Scientific Journal of Computer Science
futuristech PT. Teknologi Futuristik Indonesia
Received: May 10, 2026; Revised: July 30, 2026; Accepted: August 3, 2026; Published: August 5, 2026; Current version: August 5, 2026. *e-ISSN: 3110-3170*
Digital Object Identifier 10.64539/sjcs.v2i2.2026.507

Article

A Binary Firefly Optimized Stacking Ensemble Model for Customer Churn Prediction in the Telecommunications Industry

Abdulrashid Abdulrauf[1,*], Maaruf Mohammed Lawal[2], Oluwatoyin Omoloba[3], Omolara Busayo Abodunrin[3], Minkail Mohammad[4]

[1] Department of Computer Science, Federal Polytechnic Kaltungo, Gombe State, Nigeria; e-mail: abdulrashid.abdulrauf2022@nda.edu.ng (A. Abdulrauf).
[2] Department of Computer Science, Ahmadu Bello University, Zaria, Kaduna State, Nigeria; e-mail: mmlawal80@gmail.com (M. M. Lawal).
[3] Department of Computer Science, Federal School of Surveying Oyo, Oyo State, Nigeria; e-mail: oluwatoyin.omoloba@gmail.com (O. Omoloba), busayomi5@gmail.com (O. B. Abodunrin).
[4] Department of Mathematics and Statistics, Loyola University Chicago, Illinois State, United States; e-mail: mmuhammad2@luc.edu (M. Mohammad).
* Correspondence

The authors received no financial support for the research, authorship, and/or publication of this article.

Abstract: Customer churn prediction remains a critical challenge in the telecommunications industry because subscriber attrition directly affects revenue generation, customer retention, and long-term business sustainability. Although numerous studies have employed machine learning and deep learning techniques for churn prediction, many existing approaches rely on standalone models that have limited ability to simultaneously capture complex nonlinear customer behaviour, exploit model diversity, and eliminate redundant features. Building upon our previous study, this research proposes an enhanced heterogeneous stacking ensemble framework that integrates Binary Firefly Algorithm (BFA)-based feature optimization with Deep Neural Network (DNN), Support Vector Machine (SVM), and Random Forest (RF) as base learners, while Logistic Regression serves as the meta-learner for final churn classification. To ensure a fair and controlled comparison, the same Maven Analytics Telecom Customer Churn dataset and preprocessing strategy adopted in the previous study were retained, including data cleaning, feature transformation, stratified data partitioning, and normalization. Model development further incorporated 5-fold cross-validation on the training dataset, while BFA was introduced to identify the most informative pre-churn features. Experimental results demonstrate that the proposed framework achieved strong and balanced predictive performance and demonstrated improvements in several classification metrics compared with the previously developed BFA-based Hybrid TabNet-DNN model. Furthermore, its performance was competitive with the Random Forest baseline, which exhibited comparable classification effectiveness. These findings show that optimized feature selection and heterogeneous ensemble learning improve prediction stability and generalization. The proposed framework provides an effective decision-support tool for proactive customer retention in the telecommunications industry.

Keywords: Churn; Binary; Ensemble; Firefly; Prediction; Stacking.

Copyright: © 2026 by the authors. This is an open-access article under the CC-BY-SA license.

1. Introduction

Customer retention has become one of the most important determinants of profitability and long-term sustainability in the modern telecommunications industry. Rapid technological advancements, increasing market competition, widespread smartphone adoption, and the availability of multiple service providers have considerably lowered customer switching barriers, making customer churn a persistent challenge for telecommunications operators worldwide [1]–[5]. Customer churn, defined as the discontinuation of services by subscribers, directly affects organizational revenue, market share, and customer acquisition costs. Since acquiring new customers is considerably more expensive than retaining existing ones, accurately identifying customers with high churn propensity has become a strategic priority for telecommunications companies seeking to improve customer loyalty and operational efficiency [2], [3].

To address this challenge, predictive analytics has emerged as an essential component of modern customer relationship management. By combining statistical analysis, data mining, and machine learning techniques, predictive models can uncover hidden behavioral patterns from large volumes of customer data and estimate the likelihood of customer attrition before it occurs [6], [7]. Such predictive capabilities enable telecommunications operators to implement proactive retention strategies, including personalized promotions, improved customer support, targeted marketing campaigns, and loyalty incentives for customers identified as being at risk of leaving [8]. Consequently, customer churn prediction has become one of the most extensively studied applications of machine learning, with researchers exploiting demographic characteristics, subscription information, billing history, service utilization patterns, and network quality indicators to improve prediction accuracy [9]. Despite these advances, accurately modelling customer churn remains challenging because customer behavior is inherently dynamic, nonlinear, and influenced by numerous interacting factors. Furthermore, issues such as redundant features, noisy data, high-dimensional feature spaces, and varying data quality continue to limit the effectiveness and generalization ability of many prediction models [10].

Existing customer churn prediction studies have investigated a wide range of machine learning and deep learning techniques, including Logistic Regression, Decision Trees, Random Forest, Support Vector Machine (SVM), Artificial Neural Networks, Deep Neural Networks (DNN), and gradient boosting algorithms [11]. While these approaches have achieved promising predictive performance, individual learning algorithms often exhibit inherent limitations. Traditional machine learning models may struggle to capture complex nonlinear relationships, whereas deep learning models generally require large volumes of data and careful hyperparameter optimization to achieve stable performance. Similarly, tree-based ensemble methods provide strong predictive capability but remain constrained by the learning characteristics of a single algorithm. These limitations suggest that relying on a single predictive model may not fully exploit the complementary strengths of different learning paradigms, thereby motivating the development of heterogeneous ensemble learning approaches.

Among ensemble learning techniques, stacking has emerged as an effective strategy for improving predictive performance by combining the outputs of multiple diverse learning algorithms through a meta-learner. Unlike bagging and boosting methods, which rely on homogeneous ensembles or sequential learning, stacking integrates heterogeneous classifiers capable of learning different representations of the same dataset. By combining complementary decision boundaries generated by diverse base learners, stacking often produces superior prediction accuracy, enhanced robustness, and better generalization than individual models [12]–[14]. Nevertheless, the effectiveness of stacking ensembles largely depends on the quality of the input features provided to the constituent models. High-dimensional datasets containing irrelevant or redundant attributes may still degrade predictive performance and increase computational complexity, making feature selection an important component of ensemble-based churn prediction.

Metaheuristic optimization algorithms have therefore attracted considerable attention for identifying informative feature subsets capable of improving classification performance while reducing model complexity. Among these techniques, the Binary Firefly Algorithm (BFA) has demonstrated considerable effectiveness in selecting discriminative features from high-dimensional datasets through an intelligent population-based search mechanism. In our previous work, Odion *et al*. [15], the Binary Firefly Algorithm was successfully employed to optimize standalone TabNet and Deep Neural Network models for customer churn prediction, demonstrating that feature optimization significantly improved classification performance by eliminating redundant and irrelevant variables. Although these findings established the effectiveness of BFA for optimizing individual classifiers, the study did not investigate whether the optimized feature subsets could further enhance predictive performance when integrated within a heterogeneous ensemble learning architecture. Consequently, the applicability of Binary Firefly optimization beyond standalone classifiers remains largely unexplored.

This study extends our previous research by investigating the integration of Binary Firefly-based feature optimization within a heterogeneous stacking ensemble learning framework for customer churn prediction. Rather than optimizing individual classifiers, the proposed framework combines the complementary strengths of a Deep Neural Network (DNN), Support Vector Machine (SVM), and Random Forest (RF) as base learners, while Logistic Regression serves as the meta-learner responsible for generating the final prediction. The DNN is employed to learn complex nonlinear feature representations, SVM provides strong generalization through optimal margin separation in high-dimensional feature spaces, and Random Forest contributes robustness against noise and overfitting through ensemble decision trees. Logistic Regression is selected as the meta-learner because of its computational efficiency, interpretability, and effectiveness in learning how to combine the outputs of heterogeneous classifiers into a unified prediction. By integrating these complementary learning paradigms after Binary Firefly-based feature optimization, the proposed framework seeks to achieve improved predictive accuracy, robustness, and generalization beyond what can be obtained using standalone learning algorithms.

To provide a comprehensive performance evaluation, the proposed stacking ensemble is compared with four baseline models, namely DNN, SVM, Random Forest, and CatBoost, all trained using the Binary Firefly-selected feature subset. CatBoost is included as a benchmark because of its strong performance on structured datasets containing categorical variables, although it is not incorporated into the stacking architecture. Comparative evaluation against these representative machine learning and deep learning models enables a comprehensive assessment of the effectiveness of the proposed ensemble framework.

The experimental evaluation utilizes the publicly available Maven Analytics Telecom Customer Churn dataset, which has been widely adopted as a benchmark dataset for churn prediction research [12]–[14]. The dataset contains demographic information, customer tenure, service subscriptions, internet usage characteristics, billing information, payment methods, revenue-related variables, and customer churn status, providing a comprehensive representation of customer behavior suitable for predictive modelling. Although publicly available benchmark datasets cannot fully capture the operational characteristics of Nigerian telecommunications providers, they provide a reliable basis for evaluating and comparing advanced machine learning models under standardized experimental conditions.

The principal contributions of this study are summarized as follows:

i. A Binary Firefly Algorithm-based feature optimization framework is integrated with a heterogeneous stacking ensemble rather than optimizing standalone classifiers.
ii. A novel stacking architecture consisting of DNN, SVM, and Random Forest as base learners with Logistic Regression as the meta-learner is developed for customer churn prediction.
iii. A comprehensive comparative evaluation is conducted against standalone DNN, SVM, Random Forest, and CatBoost models using multiple performance metrics.
iv. The study demonstrates that Binary Firefly-optimized heterogeneous ensemble learning improves predictive robustness, generalization, and classification performance beyond individual machine learning models, thereby extending the findings reported in Odion *et al*. [15].

2. Review of Related Works

2.1. Traditional Machine Learning Approaches for Customer Churn Prediction

Customer churn prediction has been extensively investigated using traditional machine learning algorithms because of their simplicity, interpretability, and relatively low computational requirements. Ele *et al*. [16] proposed a regression-based machine learning framework for telecom customer churn prediction by evaluating nine regression algorithms using MAE, MSE, RMSE, and coefficient of determination (R^2). Among the evaluated models, Lasso Regression achieved the best performance with an MAE of 7.77E−02, MSE of 1.21E−02, RMSE of 1.10E−01, and an R^2 value of 0.9814, demonstrating strong predictive capability. Similarly, Summaila and Stephen [17] developed a hybrid analytical framework that combined Principal Component Analysis (PCA), K-Means Clustering, and Logistic Regression to investigate subscriber satisfaction and churn propensity among Nigerian GSM operators. Using Quality of Service (QoS) data collected from MTN, Airtel, Globacom, and 9mobile, the proposed framework achieved an overall prediction accuracy of 95%, highlighting the effectiveness of combining unsupervised and supervised learning techniques for churn prediction.

Although these conventional machine learning algorithms have demonstrated satisfactory predictive performance, they often struggle to model the complex nonlinear interactions and high-dimensional feature relationships commonly found in customer behavioral data. Consequently, researchers have increasingly shifted their attention toward deep learning methods capable of learning more expressive feature representations.

2.2. Deep Learning-Based Customer Churn Prediction

Deep learning techniques have gained considerable attention because of their ability to automatically extract hierarchical feature representations and model complex nonlinear relationships within customer data. Fujo *et al*. [18] proposed a Deep Backpropagation Artificial Neural Network (Deep-BP-ANN) framework incorporating Variance Thresholding, Lasso Regression-based feature selection, Random Oversampling, and early stopping to improve telecom customer churn prediction. Using IBM Telco and Cell2Cell datasets, the proposed framework achieved prediction accuracies of 88.12% and 79.38%, respectively, while 10-fold cross-validation further confirmed its robustness with an average accuracy of 86.57%.

Similarly, Saha *et al*. [19] introduced ChurnNet, a deep learning architecture integrating one-dimensional convolutional layers, residual learning, squeeze-and-excitation blocks, and spatial attention mechanisms. By combining SMOTE-based data balancing with 10-fold cross-validation, ChurnNet achieved prediction accuracies of 95.59%, 96.94%, and 97.52% across benchmark datasets, outperforming several existing deep learning approaches. Kumar *et al*. [20] further demonstrated the effectiveness of sequential deep learning by integrating Long Short-Term Memory (LSTM) networks with conventional classifiers such as Random Forest, Decision Trees, Support Vector Machines, AdaBoost, and Gradient Boosting. Their findings showed that the LSTM model achieved the highest testing accuracy of 89%, confirming the ability of recurrent neural networks to capture temporal dependencies in customer behavioral patterns.

Despite these encouraging results, deep learning models generally require careful parameter optimization, are computationally intensive, and may exhibit overfitting when trained on relatively small datasets. These limitations have motivated the exploration of ensemble learning techniques capable of combining the strengths of multiple predictive models.

2.3. Ensemble Learning and Stacking Approaches

Ensemble learning has become one of the most successful strategies for improving customer churn prediction by combining multiple classifiers to achieve higher predictive accuracy, robustness, and generalization. Chang *et al.* [12] evaluated Decision Trees, Boosted Trees, and Random Forest for telecom customer churn prediction and reported that Random Forest achieved the best performance with an accuracy of 91.66%, precision of 82.2%, and recall of 81.8%. The study further enhanced model transparency through SHAP and LIME explainability techniques.

Similarly, Lalwani *et al*. [21] employed the Gravitational Search Algorithm for feature selection before evaluating several machine learning algorithms, including Logistic Regression, Support Vector Machine, Random Forest, XGBoost, AdaBoost, and CatBoost. Their findings indicated that CatBoost produced the best predictive performance with an accuracy of 81.8%, precision of 81.2%, and recall of 82.2%. Hossain [22] compared sixteen machine learning algorithms using the Iranian Telco dataset and reported that CatBoost achieved the highest accuracy of 97.54%, while LightGBM attained an average cross-validation accuracy of 96.39%.

Recent research has increasingly focused on hybrid ensemble models. Asif *et al*. [23] proposed XAI-Churn TriBoost, an explainable ensemble integrating XGBoost, CatBoost, and LightGBM using a soft voting strategy. The proposed framework achieved an accuracy of 96.44%, precision of 92.82%, recall of 87.82%, and an F1-score of 90.25%, while SHAP and LIME improved model interpretability. Likewise, Beeharry *et al*. [24] introduced a two-layer flexible voting ensemble that significantly improved churn prediction on both balanced and imbalanced telecom datasets, demonstrating statistically significant improvements over individual classifiers.

Although these ensemble approaches consistently outperform standalone classifiers, most existing studies rely on voting or boosting strategies. Comparatively fewer studies have investigated heterogeneous stacking ensembles that integrate deep learning and traditional machine learning models through a dedicated meta-learner, particularly when combined with intelligent feature optimization techniques

2.4. Feature Selection and Metaheuristic Optimization

Feature selection plays a critical role in customer churn prediction because high-dimensional datasets often contain redundant, noisy, or irrelevant attributes that degrade model performance and increase computational complexity. Metaheuristic optimization algorithms have therefore become increasingly popular for identifying informative feature subsets.

Lalwani *et al*. [21] demonstrated that feature optimization using the Gravitational Search Algorithm improved classification performance across several machine learning models. Similarly, explainable machine learning approaches proposed by Poudel *et al*. [25] showed that identifying influential customer attributes not only improved predictive performance but also enhanced model interpretability through SHAP-based feature analysis. RetenNet, proposed by Prashanthan *et al*. [26], further integrated optimization strategies with explainable artificial intelligence, fuzzy clustering, and predictive modelling, demonstrating that optimized learning frameworks can significantly improve customer retention analysis.

In our previous work, Odion *et al*. [15] employed the Binary Firefly Algorithm (BFA) to optimize feature selection for standalone TabNet and Deep Neural Network models for customer churn prediction. The study demonstrated that Binary Firefly-based optimization substantially improved predictive performance by removing redundant and irrelevant features while enhancing model generalization. However, the investigation was limited to optimizing individual classifiers and did not explore whether Binary Firefly-selected feature subsets could further improve heterogeneous ensemble learning frameworks.

2.5. Research Gap

Recent comparative studies continue to demonstrate that no single machine learning algorithm consistently outperforms all others across different telecommunications datasets. Kumari *et al*. [27] reported that deep learning achieved the highest predictive performance among several machine learning algorithms, whereas Logistic Regression remained the most interpretable model. Toprak *et al*. [13] evaluated sixteen classification algorithms using IBM Telco, Orange Telecom, and Maven Analytics datasets and observed substantial variations in model

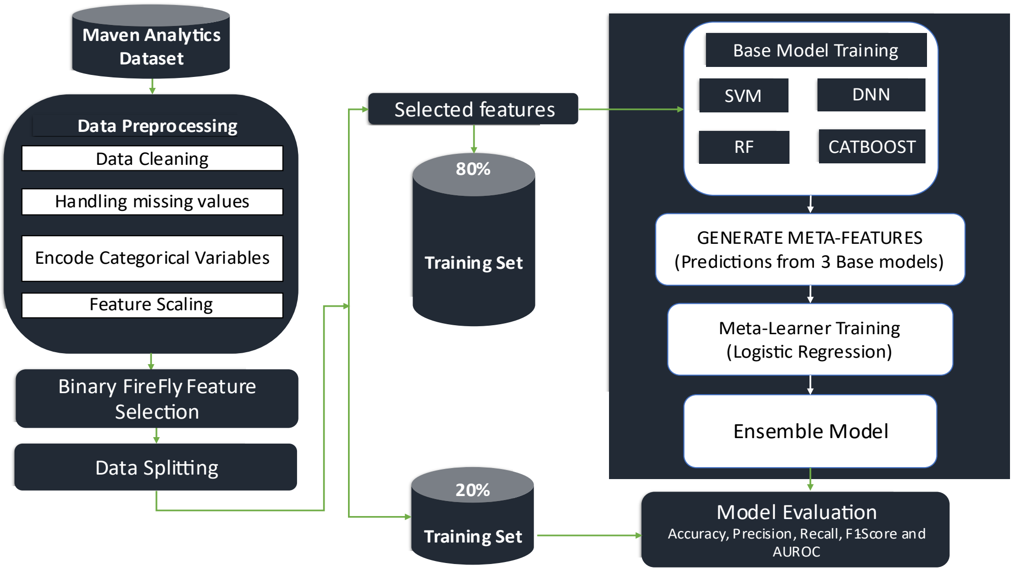

Figure 1. Proposed Process Framework.

performance across datasets. Similarly, Bansal *et al*. [28] showed that integrating ensemble learning with SMOTE balancing improved predictive accuracy to 97%, while Al-Masri *et al*. [29] highlighted that even highly accurate models remain susceptible to overfitting without appropriate preprocessing and optimization.

Collectively, the existing literature demonstrates that significant progress has been achieved through traditional machine learning, deep learning, feature optimization, and ensemble learning techniques. Nevertheless, an important research gap remains. Most previous studies have focused either on optimizing standalone classifiers or developing ensemble methods without incorporating intelligent feature optimization into heterogeneous stacking architectures. Likewise, although Binary Firefly Algorithm has demonstrated considerable effectiveness for optimizing individual customer churn prediction models [15], its integration within a heterogeneous stacking ensemble consisting of Deep Neural Network, Support Vector Machine, and Random Forest with Logistic Regression as a meta-learner has received little attention.

To address this gap, the present study proposes a Binary Firefly-optimized heterogeneous stacking ensemble that combines the complementary strengths of deep learning and conventional machine learning models. Unlike previous studies, the proposed framework simultaneously exploits intelligent feature optimization and meta-learning to improve prediction accuracy, robustness, and generalization for customer churn prediction in the telecommunications industry.

3. Method and Materials

The proposed framework consists of six sequential stages. First, the Telecom Customer Churn dataset is collected and preprocessed through data cleaning, encoding, normalization, and binary target transformation. Second, Binary Firefly Algorithm-based feature optimization is performed to identify the most informative feature subset. Third, the optimized features are supplied independently to the three heterogeneous base learners (DNN, SVM, and Random Forest). Fourth, each base learner generates probability predictions through five-fold cross-validation. Fifth, these prediction probabilities are concatenated to form a stacked feature matrix that serves as the input to the Logistic Regression meta-learner. Finally, the performance of the proposed stacking ensemble is evaluated and compared with standalone DNN, SVM, Random Forest, and CatBoost models. Details of the Proposed Process Framework are shown in Figure 1.

3.1. Dataset

To facilitate a consistent and fair evaluation with our previous study [15], the same Telecom Customer Churn dataset was adopted in this research. Maintaining an identical benchmark dataset enables a direct comparison of the proposed framework with the previous approach while ensuring that any observed performance improvements are primarily attributed to the methodological advancement, particularly the integration of Binary Firefly-based feature optimization with heterogeneous stacking ensemble learning, rather than variations in dataset characteristics.

The Maven Analytics platform dataset comprises 7,043 customer records described by 38 attributes, encompassing demographic information, geographical location, customer tenure, service subscriptions, internet usage characteristics, billing and payment information, revenue-related variables, and customer status. Representative features include customer age, gender, marital status, number of dependents, city, contract type, payment method, internet service, online security, monthly charges, total charges, total revenue, and customer churn status.

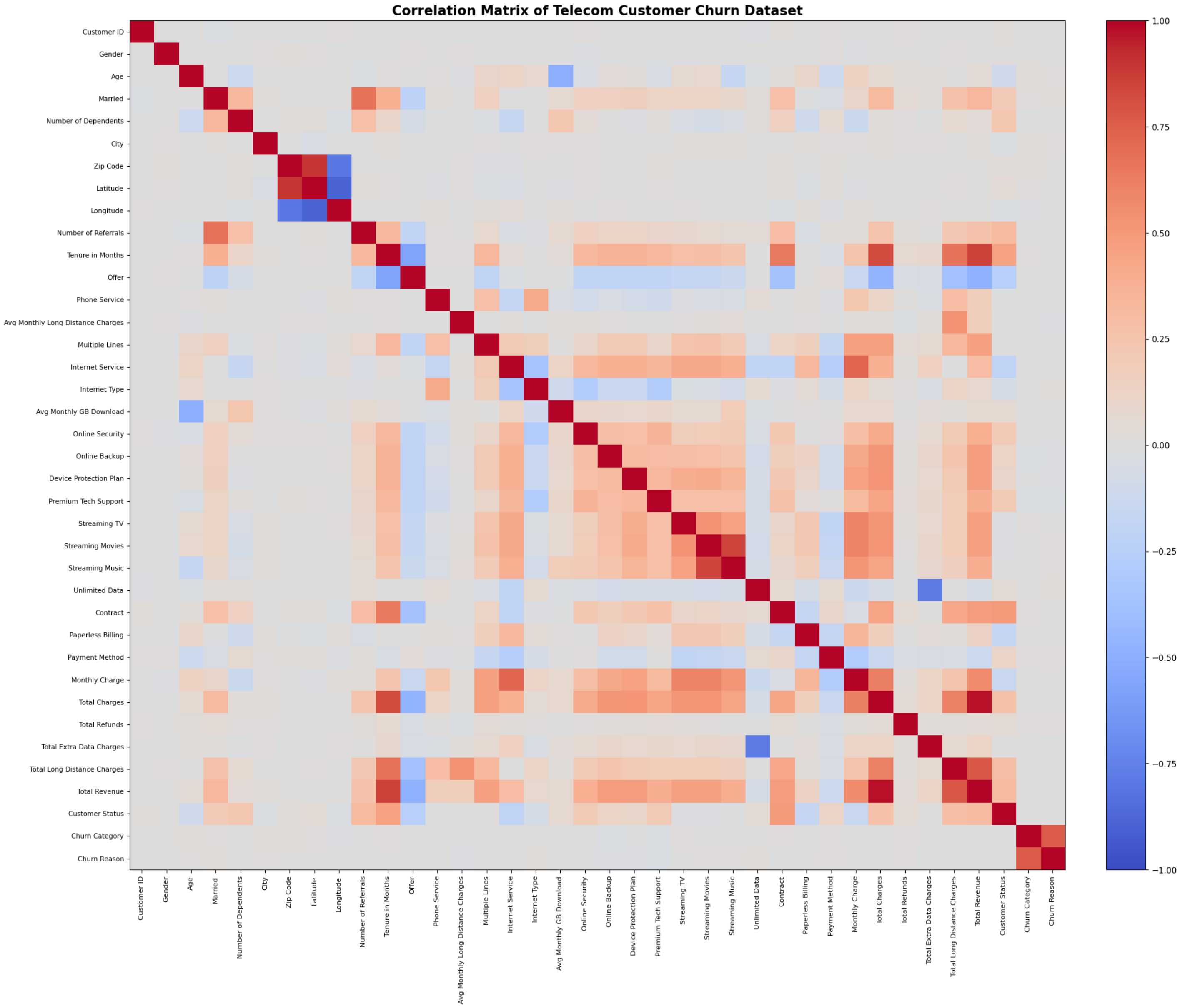

Figure 2. Correlation Heatmap.

Collectively, these variables provide a rich description of customer behavior and service utilization patterns, making the dataset suitable for developing and evaluating intelligent customer churn prediction models.

3.2. Data Preparation

To ensure a controlled and consistent comparison with our previous study [15], the data preparation procedure adopted in this research follows the same fundamental preprocessing pipeline established in the earlier work. Maintaining an identical preprocessing strategy minimizes variations arising from differences in data preparation and enables the performance improvements observed in this study to be attributed primarily to the proposed methodological advancements, particularly the integration of Binary Firefly Algorithm (BFA)-based feature optimization and the heterogeneous stacking ensemble framework. The preprocessing procedure involved multiple stages, including data cleaning, categorical feature encoding, target variable transformation, dataset partitioning, numerical feature normalization, and feature optimization, to ensure data quality, consistency, and compatibility with the developed machine learning and deep learning models.

During the data cleaning stage, the dataset was examined for missing values, inconsistent entries, and duplicate records. Missing numerical attributes were handled using median imputation, whereas missing categorical attributes were replaced using the corresponding mode values. Duplicate observations were removed to prevent redundancy and reduce potential bias during model training. Subsequently, categorical variables representing customer demographics, service usage characteristics, subscription information, and billing-related attributes were transformed into numerical representations using label encoding to enable processing by machine learning and deep learning algorithms.

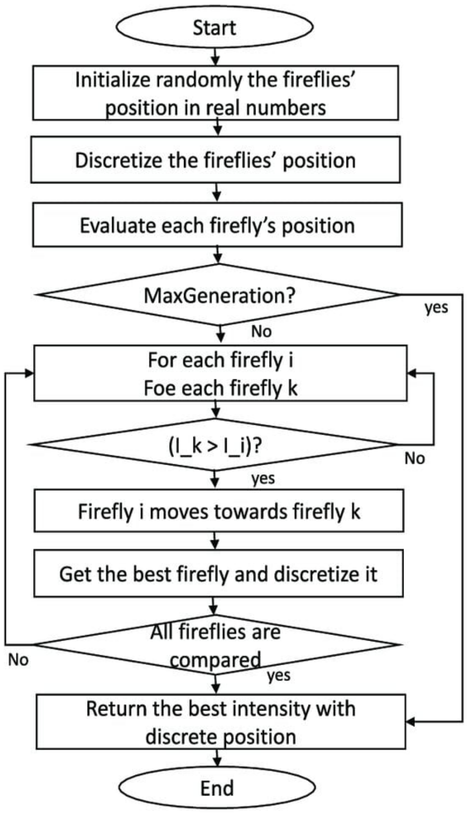

Figure 3. Flowchart of firefly algorithm [30].

Algorithm 1. Pseudocode for Binary Firefly Algorithm.

1. *Initialize the parameters:*
   - *NP - Population size*
   - *D - Number of features (problem dimension)*
   - *Max_gen - Maximum number of iterations*
2. *Define the objective function $f(x)$, where $x = (x_1, x_2, \dots, x_D)$ represents a candidate feature subset*
3. *Generate the initial population of fireflies X_i $(i = 1,2, \dots, NP)$*
4. *Define light intensity I_i for each firefly X_i*
5. *Set the light absorption coefficient*
6. *Set initial attractiveness*
7. *Set iteration counter $t = 0$*
8. *While $t < Max_{gen}$ do*
   1. *For each firefly $i = 1$ to NP:*
      1. *For each firefly $j = 1$ to NP:*
         If $I_j > I_i$, move firefly i toward firefly j
         Update attractiveness and light intensity i
         Evaluate new solutions using the objective function
         Update the light intensity I_i
      2. *End For*
   2. *End For*
   3. *Rank all fireflies and identify the current best solution*
   4. *Increment $t = t + 1$*
9. *End While*
10. *Display the best feature subset*

The original multi-class Customer Status variable, consisting of Stayed, Joined, and Churned categories, was transformed into a binary classification format. The Stayed and Joined categories were combined and encoded as 0 (non-churn), while Churned customers were encoded as 1 (churn), thereby formulating the problem as a binary classification task consistent with customer churn prediction objectives. Following these preprocessing operations, the dataset was partitioned into training and testing subsets, after which numerical feature normalization was performed using parameters estimated exclusively from the training data.

Figure 2 presents the Pearson correlation matrix of the Telecom Customer Churn dataset. The heatmap provides an overview of the relationships among the predictor variables and the target-related attributes. Most variables exhibit weak-to-moderate pairwise correlations, indicating limited multicollinearity among the predictors. Highly correlated attributes primarily occur among service usage and billing-related variables, which is expected due to their underlying business relationships. The correlation analysis also served as an initial exploratory assessment prior to feature optimization using the BFA.

3.3. Binary Firefly-Based Feature Optimization

The Binary Firefly Algorithm (BFA) was employed as a feature optimization technique to identify the most informative attributes for customer churn prediction while reducing irrelevant and redundant features. Unlike the preprocessing operations applied to the complete dataset, the BFA feature selection process was performed after the train-test partitioning and normalization stages, using only the training dataset to prevent information leakage from the unseen test data. This procedure ensures that feature selection is learned exclusively from the training samples and that the selected feature subset provides an unbiased evaluation of model generalization capability.

Following the approach adopted by Mourad and Boudour [31], the BFA represents each candidate feature subset as a binary vector, where each dimension corresponds to a dataset feature and takes a value of 1 or 0, indicating whether the feature is selected or excluded, respectively. Each candidate solution is evaluated using a fitness function based on classification performance, allowing the algorithm to identify feature combinations that maximize predictive effectiveness while minimizing unnecessary attributes. Through iterative movement and attractiveness updates among firefly solutions, the algorithm explores the feature search space and progressively converges toward an optimal subset of relevant features.

The final feature subset identified by the BFA was subsequently applied consistently to both the training and testing datasets. This approach reduced feature dimensionality, eliminated less informative attributes, improved computational efficiency, and enhanced the ability of the proposed models to learn meaningful customer churn patterns. The steps of the BFA are shown in Algorithm 1 and Figure 3.

At the completion of the optimization process, the algorithm generated an optimal subset of features, which was subsequently used to transform both the training and test datasets. The selected features were then utilized for training the predictive models, including the DNN. This feature selection approach enhanced computational efficiency, reduced dataset dimensionality, and minimized the impact of irrelevant or noisy attributes, thereby improving model performance, stability, and generalization capability.

3.4. Data Splitting

Following the data partitioning strategy adopted in our previous study [15], the preprocessed dataset was divided into training and testing subsets using an 80:20 stratified split to preserve the original distribution of churn and non-churn classes across both subsets. Maintaining the same splitting strategy ensures a fair experimental comparison and allows the observed performance differences to be attributed to the proposed methodological improvements rather than variations in data partitioning.

After the data split, numerical features were standardized using Z-score normalization. The mean and standard deviation required for scaling were computed exclusively from the training dataset and subsequently applied to both the training and testing datasets. This procedure prevented information leakage and ensured an unbiased assessment of model generalization.

The training dataset was then used for model development, where 5-fold cross-validation was performed exclusively on the training subset to improve model reliability, facilitate parameter optimization, and reduce the likelihood of overfitting. The independent testing subset was not involved in the cross-validation process and was reserved solely for the final performance evaluation of the trained models. Following normalization, the BFA was applied exclusively to the training dataset to identify the most informative feature subset while avoiding any influence from unseen test samples. The selected features were subsequently applied to both the training and testing datasets before model development and final evaluation.

3.5. Model Development

Following data preprocessing, feature optimization using the BFA, and dataset partitioning, the predictive models were developed using the optimized feature subset. Unlike the previous study [15], which focused on individual predictive models using the original feature representation, the current research introduces a heterogeneous stacking ensemble framework that integrates multiple learning paradigms to enhance predictive capability. The model development process consisted of two major stages: (i) construction of the proposed stacking ensemble model and (ii) independent training of baseline models for comparative evaluation.

3.5.1. Construction of the Proposed Stacking Ensemble Framework

The proposed framework employs a heterogeneous stacking ensemble strategy that combines the complementary strengths of Deep Neural Network (DNN), Support Vector Machine (SVM), and Random Forest (RF) as base learners, with Logistic Regression serving as the meta-learner. The optimized feature subset generated by the BFA was used as input to each base learner, enabling the models to focus on the most relevant customer attributes while reducing the influence of redundant features.

During the stacking process, each base learner was independently trained using the optimized training dataset. The probability outputs generated by the DNN, SVM, and RF models were subsequently concatenated to form a new feature representation, referred to as the meta-feature matrix. This matrix was then provided as input to the Logistic Regression meta-learner, which learned the optimal combination of predictions from the individual classifiers to generate the final churn classification output.

The motivation behind this architecture is based on the complementary characteristics of the selected algorithms. The DNN provides strong nonlinear representation learning capabilities by automatically capturing complex relationships among customer attributes. The SVM contributes effective decision boundary optimization and generalization capability, particularly in high-dimensional feature spaces. Meanwhile, RF improves model robustness through ensemble-based decision aggregation and its ability to capture nonlinear interactions among variables. By integrating these diverse learning mechanisms, the proposed stacking framework aims to achieve improved predictive accuracy, stability, and generalization compared with standalone classifiers.

Logistic Regression was selected as the meta-learner due to its computational efficiency, interpretability, and effectiveness in combining heterogeneous model predictions while minimizing additional model complexity. The configurations of proposed models are shown in Table 1.

3.5.2. Baseline Model

To evaluate the effectiveness of the proposed stacking framework, four standalone models, namely DNN, SVM, Random Forest, and CatBoost, were developed as benchmark classifiers. These models were selected because they represent different machine learning paradigms and have demonstrated effectiveness in customer churn prediction studies.

The DNN model was implemented to learn complex nonlinear patterns within customer behavioural data through multiple hidden layers, ReLU activation functions, and dropout regularization to reduce overfitting. The SVM classifier was configured with a radial basis function (RBF) kernel and probability estimation enabled, allowing effective classification in nonlinear feature

Table 1. Stacking Ensemble Configuration.

Component	Parameter / Setting	Description
Ensemble Type	Stacking Ensemble Learning	Combines multiple base learners using a meta-learner
Base Learners	DNN, SVM, Random Forest	Models used for initial predictions
Meta-Learner	Logistic Regression	Combines outputs from base learners
Cross-Validation	5-Fold Cross-Validation	Improves robustness and reduces overfitting
Stacking Method	Column-wise Probability Concatenation	Combines base learner outputs
Final Prediction Model	Logistic Regression	Generates final churn prediction

Table 2. Configuration of Baseline Classification Models.

Model	Key Parameters	Description
DNN	3 hidden layers (128–64–32), ReLU, Dropout (0.3), Adam optimizer, 100 epochs, batch size 32	Learns complex nonlinear patterns using deep feature representations
SVM	RBF kernel, probability estimation enabled	Performs high-dimensional classification using optimal hyperplanes
Random Forest	200 trees, bagging ensemble	Improves stability and reduces overfitting through multiple decision trees
CatBoost	Gradient boosting, random state = 42, verbose = 0	Handles categorical data efficiently and reduces prediction bias

spaces. Random Forest was employed due to its robustness, ability to handle complex feature interactions, and reduced susceptibility to overfitting through ensemble aggregation of multiple decision trees. CatBoost was included as an additional benchmark model because of its strong performance in structured datasets and its capability to efficiently handle categorical variables through gradient boosting optimization.

All baseline models were trained using the BFA-selected feature subset to maintain consistency with the proposed stacking framework and ensure that performance differences were primarily influenced by the learning strategies rather than variations in feature representation. The configurations of the baseline models are presented in Table 2.

3.6. Performance Evaluation

To ensure consistency and facilitate a fair comparison with our previous study [15], the same evaluation metrics adopted in the earlier work were considered for assessing the performance of the proposed churn prediction framework. Since the problem is formulated as a binary classification task involving churned and non-churned customers, the models were evaluated using Accuracy, Precision, Recall, F1-Score, and Area Under the Receiver Operating Characteristic Curve (AUROC).

Accuracy represents the overall classification effectiveness of a model by measuring the ratio of correctly predicted instances to the total number of evaluated samples. It is calculated as:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

Precision represents the ratio of true positive predictions to the total number of predicted positive instances and is calculated as:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall (also referred to as sensitivity or true positive rate) evaluates a model's ability to correctly identify actual positive instances by measuring the proportion of true positive predictions among all actual positive samples. It is expressed as:

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

The F1-score represents the harmonic mean of Precision and Recall, providing a balanced evaluation metric that considers both false positive and false negative classification errors. It is defined as:

$$F1 - Score = 2 \times \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

These metrics provide a comprehensive assessment of predictive performance by measuring overall classification correctness, the ability to correctly identify churn customers, the balance between false positives and false negatives, and the model's discriminative capability across different classification thresholds. Maintaining the same evaluation criteria ensures that any observed improvements are attributed to the proposed BFA-based feature optimization and heterogeneous stacking ensemble architecture rather than differences in performance measurement strategies.

Table 3. Selected Features Using BFA.

Feature Index	Selected Feature	Description
2	Age	Customer's age in years.
4	Number of Dependents	Number of individuals financially dependent on the customer.
5	City	Customer's city of residence.
7	Latitude	Geographic latitude of the customer's location.
11	Offer	Promotional offer or discount subscribed to by the customer.
12	Phone Service	Indicates whether the customer subscribes to phone services.
13	Avg Monthly Long-Distance Charges	Average monthly charges incurred from long-distance calls.
14	Multiple Lines	Indicates whether the customer has multiple phone lines.
16	Internet Type	Type of internet service subscribed to (e.g., DSL, Fiber Optic, Cable).
18	Online Security	Indicates whether the customer subscribes to online security services.
19	Online Backup	Indicates whether the customer subscribes to online backup services.
20	Device Protection Plan	Indicates whether the customer has device protection coverage.
21	Premium Tech Support	Indicates whether premium technical support is included in the subscription.
25	Unlimited Data	Indicates whether the customer has an unlimited data plan.
26	Contract	Type of customer contract (e.g., Month-to-Month, One-Year, Two-Year).
28	Payment Method	Method used by the customer to pay service bills.
29	Monthly Charge	Monthly subscription charge paid by the customer.
32	Total Extra Data Charges	Total charges incurred from additional data usage beyond the subscribed plan.
33	Total Long-Distance Charges	Total accumulated long-distance call charges.
34	Total Revenue	Total revenue generated from the customer during the subscription period.
35	Customer Category	Customer segmentation or classification based on subscription profile.

Figure 4. Snapshot of the selected features.

4. Results and Discussion

This section presents the results of the customer churn prediction analysis, including the features selected by the Binary Firefly Algorithm.

4.1. Feature Selection Using Binary Firefly Algorithm

The Binary Firefly Algorithm identified a subset of features, represented by their zero-based indices, which indicate their respective positions within the dataset, as illustrated in Figure 4. These indices correspond to the most informative variables retained following the feature selection process. As shown in Figure 4, the binary equivalence of the selected features is presented in Table 3.

Table 3 presents the features selected by the BFA, highlighting the most influential variables associated with customer churn across demographic, service usage, subscription, and financial dimensions. Following the preprocessing strategy adopted in our previous study [15], the original dataset attributes were carefully examined to prevent target leakage before model development. Customer Status was retained exclusively as the target variable and was not included among the candidate input features during feature selection or model training. Furthermore, post-outcome attributes, including Churn Category and Churn Reason, were excluded from the predictive feature set because these variables become available only after the churn event and have a direct relationship with the target outcome. Including such attributes would introduce data leakage and lead to misleading performance estimates. Therefore, the BFA-based feature optimization process was applied only to valid pre-churn attributes consisting of demographic characteristics, service utilization patterns, subscription information, and financial indicators.

The selected demographic attributes include Age, City, Latitude, and Number of Dependents, which describe customer characteristics that may influence retention behaviour. Service-related variables such as Phone Service, Multiple Lines, Internet Type, Online Security, Online Backup, Device Protection Plan, Premium Tech Support, and Unlimited Data represent customer engagement with available services and their potential influence on satisfaction and churn likelihood. Subscription-related attributes, including Offer, Contract, and Payment Method, capture customer commitment levels, promotional factors, and payment preferences associated with retention decisions. Furthermore, financial and usage-related variables, including Avg Monthly Long-Distance Charges, Monthly Charge, Total Extra Data Charges, Total Long-Distance Charges, and Total Revenue, provide information regarding customer spending patterns, service consumption, and perceived value.

4.2. Models Performance

The performance of the proposed stacking ensemble model and baseline models is presented in Table 4. All models were evaluated using consistent metrics, including Accuracy, Precision, Recall, F1-score, and AUROC, to ensure a fair and comprehensive comparison.

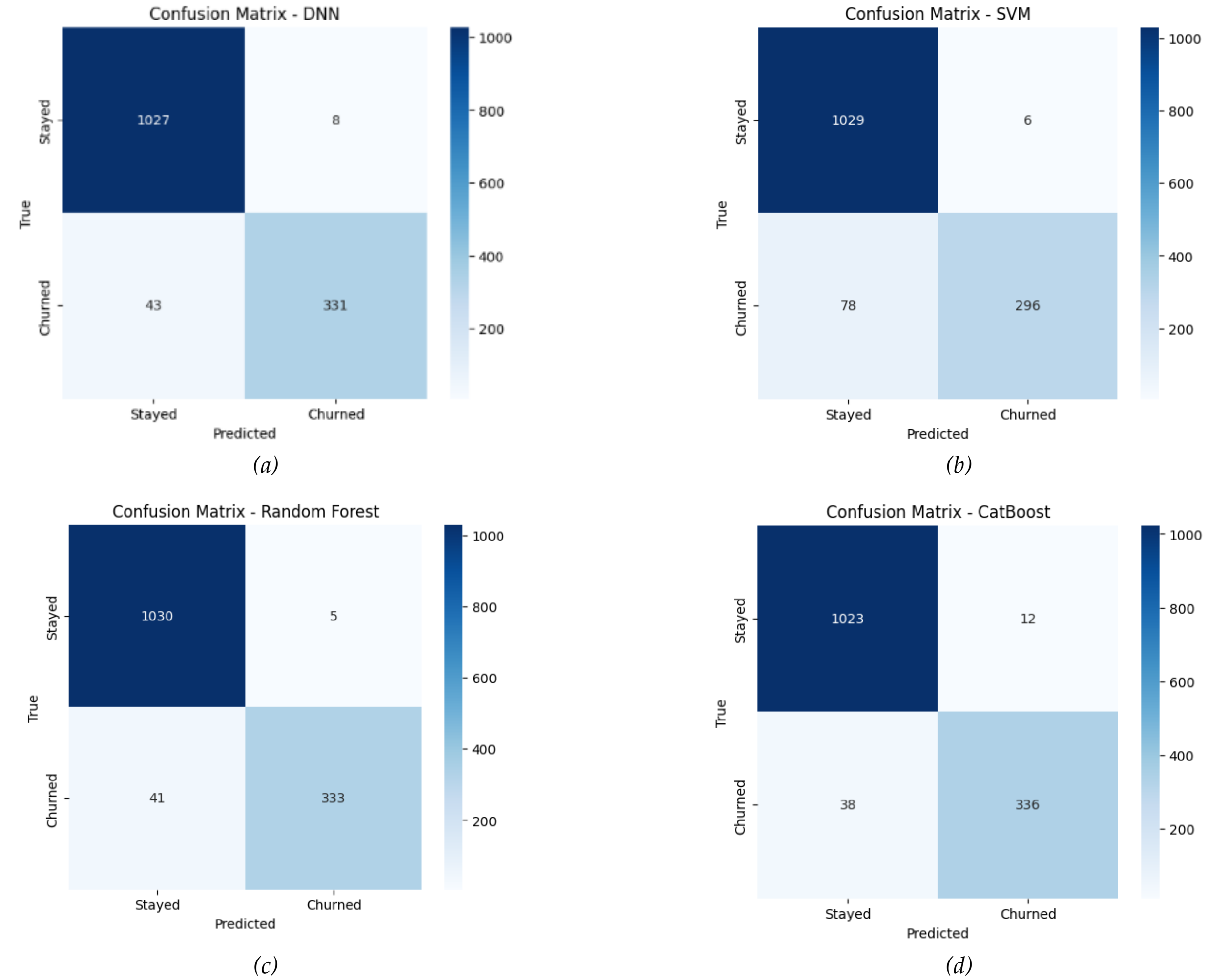

Figure 5. Confusion Matrix of the baseline models: (a) DNN; (b) SVM; (c) RF; and (d) CatBoost.

Table 4. Experimental Results of all Models the Dataset.

Model	Accuracy	Precision	Recall	F1 Score	AUROC
Proposed Stacking Ensemble	0.967353	0.985207	0.890374	0.935393	0.950984
DNN	0.963804	0.976401	0.885027	0.928471	0.985946
SVM	0.940383	0.980132	0.791444	0.875740	0.977736
RF	0.967353	0.985207	0.890374	0.935393	0.983071
CatBoost	0.964514	0.965517	0.898396	0.930748	0.987734

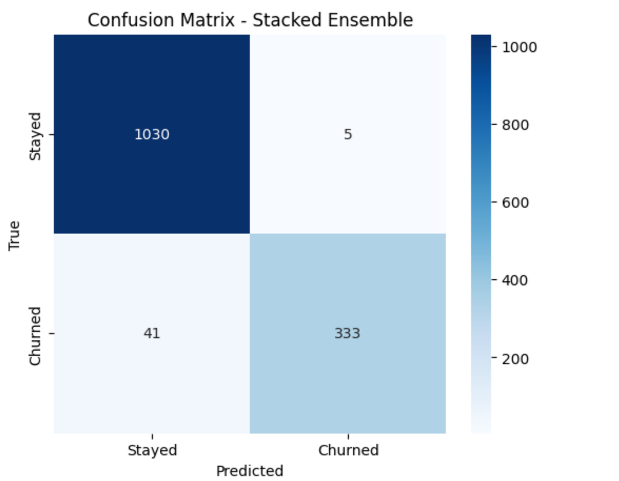

Figure 6. Confusion Matrix of the proposed hybrid stacking model.

As presented in Table 4, the proposed stacking ensemble framework demonstrates competitive predictive performance across all evaluation metrics. The model achieves an accuracy of 96.74%, precision of 98.52%, recall of 89.04%, F1-score of 93.54%, and AUROC of 95.10%. Notably, the proposed stacking ensemble achieves identical accuracy, precision, recall, and F1-score values with Random Forest, indicating that both approaches provide equivalent classification performance on the test dataset. Therefore, rather than demonstrating numerical superiority over Random Forest (see Figure 5 for details of the Confusion Matrix), the contribution of the proposed framework lies in integrating heterogeneous learning mechanisms to achieve comparable classification effectiveness while combining the complementary strengths of multiple classifiers. Further details are presented in the confusion matrix in Figure 6.

Compared with other baseline models, the proposed framework outperforms DNN and SVM in terms of accuracy, recall, and F1-score, while achieving competitive performance with CatBoost. Although CatBoost and Random Forest achieve higher AUROC values of 98.77% and 98.31%, respectively, suggesting stronger overall class-ranking capability, the proposed stacking ensemble maintains a strong balance between precision and recall, which is particularly important for identifying potential churn customers.

4.4. Discussion of the findings

This section presents a comprehensive interpretation of the experimental results by comparing the performance of the proposed BFA-optimized heterogeneous stacking ensemble framework with the individual baseline models. Since this study represents an extension of our previous work [15], which investigated a BFA-based Hybrid TabNet-DNN model using the same telecom churn dataset and evaluation criteria, the current research focuses on examining the effectiveness of integrating optimized feature selection with heterogeneous ensemble learning for customer churn prediction.

The experimental findings demonstrate that the proposed stacking ensemble framework achieves competitive and balanced predictive performance across the considered evaluation metrics. Although the proposed model does not demonstrate numerical superiority over Random Forest in all classification measures, it achieves comparable predictive capability, indicating that the integration of multiple heterogeneous learners successfully preserves the strong classification ability of individual ensemble models while introducing a more diverse decision-making mechanism. Therefore, the contribution of the proposed framework lies not only in improving predictive performance but also in providing a robust architecture that combines the strengths of different learning algorithms.

The effectiveness of the proposed framework can be attributed to the complementary characteristics of its constituent models, namely DNN, SVM, and Random Forest, whose prediction outputs are integrated through Logistic Regression as the meta-learner. The DNN contributes by capturing complex nonlinear relationships and hidden patterns within customer behavioural data, while SVM enhances classification through effective decision boundary optimization. Random Forest provides additional robustness through its ensemble-based structure and ability to model nonlinear interactions among customer attributes. By combining these diverse learning perspectives, the stacking framework reduces reliance on a single predictive mechanism and enables more flexible decision generation.

Furthermore, the incorporation of the Binary Firefly Algorithm (BFA) represents a significant methodological advancement compared with our previous study [15]. While the previous work focused on optimizing a Hybrid TabNet-DNN architecture for churn prediction, the current study extends this direction by applying BFA-based feature optimization within a heterogeneous stacking ensemble framework. The feature optimization process improves the quality of the input representation by reducing irrelevant and redundant attributes, allowing the predictive models to focus on the most informative customer characteristics. This contributes to improved model efficiency, reduced complexity, and enhanced generalization capability.

The comparative analysis indicates that the proposed stacking ensemble provides stronger overall modelling capability than several individual baseline approaches, particularly when considering the ability to combine diverse predictive behaviours. However, Random Forest demonstrates comparable classification effectiveness, highlighting its strength as an individual ensemble learner. Similarly, CatBoost exhibits strong discriminative capability, while DNN and SVM provide valuable contributions through nonlinear representation learning and decision boundary optimization, respectively. These observations demonstrate that different algorithms possess unique advantages depending on the performance criterion considered, further justifying the motivation for integrating heterogeneous models.

The confusion matrix analysis further supports the effectiveness of the proposed approach by demonstrating its ability to achieve balanced classification between churned and retained customers. Compared with individual models, the stacking framework benefits from the combined decision patterns of multiple learners, reducing dependence on the limitations of any single algorithm. This balance is particularly important in telecommunications churn prediction, where accurately identifying potential churn customers is essential for developing effective retention strategies while minimizing unnecessary intervention.

Compared with our previous BFA Hybrid TabNet-DNN model [15], the proposed framework introduces a different modelling strategy by replacing a single hybrid deep learning architecture with a heterogeneous ensemble-based decision framework. While the previous study demonstrated the effectiveness of feature optimization combined with deep learning representation, the current work advances this approach by integrating multiple complementary classifiers through stacking. This architectural extension enables the model to exploit diverse learning patterns and provides a more flexible approach for handling complex customer behaviour.

5. Comparison with State-of-the-Art Studies

To further assess the competitiveness of the proposed framework, its performance was compared with several

Table 5. Comparison of the Proposed Framework with Recent State-of-the-Art Studies.

Model	Accuracy	Precision	Recall	Specificity	F1-Score	AUC / AUROC
Reference [12]	0.8694	–	0.8547	0.8839	–	0.9500
Reference [13]	0.9900	–	–	–	–	–
Reference [32]	0.8300	–	–	–	0.7500	0.5600
Reference [15]	0.9642	0.9856	0.8779	–	0.9287	0.9885
Proposed Stacking (Ours)	0.9674	0.9852	0.8904	–	0.9354	0.9510

recent state-of-the-art customer churn prediction models reported in the literature. The comparison includes studies employing different machine learning, deep learning, and ensemble learning techniques, with the results summarized in Table 5.

The comparison demonstrates that the proposed heterogeneous stacking ensemble is highly competitive with recent customer churn prediction approaches. Although Reference [13] reported a higher classification accuracy, the evaluation was conducted under a different experimental setting and dataset, making direct numerical comparison inappropriate. Similarly, References [12] and [32] reported lower predictive performance across the available evaluation metrics.

The most meaningful comparison is with our previous study, Odion *et al*. [15], as both studies employed the consistent preprocessing strategy. While the previous study investigated BFA-optimized deep learning architectures, the present work extends that research by integrating BFA-based feature optimization with a heterogeneous stacking ensemble comprising DNN, SVM, and Random Forest, combined through a Logistic Regression meta-learner. The proposed framework demonstrated improved classification performance, particularly in terms of accuracy, recall, and F1-score. Although precision decreased slightly, the difference was negligible compared with the previous model. These findings indicate that combining diverse learning paradigms through stacking provides a more balanced and robust predictive framework than relying on a single optimized deep learning architecture.

6. Limitations

Although the proposed BFA-optimized heterogeneous stacking ensemble demonstrated strong and competitive performance in customer churn prediction, this study has some limitations. First, the experimental evaluation was conducted using the publicly available Maven Analytics Telecom Customer Churn dataset rather than operational data obtained from telecommunications providers. While this benchmark dataset facilitates reproducibility and comparison with previous studies, it may not fully capture the unique behavioural patterns, service quality issues, pricing structures, and network characteristics of specific telecommunications markets, particularly within Nigeria. Second, the proposed framework was validated on a single benchmark dataset, and additional evaluation using multiple real-world datasets would further strengthen the evidence of its robustness and generalization capability.

7. Future Work

Future research should focus on validating the proposed framework using locally sourced datasets from Nigerian telecommunications operators to better capture region-specific customer behaviour, service quality perceptions, pricing dynamics, and network-related challenges. Such datasets would improve the contextual relevance and practical applicability of customer churn prediction models for local service providers. In addition, future studies may investigate the integration of explainable artificial intelligence (XAI) techniques to improve model transparency and support decision-making, as well as explore automated hyperparameter optimization and alternative ensemble learning strategies to further enhance predictive performance, computational efficiency, and model robustness.

8. Conclusion

This study presented an enhanced customer churn prediction framework that extends our previous work [15] by introducing a BFA-optimized heterogeneous stacking ensemble architecture. While the previous study investigated the effectiveness of individual machine learning and deep learning models for telecom churn prediction using the same benchmark dataset, the present research advances this approach by integrating feature optimization and ensemble-based decision fusion to improve predictive capability and model robustness. The proposed framework combines DNN, SVM, and RF as base learners, with Logistic Regression employed as the meta-learner to generate the final churn prediction output. The model development process incorporated a consistent preprocessing strategy adopted from our previous study [15], including data cleaning, feature transformation, normalization, stratified data partitioning, and 5-fold cross-validation, thereby ensuring a fair comparison while isolating the impact of the proposed methodological improvements. Furthermore, the incorporation of BFA enabled the identification of the most informative customer attributes, reducing feature redundancy and improving the efficiency and generalization capability of the predictive models.

Experimental results showed that the proposed stacking ensemble achieved strong and balanced performance across the evaluated metrics.

Although Random Forest achieved comparable classification performance, highlighting its effectiveness as a robust individual ensemble learner, the proposed framework provides the advantage of integrating multiple heterogeneous learning mechanisms rather than relying on a single modelling approach. By combining the nonlinear representation capability of DNN, the decision boundary optimization strength of SVM, and the robustness of Random Forest, the stacking architecture provides a flexible framework capable of capturing diverse patterns within customer behaviour data. In a nutshell, the findings confirm that integrating heterogeneous learning algorithms within a stacking framework, supported by optimized feature selection, enhances predictive accuracy, robustness, and generalization for churn prediction tasks in the telecommunications industry. For future research, it is recommended that studies focus on the use of locally sourced Nigerian telecommunications datasets, as customer behavior, service quality perceptions, pricing structures, and network challenges differ significantly across regions; developing models based on local data would therefore improve contextual relevance, enhance real-world applicability, and provide more actionable insights for Nigerian telecom operators in designing effective customer retention strategies.

9. Declarations

9.1. Author Contributions

Abdulrashid Abdulrauf: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources; Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition; **Maaruf Mohammed Lawal:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition; **Oluwatoyin Omoloba:** Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft; **Omolara Busayo Abodunrin:** Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft; **Minkail Mohammad:** Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft.

9.2. Institutional Review Board Statement

Not applicable.

9.3. Informed Consent Statement

Not applicable.

9.4. Data Availability Statement

The dataset is available online via the link https://mavenanalytics.io/data-playground/telecom-customer-churn.

9.5. Acknowledgment

Not applicable.

9.6. Conflicts of Interest

The author declares no conflicts of interest.

10. 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, pp. 1–23, 2020. https://doi.org/10.3390/e22070753.

[2] S. Ouf, K. T. Mahmoud, and M. A. Abdel-Fattah, “A proposed hybrid framework to improve the accuracy of customer churn prediction in telecom industry,” *Journal of Big Data*, vol. 11, 2024. https://doi.org/10.1186/s40537-024-00922-9.

[3] N. R. Al-Molhem, Y. Rahal, and M. Dakkak, “Social network analysis in telecom data,” *Journal of Big Data*, vol. 6, no. 1, pp. 1–15, 2019. https://doi.org/10.1186/s40537-019-0264-6.

[4] A. O. J. Ibitoye, C. Onime, N. D. Zaki, and O. F. W. Onifade, "Sociotransactional impact of recency, frequency, and monetary features on customers' behaviour in telecoms' churn prediction," *Iraqi Journal of Computer Science and Mathematics*, vol. 3, no. 2, pp. 101–110, 2022. https://doi.org/10.52866/ijcsm.2022.02.01.011.

[5] M. Loukili, F. Messaoudi, and M. El Ghazi, "Supervised learning algorithms for predicting customer churn with hyperparameter optimization," *International Journal of Advanced Soft Computing and Its Applications*, vol. 14, no. 3, pp. 49–63, 2022. https://doi.org/10.15849/IJASCA.221128.04.

[6] T. Xu, Y. Ma, and K. Kim, "Telecom churn prediction system based on ensemble learning using feature grouping," *Applied Sciences*, vol. 11, no. 11, pp. 1–17, 2021. https://doi.org/10.3390/app11114742.

[7] R. Bugajev, R. Kriauzienė, O. Vasilecas, and V. Chadyšas, "The impact of churn labelling rules on churn prediction in telecommunications," *Informatica*, vol. 33, no. 2, pp. 247–277, 2022. https://doi.org/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," *Journal of Public Affairs*, vol. 20, no. 1, p. e2353, 2020. https://doi.org/10.1002/pa.2353.

[9] E. Xevelonakis and P. Som, "The impact of social network-based segmentation on customer loyalty in the telecommunication industry," *Journal of Database Marketing & Customer Strategy Management*, vol. 19, no. 2, pp. 98–106, 2012. https://doi.org/10.1057/dbm.2012.12.

[10] L. Ehrlinger and W. Wöß, "A survey of data quality measurement and monitoring tools," *Frontiers in Big Data*, vol. 5, p. 13, 2022. https://doi.org/10.3389/fdata.2022.850611.

[11] Y. Khan, S. Shafiq, A. Naeem, S. Hussain, S. Ahmed, and N. Safwan, "Customers churn prediction using artificial neural networks (ANN) in telecom industry," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 9, pp. 132–142, 2019. https://doi.org/10.14569/IJACSA.2019.0100918.

[12] V. Chang, K. Hall, Q. A. Xu, F. O. Amao, M. A. Ganatra, and V. Benson, "Prediction of Customer Churn Behavior in the Telecommunication Industry Using Machine Learning Models," *Algorithms*, vol. 17, no. 6, p. 231, 2024. https://doi.org/10.3390/a17060231.

[13] A. G. Toprak, Ö. B. Mercan, and M. S. Osmanca, "ML-Based Telecom Customer Churn Analysis," in *2025 33rd Signal Processing and Communications Applications Conference (SIU)*, Sile, Istanbul, Turkiye, 2025, pp. 1–4. https://doi.org/10.1109/SIU66497.2025.11111814.

[14] A. Manzoor, M. A. Qureshi, E. Kidney, and L. Longo, "*e-Profits*: A Business-Aligned Evaluation Metric for Profit-Sensitive Customer Churn Prediction," *International Journal of Data Science and Analytics*, vol. 22, 2026. https://doi.org/10.1007/s41060-026-01044-6.

[15] P. O. Odion, M. M. Lawal, S. S. Joshua, and A. Abdulrauf, "Binary Firefly Algorithm-Based Optimization of TabNet and DNN Models for Customer Churn Prediction," *Journal of Future Artificial Intelligence and Technologies*, vol. 3, no. 2, pp. 257–274, 2026. https://doi.org/10.62411/faith.3048-3719-376.

[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," *Journal of Advances in Information Technology*, vol. 14, no. 5, pp. 1046–1055, 2023. https://doi.org/10.12720/jait.14.5.1046-1055.

[17] A. Summaila and S. O. Stephen, "Performance analysis of QoS parameters and churn prediction model development for MNOs in Nigeria," *Journal of Systematic, Evaluation and Diversity Engineering*, vol. 11, no. 3, pp. 55–64, 2023. https://africanscholarpub.com/ajsitr/article/view/103.

[18] S. W. Fujo, S. Subramanian, and M. A. Khder, "Customer Churn Prediction in Telecommunication Industry Using Deep Learning," *Information Sciences Letters*, vol. 11, no. 1, pp. 185–198, 2022. https://doi.org/10.18576/isl/110120.

[19] 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. https://doi.org/10.1109/ACCESS.2024.3349950.

[20] A. Kumar, P. Kakkar, and M. K. Chouhan, "Enhancing Telecom Customer Churn Prediction Using Machine Learning and LSTM for Time-Series Analysis," in *2025 International Conference in Advances in Power, Signal, and Information Technology (APSIT)*, Bhubaneswar, India, 2025, pp. 1–7. https://doi.org/10.1109/APSIT63993.2025.11086177.

[21] P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi, "Customer churn prediction system: A machine learning approach," *Computing*, vol. 104, pp. 271–294, 2022. https://doi.org/10.1007/s00607-021-00908-y.

[22] M. R. Hossain, "Predicting Customer Churn in Telecommunications with Machine Learning Models," *Asian Journal of Research in Computer Science*, vol. 18, no. 1, pp. 53–66, 2025. https://doi.org/10.9734/ajrcos/2025/v18i1548.

[23] 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 in Engineering*, vol. 26, p. 104629, 2025. https://doi.org/10.1016/j.rineng.2025.104629.

[24] Y. Beeharry and R. T. Fokone, "Hybrid approach using machine learning algorithms for customers' churn prediction in the telecommunications industry," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 4, p. e6627, 2022. https://doi.org/10.1002/cpe.6627.

[25] S. S. Poudel, S. Pokharel, and M. Timilsina, "Explaining customer churn prediction in telecom industry using tabular machine learning models," *Machine Learning with Applications*, vol. 17, p. 100567, 2024. https://doi.org/10.1016/j.mlwa.2024.100567.

[26] A. Prashanthan, R. Roshan, and M. W. P. Maduranga, "RetenNet: A Deployable Machine Learning Pipeline with Explainable AI and Prescriptive Optimization for Customer Churn Management," *Journal of Future Artificial Intelligence and Technologies*, vol. 2, no. 2, pp. 182–201, 2025. https://doi.org/10.62411/faith.3048-3719-110.

[27] D. Kumari, S. K. Singh, S. S. Katira, I. V. Srinivas, and U. Salunkhe, "Telecom Customer Churn Forecasting Using Machine Learning: A Data-Driven Predictive Framework," *Metallurgical and Materials Engineering*, vol. 31, no. 4, pp. 922–929, 2025. https://doi.org/10.63278/1536.

[28] D. Bansal, B. Gupta, and S. Gupta, "Telecom Industry Customer Attrition Prediction with Machine Learning Approach," *MAIT Journal of Science Technology*, vol. 1, no. 1, pp. 1–10, 2024. [Online] Available: https://mait.ac.in/images/AcadAudit/2025/24.%20Annexure-10%20(a)-Publication%20of%20Journal.pdf.

[29] M. Al-Masri, B. Kasasbeh, and R. Qaddoura, "Mitigating Overfitting in Telecom Customer Churn Prediction Using the Orange Jordan Dataset," in *2025 1st International Conference on Computational Intelligence Approaches and Applications (ICCIAA)*, Amman, Jordan, 2025, pp. 1–6. https://doi.org/10.1109/ICCIAA65327.2025.11013157.

[30] S. L. Marie-Sainte and N. Alalyani, "Firefly Algorithm based Feature Selection for Arabic Text Classification," *Journal of King Saud University - Computer and Information Sciences*, vol. 32, no. 3, pp. 320-328, 2020. https://doi.org/10.1016/j.jksuci.2018.06.004.

[31] K. Mourad and R. Boudour, "A modified binary firefly algorithm to solve hardware/software partitioning problem," *Informatica*, vol. 45, no. 7, pp. 1–12, 2021. https://doi.org/10.31449/inf.v45i7.3408.

[32] S. Saleh and S. Saha, "Customer retention and churn prediction in the telecommunication industry: A case study on a Danish university," *SN Applied Sciences*, vol. 5, no. 7, p. 173, 2023. https://doi.org/10.1007/s42452-023-05389-6.