBM Telco Dataset, Churn-in-Telecom Dataset, and Churn-data-UCI Dataset. In performance evaluation, it obtained 95.59%, 96.94%, and 97.52% accuracy for 3 datasets respectively.

The organization of the paper is as follows. Section II shows the literature review which is followed by the methodology in Section III. Section IV presents the architecture of the proposed ChurnNet. Section V discusses the results and Section VI gives the conclusion.

II. LITERATURE REVIEW

Customer churn has been dealing with many approaches which include Machine Learning, Deep Learning, Data mining along with hybrid methodologies. Several Machine learning classifiers such as Logistic Regression (LR) [14], Random Forest (RF), K-nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) [15], Naive Bayes (NB) [16], Bayesian classifiers have been used commonly in CPP problems. Besides, ensemble Classifiers [17] such as CatBoost, XGBoost, Adaboost [18], and Gradient Boosting (GBM) have also been proposed to address the problem. Also, Deep Learning strategies are also used on the Customer Relationship Management (CRM) in TCI to predict customer churn. For prediction tasks, CRM data has been used frequently [19]. Therefore, a comprehensive review of prior work is beyond the range of this paper, but a brief relevant overview is given below.

Sana et al. [7] proposed 8 types of classifiers to deal with the CCP problem in the TCI. Along with the classifiers they applied 6 data transformation methods such as Logarithmic transformations, Z-score, Discretization, Box-cox, Weight-of-Evidence (WOE), and Rank, to see the performances of the proposed classifiers. After evaluation through cross-validation techniques on 4 publicly available datasets, it was found that the WOE data transformation method had a significant impact on the performance.

Using KNN and Pearson Correlation the author [20] made up an approach to deal with the serious issues of CCP in TCI. Using only one dataset of 7043 instances, they evaluated the proposed model. Without using any cross-validation techniques they compared their proposed model with SVM

and RF and it was visible that their proposed model performed better and got 97.78% testing accuracy.

Similarly on a telco dataset, the study [21] focused on data mining techniques. Using the cross-validation method, they applied LR, MLP, SVM, DT, and backpropagation algorithms for classification tasks. Besides, they used DB scan algorithms along with K means algorithms for the clustering tasks. Also, they used Apriori and FP Growth algorithms.

The author suggested [22] an integrated framework based on customer analytics to manage churn. Their proposed framework is dedicated not only to predicting churn but also to doing customer segmentation. They used 3 datasets where to generalize, 10-fold cross-validation was used. Besides, to mitigate the data imbalanced issues they used SMOTE technique. To predict the churn they applied the baseline machine learning classifiers along with MLP. Further, Bayesian LR was introduced for analysis and the K means algorithm was used for the customer segmentation tasks.

Amin et al. [23] revealed a Just in Time (JIT) approach to tackle the customer churn problem. They used SVM as a base classifier; they also used homogeneous and heterogeneous ensemble methods for the (JIT) approach. After evaluating the performance using 10-fold cross-validation on two publicly available telecom datasets it was demonstrated that SVM performed better when it was applied in the heterogeneous ensemble compared to homogeneous ensembles or as base classifiers.

Similarly, the authors [6] implemented 4 ensemble-based algorithms which are RF, Decision Tree, GBM, and XGBoost for classifying customer churn. But in this research, the author introduced only one evaluation metric which is Area Under Curve (AUC) value. Among these algorithms, XGBoost performed well in terms of AUC values (93%) compared to other algorithms when SNA was integrated as feature selection. Integrating SNA in the feature selection was an important aspect of this research.

The study [24] used 4 imbalanced public telco datasets to predict customer churn. They used different oversampling techniques such as MWMOTE, SMOTE, TRKNN, MTDF, ICOTE, and ADASYN to deal with the class imbalance issue. The authors used 4 different algorithms such as Genetic algorithm, Learning from Example Module, version 2 (LEM2), Rough Set Theory (RST), and covering algorithms along with cross-validation to evaluate the performance.

Wang et al. [25] proposed several machine learning approaches such as classification and regression trees (CART), partial decision trees (PART), and bagged CART for customer churn prediction in TCI. In their research, they used only one dataset which was imbalanced. To resolve this issue, they used SMOTE. After evaluation through 10-fold cross-validation, it was found that CART had 0.7643 accuracy which is the highest among all the algorithms. As telco churn is cost-sensitive, the authors [26] also introduced the CART model. After doing experimentations on real data it found that the proposed CART model performed well not only in

classification tasks but also it helped to minimize the total cost of misclassification.

Herdian and Girsang [27] applied a machine learning and deep learning-based hybrid approach for the task of CCP. In the research, they used the two public datasets of 7043 and 5000 instances respectively. They used the Decision Tree (DT), ANN, and the hybrid model (DT-ANN) for classification purposes. In the result analysis, it was found (DT-ANN) outperformed DT and ANN. Point to be noted that, without using any cross-validation techniques they evaluated the performance of the algorithms using only an evaluation metric which is AUC. Besides, this paper [28] followed an approach to predict customer churn in a cloud environment where two layers of ANN were used for feature extraction. They used the Block-Jacobi SVD algorithm to reduce the dimensions of the extracted feature. Later on, the NB algorithm was used for classification tasks.

The study [29] approached CNN for CCP. Using only a telco dataset, it obtained 86.85% accuracy without applying K-fold. In this study, [11] authors developed a model by integrating both CNN and Bidirectional LSTM with spatial

TABLE 1. Summary of literature review for customer churn prediction in telecommunication industry.

Study	Attention Mechanism used	Methods	Data Balancing Technique	Cross Validation
Mishra et al. [8]	No	Ensemble Classifiers	Not used	Not used
Amin et al. [23]	No	heterogeneous ensemble, homogeneous ensemble, SVM	Not used	10 Fold
Wang et al. [25]	No	CART, PART, bagged CART	SMOTE	10 Fold
Herdian et al. [27]	No	DT and DT-ANN were used, AUC was only used for evaluation	Under sampling	Not used
Agarwal et al. [13]	No	MLP, accuracy was only used for evaluation	Not used	Not used
Mishra et al. [29]	No	CNN	Not used	Not used
Khattak et al. [12]	No	BiLSTM-CNN	Not used	Not used
AlShourbaji et al. [30]	No	CP-EGBM	HEOMGA	Not used
Zhou et al. [31]	No	Ensemble learning	SMOTE	Not used
Jajam et al. [32]	No	Ensemble deep learning	Not required	10 Fold
Srinivasan et al. [33]	No	Decision Tree, RF	SMOTEEN	Not used
Proposed Churn-Net (This Study)	**Yes**	**1D CNN with SE Block, Residual Block, Spatial Attention**	**SMOTE, SMOTE-Tomek, SMO-TEEN**	**10 Fold**

attention mechanism to predict customer churn on bank dataset. They compared the performance of the proposed model (AttnBLSTM-CNN) with BiLSTM-CNN.In the performance evaluation, it was found that AttnBLSTM-CNN performed better.

III. METHODOLOGY

The top-level overview of the customer churn prediction is presented in Fig.1. There are several modules such as dataset collection, data preprocessing, data balancing techniques, ChurnNet classifier, and performance evaluation.
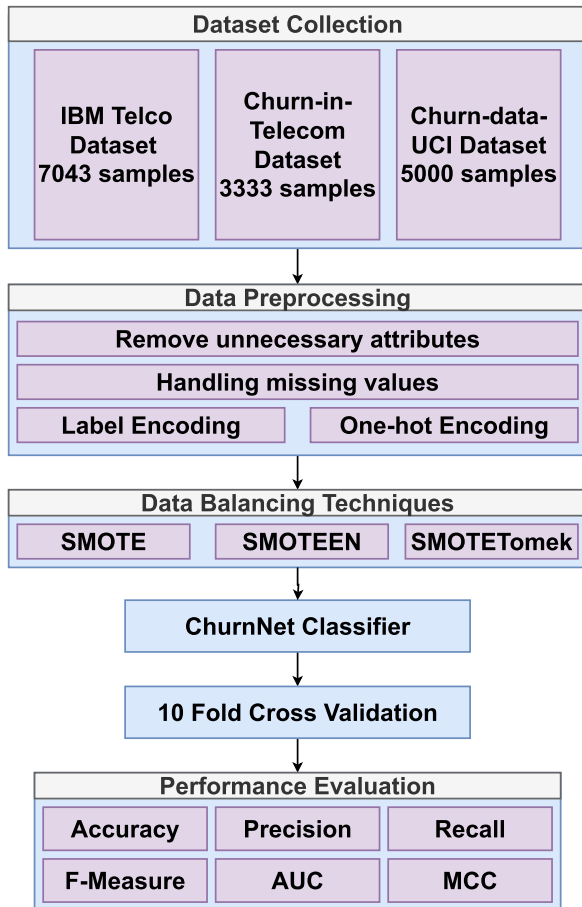

FIGURE 1. The top-level overview of the research.

A. DATASET DESCRIPTION

In the research, we have used 3 publicly available telco datasets which are IBM Telco Dataset [34], Churn-in-Telecom Dataset [35], and Churn-data-UCI Dataset [36]. Dataset descriptions are given below:

1) IBM TELCO DATASET

This dataset consists of 7043 instances where each instance refers to the customer. In each instance, it consists of 21 variables/attributes where 20 variables are independent and 1 variable is dependent. The dependent variable is namely Churn which determines whether the customer is churn or non-churn. In the IBM Telco Dataset, 26.54% samples are churn.

2) CHURN-IN-TELECOM DATASET

This dataset consists of 3333 samples where each sample defines the customer. In each sample, there are 21 attributes where 20 attributes are independent and 1 attribute is dependent. Besides, among the independent attributes, 16 attributes are numerical, and 4 attributes are discrete. In the dataset 85.51% samples are churn.

3) CHURN-DATA-UCI DATASET

The dataset contains 5000 samples of customers. Each sample consists of 20 variables where 19 variables are independent and 1 variable is dependent which is Churn. In the dataset 85.86% samples are churn.

TABLE 2. Description of the datasets.

Dataset	No of Samples	No of Attributes	Class Labels
IBM Telco Dataset	7043	21	2
Churn-in-Telecom Dataset	3333	21	2
Churn-data-UCI Dataset	5000	20	2

B. DATA PREPROCESSING

The following steps are used in preprocessing:
- Unnecessary attributes such as the attributes that indicate ID/phone numbers are ignored.
- The attributes that have numerical missing values are replaced by 0 and no categorical values are missing.
- In order to deal with categorical values Label Encoding and One Hot Encoding are performed for three datasets.

C. DATA BALANCING TECHNIQUES

Since the three employed datasets have significant data imbalance issues, it was required to use different data balancing techniques to mitigate the issue. Fig. 2, Fig. 3, and Fig. 4 portray the distribution of data after using different data balancing techniques for the three datasets.

1) SMOTE

SMOTE algorithm works by producing synthetic data points for the minority class. SMOTE initially locates the minority class. Then it chooses a data point of that class randomly and creates data points that are located near the selected data point and its KNN. This technique is repeated for the appropriate number of samples from the minority class until the class distribution is balanced.

2) SMOTETomek

SMOTETomek is a data-balancing hybrid method that combines both oversampling and undersampling. SMOTE does oversampling by creating synthetic data points of the
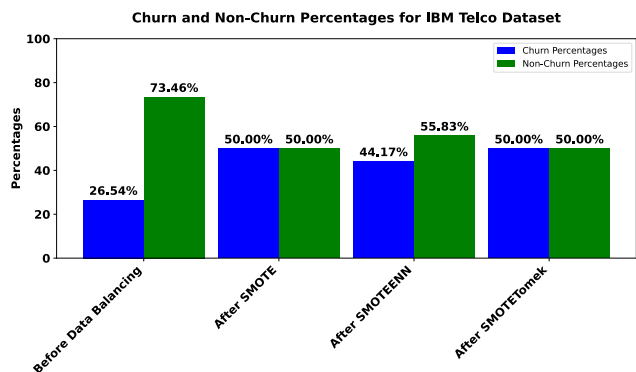
FIGURE 2. Data distribution of different data balancing techniques for IBM Telco dataset.
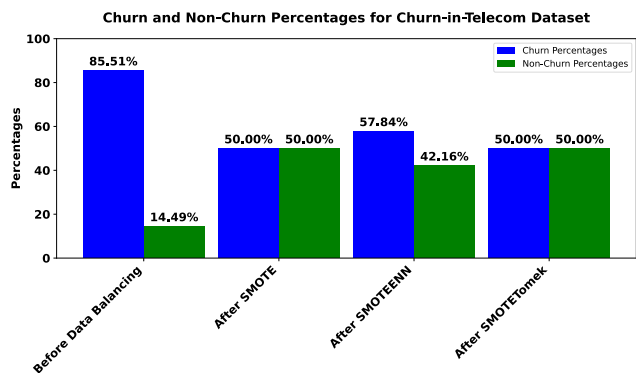

FIGURE 3. Data distribution of different data balancing techniques for Churn-in-Telecom dataset.
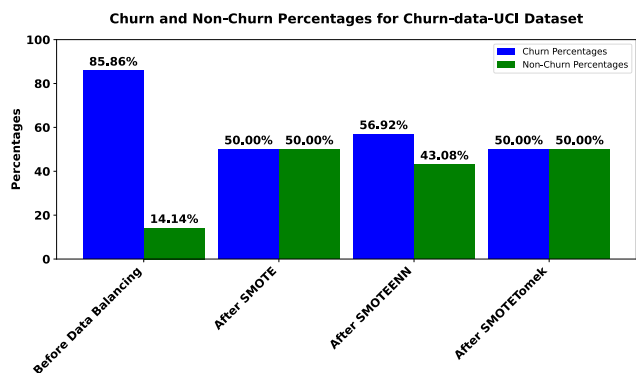

FIGURE 4. Data distribution of different data balancing techniques for Churn-data-UCI dataset.

minority class. Then the Tomek Link technique selects a random data point belonging to the majority class and if its nearest data point belongs to the minority class, Tomek link does undersampling by removing the data point (majority class).

3) SMOTEEN

SMOTEEN is a technique that is a combination of SMOTE method and Edited Nearest Neighbour (ENN). SMOTE is capable of creating synthetic data points for the minority class and ENN eliminates some observations that have been

detected to have distinct classes compared to the observation class and its K Nearest Neighbours majority class.

IV. PROPOSED CHURNNET

For the purpose of CCP, we designed a unique attention mechanism-driven deep learning model called ChurnNet. The architecture of the proposed model is different from traditional Machine Learning and Deep Learning models. In the architecture of ChurnNet shown in Fig. 5, 1D CNN [29] is incorporated with subsequent residual blocks, squeeze and excitation blocks, and spatial attention modules [11]. Residual block enables the model to upgrade the performance by residual learning which controls the flow of the gradient. Also, the Squeeze and Excitation (SE) block and Spatial attention module help to understand the interdependency of each and within the channels. To the best
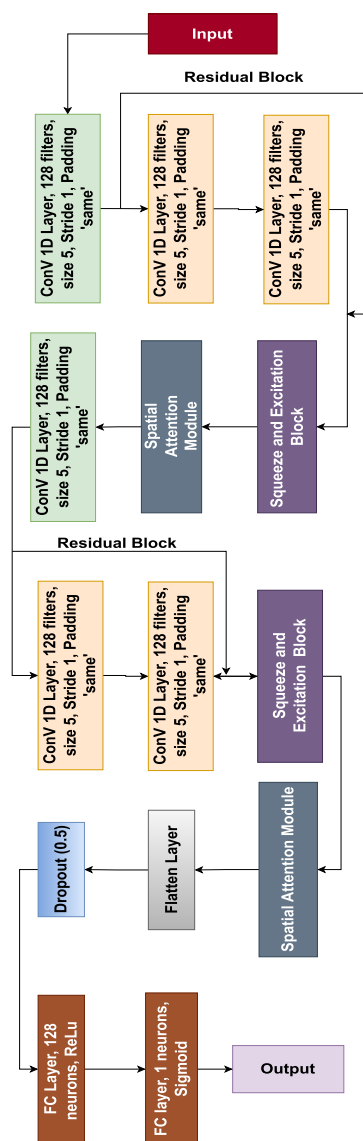

FIGURE 5. Architecture of proposed ChurnNet.

of our knowledge, no such architecture has been proposed for customer churn prediction in TCI. Also, point to mention that through extensive experiment and hyperparameter tuning the proposed ChurnNet is designed. The training configuration is shown in Table 3.

A. 1D CONVOLUTIONAL NEURAL NETWORK

1D Convolutional Neural Network(CNN) is a type of neural network that is used to process 1-dimensional data shown in Fig.6. 1D CNN works by performing convolution operation which is basically the dot products between kernel/filter and input data. Feature map is the output of the convolution operation which basically emphasizes certain features of the input data.
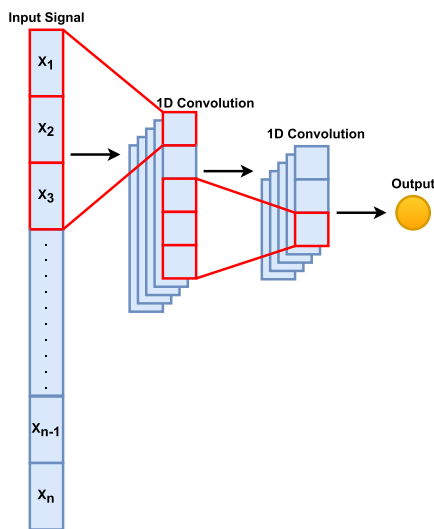
FIGURE 6. Basic structure of a 1D convolutional neural network (CNN) design with a pair of convolutional layers.

Initially, in ChurnNet, the input tensor $(1 \times W \times 1)$ (W refers to the number of attributes) is passed into the 1D convolution layer which has 128 filters having kernel size 5. Also, the stride is kept 1, and zero padding is used. This convolution layer is used mainly to extract the lower dimensional features by creating feature maps. The output tensor $(1 \times W \times 128)$ is passed to the following Residual Block. Later on in the architecture of ChurnNet, a similar configuration convolution layer is used to extract the important features.

B. RESIDUAL BLOCK

To improve the performance of the model, residual blocks are introduced in the proposed architecture which consists of two subsequent symmetric convolution layers. And the configuration of the subsequent symmetric convolution layers is similar to previous convolution layer. Besides, Rectified Linear Unit (ReLU) activation function is used to minimize the exponential computational growth of the model. Through the skip connection shown in Fig.7, the input (x), of the residual block is connected to the output (y) of the
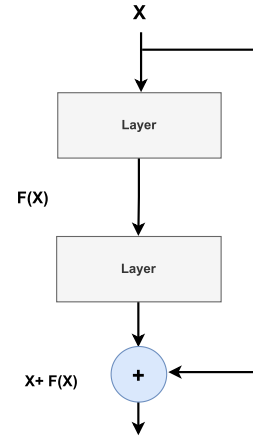
FIGURE 7. Architecture of residual block.

second convolution layer of the residual block. To mitigate the vanishing gradient problem skip connection is used [37]. The output of the residual block having tensor shape of $(1 \times W \times 128)$ is passed to the following SE Block.

$$y = x + F(x) \tag{1}$$

C. SQUEEZE AND EXCITATION BLOCK

SE block [38] is used in the proposed architecture to find effective channels of the feature maps. The output $U \in \mathbb{R}$ of the residual block is passed to the squeezer of the SE block where the Global Average Pooling (GAP) operation is performed shown in Fig.8. By performing the GAP operation, the squeezer reduced the spatial dimension, squeezed each
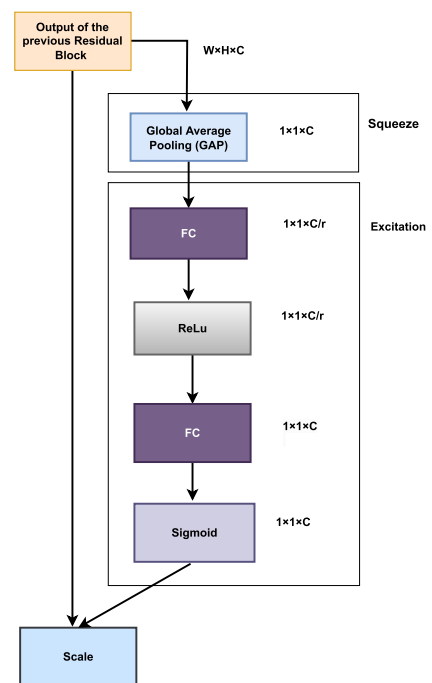
FIGURE 8. Architecture of squeeze and excitation block.

of the input channels, and converted it into a single numeric value to make the channel descriptor.

$$z_c = F_{sq}(u_c) = \frac{1}{N} \sum_{i=1}^{N} u_c(i) \quad (2)$$

Later the channel descriptor is passed to the exciter of the network which consists of two Fully Connected (FC) layers.

$$s = F_{ex}(z, W) = \sigma(W_2 \delta(W_1 z)) \quad (3)$$

ReLU activation function of the first FC layer helps the network to understand the nonlinear connection between the channels. The second FC layer helps to understand the non-mutually exclusive connections between the channels and the Sigmoid activation function helps to do so. The output of the second FC layer is multiplied by the input of the SE block and the obtained output ($1 \times W \times 128$) is passed to the next spatial attention block.

$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c u_c \quad (4)$$

D. SPATIAL ATTENTION MODULE

The spatial attention mechanism [39] is used to focus on the most important regions of each channel in the feature maps. In the proposed model, the output of the SE block is passed as an input tensor to the spatial attention module where the two types of pooling operations are performed such as max pooling and average pooling shown in Fig.9. Then the output of the pooling operation is concatenated to create a feature descriptor and later passes to the convolution layer for generating spatial attention maps multiplied with the input of the spatial attention module. The shape of the output tensor remains the same as the input tensor.

$$M_s(F) = \sigma \left(f \left([\text{AvgPool}(F); \text{MaxPool}(F)] \right) \right)$$
$$= \sigma \left(f \left([F^s\text{avg}; F^s\text{max}] \right) \right) \quad (5)$$
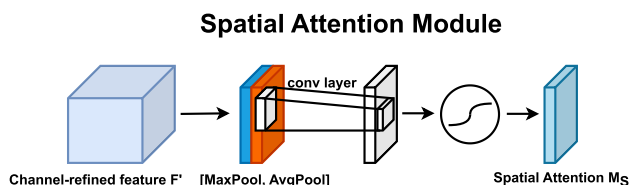
Spatial Attention Module

FIGURE 9. Architecture of spatial attention module.

E. CLASSIFICATION LAYER

The output of the last Spatial Attention module is passed to the flatten layer. Then to reduce the overfitting issues 0.5 dropout is applied. The output of the Dropout Layer is passed to the first Dense Layer which consists of 128 neurons where the ReLU activation function is used. The output of the first dense layer is then passed to the last Dense Layer which consists of 1 neuron along with a sigmoid

TABLE 3. Proposed ChurnNet training configuration.

Parameter	
Parameter	0.171 M
Optimizer	ADAM
Loss Function	Binary Cross Entropy
Learning Rate	0.001
Epoch	30
Batch Size	32

activation function to do binary classification of Churn and Non-Churn.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The performance of the proposed ChurnNet has been evaluated using several well-known performance metrics such as accuracy, precision, recall, F-Measures, AUC, and Matthew's Correlation Coefficient (MCC).

Accuracy is a measure of the overall correctness of the model. It can be computed as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Precision is a measure that identifies the model's ability to correctly classify the positive instances among the instances it labeled as positive. It can be computed as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

Recall is a measure that identifies the ability of the model to identify the positive instances from the total number of positive instances in the dataset. It can be computed as follows:

$$\text{Recall/Probability of Detection (POD)} = \frac{TP}{TP + FN} \quad (8)$$

F-measure calculates the harmonic mean of the precision and recall. It can be computed as follows:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

Probability of False Alarm (POF) defines the probability of false alarm. The value of POF should be as less as possible. It can be computed as follows:

$$\text{FPR/POF} = \frac{FP}{FP + TN} \quad (10)$$

AUC defines that area under the curve of ROC. It can be computed as follows:

$$\text{AUC} = \frac{1 + POD - POF}{2} \quad (11)$$

MCC is a metric that calculates the effectiveness of binary classification. It can be computed as follows:

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (12)$$

Here, True Positive (TP)=True Positive, False Positive (FP)=False Positive, True Negative (TN)=True Negative, False Negative (FN)=False Negative.

TABLE 4. Performance of proposed ChurnNet on IBM Telco dataset.

Encoding	Data Balancing Technique	Filter Size	Number of Filters	Flatten Layer / GAP	Channel Attention	Optimzer	Learning Rate	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	MCC (%)	AUC (%)
Label Encoding	Not Used	3	64	GAP	SE Block	ADAM	0.001	79.33	62.48	55.54	58.53	45.20	83.46
One-hot Encoding	Not Used	3	64	GAP	SE Block	ADAM	0.001	79.44	66.04	48.31	55.01	43.57	83.98
One-hot Encoding	SMOTE	3	64	GAP	SE Block	ADAM	0.001	81.22	87.05	73.46	79.61	63.29	91.01
One-hot Encoding	SMOTETomek	3	64	GAP	SE Block	ADAM	0.001	83.65	87.62	78.48	82.73	67.73	92.46
One-hot Encoding	**SMOTEEN**	3	64	GAP	SE Block	ADAM	0.001	92.87	95.28	91.66	93.41	85.77	98.33
One-hot Encoding	SMOTEEN	**5**	64	GAP	SE Block	ADAM	0.001	93.21	94.94	92.69	93.76	86.43	98.34
One-hot Encoding	SMOTEEN	5	32	GAP	SE Block	ADAM	0.001	92.60	92.60	93.54	93.33	85.29	98.13
One-hot Encoding	SMOTEEN	5	**128**	GAP	SE Block	ADAM	0.001	93.79	95.49	93.27	94.33	87.59	98.54
One-hot Encoding	SMOTEEN	5	128	**Flatten Layer**	**SE Block**	**ADAM**	**0.001**	**95.59**	**95.88**	**96.22**	**96.04**	**91.08**	**98.76**
One-hot Encoding	SMOTEEN	5	128	Flatten Layer	Basic Channel Attention	ADAM	0.001	95.35	95.97	95.67	95.81	90.60	98.71
One-hot Encoding	SMOTEEN	5	128	Flatten Layer	SE Block	NADAM	0.001	94.98	95.60	95.27	95.43	89.86	98.40
One-hot Encoding	SMOTEEN	5	128	Flatten Layer	SE Block	RMSprop	0.001	95.35	95.68	95.92	95.79	90.63	98.43
One-hot Encoding	SMOTEEN	5	128	Flatten Layer	SE Block	ADAM	0.0001	95.05	95.51	95.56	95.54	89.98	98.26
One-hot Encoding	SMOTEEN	5	128	Flatten Layer	SE Block	ADAM	0.00001	95.21	95.25	96.20	95.71	90.33	98.26

A. EXPERIMENTAL RESULTS

The proposed ChurnNet has been experimented on 3 customer churn datasets using 10-fold cross-validation. All the computations have been conducted in the Google Colaboratory platform. The source code link is given.[1]

Tables 4, 5, and 6 demonstrate the empirical result analysis of the proposed ChurnNet for 3 used datasets. For the three datasets, it was observed that the proposed ChurnNet performed reasonably well in terms of accuracy and AUC score without any data balancing techniques. It obtained 79.33%, 90.52%, and 92.86% accuracy for IBM Telco Dataset, Churn-in-Telecom Dataset, and Churn-data-UCI Dataset respectively when label encoding was performed for dealing with categorical values. An important observation from the experimentation is that the accuracy (79.44%) of the IBM telco dataset had increased slightly when one-hot encoding was performed instead of label encoding. However,

[1]https://t.ly/1Oeo5

TABLE 5. Performance of proposed ChurnNet on Churn-in-Telecom dataset.

Encoding	Data Balancimg Technique	Filter Size	Number of Filters	Flatten Layer / GAP	Channel Attention	Optimzer	Learning Rate	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	MCC (%)	AUC (%)
Label Encoding	Not Used	3	64	GAP	SE Block	ADAM	0.001	90.52	72.04	57.57	63.35	58.89	87.82
One-hot Encoding	Not Used	3	64	GAP	SE Block	ADAM	0.001	88.84	70.19	43.66	52.70	49.14	87.22
Label Encoding	SMOTE	3	64	GAP	SE Block	ADAM	0.001	90.16	91.07	89.16	90.05	80.43	96.54
Label Encoding	SMOTETomek	3	64	GAP	SE Block	ADAM	0.001	90.07	91.30	88.77	89.92	80.31	96.30
Label Encoding	**SMOTEEN**	3	64	GAP	SE Block	ADAM	0.001	93.94	94.91	94.59	94.74	87.63	97.88
Label Encoding	SMOTEEN	**5**	64	GAP	SE Block	ADAM	0.001	95.62	96.26	96.29	96.23	91.12	98.81
Label Encoding	SMOTEEN	5	32	GAP	SE Block	ADAM	0.001	94.75	95.78	95.12	95.44	89.26	98.57
Label Encoding	SMOTEEN	5	**128**	GAP	SE Block	ADAM	0.001	96.26	96.55	97.04	96.79	92.34	99.00
Label Encoding	SMOTEEN	5	128	**Flatten Layer**	**SE Block**	**ADAM**	**0.001**	**96.94**	**96.73**	**98.03**	**97.37**	**93.74**	**99.27**
Label Encoding	SMOTEEN	5	128	Flatten Layer	Basic Channel Attention	ADAM	0.001	96.34	96.08	97.65	96.85	92.51	99.03
Label Encoding	SMOTEEN	5	128	Flatten Layer	SE Block	NADAM	0.001	96.76	96.24	98.25	97.23	93.38	93.38
Label Encoding	SMOTEEN	5	128	Flatten Layer	SE Block	RMSprop	0.001	96.70	96.14	98.25	97.18	93.24	99.27
Label Encoding	SMOTEEN	5	128	Flatten Layer	SE Block	ADAM	0.0001	96.87	96.61	98.04	97.31	93.61	99.28
Label Encoding	SMOTEEN	5	128	Flatten Layer	SE Block	ADAM	0.00001	96.74	96.44	97.96	97.19	93.34	99.19

for the rest of the two datasets label encoding provided better performance.

When the data balancing techniques were introduced, the performance improvement for the 3 datasets increased significantly not only in terms of accuracy but also in other evaluation metrics. For the IBM Telco Dataset, when SMOTEEN was performed the accuracy increased to 92.87%. Also, SMOTEEN performed comparatively well in all evaluation metrics compared to SMOTE and SMOTETomek.

Similarly, SMOTEEN performed better for the Churn-in-Telecom Dataset (93.94%), and Churn-data-UCI Dataset (95.14%) in terms of accuracy compared to SMOTE and SMOTETomek.

After hyperparameter tuning, we found that the performance of the proposed ChurnNet for the three datasets had increased sharply. When the filter or kernel size increased from 3 to 5 the accuracy increased by around 1%, 2%, and 1% for IBM Telco Dataset, Churn-inTelecom Dataset, and Churn-data-UCI Dataset respectively. It showed improvements for 3 datasets when we altered the number of filters or kernels. It reached 93.79%, 96.26%, and 97.01% accuracy for IBM Telco Dataset, Churnin-Telecom Dataset, and

TABLE 6. Performance of the proposed ChurnNet for Churn-data-UCI dataset.

Encoding	Data Balancing Technique	Filter Size	Number of Filters	Flatten Layer / GAP	Channel Attention	Optimzer	Learning Rate	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	MCC (%)	AUC (%)
Label Encoding	Not Used	3	64	GAP	SE Block	ADAM	0.001	92.86	82.81	62.94	71.02	68.18	89.90
One-hot Encoding	Not Used	3	64	GAP	SE Block	ADAM	0.001	91.98	80.31	58.11	66.64	63.75	88.55
Label Encoding	SMOTE	3	64	GAP	SE Block	ADAM	0.001	92.08	95.78	88.05	91.72	84.47	97.38
Label Encoding	SMOTETomek	3	64	GAP	SE Block	ADAM	0.001	91.59	94.70	88.16	91.27	83.44	97.21
Label Encoding	**SMOTEEN**	3	64	GAP	SE Block	ADAM	0.001	95.14	96.58	94.78	95.66	90.20	98.83
Label Encoding	SMOTEEN	**5**	64	GAP	SE Block	ADAM	0.001	96.71	96.47	97.82	97.14	93.31	99.37
Label Encoding	SMOTEEN	5	32	GAP	SE Block	ADAM	0.001	95.58	96.31	95.95	96.11	91.04	99.04
Label Encoding	SMOTEEN	5	**128**	GAP	SE Block	ADAM	0.001	97.01	97.52	97.23	97.37	93.92	99.38
Label Encoding	SMOTEEN	5	128	**Flatten Layer**	**SE Block**	**ADAM**	**0.001**	**97.52**	**97.92**	**97.73**	**97.81**	**94.98**	**99.57**
Label Encoding	SMOTEEN	5	128	Flatten Layer	Basic Channel Attention	ADAM	0.001	97.42	97.15	98.36	97.75	94.74	99.48
Label Encoding	SMOTEEN	5	128	Flatten Layer	SE Block	NADAM	0.001	97.38	97.34	98.07	97.70	94.67	99.45
Label Encoding	SMOTEEN	5	128	Flatten Layer	SE Block	RMSprop	0.001	97.34	97.47	97.87	97.67	94.59	99.36
Label Encoding	SMOTEEN	5	128	Flatten Layer	SE Block	ADAM	0.0001	97.31	96.93	98.36	97.63	94.54	99.49
Label Encoding	SMOTEEN	5	128	Flatten Layer	SE Block	ADAM	0.00001	97.41	97.16	98.32	97.73	94.72	99.59

Churn-data-UCI Dataset respectively when the number of filters used was 128 instead of 64 or 32. As shown in Table 4, 5, and 6 the integration of Flatten layer in the proposed ChurnNet provided better performance instead of using the Global Average Pooling layer for the 3 used datasets. Besides, in the architecture of the proposed ChurnNet, incorporating the Basic Channel Attention module of [39] in place of the SE block provided very decent performance. However, the inclusion of the SE Block provided the best performance for the 3 datasets.

ChurnNet performed well in different optimizers (learning rate=0.001) for the 3 datasets. Among different optimizers,

ADAM obtained a better accuracy of 95.59% (IBM Telco Dataset), 96.94% (Churn-in-Telecom Dataset) and, 97.52% (Churn-data-UCI Dataset). Apart from that, NADAM also obtained 94.98%, 96.76%, and 97.38% accuracy for IBM Telco Dataset, Churn-in-Telecom Dataset, and Churn-data-UCI Dataset respectively. Besides, the RMSprop also performed well. In different metrics, the performance of RMSprop is very close to ADAM and NADAM for the 3 datasets.

The performance of ChurnNet on 0.001 learning rate (Optimizer=ADAM) provided the highest performance. From the tables, it had been found that as the learning rate

TABLE 7. Comparison with other related study methods for IBM Telco dataset.

Study	Methods	Accuracy (%)	F1 Score (%)
Saleh et al. [3]	LR	80	58
	XGBC	77	55
	DT	79	57
	ADA	70	48
	RF	82	53
Agarwal et al. [13]	ANN	80.03	-
Yang. P [44]	SVM	61	-
	RF	60	-
	GBBT	67	-
Kimura. T [45]	LR	69.7	59.6
	SVM	69.3	59.6
	RF	73.1	62.2
	XGBOOST	70.5	60.8
	LightGBM	72.8	61.7
	CatBoost	72.3	61.4
Rabah et al. [46]	Transformer	81	-
Fuji et al. [42]	Deep-BP-ANN	86.57	88.68
Pustokhina et al. [47]	(SMOTE) with optimal weighted extreme Machine Learning (ML)	92	91.8
Jain et al. [41]	LR	80.758	-
	RF	80.881	-
	SVM	82.018	-
	XGBoost	82.942	-
Amin et al. [48]	UDT	78	66.99
	LDT	70.6	61.45
AlShourbaji et al. [30]	CP-EGBM	86.5	82.1
Srinivasan et al. [33]	DT	-	93
	RF	-	95
Khattak et al. [12]	BiLSTM-CNN	81	65
This Study	**Proposed Model**	**95.59**	**96.04**

TABLE 8. Comparison with other related study methods for Churn-in-Telecom dataset.

Study	Methods	Accuracy (%)	F1 score (%)
Amin et al. [48]	UDT	84.3	58.02
	LDT	82.9	46.39
Jain et al. [40]	LR	85.24	81
	Logit Boost	85.18	80.6
Amin et al. [23]	Heterogeneous ensemble	77.27	71.42
	Homogeneous ensemble	59.24	56.73
Cenggoro et al. [49]	Vector embedding, T-SNE	89.82	81.16
Ullah et al. [50]	RF	89.6	87.6
Mishra et al. [8]	Boosting	90.3	-
This Study	**Proposed Model**	**96.94**	**97.37**

TABLE 9. Comparison with other related study methods for Churn-data-UCI dataset.

Study	Methods	Accuracy(%)	F1 Score(%)
Sagir et al. [51]	Bagging	80.8	78.4
	AdaBoost	73.9	72.06
Maldonado et al. [52]	Bagging + Deep Learning (DL)	50	33.33
	Majority Voting DL +NN+ML	66.69	67.1
	AdaBoost+ MLP	66.3	66.46
	Bagging +MLP	67.57	69.5
Kumar et al. [53]	K means + DT	84.26	90
Bilal [43]	LSTM	90.04	82.1
	CNN	89.8	82.6
	GRU	97.34	97.67
This Study	**Proposed Model**	**97.52**	**97.81**

decreased the performance of the ChurnNet also decreased. And it is valid for the 3 datasets. For IBM Telco Dataset accuracy hit 95.59% when the learning rate was 0.001. Also, the performance of other evaluation metrics was above 90%. For the Churn-in-Telecom Dataset and Churn-data-UCI Dataset, the accuracy is 96.94%, and 97.52% when the learning rate kept 0.001. Also, in other metrics performance hit above 90%. The accuracy decreased to 95.21%, 96.74%, and 97.41% when the learning rate was 0.00001 for the 3 datasets.

Table 7, Table 8, and Table 9 demonstrate the performance of the proposed ChurnNet compared to state-of-the-art methods for IBM Telco Dataset, Churn-in-Telecom Dataset, and Churn-data-UCI Dataset respectively. It is visible that the proposed ChurnNet performed better compared to Machine learning methods such as LR [40], SVM [41], and so on. Also, it performed better than deep learning methods such as MLP [42], CNN, LSTM, and GRU [43]. Apart from these, the proposed model outperformed ensemble approaches such as Bagging, Boosting [41], and so on.

VI. CONCLUSION
In TCI, customer churn prediction is a key factor from the perspective of the service provider company. To deal with this issue, this study came up with a solution called ChurnNet which is a CNN-based hybrid approach. The uniqueness of the ChurnNet architecture is residual learning, squeeze and excitation network, and an effective spatial attention module are incorporated with 1D convolution layers. The research was conducted on 3 publicly available datasets where 3 data balancing techniques (SMOTE, SMOTETomek, SMOTEEN) were used to overcome the class imbalance problem of the datasets. After the 10-fold cross-validation evaluation, it was found that our proposed ChurnNet outperformed the state-of-the-art methods applied on 3 benchmark datasets. At present, the research is conducted in the centralized data

repository. If the customer service provider aims to predict churn in a decentralized environment, then, in the future, the research can be extended through a federated learning approach. Also, different data transformation techniques can be used for empirical analysis to improve the performance of the proposed model. On top of that, as our proposed model is a black box model, so, explainable AI (XAI) techniques can be utilized to provide explainability of the produced results of the ChurnNet model.

APPENDIX
LIST OF ACRONYMS

1D	One-Dimensional
ADAM	A Method for Stochastic Optimization
AUC	Area Under Curve
CNN	Convolutional Neural Network
CRM	Customer Relationship Management
DL	Deep Learning
FC	Fully Connected
FN	False Negative
FP	False Positive
GAP	Global Average Pooling
GBM	Gradient Boosting
KNN	K-nearest Neighbors
LDA	Linear Discriminant Analysis
LR	Logistic Regression
MCC	Matthew's Correlation Coefficient
ML	Machine Learning
NADAM	Nesterov-accelerated Adaptive Moment Estimation
NB	Naive Bayes
POD	Probability of Detection
POF	Probability of False Alarm
ReLU	Rectified Linear Unit
RF	Random Forest
RMSprop	Root Mean Square Propagation
SE	Squeeze and Excitation
SMOTE	Synthetic Minority Over-sampling Technique
SMOTEEN	SMOTE with Edited Nearest Neighbors
SMOTETomek	SMOTE-Tomek Links
SVM	Support Vector Machine
TCI	Telecommunication Industry
TN	True Negative
TP	True Positive
WOE	Weight-of-Evidence

APPENDIX
LIST OF NOTATIONS

F_{ex}	Function for excitation operation
F_{scale}	Function for scale operation
F_{sq}	Function for squeeze operation
W_1	Weight matrix of first fully connected layer
W_2	Weight matrix of second fully connected layer
δ	ReLu activation function
σ	Sigmoid activation function
$\tilde{x}_c$	Output of SE block
s_c	Scaling factor
s	Output of the excitation operation
u_c	cth channel of the input
z_c	cth element of the squeezed channels
z	Squeezed input

REFERENCES

[1] M. Óskarsdóttir, C. Bravo, W. Verbeke, C. Sarraute, B. Baesens, and J. Vanthienen, "Social network analytics for churn prediction in telco: Model building, evaluation and network architecture," *Expert Syst. Appl.*, vol. 85, pp. 204–220, Nov. 2017.

[2] C.-P. Wei and I.-T. Chiu, "Turning telecommunications call details to churn prediction: A data mining approach," *Expert Syst. Appl.*, vol. 23, no. 2, pp. 103–112, Aug. 2002.

[3] S. Saleh and S. Saha, "Customer retention and churn prediction in the telecommunication industry: A case study on a Danish university," *Social Netw. Appl. Sci.*, vol. 5, no. 7, p. 173, Jul. 2023.

[4] A. Keramati, R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari, and U. Abbasi, "Improved churn prediction in telecommunication industry using data mining techniques," *Appl. Soft Comput.*, vol. 24, pp. 994–1012, Nov. 2014.

[5] N. Alboukaey, A. Joukhadar, and N. Ghneim, "Dynamic behavior based churn prediction in mobile telecom," *Expert Syst. Appl.*, vol. 162, Dec. 2020, Art. no. 113779.

[6] A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," *J. Big Data*, vol. 6, no. 1, pp. 1–24, Dec. 2019.

[7] J. K. Sana, M. Z. Abedin, M. S. Rahman, and M. S. Rahman, "A novel customer churn prediction model for the telecommunication industry using data transformation methods and feature selection," *PLoS ONE*, vol. 17, no. 12, Dec. 2022, Art. no. e0278095.

[8] A. Mishra and U. S. Reddy, "A comparative study of customer churn prediction in telecom industry using ensemble based classifiers," in *Proc. Int. Conf. Inventive Comput. Informat. (ICICI)*, Nov. 2017, pp. 721–725.

[9] A. Sharma, E. Vans, D. Shigemizu, K. A. Boroevich, and T. Tsunoda, "DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture," *Sci. Rep.*, vol. 9, no. 1, Aug. 2019, Art. no. 11399.

[10] S. F. Ahmed, M. S. B. Alam, M. Hassan, M. R. Rozbu, T. Ishtiak, N. Rafa, M. Mofijur, A. B. M. S. Ali, and A. H. Gandomi, "Deep learning modelling techniques: Current progress, applications, advantages, and challenges," *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 13521–13617, Nov. 2023.

[11] Y. Liu, M. Shengdong, G. Jijian, and N. Nedjah, "Intelligent prediction of customer churn with a fused attentional deep learning model," *Mathematics*, vol. 10, no. 24, p. 4733, Dec. 2022.

[12] A. Khattak, Z. Mehak, H. Ahmad, M. U. Asghar, M. Z. Asghar, and A. Khan, "Customer churn prediction using composite deep learning technique," *Sci. Rep.*, vol. 13, no. 1, Oct. 2023, Art. no. 17294.

[13] S. Agrawal, A. Das, A. Gaikwad, and S. Dhage, "Customer churn prediction modelling based on behavioural patterns analysis using deep learning," in *Proc. Int. Conf. Smart Comput. Electron. Enterprise (ICSCEE)*, Jul. 2018, pp. 1–6.

[14] A. De Caigny, K. Coussement, and K. W. De Bock, "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees," *Eur. J. Oper. Res.*, vol. 269, no. 2, pp. 760–772, Sep. 2018.

[15] S. Renjith, "B2C E-commerce customer churn management: Churn detection using support vector machine and personalized retention using hybrid recommendations," *Int. J. Future Revolution Comput. Sci. Commun. Eng.*, vol. 3, no. 11, pp. 34–39, 2017.

[16] C. Kirui, L. Hong, W. Cheruiyot, and H. Kirui, "Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining," *Int. J. Comput. Sci. Issues*, vol. 10, no. 2, p. 165, 2013.

[17] F. E. Usman-Hamza, A. O. Balogun, L. F. Capretz, H. A. Mojeed, S. Mahamad, S. A. Salihu, A. G. Akintola, S. Basri, R. T. Amosa, and N. K. Salahdeen, "Intelligent decision forest models for customer churn prediction," *Appl. Sci.*, vol. 12, no. 16, p. 8270, Aug. 2022.

[18] A. Idris, A. Iftikhar, and Z. U. Rehman, "Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling," *Cluster Comput.*, vol. 22, no. S3, pp. 7241–7255, May 2019.

[19] B. Q. Huang, T.-M. Kechadi, B. Buckley, G. Kiernan, E. Keogh, and T. Rashid, "A new feature set with new window techniques for customer churn prediction in land-line telecommunications," *Expert Syst. Appl.*, vol. 37, no. 5, pp. 3657–3665, May 2010.

[20] N. Sjarif, M. Rusydi, M. Yusof, D. Hooi, T. Wong, S. Yaakob, R. Ibrahim, and M. Osman, "A customer churn prediction using Pearson correlation function and K nearest neighbor algorithm for telecommunication industry," *Int. J. Advance Soft Comput. Appl*, vol. 11, no. 2, pp. 1–14, 2019.

[21] I. M. M. Mitkees, S. M. Badr, and A. I. B. ElSeddawy, "Customer churn prediction model using data mining techniques," in *Proc. 13th Int. Comput. Eng. Conf. (ICENCO)*, Dec. 2017, pp. 262–268.

[22] A. Amin, A. Adnan, and S. Anwar, "An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and Naïve Bayes," *Appl. Soft Comput.*, vol. 137, Apr. 2023, Art. no. 110103.

[23] A. Amin, F. Al-Obeidat, B. Shah, M. A. Tae, C. Khan, H. U. R. Durrani, and S. Anwar, "Just-in-time customer churn prediction in the telecommunication sector," *J. Supercomput.*, vol. 76, no. 6, pp. 3924–3948, Jun. 2020.

[24] A. Amin, S. Anwar, A. Adnan, M. Nawaz, N. Howard, J. Qadir, A. Hawalah, and A. Hussain, "Comparing oversampling techniques to handle the class imbalance problem: A customer churn prediction case study," *IEEE Access*, vol. 4, pp. 7940–7957, 2016.

[25] C. Wang, R. Li, P. Wang, and Z. Chen, "Partition cost-sensitive CART based on customer value for telecom customer churn prediction," in *Proc. 36th Chin. Control Conf. (CCC)*, Jul. 2017, pp. 5680–5684.

[26] K. Mishra and R. Rani, "Churn prediction in telecommunication using machine learning," in *Proc. Int. Conf. Energy, Commun., Data Anal. Soft Comput. (ICECDS)*, Aug. 2017, pp. 2252–2257.

[27] R. H. Herdian and A. S. Girsang, "The implementation of hybrid methods in data mining for predicting customer churn in the telecommunications sector," *Jurnal Mantik*, vol. 7, no. 1, pp. 216–228, 2023.

[28] S. A. Panimalar and A. Krishnakumar, "Customer churn prediction model in cloud environment using DFE-WUNB: ANN deep feature extraction with weight updated tuned Naïve Bayes classification with block-Jacobi SVD dimensionality reduction," *Eng. Appl. Artif. Intell.*, vol. 126, Nov. 2023, Art. no. 107015.

[29] A. Mishra and U. S. Reddy, "A novel approach for churn prediction using deep learning," in *Proc. IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC)*, Dec. 2017, pp. 1–4.

[30] I. AlShourbaji, N. Helian, Y. Sun, A. G. Hussien, L. Abualigah, and B. Elnaim, "An efficient churn prediction model using gradient boosting machine and metaheuristic optimization," *Sci. Rep.*, vol. 13, no. 1, p. 14441, Sep. 2023.

[31] Y. Zhou, W. Chen, X. Sun, and D. Yang, "Early warning of telecom enterprise customer churn based on ensemble learning," *PLoS ONE*, vol. 18, no. 10, Oct. 2023, Art. no. e0292466.

[32] N. Jajam, N. P. Challa, K. S. L. Prasanna, and C. H. V. S. Deepthi, "Arithmetic optimization with ensemble deep learning SBLSTM-RNN-IGSA model for customer churn prediction," *IEEE Access*, vol. 11, pp. 93111–93128, 2023.

[33] R. Srinivasan, D. Rajeswari, and G. Elangovan, "Customer churn prediction using machine learning approaches," in *Proc. Int. Conf. Artif. Intell. Knowl. Discovery Concurrent Eng. (ICECONF)*, Jan. 2023, pp. 1–6.

[34] *Telco Customer Churn—Kaggle.com*. Accessed: Jul. 15, 2023. [Online]. Available: https://www.kaggle.com/datasets/blastchar/telco-customer-churn

[35] *Churn in Telecom's Dataset—Kaggle.com*. Accessed: Jul. 15, 2023. [Online]. Available: https://www.kaggle.com/datasets/becksddf/churn-in-telecoms-dataset

[36] *CHURN—Dataset by Earino—Data World*. Accessed: Jul. 15, 2023. [Online]. Available: https://data.world/earino/churn

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[38] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[39] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[40] H. Jain, A. Khunteta, and S. Srivastava, "Churn prediction in telecommunication using logistic regression and logit boost," *Proc. Comput. Sci.*, vol. 167, pp. 101–112, Jan. 2020.

[41] H. Jain, G. Yadav, and R. Manoov, "Churn prediction and retention in banking, telecom and it sectors using machine learning techniques," in *Advances in Machine Learning and Computational Intelligence*. Singapore: Springer, 2020, pp. 137–156.

[42] S. W. Fujo, S. Subramanian, and M. A. Khder, "Customer churn prediction in telecommunication industry using deep learning," *Inf. Sci. Lett.*, vol. 11, no. 1, p. 24, 2022.

[43] S. F. Bilal, A. A. Almazroi, S. Bashir, F. H. Khan, and A. A. Almazroi, "An ensemble based approach using a combination of clustering and classification algorithms to enhance customer churn prediction in telecom industry," *PeerJ Comput. Sci.*, vol. 8, p. e854, Feb. 2022.

[44] P. Yang, "Data visualization and prediction for telecom customer churn," *Highlights Sci., Eng. Technol.*, vol. 39, pp. 1080–1085, Apr. 2023.

[45] T. Kimura, "Customer churn prediction with hybrid resampling and ensemble learning," *J. Manage. Inf. Decis. Sci.*, vol. 25, no. 1, pp. 1–23, 2022.

[46] J. Rabbah, M. Ridouani, and L. Hassouni, "New approach to telecom churn prediction based on transformers," in *Proc. Int. Conf. Artif. Intell. Comput. Vis.* Cham, Switzerland: Springer, 2023, pp. 565–574.

[47] I. V. Pustokhina, D. A. Pustokhin, P. T. Nguyen, M. Elhoseny, and K. Shankar, "Multi-objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector," *Complex Intell. Syst.*, vol. 9, no. 4, pp. 3473–3485, Aug. 2023.

[48] A. Amin, F. Al-Obeidat, B. Shah, A. Adnan, J. Loo, and S. Anwar, "Customer churn prediction in telecommunication industry using data certainty," *J. Bus. Res.*, vol. 94, pp. 290–301, Jan. 2019.

[49] T. W. Cenggoro, R. A. Wirastari, E. Rudianto, M. I. Mohadi, D. Ratj, and B. Pardamean, "Deep learning as a vector embedding model for customer churn," *Proc. Comput. Sci.*, vol. 179, pp. 624–631, Jan. 2021.

[50] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A churn prediction model using random forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector," *IEEE Access*, vol. 7, pp. 60134–60149, 2019.

[51] M. Saghir, Z. Bibi, S. Bashir, and F. H. Khan, "Churn prediction using neural network based individual and ensemble models," in *Proc. 16th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2019, pp. 634–639.

[52] S. Maldonado, G. Domínguez, D. Olaya, and W. Verbeke, "Profit-driven churn prediction for the mutual fund industry: A multisegment approach," *Omega*, vol. 100, Apr. 2021, Art. no. 102380.

[53] S. Kumar and M. Kumar, "Predicting customer churn using artificial neural network," in *Proc. 20th Int. Conf. Eng. Appl. Neural Netw. (EANN)*. Xersonisos, Greece: Springer, May 2019, pp. 299–306.

SOMAK SAHA received the B.Sc. degree from the Department of Computer Science and Engineering (CSE), BRAC University, Dhaka, in 2023. He has research experience in the arena of agricultural automation. He is working on multiple research projects in applied machine learning, deep learning, and computer vision. His research interests include federated learning, adversarial machine learning, biomedical image processing, and healthcare informatics.

CHAMAK SAHA received the B.Sc. degree from the Department of Computer Science and Engineering (CSE), BRAC University, Dhaka, in 2023. He has research experience in computer vision and applied machine learning. His research interests include applied AI, particularly in agricultural automation, biomedical image processing, healthcare informatics, business analytics, federated learning, and adversarial machine learning.

MD. GOLAM RABIUL ALAM (Member, IEEE) received the B.S. degree in computer science and engineering and the M.S. degree in information technology, and the Ph.D. degree in computer engineering from Kyung Hee University, South Korea, in 2017. He was a Postdoctoral Researcher with the Computer Science and Engineering Department, Kyung Hee University, from March 2017 to February 2018. He is currently a Full Professor of computer science and engineering with the Department of Computer Science and Engineering, BRAC University, Bangladesh. His research interests include healthcare informatics, mobile cloud and edge computing, ambient intelligence, and persuasive technology. He is a member of the IEEE IES, CES, CS, SPS, CIS, KIISE, and IEEE ComSoc. He received several best paper awards at prestigious conferences.

MD. MAHIDUL HAQUE received the B.Sc. degree in computer science and engineering from BRAC University, in January 2023. He focused on agricultural automation, specializing in fruit and vegetable sorting using convolutional neural networks. He is currently an accomplished Independent Researcher in applied machine learning, deep learning, and computer vision. His research interests include federated learning, adversarial machine learning, and agricultural image processing. He envisions AI-driven advancements enhancing agricultural practices and is dedicated to making meaningful contributions to the field.

ASHIS TALUKDER (Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science and engineering from the University of Dhaka, Bangladesh, and the Ph.D. degree in computer engineering from Kyung Hee University, South Korea. He is currently an Associate Professor with the Department of Management Information Systems (MIS), Faculty of Business Studies, University of Dhaka. His research interests include social networks, influence maximization, network optimization, misinformation blocking in social networks, and data mining. He served as a PC Member for conferences, such as EUSPN and IWCMC-Wireless Sensors. He is a member of the Association for Information Systems (AIS), Bangladesh Chapter, and the Internet Society (ISOC), Bangladesh Chapter. He received the Best Paper Award from several prestigious conferences, including IEEE ICOIN 2019.

• • •