Journal of Information Systems Engineering and Management
2025, 10 (31s)
e-ISSN: 2468-4376
https://jisem-journal.com/

Research Article

Customer Churn Prediction in Banking Sectors Using a Hyperparameter-Tuned Deep Learning Model

B. Thenmozhi[1], Dr. C. Jeyabharathi[2], Dr. S. Vimala[3]

1.Ph.D Research Scholar ,

Mother Teresa Women's University,Kodaikanal, Tamilnadu, India.

phdcs19p614bt@gmail.com

2. Computer Instructor Grade -I (PG),

Govt. Higher Secondary School,Thalaiyuthu, Palani,Tamilnadu, India. bharathi_guhan@yahoo.com

3. Associate Professor, Dept. of Computer Science,

Mother Teresa Women's University, Kodaikanal,Tamilnadu, India.

vimalaharini@gmail.com

ARTICLE INFO

Received: 22 Dec 2024

Revised: 22 Feb 2025

Accepted: 27 Feb 2025

ABSTRACT

The number of service providers is growing significantly across all industries. Customers have several options when determining where to put their hard-earned cash in the banking sector. Thus, customer retention and churn have become the banks' main priorities. Previously, supervised machine learning (ML) classifiers have been utilized in research, but feature engineering for these classifiers is labor-intensive, resulting in an incomplete and overly specific feature selection. So, the proposed system proposes a hyperparameter-tuned deep learning (DL) model for customer churn prediction (CCP) in banking sectors. The proposed system mainly comprises '3' phases: data preprocessing, data balancing, and CCP. Data preprocessing, such as data cleaning and normalization, is performed on the collected dataset. After that, the data balancing is done with the help of an improved synthetic minority oversampling technique (ISMOTE) to balance the preprocessed dataset. Finally, the CCP uses hyperparameter tuned and soft plus activation based on deep multi-layer perceptron (HTSADMLP). The proposed system is tested using a churn dataset of banking customers, and the empirical results demonstrate that the proposed work outperformed conventional methods with 97.81% accuracy.

Keywords: Customer Churn Prediction, Data Preprocessing, Data Balancing, Deep Learning, and Machine Learning.

INTRODUCTION

There are many service providers available today in every industry. Any alternative has an abundance of customers. The banking industry offers a variety of choices for people who want to keep their money safe [1]. In the next three years, 66.8% of existing banking clients plan to or have already utilized a bank account from a non-traditional business (big tech or fintech), according to the World Retail Banking Report 2019. This varied competition environment makes it harder for traditional banks to retain their current customer base [2]. Customer churn is critical for financial organizations since it significantly affects a bank's earnings and reputation [3, 4]. The percentage of customers who choose not to use or subscribe to a good or service offered by a business or organization is known as customer churn or customer attrition. Managing client churn is very important in the banking sector, where massive amounts of data are analyzed to extract information for effective and lucrative actions [5]. The employer may offer rewards and gifts to keep those clients once a prospective churn customer has been identified [6]. Over the past few decades, technology development has allowed banks and many other service providers to gather and preserve information on their clients and categorize them as either churners or non-churners.

Banks consider customer relationship management (CRM) crucial in retaining current customers and lowering customer churn. To keep current customers, CRM analysts must foresee which customers will leave and investigate the causes of customer churn. The prediction techniques used by CRM analysers must be accurate, but [7]—Consequently, a technique known as data mining is used to extract useful information from data. Data mining's ML component enables businesses to examine consumer behavior, especially churn [8]. Traditional learning models like logistic regression (LR), support vector machines (SVM), decision trees (DT), random forests (RF), etc., are primarily used for detecting churn in customers [9]. However, it necessitates a lot of manually designed features (hand-crafted features) and has poor simulation outcomes with vast quantities of data on the other side [10]. Some writers have recently employed DL techniques to forecast client turnover in the banking industry [11]. Although it can quickly and precisely analyze vast amounts of bank churn data, produce insightful conclusions, and effectively resolve complex problems, further development is still required to produce more effective outcomes. In this bank CCP, the class imbalance issue is also crucial. An unbalanced dataset produces inaccurate findings, and system training takes an extended period. This research suggests a novel DL classifier with hyperparameter tuning for banking sectors to forecast customer turnover to address these problems. The main objectives of the proposed system are listed as follows:

- Data preprocessing, such as data cleaning and normalization, is performed to preprocess the collected dataset, which improves the prediction accuracy and reliability of the model.
- We are employing ISMOTE to balance the pre-processed dataset, which decreases the computational complexity and helps to attain more generalized results.
- Utilizing the HTSADMLP model for CCP, the weights and biases of the DMLP are tuned using an enhanced butterfly optimization algorithm (EBOA) to get the optimal prediction results.

The remaining parts of the paper are described as follows: Related works regarding the proposed system are given in part 2. Part 3 deeply explains the phases of the proposed system for CCP. The outcome comparison of the proposed and existing schemes is given in section 4, and finally, section 5 describes the conclusion and future research of the current research model.

1. RELATED WORK

Stephane C. K. Tekouabou *et al.* [12] presented an ensemble of ML models with a data-balancing approach for banking churn prediction. The data for analysing the system was gathered from https://www.kaggle.com/datasets/shrutimechlearn/churn-modelling. The collected data were reprocessed, and the pre-processed dataset was balanced by applying SMOTE. The balanced dataset was passed to the ensemble of ML frameworks to perform churn prediction. The system attained higher results for the RF algorithm with an accuracy of 0.86. **Leonardo Jose Silveira *et al.* [13]** presented an ML model to perform CCP. The system used the rank widget model to select the more informative features from the pre-processed dataset, in which the features were selected based on the correlation between them. The selected features were fed into the classifiers such as RF, artificial neural network (ANN), DT, and LR. Comparing all, the RF model proved to be a better classification system for CCP with 82% accuracy.

Vijayakumar Bharathi S *et al.* [14] presented an ML model called extra tree classifier (ETC) to perform CCP of retail banks. The data of 7000 Indian respondents were collected and fed into the ETC to perform the CCP. The method attained an accuracy of 92% and an AUC of 91.88%, superior to the existing schemes. **Masoud Alizadeh *et al.* [15]** presented a hard and soft data fusion model for performing the CCP of the banking sector. The system collected the data from an openly available repository and performed preprocessing on the collected data to improve prediction accuracy. The mining of the pre-processed data was done using hard and soft data fusion models, in which hard modelling was carried out using change mining, and soft modelling was done based on the dempster-shafer theory. The method attained the maximum accuracy of 0.86% for CCP. **R. Yahaya *et al.* [16]** presented an ANN for CCP by incorporating a hybrid genetic algorithm (GA) and k-means clustering (KMC) approach. The dataset was obtained from the Kaggle repository to initiate the training process. After preprocessing, the pre-processed data was filtered using the GA and KMC models. Finally, ANN was utilized to perform the CCP. The model attained better results when the noise was filtered from the raw data.

Customer churn is a severe issue in the banking industry that happens more frequently. The surveys listed above employ ML approaches to forecasting client turnover in the banking sector. ML techniques allow for analyzing customer behavior, calculating their likelihood of leaving and making accurate predictions. However, while tackling the challenging issue of massive volumes of data, current ML algorithms still need to produce satisfactory results. Additionally, they require a manual feature extraction method to train the system's prediction mechanism effectively. So far, DL algorithms have produced excellent results in predicting customer turnover, but the banking sector has yet to see much development. Also, a class imbalance in the collected dataset is a significant classification performance problem. Considering these drawbacks in mind, this paper solves the class imbalance problem in the collected dataset using an efficient oversampling approach. It proposes a novel DL model for accurately predicting customer churn in banking sectors.

2. PROPOSED METHODOLOGY

Figure 1 shows the workflow of the proposed model. The process of CCP using the proposed system consists of three stages: data preprocessing, data oversampling, and CCP. The banking customer data was initially collected from the publicly available Kaggle dataset. Then preprocessing operations are performed on the collected data to improve model's performance. Afterward, an oversampling model called ISMOTE is utilized to solve the data imbalance issue in the preprocessed dataset. Finally, an HTSADMLP is used to perform the churn analysis of the customers, in which the parameters of the DMLP are tuned using a butterfly optimization algorithm.

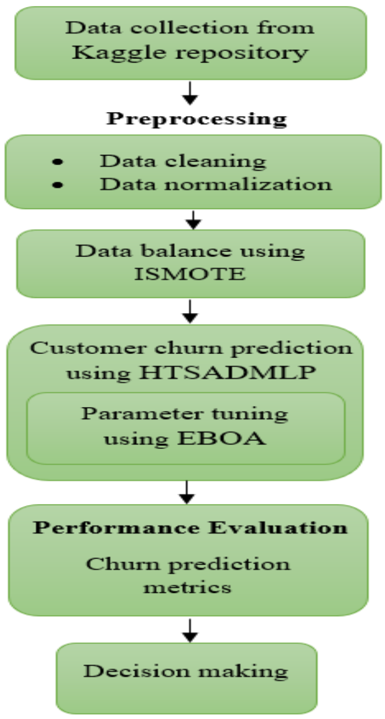

Figure 1: Workflow of the proposed methodology

2.1 Data Preprocessing

Before training the model, it is essential to preprocess the collected data to attain improved performance in CCP. This paper performs three preprocessing procedures: handling missing values, data encoding, and normalization.

Step 1: Handling missing values

This step identifies and handles the missing values in the dataset by applying the imputation approach. If any missing values are present in numerical columns like balance, age, and credit score then the mean imputation is applied that computes the mean values of the non-missing values in the particular column and fills it on the missing values. If any missing values are present in categorical columns like gender and geography, the mode imputation is applied so that the most frequent value is used to fill in the missing one.

Step 2: Encoding

This step converts the categorical data like gender and geography into numerical ones to be better apt for ML models. The one hot encoding was utilized for the geography features, whereas the label encoding was utilized for gender features (0 for female and 1 for male)

Step 3: Normalization

We use min-max normalization to normalize the dataset features that convert all values to the same scale between 0 and 1, where the smallest value of the features is normalized to 0, and the largest value is normalized to 1. So, both small and large values of the features will have equal importance in learning. The mathematical formulation is given below.

$$Min\_Max\_Norm = \left(\frac{\ddot{SI} - \ddot{SI}_{\min}}{\ddot{SI}_{\max} - \ddot{SI}_{\min}} \right) \tag{1}$$

Where, $\ddot{SI}_{min}$ and $\ddot{SI}_{\max}$ are the maximum and minimum values of the feature in the dataset.

2.2 Data Balancing

After preprocessing, data balancing is crucial because data imbalance problems negatively affect classification results by frequently classifying minority classes as the majority class. Data with significant disparities between classes, when minority classes dominate churn classes, is said to have a class imbalance problem. An ISMOTE can be used in this paper to solve the class imbalance issue. The Synthetic Minority Sampling Technique (SMOTE) is a method of oversampling that creates synthetic samples from the minority class, which creates new synthetic minority samples along the line between the minority examples and their chosen nearest neighbors. All information is recovered because the volume of data is not reduced. Despite being simple, SMOTE has a marginalization issue that could result in more significant class overlap or overgeneralization. The SMOTE algorithm may introduce marginalization when creating data. The synthetic sample points produced by the positive sample point and neighboring sample points may likewise be on this edge and become more and more marginalized if a positive (minority) sample point is close to the distribution edge of the positive sample set.Thus, to avoid this, the proposed system first computes the center points $\left(\hat{C}\hat{E}_c\right)$of negative samples $\hat{N}_s$ and positive samples $\hat{P}_s$ using equation (2).

$$\hat{C}\hat{E}_c = \frac{\hat{P}_s + \hat{N}_s}{2} \tag{2}$$

A minority class sample is selected to generate its synthetic data after computing the center points. Then, one among the $k-$nearest minority class neighbors of that sample is randomly selected.Then, one generates the synthetic sample $\dddot{D}\dddot{T}_m$ by interpolating between and as follows:

$$\dddot{D}\dddot{T}_m = \overline{\overline{SA}} + Rand\,(0,1) \times \left(\hat{C}\hat{E}_c - \overline{\overline{SA}}\right) \tag{3}$$

Where, $Rand\,(0,1)$ refers to the random number between 0 and 1 and $\overline{\overline{SA}}$ denotes a data sample in minority-class samples.

2.3 Customer Churn Prediction

Once the dataset has been balanced, the CCP uses the HTSADMLP. An ANN with multiple feeds forward called DMLP maps input vectors to output vectors. It is a connected graph with input, hidden, and output layers, among many others. Many neurons with activation functions are present in the hidden and output layers. The rate of convergence of the network on the error rate largely depends on the choice of hyperparameters, and the random hyperparameter is chosen in DMLP. Thus, the proposed system uses an enhanced butterfly optimization algorithm (EBOA) to optimally select the hyperparameter (weights and bias) to increase the prediction rate and decrease the computational efficiency.

Furthermore, the sigmoid and tanh activation functions are frequently used in DMLP because they perform well in smaller and medium-sized networks. The gradient goes extremely close to zero during backpropagation when deep networks are approached, making it the most vulnerable to the gradient vanishing problem. As a result, weight is rarely modified, which causes very slow convergence. Thus, the proposed system uses a soft plus activation function to address the vanishing gradient problem in the network. It is a smoother version of the rectified linear unit (ReLU) activation function and can constrain a machine's output to always be positive. The hyperparameter tuning and activation function modifications in conventional DMLP are termed HTSADMLP. The steps involved in the HTSADMLP are explained as follows:

Step 1: To begin, this layer's neurons receive input (balanced dataset) and pass it on to the other layers of the network. Next, compute the weighted sum of the input as follows:

$$K_m^{''} = \sum_{m=1}^{p} \dddot{\varpi}_{m,n}\, \dddot{D}\dddot{T}_m + \dddot{\mathrm{B}}_n \tag{4}$$

Where, $p-$refersto the number of neurons in the network, $\dddot{D}\dddot{T}_m$ indicates the balanced input dataset, $\dddot{\varpi}_{m,n}$ denotes the connection weight of the m^{-th} node in the input layer and the n^{-th} node in the hidden layer, and $\dddot{\mathrm{B}}_n$ represents the bias value in the n^{-th} hidden node. Herein, the weights and biases are randomly selected by the DMLP,and this random initialization affects the outcomes of the proposed approach and computation process. So, the hyperparameter is optimally selected by EBOA, which is deeply explained in step 2.

Step 2: Hyperparameter Tuning

The hyperparameters of DMLP are optimally chosen using EBOA to enhance the model performance in several ways. Because the performance of the DMLP heavily depends on the choice of hyperparameters such as learning rate, batch size, number of hidden units, activation function, number of neurons per layer, and epochs. The EBOA provides optimal combination by exploring the hyperparameter space and enables the model to learn complex patterns from the churn dataset for faster convergence, better generalization, and higher accuracy. The BOA is a swarm-based metaheuristic algorithm inspired by the butterfly's foraging behavior. The algorithm chooses the optimal set of hyperparameters by moving through the search space. However, the models fall into local optimal solutions when applied in a high-dimensional search space like DMLP because of the standard search strategy used to adjust the candidate solutions, i.e., the positions of the butterflies. So, we apply the Cauchy mutation to make large random movements in the search space that make the model move to the space ignored by the traditional search mechanism. So, this better exploration helps the model to escape from local optimal solutions and slower convergence,thereby improving optimization performance.

Step 1: Initialize the population of butterflies randomly, where each butterfly represents the set of hyperparameters for the model. Let's define the position of thei^{th}as

$$Y_i = Y_{i1}, Y_{i2}, \ldots\ldots Y_{in} \tag{5}$$

Where Y_{ij} denotes the j^{th} hyperparameter of the i^{th} butterfly, and n denotes the number of hyperparameters. Below is the range, where the hyperparameters are randomly initialized.

$$Y_{ij} \in (Ylow, Yhigh) \tag{6}$$

Where $Ylow$ & $Yhigh$ indicates the lower and upper bounds space of the hyperparameter j.

Step 2: Train the DMLP model with the current set of hyperparameters to measure model performance and evaluate the fitness of butterflies. The solutions attaining higher accuracy are chosen as the current iteration's fittest solutions, expressed as follows.

$$F(Y_i) = Accuracy\ (Y_i) \tag{7}$$

Step 3: Adjust the hyperparameters by applying Cauchy mutation, which uses Cauchy distribution to introduce the randomness. The mutation of each solution is computed below.

$$Y_{ij}(new) = Y_{ij} + \partial . \frac{K_{ij}}{1+\propto^2} \tag{8}$$

Where $K_{ij} \sim Cauchy\ (0, \partial)$ denotes the arbitrary number from the Cauchy distribution with scale parameter ∂, $\propto$ denotes the random number generally ranges from 0 to 1, ∂ controls the mutation step size to enable larger jumps in search space. By enabling the larger random jumps in search space, the Cauchy mutation allows butterflies to explore new search space regions effectively.

Step 4: Update the position of the butterflies by combining the global best butterfly position and random butterfly position, which is expressed as

$$Y_i(new) = Y_i + \mu\ (Yb - Y_i) + \rho\ (Y_i - Yr) \tag{9}$$

Where, Yb indicates the position of the butterfly having the higher fitness value, Yr indicates the random butterfly position from the population and μ & ρ indicates the coefficients that control the influence of global and random butterflies. Doing so balances exploration and exploitation.

Step 5: Retrain the DMLP model with the updated hyperparameters to re-evaluate the fitness of the butterflies. Update the final global best butterfly by comparing the new fitness.$Y_i(new)$with the best fitness ($Gbest$).

$$Gbest = \min(Gbest, F(Y_i(new)) \tag{10}$$

Update the global best solution if the new position of the butterfly obtains the best fitness.

$$Ybest = Y_i(new) \tag{11}$$

Repeat the above steps until the maximum number of iterations is reached. The algorithm output contains the best set of hyperparameters for the model.

3. RESULTS AND DISCUSSION

Here, the outcomes of the proposed CCP system using the proposed DL model are contrasted with the existing related techniques to prove its efficiency. The experiments are executed on a Windows operating system with Python 3.7 and an NVIDIA GeForce 930M graphics card. The processor is an Intel Core i7 with 8 GB of RAM. The proposed system uses anopenly available customer churn banking dataset from Kaggle: https://www.kaggle.com/datasets/radheshyamkollipara/bank-customer-churn.The dataset includes information about the bank customers and whether they churned or stayed in the bank, which can help predict which customers might leave the bank; thus, the bank tries to retain them. The attributes and their distribution of the dataset are shown in Table 1.

Table 1: Attributes and their Description

Attributes	Description/Distribution
Row number	A unique number was assigned to each customer.
Customer Id	A unique ID was assigned to each customer
Surname	Customer’s last name
Credit Score	The creditworthiness of customers generally ranges from 300 to 850.
Geography	The customer location (France, Germany and Spain)
Gender	Male or Female
Age	Age of the customer (Mostly from 30 to 50)
Tenure	The customers staying years in the bank
Balance	Account balance of the customer (most of them have low, and some have very high balances)
NumOfProducts	The number of products the customer uses, like credit cards, loans, etc.
HasCrCard	Whether the customer has a credit card or not
IsActiveMember	Whether the customer uses the banking services actively.
EstimatedSalary	Estimated salary of the customer
Exited	Whether the customer has exited the bank or retained, exited means 1, and stayed means 0.

The existing schemes utilized for comparison include the DMLP, K-Nearest Neighbor (KNN), RF, and SVM. These models are generally used in churn prediction tasks and are good baselines (use different approaches for classification) to compare with our proposed system (i.e., SMOTE with Optimal DMLP). Here is the reason why we chose these models.

- DMLP - A DL model extracts complex patterns from the input data; comparing it with our model shows how optimization and SMOTE improved the standard DMLP performance.
- KNN – It is a straightforward instance-based approach; comparison helps to show if the complex DMLP outperforms the basic strategy.
- RF – It is an ensemble approach that handles data imbalance well, so comparing it with our approach shows whether our optimal DL system with SMOTE beats this standard ensemble model.
- SVM – It is a robust baseline approach compared with our proposed model that shows how well our model deals with the complex churn data compared to SVM.

The evaluation metrics used are accuracy, precision, recall, f-measure, the area under the curve (AUC), Brier score, Kappa, and expected maximum profit of the customer churn (EMPC). These measures offer different insights into how well the model works, particularly for an imbalanced dataset like ours (i.e. non, churning customers are higher than churning customers). The metrics such as precision, recall, and accuracy show how well the model works for predictions, i.e., churned and non-churned customers. The f-score helps to analyze the false positives and negatives, and the AUC helps for imbalanced datasets to show how well it differentiates classes at different thresholds. The Brier score helps show the proposed classifier's confidence in its likelihood predictions and actual values. The Kappa shows the true prediction power of the model rather than the random choice, and finally, EMPC shows the financial benefit of churn predictions. The definition and mathematical computations of each metric are given below.

Accuracy: It measures the proportions of accurate predictions to the total number of predictions.

$$Ac = \frac{Tp+Tn}{Ts} \tag{12}$$

Precision: It measures how well the model works on optimistic predictions, i.e., the ratio of actual churned customers and the total number of customers predicted as churn.

$$Pr = \frac{Tp}{Tp+Fp} \tag{13}$$

Recall: It is defined as the ratio of the predicted churned customers to all actual churned customers.

$$Rc = \frac{Tp}{Tp+Fn} \tag{14}$$

F-measure: It balances both precision and recall, i.e., balances the significance of both false positives and false negatives.

$$Fm = 2 \times \frac{Pr \times Rc}{Pr+Rc} \tag{15}$$

AUC: It evaluates the Receiver Operating Characteristic (ROC) curve to measure the classification model's performance. The true positive and false positive rates are plotted to calculate the ROC.

Brier Score: It evaluates the accuracy of likelihood predictions. The means squared error between the predicted probability (li) and the actual outcome (ai) is measured using this metric, which is computed as

$$Bs = \frac{1}{N}\sum_{i=1}^{N}[li - ai]^2 \tag{13}$$

Kappa: This metric helps to identify how much our model is better than random guessing, particularly for imbalanced datasets. So, the agreement between the predicted and true labels is measured by kappa, and the higher value of kappa represents the better performance of the model.

$$Ka = \frac{Ao - Ae}{1 - Ae} \tag{14}$$

Where Ae & Ao indicates the expected and observed agreement.

EMPC: It evaluates the expected financial benefit of accurately identifying the customers who will churn.

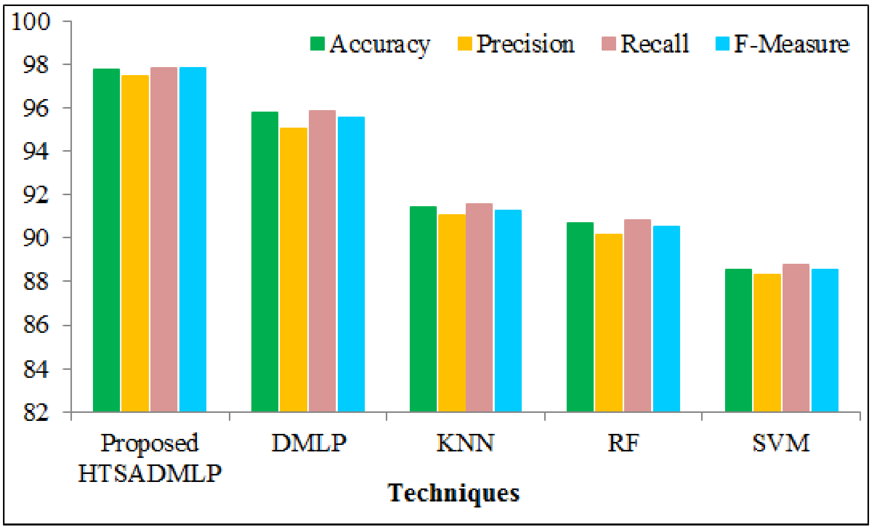

Figure 2: Performance evaluation of various churn prediction models based on four evaluation metrics: accuracy, precision, recall, and f-score

Figure 2 shows the effectiveness of the proposed HTSADMLP model with the existing methods in terms of accuracy, precision, recall, and f-measure. The proposed one attains higher performance than the existing methods. It reaches a maximum accuracy, precision, recall, and f-measure of 97.81%, 97.52%, 97.89%, and 97.85%, which are higher than the conventional methods. Thus, it concludes that the proposed one attains better outcomes than the conventional methods.

Table 2: Performance evaluation of various churn prediction models

Techniques	AUC	Kappa	Brier score	EMPC
Proposed HTSADMLP	0.967	0.7845	0.1298	6.3221
DMLP	0.942	0.6471	0.1312	5.4774
KNN	0.936	0.5311	0.1588	4.9205
RF	0.898	0.4734	0.2064	4.4107
SVM	0.864	0.3805	0.2671	3.8547

Table 2 indicates the outcomes of the proposed HTSADMLP with conventional methods based on AUC, Kappa, Brier, and EMPC metrics. Considering the AUC and Kappa, the existing SVM has 0.864 AUC and 0.3805 Kappa, the existing RF has 0.898 AUC and 0.4734 Kappa, the existing KNN has 0.936 AUC and 0.5311 Kappa, and the existing DMLP has 0.942 AUC and 0.6471 Kappa, which yields minimal outcomes for CCP. However, the proposed system offers 0.967 AUC and 0.7845 Kappa. In addition, the evaluation metrics such as the Brier score and EMPC are the most prominent metrics for CCP. The Brier is a function that demonstrates the precision of the classifier's probabilistic prediction, ranging from 0 (perfect prediction) to 1 (lousy prediction). The EMPC likewise measures the model's profitability in CCP systems. The higher rates show the top good classifier in CCP. Due to its business usefulness, it is frequently used to assess how well churn prediction systems are doing. Here also, the proposed one has a Brier and EMPC of 0.1298 and 6.3221, which are outstanding outcomes than the conventional methods. Our work uses ISMOTE to balance the dataset that generates synthetic minority samples to avoid the classifier's biased outcomes to the majority class, thus improving the model's ability for churn prediction. In addition, the EBOA usage in DMLP for optimization helps to enhance the model performance by capturing complex patterns in the data compared to traditional models like KNN, SVM, RF, and DMLP.

4. CONCLUSION

This paper proposes a CCP system using a hyperparameter-tuned novel DL classifier. The proposed system mainly comprised three phases: data preprocessing, data balancing, and CCP. The performance of the proposed work is investigated against the conventional DMLP, KNN, RF, and SVM approaches. The evaluation is done by some performance metrics: accuracy, precision, recall, f-measure, AUC, Kappa, Brier, and EMPC. The experimental results indicate that the proposed system outperformed the existing methods. For example, the proposed one achieves 97.81% accuracy, 97.52% precision, 97.89% recall, 97.85% f-measure, 0.967 AUC, 0.1298 Brier, 0.7845 Kappa, and 6.3221 EMPC, which are more significant outcomes than the conventional methodologies. Thus, the results show evidence that the proposed models can efficiently predict customer churn in banking sectors. In future work, this will be prolonged with an efficiency feature reduction approach to decreasing the dimensionality of the dataset to increase the prediction rate.

REFERENCES

[1] Kavyarshitha, Y., Sandhya, V., & Deepika, M. (2022, May). Churn Prediction in Banking using ML with ANN. In *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1191-1198). IEEE.

[2] de Lima Lemos, R. A., Silva, T. C., & Tabak, B. M. (2022). Propension to customer churn in a financial institution: A machine learning approach. *Neural Computing and Applications*, *34*(14), 11751-11768.

[3] Jain, L. C., Behera, H. S., Mandal, J. K., & Mohapatra, D. P. (Eds.). (2014). *Computational Intelligence in Data Mining-Volume 3: Proceedings of the International Conference on CIDM, 20-21 December 2014* (Vol. 33). Springer.

[4] Seid, M. H., &Woldeyohannis, M. M. (2022, November). Customer Churn Prediction Using Machine Learning: Commercial Bank of Ethiopia. In *2022 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)* (pp. 1-6). IEEE.

[5] Zakharov, G. N., Kruss, V. I., Lyubovenko, E. S., & Smirnov, S. N. (2021). Lecture Notes in Networks and Systems.

[6] Liu, R., Ali, S., Bilal, S. F., Sakhawat, Z., Imran, A., Almuhaimeed, A., ... & Sun, G. (2022). An Intelligent Hybrid Scheme for Customer Churn Prediction Integrating Clustering and Classification Algorithms. *Applied Sciences*, *12*(18), 9355.

[7] Sagala, N. T. M., &Permai, S. D. (2021, October). Enhanced Churn Prediction Model with Boosted Trees Algorithms in The Banking Sector. In *2021 International Conference on Data Science and Its Applications (ICoDSA)* (pp. 240-245). IEEE.

[8] Domingos, E., Ojeme, B., & Daramola, O. (2021). Experimental analysis of hyperparameters for deep learning-based churn prediction in the banking sector. *Computation*, *9*(3), 34.

[9] ELYUSUFI, Y., &M'hamed, A. I. T. (2022). Churn Prediction Analysis by Combining Machine Learning Algorithms and Best Features Exploration. *International Journal of Advanced Computer Science and Applications*, *13*(7).

[10]Liu, Y., Shengdong, M., Jijian, G., &Nedjah, N. (2022). Intelligent Prediction of Customer Churn with a Fused Attentional Deep Learning Model. *Mathematics*, *10*(24), 4733.

[11] Al-Darraji, S., Honi, D. G., Fallucchi, F., Abdulsada, A. I., Giuliano, R., & Abdulmalik, H. A. (2021). Employee attrition prediction using deep neural networks. *Computers*, *10*(11), 141.

[12]Tékouabou, S. C., Gherghina, Ș. C., Toulni, H., Mata, P. N., & Martins, J. M. (2022). Towards Explainable Machine Learning for Bank Churn Prediction Using Data Balancing and Ensemble-Based Methods. *Mathematics*, *10*(14), 2379.

[13]Silveira, L. J., Pinheiro, P. R., & Junior, L. S. D. M. (2021). A Novel Model Structured on Predictive Churn Methods in a Banking Organization. *Journal of Risk and Financial Management*, *14*(10), 481.

[14]Bharathi S, V., Pramod, D., & Raman, R. (2022). An ensemble model for predicting retail banking churn in the youth segment of customers. *Data*, *7*(5), 61.

[15]Alizadeh, M., Zadeh, D. S., Moshiri, B., & Montazeri, A. (2023). Development of a Customer Churn Model for Banking Industry Based on Hard and Soft Data Fusion. *IEEE Access*, *11*, 29759-29768.

[16]Yahaya, R., Abisoye, O. A., & Bashir, S. A. (2021, February). An enhanced bank customers churn prediction model using a hybrid genetic algorithm and k-means filter and artificial neural network. In *2020 IEEE 2nd International Conference on Cyberspac (CYBER NIGERIA)* (pp. 52-58). IEEE.