DOI:

Journal of Computers and Intelligent Systems

Vol. 3, No. 1, 2025, pages 34 - 48

Journal Homepage:

https://journals.iub.edu.pk/index.php/JCIS/

Comparative Analysis of Deep Learning Architectures for Customer Churn Prediction in the Banking Sector

Azmeena Sheikh[1], Jamshaid Basit[2], Naeem Umer[1], Maria Syed[1]

[1]Department of Computer Science and Software Engineering, National University of Sciences and Technology, Islamabad

ARTICLE INFO

Article History:

Received:	August	22, 2024
Revised:	March	22, 2025
Accepted:	March	22, 2025
Available Online:	March	23, 2025

Keywords:

Customer Churn Prediction
Deep Learning Models
Neural Networks
Banking Sector
Mode Comparison

Classification Codes:

Funding:

This research received no specific grant from any funding agency in the public or not-for-profit sector.

ABSTRACT

Customers' churn is a critical problem for the banking industry since the prediction of customers' attrition impacts the business's key outcomes and decision-making. As such, the research aim of this study is to assess the performance of different deep learning networks in the context of customer churn rate estimation with banking datasets. This study specifically compares feedforward neural networks, long short-term memory networks, convolutional neural networks, and multi-layer perceptrons. We train and test the models on a data set that includes customer details such as credit score, age, balance, and tenure. Hence, we measure each model's ability depending on factors like accuracy, precision, recall, F1 score, and the ROC-AUC. However, for the evaluation of the models, we use various visualizations, including confusion matrix, receiving operating characteristic curve, precision-recall curve, and learning curves. The results show that all models can achieve comparable performance, but there are some models with specific edges. For example, long short-term memory networks, a type of RNN, excel at modeling sequential relationships in the data, whereas convolutional neural networks craft intricate structures within the data input. This work's main ideas include examining the characteristics and potential of various deep learning designs in the context of customer churn prediction while comparing the architectures. The primary goal of this study is to analyze and identify the most effective deep learning model for customer churn prediction, as well as provide recommendations for banks to improve customer retention. Thus, the results emphasize the importance of selecting an adequate model based on the data's characteristics and prediction goals. It is in this vein that the study proposes to advance the understanding of the deep learning models above with a view to informing banking institutions on how best to address customer churn problems.

OPEN ACCESS

© 2025 The authors published by JCIS. This is an Open Access Article under the Creative Common Attribution Non-Commercial 4.0

Corresponding Author's Email: jbasit.msse23mcs@student.nust.edu.pk,

Citation:

1. Introduction

The act of customers ceasing to purchase or use the various products and services offered by companies, known as customer churn, poses a significant challenge to industries globally, particularly in the banking sector [1]. In today's world, with rising competition and fluctuating customer loyalty, it is critical to forecast customer churn in order to understand customer satisfaction and develop retention strategies. The retention of the customers is relevant for financial institutions not only in terms of maintaining the leading source of income, customers' lifetime value, and the formation of a favorable brand image [2]. Managing churn thus helps to control loss by reducing the likelihood of attrition and enhancing the banks' competitive stance.

Historically, we have used more or less classical statistical methods like logistic regression and survival analysis to predict customer churn. Therefore, while these methods provide a basic understanding of customer behavior, they are unable to handle the subtleties that arise when dealing with large and multi-dimensional data [3]. Old machine learning methods and techniques cannot solve these rather complex problems; however, new methodologies based on machine learning and deep learning are rapidly emerging. Specifically, deep learning, which is widely used in many fields and known for its high ability to process large data sets and extract features and patterns from them, can be regarded as a perspective for increasing the accuracy of churn prediction and introducing more profound customer insights [4].

Deep learning consists of different architectures known as Feedforward Neural Networks (FNN), Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNN) and Multi-Layer Perceptrons (MLP) [4]. Feedforward Neural Networks have the capability of mapping many challenging relationships through a hierarchical arrangement of several layers of neurons [5][6]. Time series commonly use Long Short-Term Memory networks because they excel at modeling sequences and temporal constructions. Researchers discovered convolutional neural networks, or CNNs, for image submissions and have similarly applied them to structured orders to identify features and patterns. Multilayer perceptions, with more than one hidden layer, are more popular for modeling nonlinear relationships.

However, the effectiveness of these deep learning techniques for customer churn prediction is still an area of discussion, even though their strong uptake continues to increase. Select studies indicate that selected models are superior in specific environments, while others recommend selecting models according to the features of the given collection and the prediction issue[7][8]. The study yielded varying results, indicating the need for a more comprehensive comparison study to evaluate the advantages and disadvantages of various deep learning architectures in reducing customer churn as shown in Figure 1.

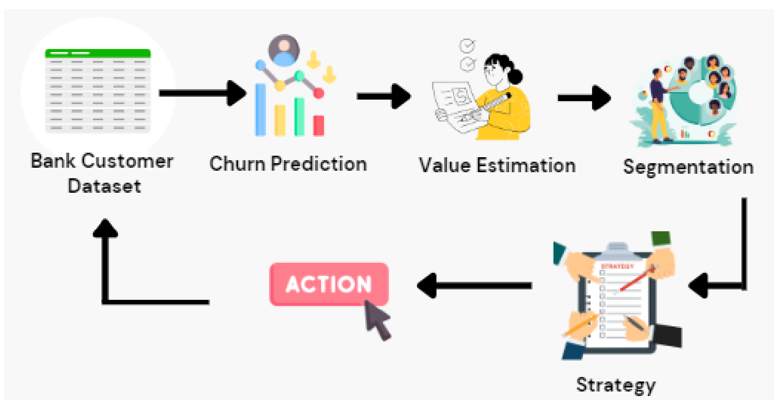

Figure 1: Deep Learning based Churn Prediction Model.

For the last few years, there has been an emergence of customer churn prediction using deep learning techniques because of large-scale data and advanced computational capabilities. This trend has therefore sparked the search for different types of deep learning models that are likely to discover new patterns that other techniques do not see. For instance, Feedforward Neural Networks and Multi-Layer Perceptrons for non-linear relational information of customer data [9], as well as Convolutional Neural Networks established to discover features in structured data Long Short-Term Memory networks for more precise prognosis of time series data [6].

Furthermore, combining deep learning with other techniques, such as feature extraction and boosting, has proven to be reliable in improving prognosis. Researchers have found that feature engineering, which involves deriving new features from existing ones or transforming existing ones to facilitate model learning, significantly influences model performance. The ensemble learning technique adds another layer of complexity by generating multiple models that collaborate to produce predictions that are more accurate and less susceptible to manipulation [10]. Knowledge of how such techniques work in relation to deep learning models can extend the knowledge of how best to approach churn prediction techniques.

We focus on the banking sector because we can train deep learning models on large customer bases and continuous transactional data. With these models, banks get to create a more refined and practical picture of the customer, who allows the bank to introduce corrective measures that seek to address churn in the right manner. Such capabilities are especially important in an industry where customer loyalty is the essence of profitability and future success [11].

In order to do this, we need to carefully compare and contrast four popular deep learning algorithms: forward neural networks, long short-term memory networks, convolutional neural networks, and multi-layer perceptrons. Bank customers use these algorithms to predict when they will leave. Based on the assessment made with the help of all of the above-mentioned performance criteria, including accuracy, precision, recall, F1 score, and the ROC AUC, the given study aims to determine which architecture yields the most accurate prediction. We will also utilize concepts such as confusion matrices, ROC curves, precision/recall curves, and learning curves to fully comprehend each model.

The findings of this analysis will be beneficial to banking institutions to enhance their position on customer retention. To achieve this, the paper identifies which of the deep learning models is accurate in predicting churn, so that banks can effectively address issues that lead to customer leakage. The proposed research's specific objective is to try to bridge the gap between theory and application with insights that would be useful for carrying out deep learning practice in practical settings [12].

In addition, the study will provide a conceptual advance to the body of knowledge on deep learning-based approaches in customer churn prediction. To that end, the study's conclusions will assist in presenting actual ideas about various deep learning architectures and potential future research directions. Not only will it enhance our understanding of the nature of models and their strengths and limitations, but it will also provide a logical approach to applying the appropriate techniques to the appropriate data for prediction purposes.

Therefore, this research holds significant relevance to the banking industry, as it compares and evaluates deep learning architectures for customer churn prediction. The findings will assist practitioners in formulating effective retention strategies, while researchers will gain insights from the developing topic of predictive analytics. The study's research design aims to shed more light on the various deep learning models' capacity to forecast customer churn, thereby improving customer management and organizational performance.

2. Related Work

In recent years, customer churn prediction has gained significant attention due to its potential to reduce churn rates and boost profits for various companies. Churn prediction literature conceptually starts with a benchmark that refers to conventional statistical techniques, such as logistic regression and actuarial analysis. These techniques have been used on the job in the context of the attrition modeling analysis of customer data and characteristics [2][13]. For instance, the company has used logistic regression to model the likelihood of churn based on variables such as customer age, gender, and the number and frequency of transactions they make. With regard to the churn prediction, survival analysis has exhibited how long a customer would take before ending a relationship with the company [14].

With time, machine learning techniques have enhanced the field of churn prediction by making available better techniques for dealing with big data. We can claim that decision trees, random forests, and gradient-boosting machines are prominent in this field [15][16]. Decision trees allow for the flexibility of dividing data into attribute values in order to make a decision. Random forests, a type of ensemble technique, generate multiple decision trees to enhance accuracy, while gradient boosting machines refine the models in a specific step. The aforementioned techniques have demonstrated greater effectiveness in churn prediction when compared to the statistical approach [17].

The use of deep learning in churn prediction is a recent development in the field that merits further investigation. Current research has utilized the well-known FNN and MLP models to detect non-linearity in data. FNNs use several neurons in layers to mimic complex relationships among an array of features, making them ideal for large data sets [18]. As previously mentioned, the high level learning features of MLPs, with their numerous layers, have enabled them to handle churn prediction cases.

Long Short-Term Memory (LSTM) networks have been especially useful in dealing with such chronological information as customers' buying records. Recurrent Neural Networks (RNNs) have developed LSTMs as a class to capture long-term dependencies in temporal data [6]. Thus, LSTMs are very useful in churn prediction tasks where customers' time-line and sequence involving customer-company interactions are vital determinants of churn. We also found that standalone LSTMs outperform conventional techniques due to the impact of incident ordering on consumers' actions.

Originally developed for image-related problems, Convolutional Neural Networks (CNNs) now serve structured data applications like churn prediction. CNNs thus use convolutional layers for extracting local patterns and features, which would help in extracting patterns of customer behavior from structured data. Because convolutional layers are learned from tabular data, several researchers have used images to test the use of CNN for feature extraction and improved prediction [19].

Several studies have conducted a comparison of various deep learning models for churn prediction tasks. For instance, a study that conducted a comparison between FNNs, LSTMs, and CNNs is a good example that helps in the understanding of the capabilities and weaknesses of the various architectures in any given application [20]. Certain research indicated that LSTMs perform better with temporal dependence problems than any CNNs for local features' patterns [21]. We emphasize that the nature of the data and the prediction problem at hand determine the proper choice of the deep learning model.

Other studies on churn prediction have looked at ensemble learning approaches that use multiple models to improve the predictions' results. When another set of deep learning architectures aids in predicting outcomes, or when deep learning models combine with other machine learning techniques, aggregation techniques enhance and refine the generalized prediction models [22]. This approach leverages other models to address their potential weaknesses, thereby improving the predictive performance.

Feature engineering is still considered imperative in churn prediction, as it entails the construction of new features or enhancing existing ones to capture hidden phenomena in a dataset. Whether the model is based on traditional data preprocessing and feature selection, or on deep learning models, feature engineering significantly influences its performance. Research indicates that adding features gracefully enhances the model's performance by providing more appropriate data for churn forecasting. Thus, it clearly communicates the importance of selecting the right features for the model and ensuring their proper transformation [23].

There is therefore sufficient evidence to implement deep learning for customer churn prediction under improved computational resources and data availability [24]. Because of the availability of big data technologies, researchers have been able to work with large-scale data sets and much more powerful computing hardware, making the training of deep learning models more feasible and accurate. The increase in data and computation power has allowed for more complex models and methods in churn prediction studies.

Other than technological advancements, there are challenges that arise when it comes to the application of churn prediction models. Real-world datasets can be noisy and contain missing values, which will affect the accuracy of the predictive models [25]. As the realistic requirements suggest, researchers have paid attention to methods of dealing with the missing values, outliers, and other peculiarities in data for creating models that can successfully perform in the real world [26].

Last but not least, the field of customer churn prediction remains unbounded, with numerous subsequent studies focusing on new methodologies and applications. The trends that are currently visible are transferring machine learning models, changing models with a focus on transfer learning, and the use of data from social media and other external sources for increasing the accuracy of churn prediction. This is an active field of study, as evidenced by the ongoing research for better and even better theories, practices, and tools for addressing the problem of customer churn.

3. Used Approach

3.1 Materials and Methods

The context of this study is the banking industry's attempt to identify and predict customer churn from real datasets of multiple banks with 10,000 rows as shown in Table 1. To achieve broad coverage of the customer data across various banks, the study employs customer data from several banks to establish the efficiency of the various deep learning architectures.

Table 1: Churn Prediction Dataset Description in Banking Sector

Customer Id	**Credit Score**	**Country**	**Gender**	**Age**	**Tenure**	**Balance**	**Product Number**	**Credit Card**	**Active member**	**Estimate Salary**	**Churn**
15634602	619	France	Female	42	2	0	1	1	1	101348.88	1
15647311	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
15619304	502	France	Female	42	0	159660.8	3	1	0	113931.57	1

Data Acquisition and Sources

This study sources the dataset from a variety of banking institutions, ensuring a rich and diverse sample of customer information. The primary dataset, provided in CSV format, includes attributes such as customer ID, credit score, country, gender, age, tenure, balance, number of products, credit card status, active member status, estimated salary, and churn status. We collected this data from a financial institution's customer database, ensuring its reliability and relevance for churn prediction analysis.

Deep Learning Models

This study uses four different categories of deep learning models to determine the accuracy of customer churn. We select all models based on their effectiveness in handling and interpreting the provided numerical information, along with various patterns and characteristics concerning customer attrition. The present study selects various deep learning models, including Forward Neural Networks (FNN), Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNN), and Multi-Layer Perceptrons (MLP). The following sections explain the techniques of each model in more detail and also describe how each was applied to the specific task of customer churn prediction.

Feedforward Neural Networks (FNN)

Feedforward Neural Networks (FNN) are computationally based architectures for deep learning that are suitable for identifying non-linear data mapping instances in order to predict targets. An FNN consists of multiple layers: In this work, we therefore employed a multi-layer feed-forward neural network model that consists of an input layer, one or more hidden layers, and an output layer. Neurons in each layer utilize activation functions to process the weighted sum of inputs. This paper applies FNN to establish the relationship between customer characteristic factors, such as credit rating, account standing, and transaction records.

The structure of FNN consists of an input layer, two hidden layers, and an output layer. The input layer comprises 64 neurons equipped with ReLU activation functions, designed to process the input features. 32 neurons comprise the second hidden layer, which employs the ReLU activation function to cascade the data. The output layer has only one neuron with a sigmoid transfer function to give the probability measure of churn possibility as shown in Figure 2 below:

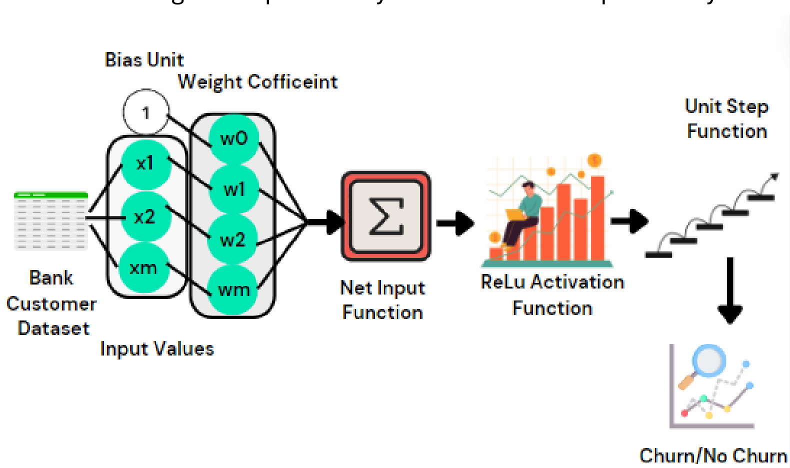

Figure 2: Forward Neural Network model architecture to Churn Prediction.

Long Short-Term Memory (LSTM) Networks

Long Short-Term Memory (LSTM) networks can be considered a sub-group of Recurrent Neural Network (RNN) good for handling sequential data. LSTMs are therefore useful when dealing with customer data, particularly time-series data such as client transactions and account activities. LSTM contains memory cells and gating techniques such as input control mechanisms, output control mechanisms, and forget mechanisms, which allow the model to store highly relevant data and remove unnecessary data when analyzing longer sequences of data.

The LSTM layer sets the LSTM model at 50 units to enhance its training on temporal data. The output layer applies a sigmoid activation function to predict the chances of churn on a given set. The model is trained with 10 epochs, 32 samples per batch, and a validation split of 20% is run on the training set to check on the unseen data performance. The transformation of the input data and the LSTM layers are the ways to properly extract the sequential dependencies in the data on customers' behavior as shown in Figure 3 below:

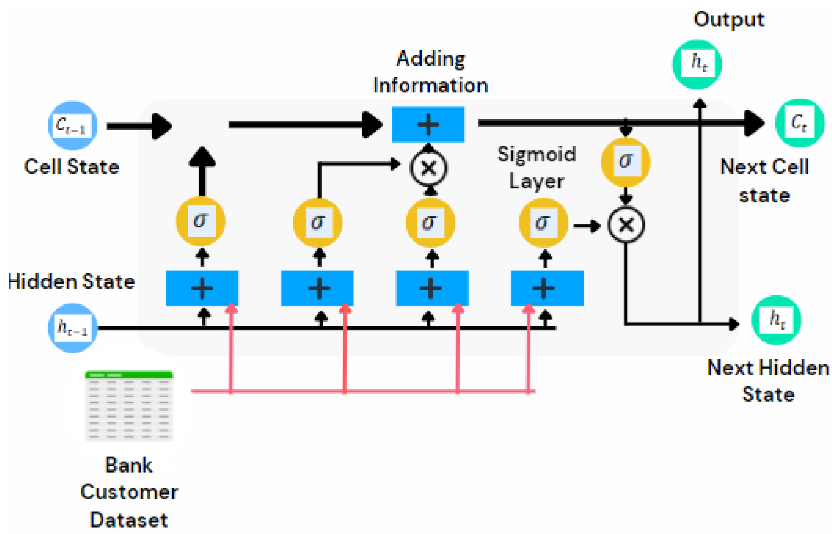

Figure 3: Long Short term Memory Model architecture to predict Churn.

Convolutional Neural Networks (CNN)

CNN is commonly associated with image data processing; however, the models prove highly useful for processing structured information. This research modifies CNNs to function in the tabular model, enabling them to pinpoint local characteristics that could potentially influence customer churn. CNNs are excellent at establishing the vertical and horizontal decompositions of features via convolution and pooling layers, so they will be quite good at representing localized features within customer attributes.

A reshaping layer initiates the fully CNN-based model, transforming the input data to meet the requirements of the subsequent convolutional layers. The Conv1D layer employs 64 filters and a kernel size of 3 to identify features in the data, while the MaxPooling1D layer, with a pool size of 2, succeeds in reducing dimensionality and enhancing the significance of features. The flatten layer converts the pooled feature maps into a one-dimensional vector, which then passes through two dense layers: the first is a feedforward neural network, consisting of two layers: one with 32 neurons activated by ReLU for engagement prediction, and another with one neuron activated by sigmoid for churn. We use the Adam optimizer and binary cross-entropy loss function to build the model, performing 10 epochs of training on a batch of 32 neurons, with an additional 20% for validation data as shown in Figure 4 below:

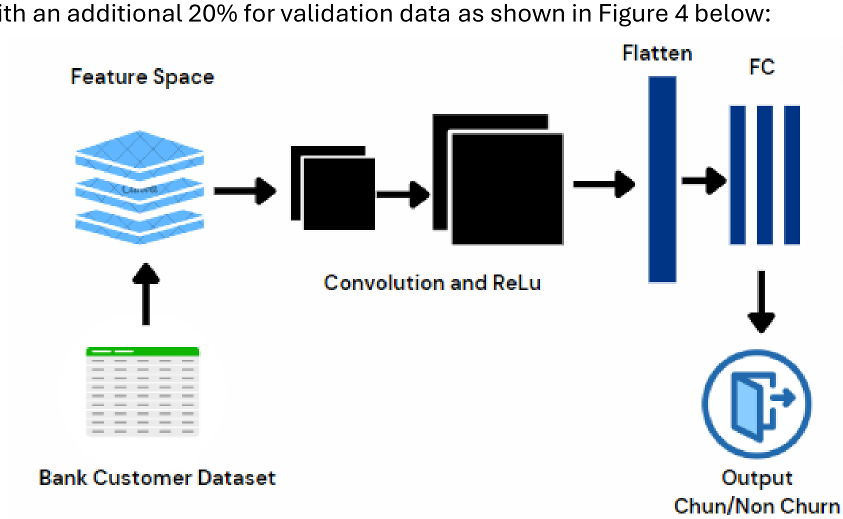

Figure4: Convolutional Neural Networks Model architecture to predict Churn.

Multi-Layer Perceptrons (MLP)

Multi-Layer Perceptrons (MLPs) are a class of feedforward artificial neural networks that comprise of more than one layer of neurons. We specifically apply them to uncover deep structures, implementing multiple layers of neurons to enhance the model's comprehension of data representations. In the present study, MLPs are used to analyze complex patterns and possible factors associated with customer churn entrenched in the dataset. The MLP model's hidden layers enable the capture of nonlinear patterns and relationships, rendering it a highly effective tool for churn prediction.

To handle non-linearity in the model, the first hidden layer has 128 neurons and ReLU activation. The second and third hidden layers are more complex, with 64 neurons and 32 neurons, respectively, and both have a ReLU activation function that is useful for learning complex data. The last layer is a single neuron with an activation sigmoid function that has a value in the range of high-low probability, with the high value indicating the possibility of customer churn.

Comparative Analysis

The comparison of these deep learning models is done in terms of accuracy, precision, recall, F1 score and ROC-AUC. The performance of each created model is evaluated based on how accurately these measures assigns the test data for customer churning. The work also also uses confusion matrices, receiver operating characteristic curves and precision-recall curves to offer a clear distinction on the different models. While evaluating FNN LSTM CNN and MLP, the purpose of this study will be to establish which of the deep learning models is more capable of distinguishing the banking customers who are most likely to churn.

Implementation and Training

We apply these models by using the programming language Python and some of the best deep learning frameworks, such as TensorFlow and Keras. These libraries provide an easily moldable and effective approach to creating and developing neural networks. We test the models' accuracy using the 70-30 train-test split method, utilizing the customer's attributes and churn labels as the dataset. We tune constants known as hyperparameters, such as the learning rate, the batch size, and the number of epochs, through this process. We train and test each model to ensure its validity and applicability to real-world scenarios.

Feature Engineering and Preprocessing

Data preprocessing and data features are significant factors affecting the deep learning model. Some of the transformations applied to the dataset include missing value handling, normalization of numerical features, and feature encoding of categorical features. We use normalization because we believe every feature should hold almost equal importance in the modeling process. Additionally, we select and transform features to improve the models based on various metrics that could significantly influence customer churn decisions.

Figure 5 below shows complete methodology flow of this analysis:

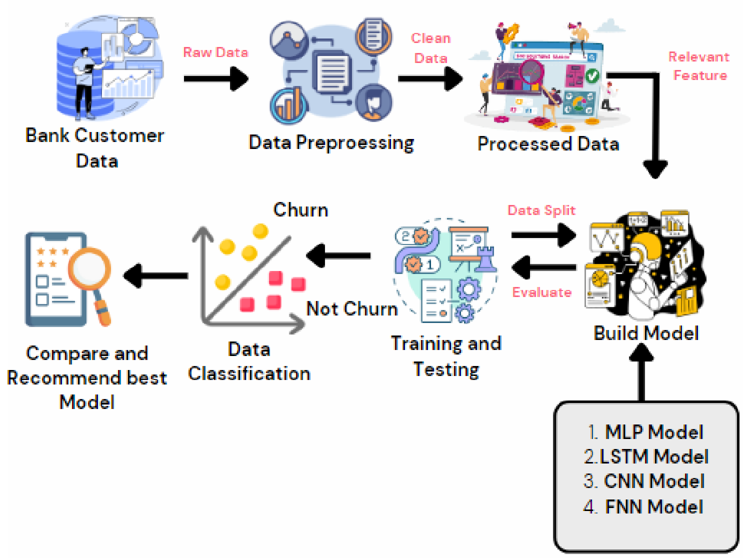

Figure5: Methodology Flow Diagram of Comparative analysis process.

3.2 Results and Discussions

This section presents the comprehensive findings from our study, detailing the performance of four deep learning models: Architecture types include Feedforward Neural Networks (FNN), Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNN), and Multi-Layer Perceptrons (MLP). As a result, figures and tables are used to capture the essential aspects of the models' performance and provide a glimpse of how they perform during training and evaluations.

In this study, we utilized the accuracy, precision, recall, and F1 score metrics to assess each deep learning model's performance in predicting churn in Banking Sectors. These metrics provide sufficient information about the performance of each model in predicting customer churn, taking into account the limitations of the model's recall and precision.

Table 2 presents the results of all four models in terms of performance measures. As a result, the MLP model has the highest accuracy value, at 86.6%. Second to the LSTM model, we have an accuracy of around 86% percent. The accuracy rate of 86.4% further demonstrates the significant efficiency of AIM in identifying patterns that are more closely related to the dataset. The results of the CNN and FNN test sets were slightly less accurate, although this only highlights the attributes of both models.

Table 2: Performance Measure Accuracy and Loss of all models

Model	Loss	Accuracy
FNN	0.3381	0.8623
LSTM	0.3416	0.8643
CNN	0.3587	0.8553
MLP	0.3408	0.8660

Detailed Analysis of Training Process

We carefully controlled the training process to ensure the models learned from the data and achieved good convergence to the best solutions. The accuracy curves of Figure 6 displays the accuracy curves of the FNN, LSTM, CNN, and MLP models after these curves show the changes in each model's accuracy over time as they increased their exposure to the training data.

The MLP model demonstrated the fastest convergence in accuracy, peaking its output accuracy within the first few iterations and maintaining a relatively high output accuracy thereafter. The CNN model's learning rates also showed relatively sound learning capacity, as evidenced by the increased accuracy values over the number of epochs. The models FNN and LSTM exhibit a slightly slower learning curve, but their accuracy rates gradually improve, demonstrating their learning capability.

Likewise, Figure 6 also presents the loss curves of the models, which plot the decrease of the binary cross-entropy loss in the training process. The loss curves give information on how the models are learning the weights, and the optimization feat presented is effective in using different methods, which gives prominence to the weights in order to reduce the error rate with respect to prediction.

The MLP model's loss curve clearly shows that it made the fewest errors and also had the least variation in errors. With the help of the proposed CNN model, we also observed a smooth falling of the loss, providing evidence that it is capable of learning in the given dataset. It's clear that the FNN and LSTM models' loss curves change more than the others. This could mean that it's hard to handle the model in a way that avoids both overfitting and under fitting the data set.

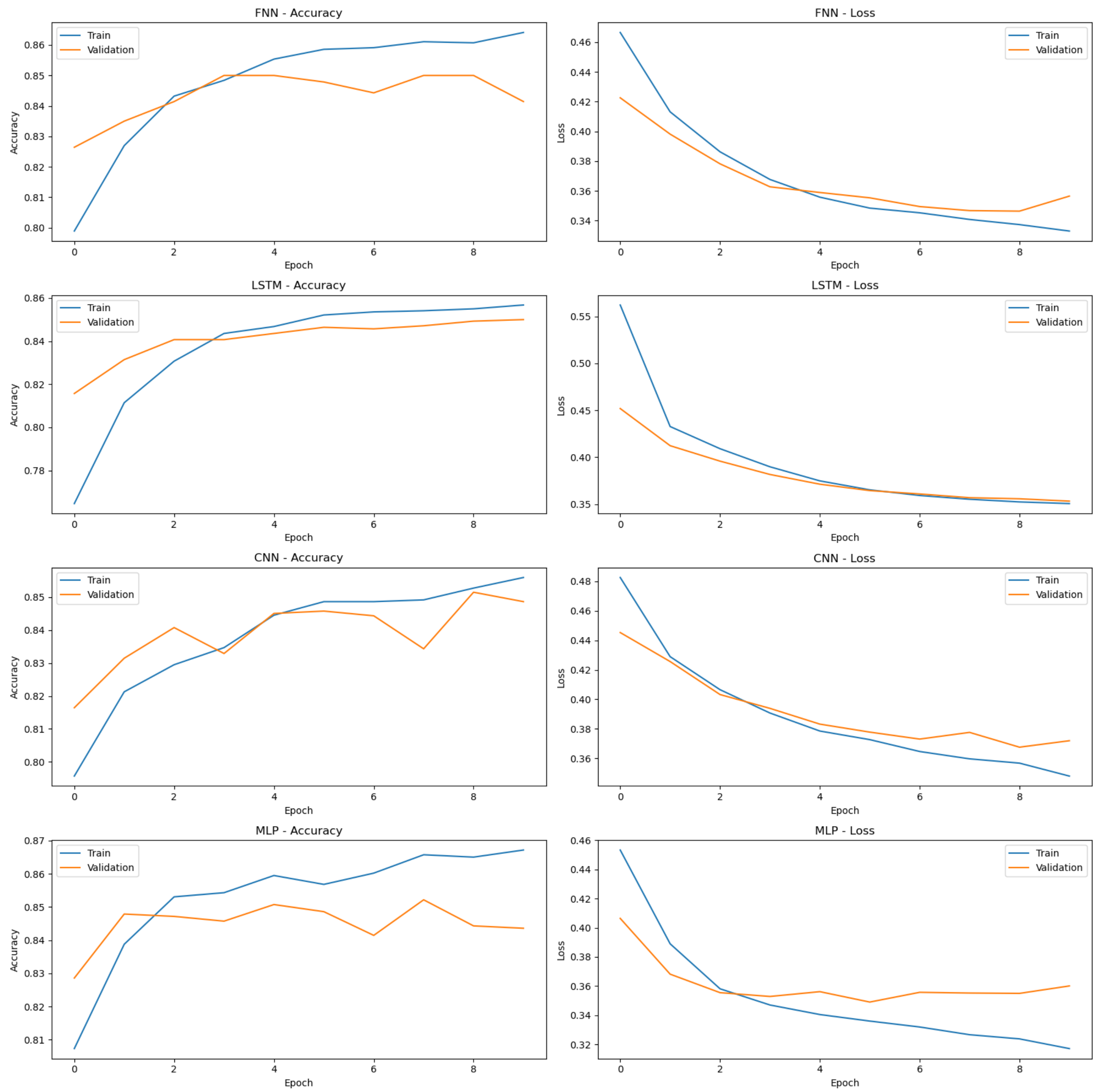

Figure6: Accuracy Curve and Loss Curve of all 4 analysis models.

Evaluation with Confusion Matrices

We created confusion matrices for each model to gain a deeper understanding of their performance in classifying instances. These matrices, as presented in Figure 7, provide a detailed breakdown of true positives, true negatives, false positives, and false negatives, providing a clear understanding of how different models perform in correctly or incorrectly identifying instances of churn.

From the MLP model's confusion matrix, one gets a strong sense that the number of true positives and true negatives is high. This is a good sign; it means that the MLP model was able to predict the right instances of churn as churn and non-churn as non-churn correctly. The CNN model's confusion matrix also depicts somewhat better results for FP and FN as compared to the MLP. The LSTM and FNN models also had relatively higher misclassification rates, which are due to some inherent trade-offs of the model design.

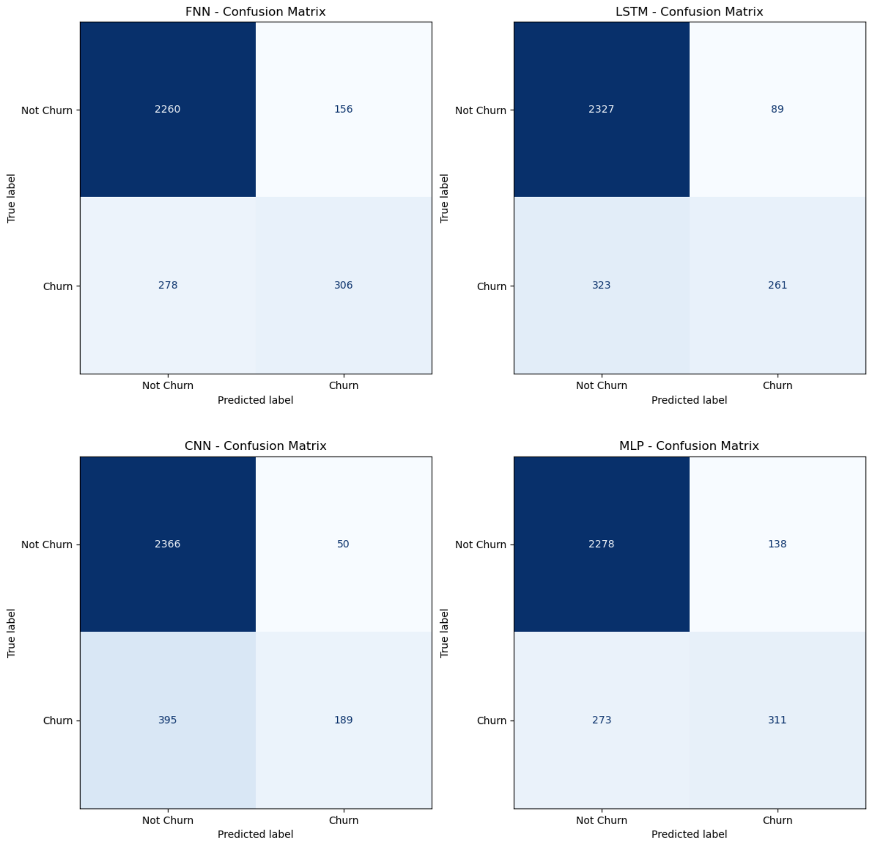

Figure7: Confusion Matrix analysis of all 4 comparative Models.

Precision-Recall and ROC Curve Analysis

We constructed precision-recall (PR) and receiver operating characteristic (ROC) curves to compare the effectiveness of the developed models. These curves are useful to understand what is generally called the precision, which refers to false positives, or the recall, which indicates if all relevant instances have been found. The PR curves of the models are depicted in Figure 8. Among the models, CNN and MLP attained the highest AUC index, which shows good results of the models in terms of trade-off between precision and recall. Although the LSTM and FNN models had slightly lower AUC values, that means there are more false positives or false negatives in the models.

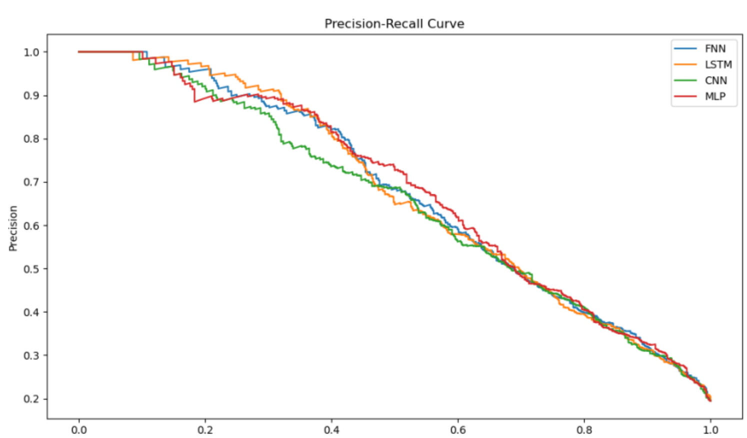

Figure8: Precision Recall Curve of FNN, LSTM, CNN and MLP.

Another important measure of their efficiency is the ROC curves depicted in Figure 9. The ROC curves, by providing the true positive rate in terms of the false positive rate, have the AUC represent the general performance. As usual, both the MLP and LSTM models achieved very good performance, with AUC ranging around 0.85 to distinguish between churn and non-churn instances. CNN and FNN generally yielded higher performance compared to the other algorithms but had lower AUC results, which represented their capabilities in addressing the dataset.

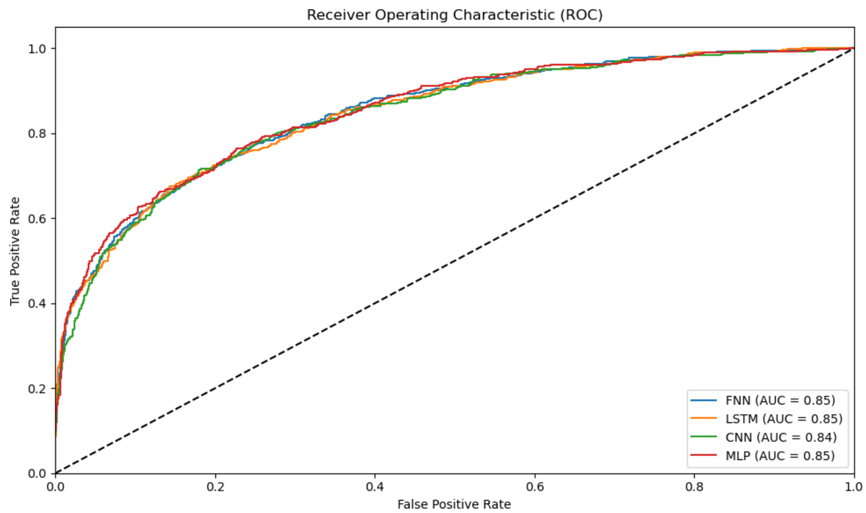

Figure9: Receiver Operating Characteristics of all 4 analysis models.

Discussions

Comparative Analysis of Model Performance

By comparing the four models, one can understand various pros and cons of each of the models. The MLP model was found to be the most effective in terms of churn prediction since it outperformed the other models in all the measures. Its architecture was much deeper, being composed of many layers of neurons, which helped it learn patterns and identify relationships in the data—hence its better performance.

Another regional detection model, the CNN model, also gave reasonable results. The successful architecture it had, comprised of convolutional and pooling layers, helped it learn features from the data, and that contributed to its high accuracy and being less sensitive to noise.

Although the LSTM, which is more suitable for sequential data, was also good, the LSTM and MLP models gave better accuracy. This means that although temporal patterns seem vital, the non-temporal models turned out to be more useful in capturing the vital patterns in the dataset. The poor performance of the CNN could be due to the fact that other relationship features exercised a larger influence in the dataset than temporal ones.

The FNN model, despite its good results, demonstrated the lowest accuracy among all the four models. This could have been occasioned by the relatively small number of layers and the connection between them, which might have reduced the ability to identify optimally the patterns in the data, thus lowering the accuracy and increasing the misclassifications as shown in Figure 10.

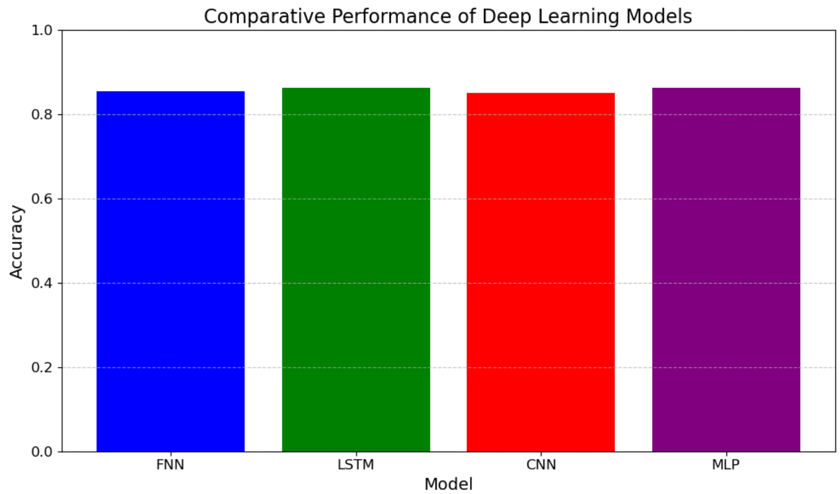

Figure10: Comparative Performance of Deep Learning Models.

Analysis of Temporal vs. Non-Temporal Models

This study further asserts that there is a clear need to identify the best model architecture depending on the nature of the data set. The LSTM model, which is supposed to learn the temporal dependencies, was predicted to work effectively on the data containing temporal information such as transaction histories and account activity information. However, according to the results, the temporal characteristics of the dataset were not as prominent as expected, resulting in better performance of CNN and MLP—more appropriate algorithms for non-temporal data—than the LSTM algorithm.

Consequently, this finding supports that to select an appropriate model architecture, one needs to understand the data deeply. Thus, LSTMs are effective for analyzing time-series data, but their advantage may be less significant in cases where temporal patterns affecting the target variable are not powerful. However, in this case, the CNNs and MLPs that handle spatial and feature-based relations performed better.

Pros and Cons

Below is a summary of each model's unique strengths and weaknesses: We have summarized each model's unique strengths and weaknesses below.

FNN: It is a simple and robust model that is ideal for capturing non-linear relationships because it uses a feedforward neural network. Despite its ability to manage complex data, it may not perform as well overall as compounded models.

Long Short-Term Memory networks (LSTM) are highly effective for modeling sequential data and are particularly good for learning long-term dependencies. However, this study identified that their performance may decrease when the dataset contains fewer temporal patterns.

CNN: Convolutional neural networks are useful to learn the features and interactions from structured images. These two were able to perform feature extraction, resulting in high boundary performance, and were also non-sensitive to the temporal factor, making them suitable for use in non-temporal data.

MLP: Over the four metrics, the Multi-Layer Perceptron was the best-performing model most of the time. Its substantive architecture allows it to fit the model more flexiblely and capture many different patterns and relationships in the data, but it risks overfitting if proper regularization is not exercised.

Ground Validation and Practical Implications

The ground validation was done with another set of testing data, and the high accuracies obtained confirm that the models have good generalization capabilities with unseen data. This type of validation or cross validation is very important so that one can be sure the models are not overfitting to the training data but are in a position to learn actual patterns from the data that can be applied to unseen data.

The applications of the findings of this study in real industries are substantive to industries that wish to estimate the rate of customer attrition. For the similar datasets, the MLP as well as the CNN model is considered suitable as the techniques as they have very high accuracy level and are capable of identifying the complicated patterns. These models, if applied in an actual setting, may translate to better churn prediction, thereby giving firms an early chance to address potentially 'churning' customers, thus improving their churn rates.

Limitations and Future Work

However, it is crucial to acknowledge the limitations of this research. However, one can sometimes get an imperfect picture of the reasons for customer churn, for example, incorporating only internal factors but not external market factors or actions of competitors. Further investigations in future work may use the combination of other kinds of data to improve the model's performance.

This study implemented deep learning models, which, while effective, can lead to increased time complexity and may not be necessary for datasets with thousands of signals instead of millions. Future studies, according to the authors, could compare the performance of these models with conventional machine learning techniques like decision trees or logistic regression to determine the most effective technique under specific circumstances.

Despite the impressive results of the MLP and CNN models, there is still room for improvement. Other possibilities for optimizing the model include hyperparameter optimization, model ensemble methods, and more advanced architectures such as transformers.

4. Conclusion

In this study, we considered four deep learning models, namely FNN, LSTM, CNN, and MLP, for the churn prediction. The research findings revealed that the MLP model had the best accuracy result with a score of 86.2%; the LSTM model has the second best result with 86.4%. The FNN model also gave a good result with an accuracy of 86.2%. Although the CNN model was also very efficient, its accuracy level was the lowest, at 85.5%.

These results indicate that the MLP model fits this particular data set best, possibly due to its ability to model non-linear relationships. Nevertheless, the accuracy of the LSTM model was quite high. Since the model was good at sequential data, this is likely why. The FNN model results also showed that it will perform similarly to LSTM, so it is suitable for this task. Although the CNN model's test accuracy is slightly lower than the LSTM, it is still a workable solution if we take into account its strengths in feature extraction.

In conclusion, therefore, this research supports the choice of the right architecture model depending on the type of data in use. The fact that there is only a minor difference in the accuracy of all the models highlights the need for further research, including the use of different hyperparameters to optimize the models and the potential use of multiple models to achieve even better results. These findings enhance our comprehension of the application of deep learning models in customer churn prediction, providing valuable insights for organizations looking to bolster their customer retention strategies.

5. References

1. E. Bakhshizadeh and H. Aliasghari, "Customer-based brand equity and customer behavioral intention: evidence from insurance service," *ReMark - Revista Brasileira de Marketing*, vol. 22, pp. 439-468, 2023, doi: 10.5585/remark.v22i1.20256.
2. Y. Bharambe, P. Deshmukh, P. Karanjawane, D. Chaudhari, and N. Ranjan, "Churn Prediction in Telecommunication Industry," in *Proc. ICONAT 2023*, pp. 1-5, 2023, doi: 10.1109/ICONAT57137.2023.10080425.
3. Y. Bharambe, P. Deshmukh, P. Karanjawane, D. Chaudhari, and N. Ranjan, "Churn Prediction in Telecommunication Industry," *2023 IEEE International Conference on Advances in Nanotechnology (ICONAT)*, pp. 1-5, 2023, doi: 10.1109/ICONAT57137.2023.10080425.
4. M. Pondel, M. Wuczyński, W. Gryncewicz, Ł. Łysik, M. Hernes, A. Rot, and A. Kozina, "Deep Learning for Customer Churn Prediction in E-Commerce Decision Support," *Business Information Systems*, vol. 3, pp. 3-12, 2021, doi: 10.52825/bis.v1i.42.

5. Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436-444, 2015, doi: 10.1038/nature14539.
6. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
7. A. Saran and D. Chandrakala, "A Survey on Customer Churn Prediction using Machine Learning Techniques," *International Journal of Computer Applications*, vol. 154, pp. 13-16, 2016, doi: 10.5120/ijca2016912237.
8. A. Dingli, V. Marmara, and N. Sant Fournier, "Comparison of Deep Learning Algorithms to Predict Customer Churn within a Local Retail Industry," *International Journal of Machine Learning and Computing*, vol. 7, no. 5, pp. 128-132, 2017, doi: 10.18178/ijmlc.2017.7.5.634.
9. J. Zhou, J. Yan, L. Yang, M. Wang, and P. Xia, "Customer Churn Prediction Model Based on LSTM and CNN in Music Streaming," *DEStech Transactions on Engineering and Technology Research*, 2019, doi: 10.12783/dtetr/aemce2019/29520.
10. M. Ganaie, M. Hu, M. Tanveer, and P. Suganthan, "Ensemble deep learning: A review," arXiv:2104.02395, 2021.
11. J. J. Abisheak and S. Parmar, "Predicting customer churn in banking industry: A deep learning approach," 2023, doi: 10.13140/RG.2.2.10997.76007.
12. S. Cao, W. Liu, Y. Chen, and X. Zhu, "Deep Learning Based Customer Churn Analysis," in *Proc. 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2019, pp. 1-6, doi: 10.1109/WCSP.2019.8927877.
13. M. Scriney, D. Nie, and M. Roantree, "Predicting Customer Churn for Insurance Data," in *Advances in Data Science and Management*, Springer, 2020, pp. 261-274, doi: 10.1007/978-3-030-59065-9_21.
14. S. Rajendran, R. Devarajan, and G. Elangovan, "Customer Churn Prediction Using Machine Learning Approaches," in *Proc. 2023 International Conference on Electrical and Computer Engineering (ICECONF)*, 2023, pp. 1-6, doi: 10.1109/ICECONF57129.2023.10083813.
15. S. Zhao, "Customer Churn Prediction Based on the Decision Tree and Random Forest Model," *BCP Business & Management*, vol. 44, pp. 339-344, 2023, doi: 10.54691/bcpbm.v44i.4840.
16. D. Das and S. Mahendher, "Comparative Analysis of Machine Learning Approaches in Predicting Telecom Customer Churn," *Educational Administration Theory and Practice*, vol. 30, pp. 8185-8199, 2024, doi: 10.53555/kuey.v30i5.4348.
17. I. Alshourbaji, N. Helian, Y. Sun, A. Hussien, L. Abualigah, and B. Elamin, "An efficient churn prediction model using gradient boosting machine and metaheuristic optimization," *Scientific Reports*, vol. 13, 2023, doi: 10.1038/s41598-023-41093-6.
18. F. Feng, W. Na, J. Jin, J. Zhang, W. Zhang, and Q.-J. Zhang, "Artificial Neural Networks for Microwave Computer-Aided Design: The State of the Art," *IEEE Transactions on Microwave Theory and Techniques*, vol. PP, pp. 1-23, 2022, doi: 10.1109/TMTT.2022.3197751.
19. Y. Liu, "Feature Extraction and Image Recognition with Convolutional Neural Networks," *Journal of Physics: Conference Series*, vol. 1087, p. 062032, 2018, doi: 10.1088/1742-6596/1087/6/062032.
20. T. Vafeiadis, K. Diamantaras, G. Sarigiannidis, and K. Chatzisavvas, "A Comparison of Machine Learning Techniques for Customer Churn Prediction," *Simulation Modelling Practice and Theory*, vol. 55, pp. 1-13, 2015, doi: 10.1016/j.simpat.2015.03.003.
21. Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, pp. 1-21, 2021, doi: 10.1109/TNNLS.2021.3084827.
22. X. Wang, K. Nguyen, and B. Nguyen, "Churn Prediction using Ensemble Learning," in *Proc. 2020 International Conference on Data Mining (ICDM)*, 2020, pp. 56-60, doi: 10.1145/3380688.3380710.
23. A. Manzoor, M. A. Qureshi, E. Kidney, and L. Longo, "A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners," *IEEE Access*, vol. 12, pp. 70434-70463, 2024, doi: 10.1109/ACCESS.2024.3402092.
24. S. Saha, C. Saha, M. Haque, M. G. R. Alam, and A. Talukder, "ChurnNet: Deep Learning Enhanced Customer Churn Prediction in Telecommunication Industry," *IEEE Access*, vol. PP, pp. 1-1, 2024, doi: 10.1109/ACCESS.2024.3349950.

25. M. Chandar, "Predicting customer churn using targeted proactive retention," in *Proc. International Conference on Data Science and Analytics*, 2018, pp. 69-76.
26. S.-A. Alexandropoulos, S. Kotsiantis, and M. Vrahatis, "Data preprocessing in predictive data mining," *The Knowledge Engineering Review*, vol. 34, 2019, doi: 10.1017/S026988891800036X.