

Research Article

Binary Firefly Algorithm-Based Optimization of TabNet and DNN Models for Customer Churn Prediction

Philip O. Odion ¹, Maaruf M. Lawal ², Sonia S. Joshua ³, and Abdulrashid Abdulrauf ^{4,*}

¹ Department of Computer Science, Nigerian Defence Academy, Kaduna 800281, Kaduna State, Nigeria; e-mail : podion2012@nda.edu.ng

² Department of Computer Science, Ahmadu Bello University, Zaria 810006, Kaduna State, Nigeria; e-mail : mmlawal80@gmail.com

³ Department of Computer Education, Kaduna State College of Education, Gidan Waya 801001, Kaduna State, Nigeria; e-mail : simjosh247@gmail.com

⁴ Department of Computer Science, Federal Polytechnic Kaltungo, Kaltungo 770101, Gombe State, Nigeria; e-mail : abdulrashid.abdulrauf2022@nda.edu.ng

* Corresponding Author : Abdulrashid Abdulrauf

Abstract: Customer churn remains a major challenge in the telecommunications industry, where intense market competition and evolving customer expectations significantly affect customer retention. Accurate churn prediction is therefore essential for developing effective retention strategies. However, conventional machine learning approaches often exhibit limited capability in capturing complex non-linear relationships and may be sensitive to high-dimensional, noisy, and redundant features commonly found in telecom datasets. To address these challenges, this study investigates the effectiveness of Binary Firefly Algorithm (BFA)-based feature optimization for deep learning-driven customer churn prediction. BFA is employed to identify the most informative feature subset, reducing feature redundancy and improving model efficiency. Three deep learning architectures are evaluated using the Maven Analytics Telecom Customer Churn dataset, namely a Deep Neural Network (DNN), TabNet, and a hybrid TabNet–DNN model. The DNN is designed to learn complex nonlinear patterns, whereas TabNet utilizes a sequential attention mechanism to perform adaptive feature representation learning for tabular data. Experimental results demonstrate that BFA-based feature optimization substantially improves predictive performance across all evaluated models. Among the investigated architectures, TabNet achieved the best overall performance, obtaining an accuracy of 96.42% and an AUROC of 0.9885. Although the hybrid TabNet–DNN model achieved competitive results, it did not consistently outperform the standalone TabNet model. These findings suggest that effective feature optimization contributes more significantly to predictive performance than increased architectural complexity in structured telecom churn prediction tasks. The study further highlights the potential of BFA-enhanced deep learning models as a robust framework for customer churn analytics in telecommunications.

Keywords: Binary Firefly Algorithm; Customer Churn Prediction; Deep Learning; Feature Selection; Predictive Analytics; TabNet; Telecommunications; Tabular Data Learning.

Received: April, 24th 2026

Revised: June, 17th 2026

Accepted: June, 19th 2026

Published: June, 23rd 2026

Curr. Ver.: June, 23rd 2026



Copyright: © 2026 by the authors.
Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>)

1. Introduction

In today's highly competitive and rapidly evolving telecommunications industry, customer retention has become a critical determinant of profitability and long-term business sustainability [1]. Telecommunications providers face substantial infrastructure investments, intense market competition, and high customer acquisition costs, making customer churn a major operational and financial concern [2]. Customer churn, defined as the discontinuation of service by subscribers, directly affects revenue generation, market share, and business growth [3]. The widespread adoption of smartphones, high-speed internet services, and digital platforms has further intensified competition, enabling customers to switch providers more easily than ever before. Consequently, telecommunications operators increasingly rely

on data-driven strategies to identify at-risk customers and implement proactive retention measures [4], [5].

Within this context, predictive analytics has emerged as an essential tool for customer churn management. By leveraging customer demographics, service usage patterns, billing records, and behavioral data, predictive models can identify churn risk and support targeted retention initiatives [6]. These models enable organizations to deploy personalized incentives, improve customer engagement, and optimize resource allocation for retention campaigns [7]. As a result, customer churn prediction has attracted considerable attention from both academia and industry, leading to the development of numerous machine learning and data mining approaches [8]. Recent advances in big data analytics have further expanded the availability of customer-related information, creating opportunities for more accurate and scalable churn prediction systems [9]. Nevertheless, churn prediction remains a challenging task due to the heterogeneous nature of customer behavior, data quality issues, class imbalance, and the presence of high-dimensional feature spaces [10].

Traditional machine learning approaches, including Logistic Regression, Decision Trees, K-Nearest Neighbors, Support Vector Machines, and ensemble-based techniques, have been widely applied to churn prediction problems [11]. Although these methods often provide satisfactory performance, their effectiveness may be limited when modeling complex nonlinear interactions among customer attributes. Consequently, deep learning approaches have gained increasing attention due to their ability to learn hierarchical feature representations and capture intricate nonlinear relationships within large-scale datasets. Deep Neural Networks (DNNs), attention-based architectures, and other advanced neural learning frameworks have demonstrated strong predictive performance across various customer analytics applications, including churn prediction.

Despite these advances, the application of deep learning to structured tabular telecommunications data presents several challenges. Unlike image and text data, telecom datasets frequently contain redundant, correlated, and weakly informative attributes originating from demographic profiles, service subscriptions, billing histories, and customer interaction records. The presence of irrelevant features may increase model complexity, reduce learning efficiency, and negatively affect generalization performance. Therefore, feature optimization has become an important research direction for improving the effectiveness of deep learning models on structured tabular datasets [12]. To address this challenge, this study investigates the use of the Binary Firefly Algorithm (BFA) as a feature optimization strategy for customer churn prediction. BFA is employed to identify the most informative subset of features, thereby reducing feature redundancy and improving the quality of the input space provided to downstream predictive models [13]. The central premise of this study is that effective feature optimization may contribute more substantially to predictive performance than increasing model complexity alone.

While numerous deep learning architectures have been proposed for predictive analytics, they differ substantially in their mechanisms for representation learning and feature utilization. Conventional Deep Neural Networks (DNNs) rely on fully connected layers to learn complex nonlinear relationships from data and have demonstrated strong predictive capability in customer churn prediction tasks [14]. However, DNNs do not inherently perform feature selection and may be sensitive to noisy or redundant inputs. In contrast, TabNet was specifically developed for structured tabular data and employs a sequential attention mechanism that performs adaptive feature selection and representation learning during training, enabling both interpretability and efficient learning from informative features [15]. These complementary characteristics make DNN and TabNet suitable candidates for investigating the interaction between feature optimization and deep representation learning. Furthermore, a hybrid TabNet–DNN architecture is included to examine whether attention-guided representations learned by TabNet can provide additional predictive value when transferred to a downstream neural classifier. Accordingly, this study evaluates DNN, TabNet, and Hybrid TabNet–DNN architectures within a common BFA-enhanced framework, enabling a systematic assessment of how different deep learning paradigms respond to feature optimization in customer churn prediction.

The primary objective of this study is to examine the impact of BFA-based feature optimization on the predictive performance of deep learning models for telecommunications churn prediction. Unlike studies that focus on a single predictive architecture, this work adopts a comparative deep learning perspective by systematically evaluating DNN, TabNet,

and hybrid TabNet–DNN models under a unified feature optimization framework. The contributions of this study are threefold: (1) the application of BFA-based feature optimization to reduce feature redundancy in telecommunications churn data, (2) a comparative evaluation of multiple deep learning architectures under identical optimized feature conditions, and (3) an empirical analysis of the relative roles of feature optimization and architectural complexity in customer churn prediction. These contributions provide practical insights into the development of robust and efficient deep learning frameworks for structured telecommunications analytics.

2. Related Work

Customer churn prediction has attracted significant attention in both academia and industry due to its direct impact on customer retention, revenue generation, and long-term business sustainability. Existing studies have explored a wide range of predictive approaches, including traditional machine learning, ensemble learning, explainable artificial intelligence (XAI), and deep learning methods, with the common objective of improving predictive accuracy and supporting effective retention strategies.

Early studies primarily relied on conventional machine learning techniques. Ele et al. [16] investigated multiple regression-based models for churn prediction and reported strong predictive performance using Lasso Regression, demonstrating the effectiveness of feature regularization in customer analytics. Similarly, Fujo et al. [14] proposed a Deep Backpropagation Artificial Neural Network (Deep-BP-ANN) framework that integrated feature selection, oversampling, and early stopping mechanisms to address class imbalance and overfitting. Their results showed consistent improvements over traditional classifiers, including Logistic Regression, Naïve Bayes, K-Nearest Neighbors, and XGBoost. These findings highlight the growing importance of representation learning techniques in churn prediction tasks.

Ensemble learning has also emerged as a dominant research direction due to its ability to combine multiple predictive models and improve generalization performance. Chang et al. [17] evaluated Decision Trees, Boosted Trees, and Random Forest models, reporting superior performance from Random Forest while also emphasizing model interpretability through SHAP and LIME explanations. Similarly, Lalwani et al. [18] combined feature selection using a gravitational search algorithm with multiple machine learning classifiers and demonstrated the effectiveness of boosting-based approaches, particularly CatBoost and XGBoost. Asif et al. [19] proposed the XAI-Churn TriBoost framework, integrating XGBoost, CatBoost, and LightGBM with SHAP-based explanations to achieve strong predictive performance on large-scale churn datasets. Collectively, these studies demonstrate that ensemble learning remains one of the most effective paradigms for churn prediction, particularly when combined with feature engineering and explainability techniques.

Beyond predictive accuracy, recent studies have increasingly focused on integrating explainability and decision support capabilities into churn management systems. Poudel et al. [20] employed Gradient Boosting Machines together with SHAP explanations to identify key churn drivers and support business decision-making. RetenNet, proposed by Prashanthan et al. [21], extended this direction by integrating machine learning, explainable AI, prescriptive optimization, and Large Language Models within a unified churn management framework. Similarly, Ndabarishye and Singh [22] developed a heterogeneous stacking ensemble that combined imbalance handling, threshold optimization, and SHAP-based interpretation to improve both predictive performance and decision transparency. These studies reflect a growing trend toward interpretable and actionable churn analytics rather than purely predictive solutions.

Recent advances in deep learning have introduced more sophisticated architectures capable of learning complex nonlinear customer behavior patterns. Sharmila et al. [23] demonstrated that Random Forest can achieve exceptionally high predictive performance on telecom churn datasets, highlighting the continuing competitiveness of optimized machine learning models. However, deep learning approaches have increasingly shown advantages when dealing with large-scale and highly complex customer data. Saha et al. [24] proposed ChurnNet, a deep architecture incorporating convolutional layers, residual connections, squeeze-and-excitation modules, and attention mechanisms, achieving strong performance across multiple churn datasets. Likewise, Kumari et al. [25] compared conventional machine learning and deep learning approaches and reported that deep learning models generally achieved

superior predictive accuracy, although at the cost of reduced interpretability. These findings suggest that deep learning architectures are becoming increasingly attractive for customer churn prediction, particularly when complex feature interactions must be captured.

Despite these advancements, several challenges remain. First, many studies continue to rely on either traditional machine learning models or standalone deep learning architectures without systematically investigating how feature optimization influences model behavior. Second, telecom datasets often contain redundant, correlated, and weakly informative variables, yet feature optimization is frequently treated as a preprocessing step rather than as a central component of model development. Third, existing studies rarely provide a comparative analysis of different deep learning paradigms under a common feature optimization framework, making it difficult to assess the relative contributions of feature quality and architectural complexity to predictive performance.

To address these gaps, this study investigates customer churn prediction using BFA-based feature optimization combined with multiple deep learning architectures. Unlike previous studies that primarily focus on a single predictive model, the proposed framework evaluates DNN, TabNet, and a hybrid TabNet–DNN architecture under identical BFA-optimized feature conditions. This design enables a systematic assessment of how feature optimization influences different deep learning paradigms while providing deeper insight into the relationship between feature relevance, representation learning, and churn prediction performance in structured telecommunications datasets.

3. Proposed Method

This section describes the overall methodological framework adopted for customer churn prediction, including dataset selection, data preprocessing, feature optimization, model development, and performance evaluation. The proposed framework integrates Binary Firefly Algorithm (BFA)-based feature optimization with multiple deep learning architectures to investigate the impact of optimized feature representations on churn prediction performance. A schematic overview of the proposed workflow is presented in Figure 1.

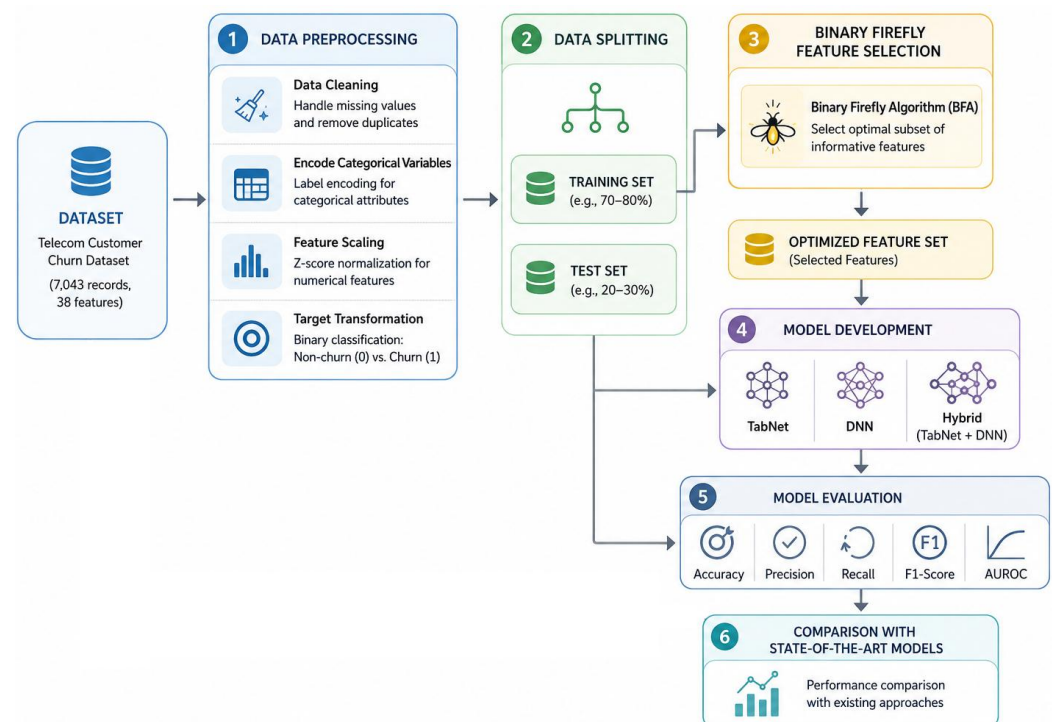


Figure 1. Proposed framework for BFA-enhanced deep learning-based customer churn prediction.

3.1. Dataset

This study utilizes the Telecom Customer Churn dataset obtained from the Maven Analytics platform. The dataset was selected because it provides a sufficiently large and representative collection of telecommunications customer records while containing a diverse set

of demographic, behavioral, service usage, and billing-related attributes commonly employed in churn prediction research. Furthermore, the dataset has been widely adopted as a benchmark dataset in customer churn analytics, facilitating comparison with existing studies.

The dataset consists of 7,043 customer records and 38 attributes describing customer characteristics, subscription behavior, service utilization, and financial information. These attributes capture multiple dimensions of customer activity, enabling a comprehensive analysis of factors associated with churn behavior. The target variable represents customer status and is subsequently transformed into a binary churn classification problem during preprocessing.

3.2. Data Preprocessing

Data preprocessing was performed to ensure data quality, improve model compatibility, and establish a consistent input representation for feature optimization and deep learning analysis. The preprocessing pipeline consisted of data cleaning, categorical encoding, feature scaling, and target variable transformation. These steps were applied sequentially prior to feature selection and model development to minimize noise and improve learning efficiency.

3.2.1. Data Cleaning

The dataset was initially examined for missing values, inconsistencies, and duplicate records. Numerical attributes containing missing values were imputed using the median of the corresponding feature, while categorical attributes were imputed using the most frequent category. Duplicate observations, when present, were removed to prevent potential bias arising from repeated customer records. These procedures ensured data completeness and consistency while preserving the overall distribution of the dataset.

3.2.2. Categorical Feature Encoding

Since the dataset contained a mixture of numerical and categorical variables, categorical predictor variables were transformed into numerical representations prior to model training. Examples of categorical predictor variables include Gender, Married, Offer, Phone Service, Internet Type, Online Security, Device Protection Plan, Premium Technical Support, Contract, and Payment Method. Numerical variables comprised customer demographics, service usage statistics, billing information, and tenure-related attributes. The Customer Status attribute was treated separately as the target variable and was therefore excluded from the predictor feature set during the encoding process.

Label encoding was applied to all categorical predictor variables by assigning a unique integer value to each category. This approach ensured compatibility with both TabNet and DNN architectures while preserving categorical distinctions within the dataset. The encoded variables were subsequently incorporated into the unified feature space used for feature optimization and model development.

3.2.3. Feature Scaling

To improve numerical stability and facilitate efficient optimization, all numerical variables were standardized using Z-score normalization. The scaling parameters were estimated exclusively from the training data and subsequently applied to both training and testing subsets to prevent information leakage. Standardization ensured that features with larger numerical ranges did not dominate the learning process and contributed to more stable convergence during deep learning training.

The standardized variables included customer demographics, geographic information, referral statistics, tenure-related attributes, service usage metrics, and billing-related variables such as Monthly Charges, Total Charges, and Total Revenue. This normalization step was particularly important for neural network-based models, which are sensitive to variations in feature magnitude.

3.2.4. Target Variable Transformation

The Customer Status attribute was designated as the target variable for customer churn prediction. Originally, it consisted of three categories: Stayed, Joined, and Churned. To formulate the problem as a binary classification task, the target variable was transformed into two classes, where Stayed and Joined were assigned a value of 0 (non-churn), while Churned was assigned a value of 1 (churn).

This transformation reflects the practical objective of churn prediction, namely identifying customers who discontinue the service versus those who remain active. The resulting

binary representation is consistent with the evaluation framework adopted in this study and facilitates direct comparison with prior customer churn prediction research.

3.3. Binary Firefly Algorithm-Based Feature Optimization

Binary Firefly Algorithm (BFA) serves as the central optimization component of the proposed framework. Unlike conventional filter-based or statistical feature selection techniques, BFA performs population-based global optimization in high-dimensional search spaces, enabling the identification of informative feature subsets while simultaneously eliminating redundant and weakly informative variables. Given the heterogeneous and highly correlated nature of telecommunications customer data, feature optimization is particularly important for reducing dimensionality, improving learning efficiency, and enhancing the generalization capability of deep learning models.

Prior to model development, an initial exploratory feature analysis was conducted using the complete feature space. During this process, variables such as Churn Category and Churn Reason were identified as post-outcome attributes that contain information directly derived from the target variable. Including such variables would introduce data leakage and lead to unrealistically optimistic performance estimates. Consequently, these attributes were excluded from the final predictive feature space before model training and evaluation. The resulting feature set consisted exclusively of pre-churn demographic, behavioral, service-related, and financial variables, ensuring a realistic and practically deployable churn prediction framework.

Following data preprocessing and train–test partitioning, BFA was applied to the training subset to identify the most informative features for customer churn prediction. The algorithm was adopted based on the work of Mourad and Boudour [26] and was implemented as a wrapper-based optimization approach. In this framework, each candidate solution is represented as a binary vector, where a value of 1 indicates feature inclusion and a value of 0 indicates feature exclusion. The quality of each candidate feature subset is evaluated using predictive performance as the optimization objective, enabling the algorithm to directly search for feature combinations that maximize classification effectiveness.

Algorithm 1. Binary Firefly Algorithm (BFA) for Feature Optimization

INPUT: Population size NP ; Number of features D ; Maximum number of iterations

Max_{gen}

OUTPUT: Best feature subset

- 1: Define objective function $f(x)$, where $x (x_1, x_2, \dots, x_D)$ represents a candidate feature subset
 - 2: Generate initial population of fireflies X_i , $i = 1, 2, \dots, NP$
 - 3: Compute light intensity I_i for each firefly X_i
 - 4: Initialize light absorption coefficient γ
 - 5: Initialize attractiveness β_0
 - 6: Set iteration counter $t = 0$
 - 7: While $t < Max_{gen}$ do
 - 8: For each firefly $i = 1$ to NP do
 - 9: For each firefly $j = 1$ to NP do
 - 10: If $I_j > I_i$ then
 - 11: Move firefly X_i toward X_j
 - 12: Update attractiveness β
 - 13: Update light intensity I
 - 14: Evaluate new solution using $f(x)$
 - 15: End If
 - 16: End For
 - 17: End For
 - 18: Rank all fireflies
 - 19: Identify current best solution
 - 20: $t = t + 1$
 - 21: End While
 - 22: Return best feature subset
-

The optimization process begins with the initialization of a population of fireflies representing candidate feature subsets. Each firefly is assigned a fitness value corresponding to the predictive performance achieved by the selected features. During successive iterations, less attractive solutions probabilistically move toward more attractive ones according to their relative fitness, while random perturbations are introduced to maintain population diversity and avoid premature convergence. Through this iterative search process, BFA balances exploration and exploitation, enabling efficient discovery of high-quality feature subsets within the combinatorial search space. The optimization procedure employed by the Binary Firefly Algorithm is summarized in Algorithm 1. The algorithm iteratively searches for an optimal feature subset by balancing exploration and exploitation within the feature search space while maximizing predictive performance. Upon convergence, BFA produces an optimized subset of predictive features that is subsequently applied to both the training and testing datasets. The selected features are then used consistently across all evaluated architectures, including DNN, TabNet, and the hybrid TabNet–DNN model. By reducing feature redundancy and eliminating irrelevant variables, the optimized feature space improves computational efficiency while enabling a more reliable assessment of how different deep learning architectures respond to feature optimization. This design further supports the central objective of the study, namely evaluating the relative contributions of feature optimization and architectural complexity to customer churn prediction performance.

3.4. Data Splitting

To ensure a reliable and unbiased evaluation of predictive performance, the preprocessed dataset was partitioned using a stratified train–test strategy while preserving the original class distribution of churn and non-churn customers. A hold-out split of 75:25 was adopted, resulting in 5,282 training samples and 1,761 testing samples. This proportion is consistent with previous churn prediction studies and provides an appropriate balance between model training and independent evaluation [17]. In addition to the hold-out evaluation, stratified 5-fold cross-validation was performed on the training subset to assess model stability and generalization performance. Under this procedure, the training data were divided into five stratified folds, where four folds were used for model training and the remaining fold was used for validation. The process was repeated five times such that each fold served as the validation subset once. The final cross-validation performance was computed as the average across all folds, providing a more robust estimate of model behavior while reducing the influence of random data partitioning.

Feature optimization was performed exclusively on the training data to prevent information leakage. The feature subset identified by the Binary Firefly Algorithm (BFA) was subsequently applied to both the training and testing partitions, ensuring that all evaluated models were trained and tested using an identical optimized feature space. This design enables a fair comparison among the investigated architectures while preserving the integrity of the evaluation process. Following preprocessing, feature optimization, and data partitioning, the resulting dataset provided a standardized and high-quality input space for deep learning model development. The adopted workflow improves reproducibility, minimizes the influence of irrelevant attributes, and facilitates a rigorous assessment of the impact of feature optimization on customer churn prediction performance.

3.5. Model Development

To systematically investigate the impact of feature optimization on deep learning performance, three architectures were evaluated using the same BFA-optimized feature subset: Deep Neural Network (DNN), TabNet, and a hybrid TabNet–DNN model. Employing a common optimized feature space ensures a fair comparison among architectures and enables a comprehensive analysis of how different representation-learning paradigms respond to feature optimization in telecommunications churn prediction.

3.5.1. TabNet Model

TabNet was adopted as the first standalone architecture due to its ability to process structured tabular data through sequential attention-based feature selection and representation learning. Unlike conventional neural networks that process all features uniformly, TabNet dynamically focuses on the most informative variables at each decision step, thereby improving both interpretability and predictive performance.

The model was trained using the BFA-optimized feature subset generated during the feature optimization stage. Training was performed using gradient-based optimization with early stopping to mitigate overfitting and improve generalization. Through its sequential attention mechanism, TabNet learns complex feature interactions while simultaneously generating interpretable feature importance representations. Consequently, TabNet serves as an attention-based benchmark for evaluating the effectiveness of deep learning on optimized telecommunications churn data. The hyperparameter configuration adopted for TabNet is summarized in Table 1.

Table 1. TabNet configuration

Parameter	Value / Setting	Description
Maximum Epochs	100	Maximum number of training epochs
Patience	10	Early stopping patience
Batch Size	256	Number of samples per training batch
Virtual Batch Size	128	Sub-batch size used for virtual batch normalization
Feature Input	BFA-optimized features	Features selected through BFA optimization
Optimization	Gradient-based	Parameter optimization using gradient descent
Attention Mechanism	Sequential attention	Adaptive feature selection at each decision step
Regularization	Sparse attention + early stopping	Overfitting prevention mechanism

As shown in Table 1, the TabNet configuration emphasizes stable training and efficient feature utilization through a combination of sequential attention and regularization mechanisms.

3.5.2. Deep Neural Network Model

A Deep Neural Network (DNN) was developed as the second standalone architecture to model nonlinear relationships and higher-order interactions among customer attributes. Unlike TabNet, which incorporates an attention-based feature selection mechanism, the DNN relies on fully connected layers to learn hierarchical feature representations directly from the optimized feature space. This makes the DNN a suitable baseline for assessing the contribution of deep representation learning when feature redundancy has already been reduced through BFA. The network consisted of three fully connected hidden layers with Rectified Linear Unit (ReLU) activation functions. Dropout regularization was applied after the first and second hidden layers to improve generalization and reduce overfitting. The final output layer employed a sigmoid activation function to generate binary churn probabilities. Model training was performed using the Adam optimizer and binary cross-entropy loss function, which are widely adopted for binary classification tasks. The architectural configuration of the DNN is summarized in Table 2.

Table 2. DNN configuration

Layer / Component	Units / Setting	Activation	Dropout	Description
Input Layer	Number of BFA-optimized features	–	–	Receives optimized feature subset
Dense Layer 1	128	ReLU	0.3	First hidden layer
Dense Layer 2	64	ReLU	0.3	Second hidden layer
Dense Layer 3	32	ReLU	–	Third hidden layer
Output Layer	1	Sigmoid	–	Binary churn prediction
Optimizer	Adam	–	–	Gradient-based optimization
Loss Function	Binary Cross-Entropy	–	–	Binary classification objective
Maximum Epochs	100	–	–	Training iterations
Batch Size	32	–	–	Samples per training batch

As summarized in Table 2, the DNN architecture was designed to provide sufficient representational capacity while maintaining model simplicity and training stability. The standalone DNN serves as a complementary benchmark to TabNet, enabling a direct comparison between attention-based tabular learning and conventional deep representation learning under the same BFA-optimized feature space.

3.5.3. Hybrid TabNet–DNN Model

The hybrid TabNet–DNN model was introduced as part of a comparative investigation to determine whether combining attention-based tabular learning with conventional deep representation learning provides additional predictive benefits beyond standalone architectures. Rather than employing an ensemble strategy such as voting, bagging, or stacking, the proposed hybrid architecture follows a sequential feature representation learning paradigm in which TabNet acts as a feature representation extractor and a downstream DNN performs the final classification task. The overall architecture of the hybrid model is illustrated in Figure 2, while the corresponding hyperparameter configuration is presented in Table 3.

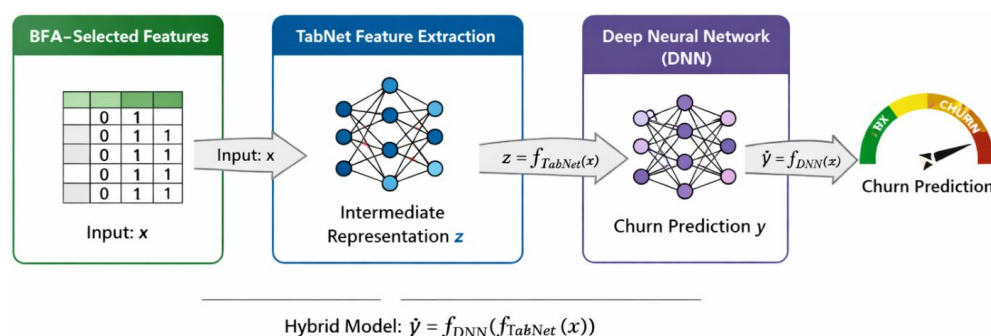


Figure 2. Hybrid TabNet–DNN architecture

Table 3. Hybrid TabNet–DNN configuration

Component	Parameter / Setting	Description
TabNet	Maximum Epochs = 100	Maximum training epochs
	Patience = 10	Early stopping patience
	Batch Size = 256	Number of samples per training batch
	Virtual Batch Size = 128	Virtual batch normalization size
	Feature Input = BFA-optimized features	Optimized features selected by BFA
TabNet Output	Latent feature representation	Learned embedding space for DNN input
DNN Layer 1	64 neurons, ReLU	First hidden layer
Dropout	0.3	Applied after the first hidden layer
DNN Layer 2	32 neurons, ReLU	Second hidden layer
DNN Output	1 neuron, Sigmoid	Binary churn probability
Optimizer	Adam	Gradient-based optimization
Loss Function	Binary Cross-Entropy	Binary classification objective
Epochs	20	Number of training iterations
Batch Size	32	Samples per training batch

In the first stage, TabNet was trained using the BFA-optimized feature subset with a maximum of 100 training epochs, a batch size of 256, and a virtual batch size of 128. Early stopping with a patience value of 10 was employed to reduce overfitting and improve generalization. Through its sequential attention mechanism, TabNet learned both feature selection and high-level feature representations, producing optimized latent embeddings that capture complex interactions among customer attributes. After training, the TabNet model was used as a fixed feature extractor. Specifically, the latent representations generated by the TabNet encoder were obtained through a forward pass of the trained network. These representations correspond to the final embedding space learned by TabNet and can be expressed as an $n \times d$

feature matrix, where n denotes the number of samples and d represents the learned embedding dimension. The resulting embeddings encode attention-weighted feature interactions and provide a compact representation of the original customer data.

Importantly, the TabNet parameters were frozen after the representation learning stage and were not further updated during DNN training. Consequently, the hybrid architecture follows a feature-extraction–then-classification strategy rather than an end-to-end trainable framework. This design allows the contribution of the learned TabNet representations to be evaluated independently from the attention-learning process.

In the second stage, a Deep Neural Network (DNN) was trained using the extracted TabNet embeddings as input features. The DNN consisted of two hidden layers with 64 and 32 neurons, respectively, employing ReLU activation functions and dropout regularization with a rate of 0.3. The output layer utilized a sigmoid activation function to generate binary churn probabilities. Model training was performed using the Adam optimizer and binary cross-entropy loss function for 20 epochs with a batch size of 32. The hybrid architecture therefore combines attention-guided representation learning and conventional deep neural classification within a unified two-stage framework. This design enables an empirical assessment of whether latent representations learned by TabNet can be further exploited by a downstream neural classifier to improve customer churn prediction performance.

3.5.4. Model Hyperparameter Configuration and Reproducibility Settings

To ensure transparency and reproducibility, all major hyperparameters used throughout the experimental framework were explicitly defined. The same configuration was consistently applied across all experiments unless otherwise stated. For the Binary Firefly Algorithm (BFA), the population size was set to 20 fireflies, while the maximum number of iterations was fixed at 20. Each firefly represented a candidate feature subset encoded as a binary vector. Candidate solutions were evaluated using classification performance as the fitness criterion, and a Random Forest classifier with default Scikit-learn parameters was employed as the evaluation model. A fixed random seed was used to ensure reproducibility of the optimization process.

For the TabNet architecture, training was performed using a maximum of 100 epochs with an early stopping patience of 10 epochs. The batch size and virtual batch size were set to 256 and 128, respectively. Optimization was performed using the Adam optimizer. Architectural parameters, including the number of decision steps (n_{steps}), decision dimension (n_d), attention dimension (n_a), and relaxation coefficient (γ), were not manually tuned and therefore retained the default values provided by the TabNet implementation library. This configuration allows the evaluation to focus on the impact of feature optimization rather than extensive hyperparameter tuning.

For the DNN architecture, training was conducted using the Adam optimizer with a learning rate of 0.001 and binary cross-entropy loss. The batch size was fixed at 32, and dropout regularization with a rate of 0.3 was applied to mitigate overfitting. The same training configuration was employed for both the standalone DNN and the DNN classifier component of the hybrid architecture to ensure a fair comparison among models. Overall, the adopted hyperparameter settings provide a reproducible experimental framework while maintaining consistency across the evaluated architectures. This configuration enables the comparative analysis to focus primarily on the effects of BFA-based feature optimization and architectural design choices rather than differences arising from extensive hyperparameter optimization.

3.6. Performance Evaluation

The predictive performance of the proposed framework was assessed using multiple evaluation metrics to provide a comprehensive analysis of classification effectiveness. Since customer churn prediction is formulated as a binary classification problem, where customers are categorized as either churn or non-churn, model performance was evaluated using Accuracy, Precision, Recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUROC).

Accuracy measures the overall proportion of correctly classified instances, whereas Precision evaluates the reliability of churn predictions by quantifying the proportion of correctly identified churn cases among all predicted churn instances. Recall measures the ability of a model to identify actual churned customers and is particularly important in customer

retention applications, where failing to detect at-risk customers may lead to revenue loss. The F1-score provides a balanced measure of Precision and Recall, making it suitable for evaluating performance under class imbalance. Finally, AUROC assesses the discriminative capability of a classifier across different decision thresholds and is widely regarded as a robust indicator of classification performance. Collectively, these metrics provide complementary perspectives on model behavior and enable a comprehensive comparison of the evaluated deep learning architectures under the same BFA-optimized feature space.

4. Results and Discussion

This section presents the experimental results obtained from the proposed customer churn prediction framework. The analysis begins with an examination of the features selected by the Binary Firefly Algorithm (BFA), followed by an assessment of the effectiveness of feature optimization. Subsequently, the predictive performance of the evaluated deep learning architectures is analyzed using Accuracy, Precision, Recall, F1-score, and AUROC to determine the impact of feature optimization on customer churn prediction.

4.1. Feature Optimization Using Binary Firefly Algorithm

The Binary Firefly Algorithm (BFA) reduced the original feature space from 38 variables to 18 retained features, corresponding to a dimensionality reduction of approximately 52.63%. This reduction demonstrates the effectiveness of the optimization process in eliminating redundant and weakly informative attributes while preserving the variables that contribute most significantly to customer churn prediction. Reducing the dimensionality of the input space is beneficial for improving computational efficiency, mitigating overfitting, and enhancing the generalization capability of deep learning models.

The final feature subset selected by BFA and retained for model development is presented in Table 4. The selected features span multiple dimensions of customer behavior, including demographic characteristics, service subscriptions, usage patterns, customer engagement indicators, and financial attributes. This distribution suggests that customer churn is influenced by a combination of behavioral, service-related, and economic factors rather than by any single category of variables.

Table 4. Features selected by the Binary Firefly Algorithm

Feature Index	Feature Name	Importance Score
1	Gender	0.102961
2	Age	0.075220
3	Married	0.062920
9	Number of Referrals	0.060829
11	Offer	0.059280
12	Phone Service	0.056728
13	Avg Monthly Long Distance Charges	0.054344
16	Internet Type	0.046312
18	Online Security	0.041727
20	Device Protection Plan	0.040630
21	Premium Tech Support	0.040412
22	Streaming TV	0.040231
24	Streaming Music	0.039138
25	Unlimited Data	0.038131
26	Contract	0.033550
29	Monthly Charge	0.028573
30	Total Charges	0.020365
33	Total Long Distance Charges	0.017825

As shown in Table 4, the retained features encompass variables related to customer demographics, customer engagement, service adoption, usage behavior, and billing information. Among the selected variables, Gender, Age, Married status, Number of Referrals, and Offer

received the highest importance scores, indicating that demographic and engagement-related factors play an important role in distinguishing churned and non-churned customers. These variables may capture differences in customer loyalty, responsiveness to promotional programs, and long-term service commitment.

In addition to demographic and engagement-related attributes, several service and financial variables were retained by the optimization process, including Phone Service, Internet Type, Online Security, Premium Tech Support, Device Protection Plan, Streaming TV, Streaming Music, Unlimited Data, Monthly Charge, Total Charges, and Average Monthly Long Distance Charges. The inclusion of these variables suggests that service dependency, perceived service value, and customer spending behavior provide complementary information for churn prediction. From a telecommunications perspective, churn decisions are often influenced by a combination of service experience, subscription characteristics, and financial considerations rather than a single determinant.

It should be noted that post-outcome variables such as Customer Reason and Churn Category were intentionally excluded from the final predictive feature space prior to model training and evaluation. Although these variables appeared during an initial exploratory feature selection stage, they were removed because they contain information directly related to the churn outcome. Retaining such variables would introduce data leakage and potentially lead to overly optimistic performance estimates. Consequently, only pre-churn demographic, behavioral, service-related, and financial attributes were retained for subsequent model development.

The feature optimization results demonstrate that a relatively compact subset of variables is sufficient to capture the key characteristics associated with customer churn. By reducing the feature space from 38 to 18 variables while preserving informative predictors, BFA improves data representation efficiency and establishes a consistent input space for all evaluated deep learning architectures. This optimized feature subset serves as the foundation for the comparative model analysis presented in the following sections.

4.2. Model Performance Analysis

To evaluate the effectiveness of BFA-enhanced feature optimization across different deep learning paradigms, three architectures—DNN, TabNet, and Hybrid TabNet–DNN—were trained and evaluated using the same optimized feature subset. This experimental design ensures a fair comparison by controlling the input feature space and enables the analysis to focus on architectural differences rather than variations in feature availability. The evaluated models represent three distinct learning strategies: conventional deep representation learning (DNN), attention-based tabular learning (TabNet), and a hybrid representation-learning framework that combines both approaches. The comparative performance of the evaluated models is presented in Table 5.

Table 5. Comparative performance of DNN, TabNet, and Hybrid TabNet–DNN

Model	Accuracy	Precision	Recall	F1-Score	AUROC
DNN	0.961953	0.965116	0.888651	0.925307	0.985914
TabNet	0.964225	0.985577	0.877944	0.928652	0.988549
Hybrid TabNet–DNN	0.961386	0.969412	0.882227	0.923767	0.981814

As shown in Table 5, all evaluated architectures achieved strong predictive performance, with accuracy values exceeding 96% and AUROC values above 0.98. These results demonstrate that the BFA-selected feature subset effectively preserves the most informative customer attributes while eliminating redundant variables, enabling robust churn prediction across all evaluated deep learning models. Among the three architectures, TabNet achieved the strongest overall performance, obtaining the highest Accuracy (96.42%), Precision (98.56%), F1-score (92.87%), and AUROC (0.9885). These results indicate that TabNet's sequential attention mechanism effectively captures informative feature interactions and provides superior class discrimination on structured telecommunications data. The consistently strong performance across multiple metrics suggests that attention-guided representation learning is particularly well suited for churn prediction tasks involving heterogeneous tabular features.

In contrast, the DNN achieved the highest Recall (88.87%), indicating a stronger ability to identify actual churn cases and minimize false negatives. From a customer retention perspective, this characteristic is valuable because undetected churners may represent lost intervention opportunities. Although its overall Accuracy and AUROC were slightly lower than those of TabNet, the DNN demonstrated greater sensitivity to churn-related patterns within the optimized feature space. The Hybrid TabNet–DNN model produced competitive results, achieving an Accuracy of 96.14%, F1-score of 92.38%, and AUROC of 0.9818. However, the hybrid architecture did not outperform the standalone TabNet model across any evaluation metric. This observation suggests that the feature representations learned by TabNet already capture most of the discriminative information contained in the dataset, limiting the additional benefit obtained from subsequent processing by a downstream neural classifier.

To further investigate model behavior beyond aggregate performance metrics, confusion matrix analysis was conducted. The confusion matrices of the evaluated architectures are presented in Figure 3, where Figure 3(a), Figure 3(b), and Figure 3(c) correspond to the Hybrid TabNet–DNN, TabNet, and DNN models, respectively.

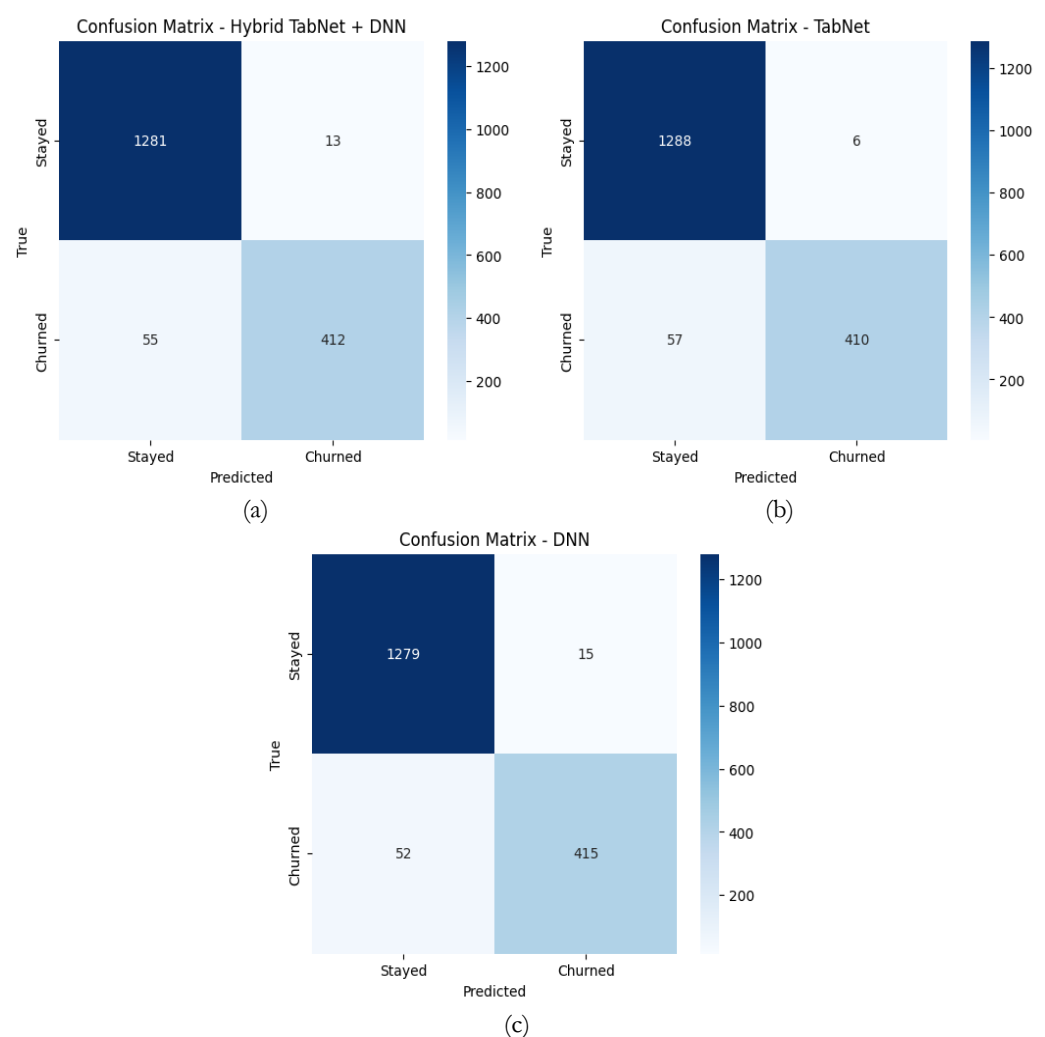


Figure 3. Confusion matrices of evaluated deep learning architectures: (a) Hybrid TabNet–DNN; (b) TabNet; and (c) DNN.

The confusion matrix analysis provides additional insight into the classification characteristics of each model. TabNet recorded the lowest number of false positives (6), which directly contributed to its superior Precision. Conversely, the DNN produced the lowest number of false negatives (52), explaining its higher Recall and stronger capability to identify customers at risk of churn. The Hybrid TabNet–DNN model yielded intermediate values, with 13 false positives and 55 false negatives, reflecting a balanced trade-off between Precision and Recall. Overall, the confusion matrix results support the metric-based findings presented

in Table 5. The results indicate that TabNet provides the most stable and effective overall classification performance, whereas the DNN offers improved churn detection sensitivity. Furthermore, the absence of measurable gains from the Hybrid TabNet–DNN architecture suggests that increasing architectural complexity alone does not necessarily improve predictive performance when attention-based feature selection and representation learning are already effectively captured by TabNet. These findings reinforce the broader observation of this study that feature optimization plays a more substantial role than architectural hybridization in improving customer churn prediction performance on structured telecommunications datasets.

4.3. Discussion

The experimental results indicate that feature optimization plays a decisive role in improving customer churn prediction performance on structured telecommunications datasets. Across all evaluated architectures, the application of the BFA consistently resulted in substantial improvements in classification performance, suggesting that the quality and relevance of input features exert a greater influence on predictive outcomes than architectural complexity alone. This observation highlights the importance of feature optimization as a prerequisite for effective deep learning on tabular telecom data. To quantify the contribution of BFA, an ablation study was conducted by comparing model performance with and without feature optimization. The results are summarized in Table 6.

Table 6. Ablation study of BFA-based feature optimization

Model	BFA	Accuracy	Precision	Recall	F1-Score	AUROC
DNN	Yes	0.96195	0.96512	0.88865	0.92531	0.98591
DNN	No	0.82453	0.70681	0.57816	0.63604	0.88527
TabNet	Yes	0.96422	0.98558	0.87794	0.92865	0.98855
TabNet	No	0.82112	0.66170	0.66595	0.66382	0.88046
Hybrid TabNet–DNN	Yes	0.96139	0.96941	0.88223	0.92377	0.98181
Hybrid TabNet–DNN	No	0.82567	0.67857	0.65096	0.66448	0.88033

As shown in Table 6, the application of BFA consistently improved the performance of all evaluated architectures. Accuracy increased from approximately 82% to over 96%, while AUROC improved from around 0.88 to above 0.98 across DNN, TabNet, and Hybrid TabNet–DNN. Similar improvements were observed in Precision, Recall, and F1-score, indicating that the benefits of feature optimization were not limited to a single evaluation criterion. The consistency of these improvements across different architectures suggests that the quality of the input feature space plays a critical role in determining churn prediction performance.

The substantial performance gains can be attributed to the ability of BFA to identify informative features while eliminating redundant and weakly predictive variables. Telecommunications churn datasets often contain highly correlated information related to customer demographics, service subscriptions, billing behavior, and usage patterns. Such redundancy may increase model complexity and hinder generalization when all available variables are used simultaneously. By retaining only the most relevant predictors, BFA reduces noise in the input space and enables the learning algorithms to focus on patterns that are more strongly associated with churn behavior.

The comparative results also reveal meaningful differences among the evaluated deep learning architectures. TabNet achieved the strongest overall performance, recording the highest Accuracy (96.42%), Precision (98.56%), F1-score (92.87%), and AUROC (0.98855). These results suggest that TabNet is particularly well suited for structured telecommunications data, as its sequential attention mechanism dynamically emphasizes informative features while suppressing less relevant information. When combined with the optimized feature subset generated by BFA, this capability enables TabNet to learn highly discriminative representations that support robust classification performance.

In contrast, the DNN achieved the highest Recall (88.87%), indicating greater sensitivity in identifying churned customers. This behavior suggests that the model favors broader decision boundaries, allowing it to detect a larger proportion of positive churn cases. From a customer retention perspective, high Recall is particularly valuable because undetected

churners represent missed intervention opportunities. Although its overall performance was slightly lower than that of TabNet, the DNN demonstrated strong capability in capturing churn-related patterns after feature optimization.

The Hybrid TabNet–DNN model produced competitive results but did not outperform the standalone TabNet architecture. Despite combining attention-based representation learning with a downstream neural classifier, the hybrid model achieved slightly lower Accuracy, F1-score, and AUROC than TabNet. This outcome suggests that the representations learned by TabNet are already highly informative and that additional nonlinear processing through a DNN provides limited incremental benefit. Consequently, increasing architectural complexity does not necessarily result in improved predictive performance when the feature space has already been effectively optimized.

Collectively, these findings indicate that feature optimization exerts a greater influence on predictive performance than architectural variation, although architectural choice still affects the balance between Precision and Recall. Once irrelevant and redundant attributes were removed through BFA, all evaluated models exhibited substantial improvements, highlighting the importance of feature quality in churn prediction tasks involving structured tabular data.

A comparison with previous machine learning-based studies further demonstrates the effectiveness of the proposed framework. Traditional approaches such as Random Forest, boosting methods, and other conventional classifiers generally rely on fixed feature representations and limited representation learning capability. In contrast, the integration of BFA-based feature optimization with deep learning architectures enables learning from a more informative and less redundant feature space. The resulting performance indicates that BFA-enhanced deep learning models can achieve highly competitive results for customer churn prediction on structured telecommunications datasets.

The experimental evidence therefore suggests that feature optimization contributes more substantially to predictive performance than architectural hybridization in this application domain. Among the evaluated architectures, TabNet delivered the strongest overall performance, whereas the Hybrid TabNet–DNN architecture should be viewed as an exploratory extension whose effectiveness may depend on dataset characteristics and does not necessarily exceed that of a well-optimized standalone TabNet model.

5. Comparison with Previous Studies

To further assess the effectiveness of the proposed BFA-enhanced deep learning framework, the performance of the evaluated models was compared with several existing customer churn prediction studies reported in the literature. The comparison is presented in Table 7, where the best-performing model identified in this study (TabNet and the Hybrid TabNet–DNN) model are evaluated alongside representative machine learning-based approaches.

Table 7. Performance comparison with previous studies

Model	Accuracy	Precision	Recall	Specificity	F1-Score	AUC / AUROC
Reference [17]	0.8694	–	0.8547	0.8839	–	0.95
Reference [27]	0.99	–	–	–	–	–
Reference [28]	0.83	–	–	–	0.75	0.56
TabNet (Ours)	0.96422	0.98558	0.87794	–	0.92865	0.98855
Hybrid TabNet–DNN (Ours)	0.96139	0.96941	0.88223	–	0.92377	0.98181

As shown in Table 7, the proposed BFA-enhanced deep learning models achieve highly competitive performance relative to previously reported churn prediction approaches. In particular, the standalone TabNet model achieved the highest AUROC (0.98855) and the strongest overall balance between Precision, Recall, and F1-score, indicating excellent discriminative capability and robust classification performance. The Hybrid TabNet–DNN model also produced competitive results, achieving an Accuracy of 96.14% and an AUROC of 0.98181.

Compared with the Random Forest-based framework reported in [17], both TabNet and Hybrid TabNet–DNN achieved substantially higher Accuracy and AUROC values. The results suggest that combining feature optimization with deep representation learning provides

a meaningful advantage in modeling complex customer churn behavior. Furthermore, the high Precision values obtained by the proposed models indicate a reduced tendency to generate false churn predictions while maintaining strong churn detection capability.

It should be noted that direct comparison across studies should be interpreted cautiously because differences in datasets, preprocessing procedures, class distributions, and evaluation protocols may influence the reported performance. Nevertheless, the results demonstrate that BFA-enhanced deep learning architectures are highly competitive with existing approaches and represent an effective framework for customer churn prediction on structured telecommunications datasets.

6. Conclusions

This study investigated the impact of Binary Firefly Algorithm (BFA)-based feature optimization on customer churn prediction using multiple deep learning architectures, including DNN, TabNet, and Hybrid TabNet–DNN. The experimental results demonstrate that feature optimization plays a central role in improving predictive performance by reducing feature redundancy and preserving the most informative customer attributes. Across all evaluated models, the incorporation of BFA resulted in substantial improvements in Accuracy, Precision, Recall, F1-score, and AUROC, confirming the importance of feature quality in structured telecommunications datasets. Among the evaluated architectures, TabNet achieved the strongest overall performance, obtaining the highest Accuracy, Precision, F1-score, and AUROC. These results indicate that TabNet is particularly effective for churn prediction on tabular telecommunications data, owing to its ability to combine attention-based feature selection and representation learning within a unified framework. When coupled with BFA-based feature optimization, TabNet was able to exploit the most informative features and generate highly discriminative representations for churn classification.

The Hybrid TabNet–DNN architecture produced competitive results but did not outperform the standalone TabNet model. This finding suggests that increasing architectural complexity does not necessarily lead to improved predictive performance once the feature space has been effectively optimized. The results further indicate that the representations learned by TabNet already capture most of the predictive information contained within the selected feature subset, thereby limiting the additional benefit obtained from subsequent DNN-based processing.

The findings of this study demonstrate that feature optimization contributes more substantially to predictive performance than architectural hybridization in customer churn prediction tasks. More broadly, the results highlight the importance of integrating intelligent feature selection strategies with deep learning models when dealing with high-dimensional structured data. The proposed BFA-enhanced framework achieved highly competitive performance compared with existing churn prediction studies, confirming its effectiveness as a practical approach for telecommunications analytics. Future research may investigate ensemble and stacking-based frameworks that combine optimized deep learning and conventional machine learning models through meta-learning strategies. Additional studies involving larger datasets, cross-domain validation, and advanced explainable artificial intelligence techniques may further improve model robustness, generalization capability, and practical applicability in customer churn management.

Author Contributions: P.O.O., M.M.L., S.S.J., and A.A.; Methodology: S.S.J., and A.A.; Software: A.A.; Validation: P.O.O., M.M.L., S.S.J., and A.A.; Formal analysis: A.A.; Investigation: P.O.O., M.M.L., S.S.J., and A.A.; Resources: P.O.O., M.M.L., S.S.J., and A.A.; Data curation: A.A.; Writing—original draft preparation: S.S.J., and A.A.; Writing—review and editing: P.O.O., M.M.L., S.S.J., and A.A.; Visualization: A.A.; Supervision: P.O.O., and M.M.L.; Project administration: P.O.O., M.M.L., S.S.J., and A.A.; Funding acquisition: P.O.O., M.M.L., S.S.J., and A.A.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset is publicly available online and can be accessed via the link: <https://mavenanalytics.io/data-playground/telecom-customer-churn>.

Acknowledgments: The authors acknowledge the use of generative artificial intelligence tools to assist with language editing, grammar refinement, and manuscript organization. The authors reviewed and validated all generated content and assume full responsibility for the accuracy, originality, and integrity of the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

- [1] S. M. Kostić, M. I. Simić, and M. V. Kostić, "Social Network Analysis and Churn Prediction in Telecommunications Using Graph Theory," *Entropy*, vol. 22, no. 7, p. 753, Jul. 2020, doi: 10.3390/e22070753.
- [2] M. J. Nur, D. R. I. Moses Setiadi, A. A. Ojugo, and M. T. Nguyen, "Improving Customer Churn Prediction Using Domain-Driven Feature Engineering, Resampling, and CatBoost with Explainability Extensions," in *2025 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Sep. 2025, pp. 493–499. doi: 10.1109/ISemantic67418.2025.11291801.
- [3] N. R. Al-Molhem, Y. Rahal, and M. Dakkak, "Social network analysis in Telecom data," *J. Big Data*, vol. 6, no. 1, p. 99, Dec. 2019, doi: 10.1186/s40537-019-0264-6.
- [4] M. Loukili, "Supervised Learning Algorithms for Predicting Customer Churn with Hyperparameter Optimization," *Int. J. Adv. Soft Comput. its Appl.*, vol. 14, no. 3, pp. 50–63, Dec. 2022, doi: 10.15849/IJASCA.221128.04.
- [5] A. Prashanthan, "An Integrated Framework for Optimizing Customer Retention Budget using Clustering, Classification, and Mathematical Optimization," *J. Comput. Theor. Appl.*, vol. 3, no. 1, pp. 45–63, Jul. 2025, doi: 10.62411/jcta.13194.
- [6] T. Xu, Y. Ma, and K. Kim, "Telecom Churn Prediction System Based on Ensemble Learning Using Feature Grouping," *Appl. Sci.*, vol. 11, no. 11, p. 4742, May 2021, doi: 10.3390/app11114742.
- [7] A. Bugajev, R. Kriauzienė, O. Vasilecas, and V. Chadyšas, "The Impact of Churn Labelling Rules on Churn Prediction in Telecommunications," *Informatica*, pp. 247–277, May 2022, doi: 10.15388/22-INFOR484.
- [8] D. K. Sahoo and P. K. Sahoo, "Efficiency, productivity dynamics and determinants of productivity growth in Indian telecommunication industries: An empirical analysis," *J. Public Aff.*, vol. 22, no. 1, Feb. 2022, doi: 10.1002/pa.2353.
- [9] E. Xevelonakis and P. Som, "The impact of social network-based segmentation on customer loyalty in the telecommunication industry," *J. Database Mark. Cust. Strateg. Manag.*, vol. 19, no. 2, pp. 98–106, Jun. 2012, doi: 10.1057/dbm.2012.12.
- [10] R. E. Ako et al., "Effects of Data Resampling on Predicting Customer Churn via a Comparative Tree-based Random Forest and XGBoost," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 86–101, Jun. 2024, doi: 10.62411/jcta.10562.
- [11] Y. Khan, S. Shafiq, A. Naeem, S. Ahmed, N. Safwan, and S. Hussain, "Customers Churn Prediction using Artificial Neural Networks (ANN) in Telecom Industry," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 9, 2019, doi: 10.14569/IJACSA.2019.0100918.
- [12] J. Nagaraju and J. Vijaya, "Boost customer churn prediction in the insurance industry using meta-heuristic models," *Int. J. Inf. Technol.*, vol. 14, no. 5, pp. 2619–2631, Aug. 2022, doi: 10.1007/s41870-022-01017-5.
- [13] V. P. Malikireddy and M. Kasa, "Customer Churns Prediction Model Based on Machine Learning Techniques: A Systematic Review," in *Proceedings of the 3rd International Conference on Integrated Intelligent Computing Communication & Security (ICIIC 2021)*, 2021. doi: 10.2991/ahis.k.210913.021.
- [14] 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, pp. 185–198, Jan. 2022, doi: 10.18576/isl/110120.
- [15] S. Ö. Arik and T. Pfister, "TabNet: Attentive Interpretable Tabular Learning," *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 8, pp. 6679–6687, May 2021, doi: 10.1609/aaai.v35i8.16826.
- [16] S. I. Ele, U. R. Alo, H. F. Nweke, and O. A. Ofem, "Regression-Based Machine Learning Framework for Customer Churn Prediction in Telecommunication Industry," *J. Adv. Inf. Technol.*, vol. 14, no. 5, pp. 1046–1055, 2023, doi: 10.12720/jait.14.5.1046-1055.
- [17] V. Chang, K. Hall, Q. Xu, F. Amao, M. Ganatra, and V. Benson, "Prediction of Customer Churn Behavior in the Telecommunication Industry Using Machine Learning Models," *Algorithms*, vol. 17, no. 6, p. 231, May 2024, doi: 10.3390/a17060231.
- [18] P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi, "Customer churn prediction system: a machine learning approach," *Computing*, vol. 104, no. 2, pp. 271–294, Feb. 2022, doi: 10.1007/s00607-021-00908-y.
- [19] D. Asif, M. S. Arif, and A. Mukheimer, "A data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry," *Results Eng.*, vol. 26, p. 104629, Jun. 2025, doi: 10.1016/j.rineng.2025.104629.
- [20] S. S. Poudel, S. Pokharel, and M. Timilsina, "Explaining customer churn prediction in telecom industry using tabular machine learning models," *Mach. Learn. with Appl.*, vol. 17, p. 100567, Sep. 2024, doi: 10.1016/j.mlwa.2024.100567.
- [21] A. Prashanthan, R. Roshan, and M. Maduranga, "RetenNet: A Deployable Machine Learning Pipeline with Explainable AI and Prescriptive Optimization for Customer Churn Management," *J. Futur. Artif. Intell. Technol.*, vol. 2, no. 2, pp. 182–201, Jun. 2025, doi: 10.62411/faith.3048-3719-110.
- [22] P. Ndarishye and A. K. Singh, "Understanding Customer Churn in Retail Banking through Explainable Predictive Analytics: Evidence of a Product Paradox," *J. Comput. Theor. Appl.*, vol. 3, no. 4, pp. 473–486, Apr. 2026, doi: 10.62411/jcta.15870.
- [23] S. K. Wagh, A. A. Andhale, K. S. Wagh, J. R. Pansare, S. P. Ambadekar, and S. H. Gawande, "Customer churn prediction in telecom sector using machine learning techniques," *Results Control Optim.*, vol. 14, p. 100342, Mar. 2024, doi: 10.1016/j.rico.2023.100342.
- [24] S. Saha, C. Saha, M. M. Haque, M. G. R. Alam, and A. Talukder, "ChurnNet: Deep Learning Enhanced Customer Churn Prediction in Telecommunication Industry," *IEEE Access*, vol. 12, pp. 4471–4484, 2024, doi: 10.1109/ACCESS.2024.3349950.
- [25] Deepika Kumari, Santosh Kumar Singh, Sanjay Subhash Katira, Inumarthi V Srinivas, and Uday Salunkhe, "Telecom Customer Churn Forecasting Using Machine Learning: A Data-Driven Predictive Framework," *Metall. Mater. Eng.*, vol. 31, no. 4, pp. 922–929, Apr. 2025, doi: 10.63278/1536.

-
- [26] K. Mourad and R. Boudour, "A Modified Binary Firefly Algorithm to Solve Hardware/Software Partitioning Problem," *Informatica*, vol. 45, no. 7, Nov. 2021, doi: 10.31449/inf.v45i7.3408.
 - [27] A. G. Toprak, Ö. Berfin Mercan, and M. S. Osmanca, "ML-Based Telecom Customer Churn Analysis," in *2025 33rd Signal Processing and Communications Applications Conference (SIU)*, Jun. 2025, pp. 1–4. doi: 10.1109/SIU66497.2025.11111814.
 - [28] S. Saleh and S. Saha, "Customer retention and churn prediction in the telecommunication industry: a case study on a Danish university," *SN Appl. Sci.*, vol. 5, no. 7, p. 173, Jul. 2023, doi: 10.1007/s42452-023-05389-6.