Journal of Theoretical and Applied Information Technology
15th March 2025. Vol.103. No.5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

BOOSTING STUDENT PERFORMANCE PREDICTION IN E-LEARNING: A HYBRID FEATURE SELECTION AND MULTI-TIER ENSEMBLE MODELLING FRAMEWORK WITH FEDERATED LEARNING

N S KOTI MANI KUMAR TIRUMANADHAM[1*], THAIYALNAYAKI S[2], NIRUPA V[3], M MADHAVI[4], PERURI VENKATA ANUSHA[5], V S PAVAN KUMAR[6], VAHIDUDDIN SHARIFF[7]

[1*]Research Scholar, Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India.
[2]Associate Professor, Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India
[3]Assistant Professor, Department of CSE (AI), Madanapalle Institute of Technology & Science, Madanapalle, India
[4]Assistant Professor, Department of Artificial Intelligence, Anurag University, Ghatkesar, Telangana, India.
[5]Assistant Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India.
[6]Assistant Professor, Department of CSE-AIML, MLR Institute of Technology, Hyderabad, India
[1,7]Assistant Professor, Department of CSE, Sir C R Reddy College of Engineering, Eluru, Andhra Pradesh, India.

[1]manikumar1248@gmail.com, [2]thaiyalnayaki.cse@bharathuniv.ac.in, [3]vakkalanirupa@gmail.com, [4]madurimadhavi@gmail.com, [5]anushaperuri@gmail.com, [6]pavankumar620@gmail.com, [7]shariff.v@gmail.com

ABSTRACT

This research introduces a novel and advanced methodology for predictive modeling using federated learning, addressing critical challenges such as data privacy, class imbalance, and model performance. Unlike traditional centralized approaches, our work ensures data privacy through federated learning, enabling high-performance models without exposing sensitive data. The novelty of our approach lies in the integration of advanced preprocessing techniques, such as the Synthetic Minority Oversampling Technique (SMOTE) for class imbalance, hybrid feature selection by the combination of Boruta algorithm and L2 regularization's (Boruta-L2) for robust feature selection, and a 3-tier ensemble model with cutting-edge hyperparameter tuning techniques, including Bayesian Optimization, Random Search, and Particle Swarm Optimization (PSO). As a result, our global model achieves an accuracy of 98.90%, significantly outperforming previous methodologies. The advancements in our work are highlighted by the superior model performance, scalability, and privacy preservation, making it a significant contribution to federated learning. This research provides a comprehensive, efficient, and privacy-preserving solution for distributed predictive tasks, setting a new benchmark in machine learning applications across various domains.

Keywords: *Boruta, L2 Regularization, Particle Swarm Optimization (PSO), E –Learning, Federated Learning, Hyperparameter Tuning*.

1. INTRODUCTION

The rapid digital transformation of education has led to the widespread adoption of e-learning platforms, providing students with flexible and accessible learning opportunities. These platforms, including Massive Open Online Courses (MOOCs) and personalized learning environments, allow learners to access educational content anytime and anywhere [1]. However, despite their benefits, challenges such as data privacy concerns, class imbalance, and model generalization issues hinder their full potential. Traditional machine learning models for student performance prediction rely on centralized data collection, raising privacy risks and security concerns, especially in sensitive domains like education and healthcare [2]. Additionally, the heterogeneity of data across different institutions and learning environments makes it difficult to develop a single model that performs well across diverse datasets.

Federated Learning (FL) has emerged as a promising approach to address these challenges by enabling collaborative model training without exposing raw data. Unlike conventional centralized learning, FL ensures data privacy by keeping training data on local devices and only sharing model updates. This decentralized approach enhances security and complies with data protection regulations. However, FL still faces key challenges such as data heterogeneity, class imbalance, adversarial attacks, and secure aggregation that impact model performance and scalability. Effective strategies are required to mitigate these issues and ensure highly accurate, fair, and privacy-preserving predictive models for e-learning environments.[3]

Previous research has explored various machine learning techniques for student performance prediction, but several limitations remain. Beaulac et al. (2019) used Random Forest models for student grade prediction, achieving 78% accuracy, but their models lacked generalizability due to dataset constraints. Enughwure et al. (2020) and Ashfaq et al. (2020) addressed class imbalance using SMOTE and logistic regression, yet their models struggled with scalability and privacy in real-world applications. Gupta et al. (2023) employed hyperparameter-tuned ML models for diabetes prediction, reaching 88.61% accuracy, but their centralized approach compromised data security. Tariq (2023) investigated oversampling techniques to improve fairness in multi-class educational datasets, yet their work did not focus on privacy-preserving solutions. These studies indicate progress in predictive modeling but do not fully resolve the challenges of privacy, scalability, and model optimization in decentralized learning environments. In contrast, this study introduces a privacy-preserving, scalable, and high-accuracy FL framework designed for e-learning environments. Unlike previous works, this research integrates Federated Learning to enable decentralized training while ensuring data privacy and security compliance [4]. It also employs Boruta-L2 hybrid feature selection, enhancing model interpretability and reducing overfitting by selecting only the most relevant features. To further improve predictive accuracy, the study implements a 3-tier ensemble model incorporating Random Forest, Support Vector Machine (SVM), and Gradient Boosting, optimized using Bayesian Optimization, Random Search, and Particle Swarm Optimization (PSO). This novel approach achieves 98.90% accuracy, significantly outperforming previous models while ensuring class balance, fairness, and robustness in student performance prediction.

By addressing key limitations in existing methods, this research establishes a new benchmark in privacy-preserving predictive modeling for distributed e-learning environments. The following sections further elaborate on the research gaps, proposed methodology, and experimental results, demonstrating the effectiveness of the FL-based ensemble learning approach in enhancing predictive accuracy while maintaining privacy and scalability.

1.1. Research Gap

Despite advancements in machine learning for education and healthcare, ensuring reliable, secure, and scalable Federated Learning (FL) frameworks remains a challenge. Traditional centralized models pose privacy risks by requiring data aggregation, whereas FL decentralizes training to enhance privacy. However, issues such as class imbalance, feature selection, and model generalization affect performance. Many e-learning datasets suffer from class imbalance, leading to biased predictions, which necessitates the use of SMOTE for balanced learning. Additionally, high-dimensional data increases overfitting risks, making Boruta-L2 hybrid feature selection crucial for optimal feature extraction. Handling heterogeneous, non-IID datasets further requires a 3-tier ensemble model optimized with Bayesian Optimization, Random Search, and PSO to improve accuracy and scalability. While FL enhances privacy, challenges such as adversarial attacks, secure aggregation, and update security persist, along with limited dataset diversity, affecting model applicability. To address these gaps, future research must focus on developing robust FL frameworks that integrate advanced preprocessing, security protocols, and bias mitigation techniques to enhance privacy-preserving machine learning in real-world applications.

1.2. Research Questions

RQ1. How does Federated Learning (FL) improve privacy and security in student performance prediction compared to traditional centralized learning methods?

RQ2. What advantages does a 3-tier ensemble model with Bayesian Optimization, Random Search, and PSO offer over traditional machine learning models?

1.3. Contributions

This study introduces a privacy-preserving Federated Learning (FL) framework for student performance prediction in e-learning, addressing key limitations in existing research. The main contributions are:

1. **Privacy-Preserving Learning:** Unlike centralized models, FL enables collaborative

training without sharing raw data, ensuring security and compliance.
2. **Improved Feature Selection:** The Boruta-L2 hybrid approach enhances feature selection, reducing overfitting and improving model interpretability.
3. **High-Accuracy Ensemble Model:** A 3-tier ensemble (Random Forest, SVM, Gradient Boosting) optimized with Bayesian Optimization, Random Search, and PSO achieves 98.90% accuracy, surpassing previous approaches.
4. **Fairer Predictions:** SMOTE balances class distributions, reducing bias in student performance predictions.
5. **Significance in E-Learning:** This framework enhances scalability, fairness, and privacy, making it a robust solution for real-world educational applications.

These advancements bridge key gaps in privacy, fairness, and model performance, providing a scalable and secure predictive modeling approach for e-learning and beyond.

2. LITERATURE REVIEW

The research by Beaulac et al. [5] in 2019 made use of random forests to analyze grades from two semesters of Canadian university students which resulted in successful degree completion prediction (78% accuracy) and student major prediction (47.41% accuracy). The key indicators for successful predictions came from courses graded as low by Mathematics Economics and Finance departments. The research findings showed that random forests successfully analyzed educational datasets from large datasets while achieving better accuracy results than conventional models while also providing improved variable importance understanding. The research team proposed upcoming advancements should include methods to handle missing data and address multi-label classification because they would improve predictive accuracy of student majors.

Enughwure et al. [6] used SMOTE to balance unequal class distributions when they studied engineering drawing course outcomes predictions with decision trees and logistic regression models in 2020. The predictive models delivered between 67% to 78% accurate results as logistic regression demonstrated highest performance. The use of SMOTE allowed creators to develop synthetic minority class data which boosted prediction reliability. Engineering departments provided questionnaire data which showed machine learning has opportunities to address performance issues in essential subjects. The method enabled focused intervention strategies which demonstrated the worth of cutting-edge methods for helping students in critical engineering subjects.

Ashfaq et.al investigated educational dataset balancing techniques via oversampling techniques as well as undersampling approaches along with hybrid strategies in 2020 for predictive analytics improvement [7]. Performance of the model improved substantially after balancing the dataset because it better detected students at risk for timely intervention. Ashfaq stressed that education data mining needs equal treatment between classes to create meaningful predictions which can be trusted by all stakeholders. The research established that predictive analytics delivers educational data tools which lead educators toward performance-based decision making that improves student achievement results. The research project propelled artificial intelligence technology towards better analysis methods for heterogeneous educational datasets that exist in uneven distributions.

The research by Gupta et al. [8] in 2023 explored diabetes prediction models based on hyperparameter optimization through PIMA Indian Diabetes dataset analysis. Research focused on machine learning classifiers through the evaluation of K-Nearest Neighbors, Decision Trees, Random Forests and Support Vector Machines classifiers. The Random Forest model achieved its best results with 88.61% accuracy while sustaining 75.68% F1-score through applying preprocessing techniques for handling missing values. This research underlined the necessity of preprocessing and model hyperparameter adjustment for creating more accurate and reliable prediction models. Optimization methods helped scientists demonstrate the importance of developing robust predictive healthcare tools.

Tariq [9] performed a study in 2023 about how oversampling methods affect multi-class educational datasets through SMOTE ADASYN and random oversampling. The research validated substantial enhancements in model accuracy as well as precision and recall statistics through the handling of class imbalance problems. The optimal results require choosing oversampling techniques which fit the specific characteristics of datasets and their classification objectives according to Tariq. The study proved preprocessing techniques to have a decisive effect on enhanced prediction capabilities by machine learning technology in educational applications. This work addressed common data quality difficulties to drive machine learning

forward through the generation of useful information in educational contexts that involve multiple classes.

The review investigates recent developments in machine learning applications for education and healthcare data with a specific focus on accuracy enhancement and disproportioned class handling along with operational optimization. Random forests combined with logistic regression have achieved positive results yet researchers need to address three main problems which include imbalanced heterogeneous data and multi-label classification and data points with missing values. The research demonstrates the necessity of using secure preprocessing methods as well as adaptable frameworks to improve prediction projection and security levels. The research discoveries provide essential knowledge that future work should use to develop more precise methods delivering dependable actionable outcomes within privileged data settings.

3. BASIC CONCEPTS

In this section, a detailed overview of the fundamental principles and key concepts required to understand the proposed method is presented.

3.1. Federated Learning

The Federated Learning (FL) [10] architecture shown in Fig.1 is a decentralized framework that facilitates collaborative model training while ensuring data privacy. Clients, such as organizations or devices, train local models on their private datasets and share only model updates (e.g., weights or gradients) with a central server. The server aggregates these updates using Federated Averaging (FedAvg) to create a unified global model, which is redistributed to clients in iterative rounds. This approach eliminates the need to share raw data, preserves privacy, and ensures compliance with regulations. FL employs encryption, secure aggregation, and techniques like differential privacy to enhance security to protect individual data points. It also addresses challenges like non-IID data and resource variability through weighted aggregation and optimized communication protocols. FL is particularly valuable in privacy-sensitive domains, enabling organizations to collaboratively train robust models without compromising data ownership or confidentiality.

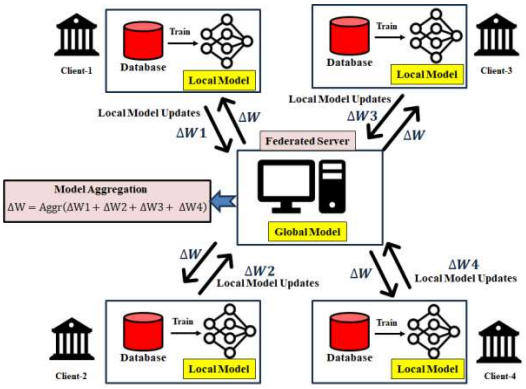


Figure 1: Architecture of Federated Learning

3.1.1. Decentralized Data and Local Objective

In Federated Learning (FL), each client has its local dataset, and the objective is to minimize a global loss. Each client computes the local loss for its data and updates its model based on this loss, ensuring privacy by not sharing raw data. Each client k owns a private dataset D_k with $|D_k|$ samples. The global objective in FL is to minimize the aggregated loss across all clients shown in Equation (1):

$$L(w) = \sum_{k=1}^{N} p_k L_k(w) \quad (1)$$

where,

- w is the global model parameter.
- $\mathcal{L}_k(w) = \frac{1}{|D_k|} \sum_{i \epsilon D_k} l(f(x_i; w), y_i)$ s the local loss function for client k,
- $p_k = \frac{|D_k|}{\sum_{j-1}^{N} |D_j|}$ is the relative weight of client $k's$ contribution.

3.1.2. Local Model Training

Clients train their own model locally using their private data. They compute the gradient of the local loss function and update the model parameters using gradient descent. This allows for learning to occur without the need to send raw data to a central server. Each client updates the global model w_t by performing local optimization using gradient descent. The update equation for client k after E local epochs are represented in Equation (2):

$$w_t^k \leftarrow w_t - \eta \Delta \mathcal{L}_k(w) \quad (2)$$

where:

- ω_t^k is the locally updated model after training on client k.
- η is the learning rate.
- ∇L_k(w) is the gradient of the local loss with respect to w.

3.1.3. Global Model Aggregation (FedAvg)

Once clients have updated their models, the server aggregates all the local model updates using a weighted average. This process combines the knowledge from all clients to update the global model, ensuring that each client's data is taken into account proportionally based on the dataset size. After local updates, the central server aggregates the client models to update the global model. Using Federated Averaging (FedAvg), the global model update are shown in Equation (3):

$$w_{t+1} = \sum_{k=1}^{N} p_k w_t^k \quad (3)$$

where w_t^k is the model update from client k, and p_k is the weighting factor based on the client's dataset size.

3.1.4. Communication Efficiency

To reduce the cost of communication, updates can be compressed or sparse. Instead of sending full model updates, only the most important parts of the updates are shared, or they are quantized to reduce the data sent between clients and the server.

$$\Delta w_t^k = Top_m\left(w_t^k - w_t\right) \quad (4)$$

where Top_m retains the top m elements of the update with the largest magnitudes.
Alternatively, updates can be compressed using quantization showed in Equation (4) & Equation (5):

$$Q\left(\Delta w_t^k\right) = round(\Delta w_t^k . s)/s \quad (5)$$

where s is a scaling factor.

3.1.5. Secure Aggregation

FL ensures data privacy by using techniques like encryption or adding noise to the model updates. To ensure privacy, individual updates are encrypted or perturbed before aggregation.

For example, differential privacy adds noise ϵ to updates Equation (6):

$$\Delta w_t^k \leftarrow \Delta w_t^k + \mathcal{N}(0, \sigma^2) \quad (6)$$

where $\mathcal{N}(0, \sigma^2)$ is Gaussian noise with variance σ^2.

3.1.6. Handling Non-IID Data

In real-world scenarios, client data might be heterogeneous (non-IID). To handle this, personalized models can be used, where each client's model is fine-tuned by blending the global model with the local one, ensuring better performance even with diverse data. FL addresses non-IID data distributions by adjusting the aggregation process. Personalized models can be fine-tuned for each client represented in Equation (8):

$$w_t^k \leftarrow \propto w_t + (1-\propto) w_t^k \quad (8)$$

where α balances the global and local models.

3.1.7. Convergence of FL

The goal of FL is to minimize the global loss function by aggregating local updates. The convergence is achieved when the global model reaches a state where all local models are aligned and the global loss is minimized, ensuring that the model is accurate across all clients. The convergence of FL is defined as the minimization of the global objective $\mathcal{L}$ (w). Using gradient aggregation shown in Equation (9):

$$w_{t+1} = w_t - \eta \sum_{k=1}^{N} p_k \nabla \mathcal{L}_k(w) \quad (9)$$

The convergence rate depends on factors such as local updates, learning rate, and data heterogeneity.

4. METHODOLOGY

The proposed method implements sophisticated preprocessing methods combined with resistant feature selection techniques alongside a multi-level ensemble model which optimizes hyperparameters to achieve maximum performance output without compromising privacy standards. At the beginning of the process every client applies data preprocessing to their local data by cleaning data to eliminate inconsistencies then using SMOTE [11] on class imbalance data while coping with outliers by IQR methods and finalizing with Z-score normalization. The Boruta-L2 hybrid approach serves for both improving model interpretability and minimizing complexity when selecting features. Feature importance ranking from Boruta algorithm [12] undergoes L2 regularization [13] processing to achieve stability and prevent overfitting in the selection of relevant features.

Each model in the 3-tier ensemble uses Random Forest [14], Support Vector Machine (SVM) [15], Gradient Boosting [16] with their dedicated hyperparameter optimizers Bayesian Optimization [17], Random Search [18], Particle Swarm Optimization (PSO) [19] respectively. The ensemble model utilizes Voting Classifier to aggregate predictions from individual models for creating a strong predictive framework for each client. After completing local training the server receives model parameters while privacy regulations are maintained through federated learning data protection principles.

The central server generates a global model by uniting local models through averaging or weighted averaging computation and by bestowing weights based on client-submitted performance assessment. The aggregated global model includes all knowledge shared by clients to produce a sturdy generalized

model. The aggregated model performs evaluations against independent test data in order to determine its general performance. The combination of feature selection with ensemble learning and a complex data preprocessing method through federated learning enables both data privacy preservation and model performance improvement and effective computations to address model overfitting and heterogeneous data. The diagram illustrating the proposed method is displayed in Fig.2.

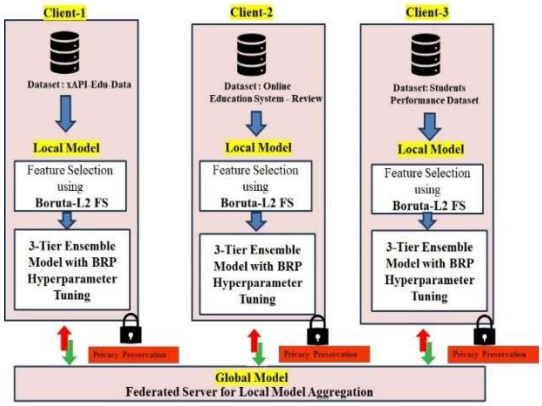


Figure 2. The Proposed Network Structure

4.1. Data Collection

XAPI-EDu-Data tracks student information through demographic data which combines academic metrics along with classroom participating levels. The system contains features for storing gender information along with national origin data and educational level and parent involvement metrics and counts of student engagement behaviors including hand raises and resource exploration and discussion participation. This database integrates two additional data components regarding parental happiness and child absenteeism to give a complete analysis of academic success effects from student actions. This dataset offers essential insights into classroom settings through its ability to expose performance-influencing variables. The localized nature of the dataset protects privacy as it provides essential data input that supports collaborative model training in federated learning systems.

The Online Education System Review dataset delivers an extensive review relating to students' experiences in virtual classrooms. The data includes population statistics about age groups and residential areas together with behavioral characteristics including how students spend their time and which devices they use as well as their sleeping habits and their assessment of online learning platforms. This analysis benefits from additional elements which include the involvement of team study groups and online connectivity together with historical education results. The data collection examines various influences which determine student achievement while studying virtually. The client can protect confidential data throughout their device while federated learning enables their inputs to support enhanced global online learning models.

Through the Students Performance Dataset researchers measure different elements that determine student academic performance levels. The dataset features population statistics including student ages and genders and educational information about parents together with weekly study sessions and grade point averages. The data contains specific attributes which allow researchers to predict how external influences affect student academic outcomes. The offered dataset presents an extensive basis for evaluating numerous elements contributing to educational results. The client maintains data ownership during federated learning but receives model improvements from decentralized processing that utilizes the exclusive features of each dataset. All essentials regarding the client data sets appear in Table.1.

Table 1: Client Dataset Summary for Federated Learning

Client	Dataset Name	Size	Rows	Columns	Key Features
Client-1	xAPI-EDu-Data	37.13KB	480	17	Demographics, academic performance, engagement metrics (raised hands, visited resources), parental involvement
Client-2	Online Education System Review	106.71KB	1033	22	Demographics, study time, device usage, sleep patterns, satisfaction with online education
Client-3	Students Performance Dataset	162.99KB	2392	14	Demographics, weekly study time, absenteeism, extracurricular activities, GPA

4.2. Data Preprocessing

4.2.1. Data Cleaning

The collected datasets generate essential educational findings through their examination of student engagement along with internet learning experiences together with academic results. The dataset managed by Client-1 monitors 480 students through recording demographic statistics along with parental involvement and performance evaluation including participation and attendance data. With

Journal of Theoretical and Applied Information Technology
15th March 2025. Vol.103. No.5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

1033 entries Client-2 provides details about online learning by analyzing student technology use, homework practices and user satisfaction towards digital learning. Client-3 includes a total number of 2392 students who have been assessed based on GPA alongside weekly study hours and extracurricular activities as well as parental support levels to create a wide-ranging academic success overview. The analysis used complete data sets which showed no null values for maintaining consistent reliability throughout the research. The extensive data source offers a superior platform to analyze educational patterns in various classrooms and discover student engagement patterns while measuring the effects of virtual learning and traditional education on academic results.

4.2.2. Handling Imbalanced Dataset

The mismatch of instance numbers between classes creates bias in model predictions so this phenomenon is designated as class imbalance. The primary class in machine learning models receives favoured treatment which leads to substandard predictions for minority classes. The minority class balancing method used frequently in practice is Synthetic Minority Over-sampling Technique (SMOTE). The minority class synthetic samples created by SMOTE start with selecting an instance then generate new examples through interpolating the chosen instance with its nearest neighbours. SMOTE creates new examples to equalize class proportions thereby improving model learning effectiveness of minority instances. SMOTE allows the model to perform more accurately for minority class predictions by maintaining unbiased accuracy toward majority classes. The rating of recall along with precision and F1-score performance improves better because of this approach on imbalanced datasets as shown in Table.2.

Table 2: Class Distribution Before and After SMOTE for Client Datasets

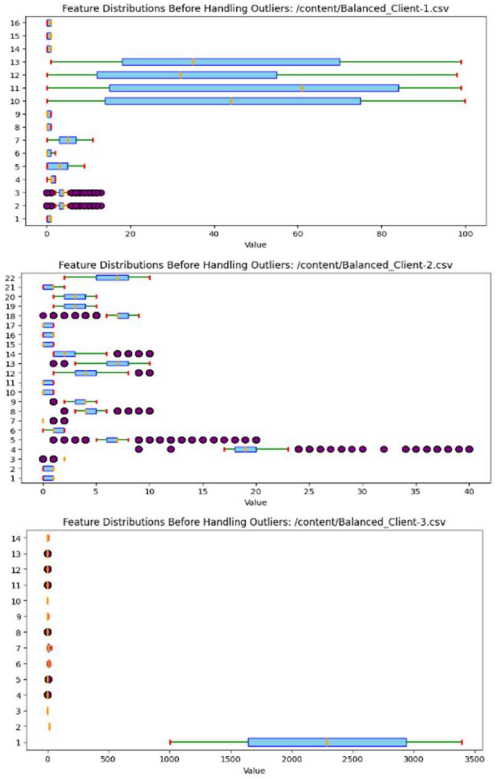
Dataset	Before applying SMOTE		After applying SMOTE	
	Class	Frequency	Class	Frequency
Client-1	0	142	0	211
	1	127	1	211
	2	211	2	211
Client-2	Class	Frequency	Class	Frequency
	0	541	0	541
	1	241	1	541
	2	251	2	541
Client-3	Class	Frequency	Class	Frequency
	0	107	0	1211
	1	269	1	1211
	2	391	2	1211
	3	414	3	1211
	4	1211	4	1211

4.2.3. Handling Outliers

An effective model performance improvement technique based on Interquartile Range (IQR) can minimize extreme values to enhance the accuracy of predictions. The IQR [21] represents the numeric distance between Q1 and Q3 which are the 25th percentile and 75th percentile values in the dataset. Data points that exceed Q1−1.5×IQR and Q3+1.5× IQR identify as outliers in this method of analysis. Outlier detection enables researchers to reduce their impact through removal or limiting values or value transformation according to Table.3 and Fig.3. Extreme values have no influence on models through this approach which leads to higher prediction accuracy and generalization and model stability.

Table 3: Outlier Count Before and After Handling for Client Datasets

Dataset	Before IQR	After IQR
Client-1	255	155
Client-2	1323	876
Client-3	3964	2884



Journal of Theoretical and Applied Information Technology
15th March 2025. Vol.103. No.5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

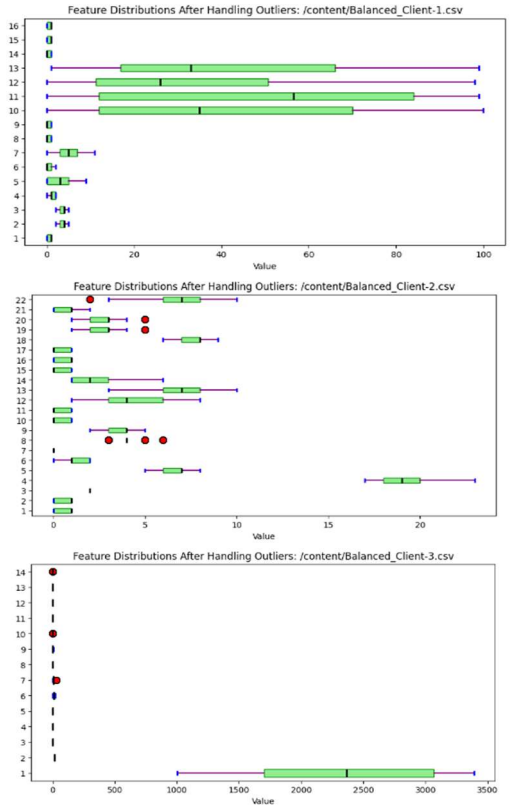


Figure 3: A Box Plots Across Clients for Before and After Handling Outlier

4.2.4. Normalizing Dataset

The preprocessing method known as normalization transforms data to create a standardized format for uniform analysis and modeling purposes. The data transformation process creates a consistent measurement scale which maintains actual differences between values. Z-score normalization stands out as one of various normalization techniques. The normalization process locates data at the mean point of zero while stretching it to achieve a standard deviation value of one because it enables features to be comparable across datasets. Standardization proves essential when working with datasets which include variables from differing measurement scales and have extreme data points. Support vector machines and neural networks along with principal component analysis work better through normalization since they require equal feature scales according to Fig.4. Model reliability strengthens and bias reduction from learning processes and accurate outcomes emerge in data-driven applications because of normalization techniques.

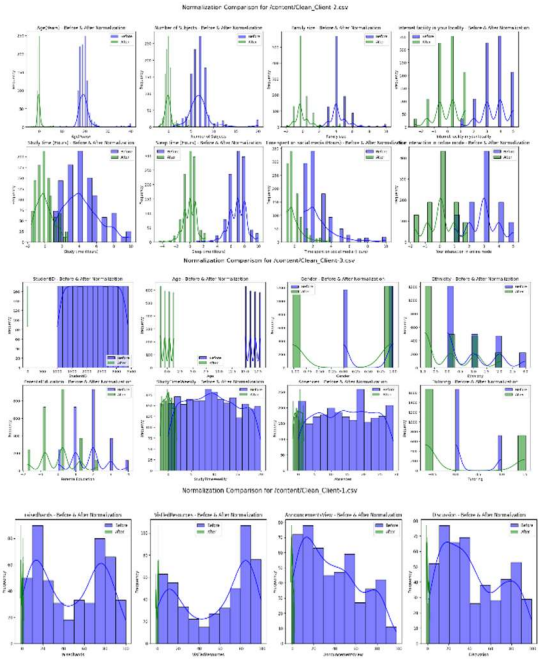


Figure 4: Feature Distributions Across Clients Before and After Normalization

4.3. Feature Selection using Boruta-L2

This section delivers an in-depth review of essential principles and main concepts regarding the proposed feature selection strategy that utilizes the hybrid Boruta-L2 approach. Model performance excellence depends on feature selection methods that find the most important features since it simplifies model complexity while boosting computational performance. Random Forest-based Boruta features all important variables by comparing actual features against random versions for identification. The implementation of L2 regularization in prediction strengthens the selection process of features that demonstrate maximal contribution to model prediction accuracy. Boruta [23] robust feature importance ranking when combined with L2 regularization delivers an approach for precise selection of relevant features alongside overfitting control. Using this dual strategy produces predictive models that should demonstrate enhanced interpretability along with computational effectiveness and reduced susceptibility to overfitting numbers.

4.3.1. Boruta

The Boruta algorithm is a wrapper-based feature selection method that identifies all features strongly correlated with the target variable. It works by creating shadow features—duplicates of the original features with shuffled values—to act as a baseline for feature importance. Features with importance higher than shadow features are deemed significant. Boruta uses Random Forest as a base model, leveraging its feature importance scores for selection.

The feature importance is derived as showed in Equation (10):

$$I(f_i) = \frac{1}{n_{trees}} \sum_{t=1}^{n_{trees}} G_t(f_i) \qquad (10)$$

where $I(f_i)$ is the importance of feature f_i , and G_t represents the Gini impurity decrease in tree t.

4.3.2. L2 regularization

L2 regularization [24] is a shrinkage technique that penalizes large feature coefficients to reduce multicollinearity and overfitting. It modifies the loss function by adding the squared magnitude of coefficients as a penalty term. This encourages smaller, yet non-zero coefficients, retaining all relevant features.

The penalized objective function for linear models is shown in Equation (11):

$$L = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \qquad (11)$$

where L is the loss, λ is the regularization strength, β_j are the coefficients, and p is the number of features.

Algorithm for Boruta-L2 Regularization

Input: Dataset (X, y), Regularization Parameter (λ)
Output: Selected Features $F_{selected}$
1. Initialize
   - Load dataset with features (X) and target variable (y).
   - Set hyperparameters for Boruta and L2 regularization.
2. Step 1: Feature Importance via Boruta
   - Create shadow features by shuffling original feature values.
   - Train a Random Forest model on (X, y) including shadow features.
   - Calculate feature importance scores for original and shadow features.
   - Compare importance of original features to the maximum shadow feature importance:
   - Mark features with higher importance as significant.
   - Mark features with lower importance as irrelevant.
   - Retain tentatively important features for further evaluation.
3. Step 2: Apply L2 Regularization
   - Subset the dataset X_{boruta} to only Boruta-selected features.
   - Train a linear model with L2 regularization (e.g., Ridge Regression):
   - Minimize the penalized loss:

$$L = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

   - Retain features with non-zero coefficients after regularization.
4. Step 3: Final Feature Selection
   - Combine results from Boruta and L2:
   - Finalize features retained by L2 regularization.
5. Output Selected Features
   - Return the selected features F_selected .

End

4.4. 3–Tier Ensemble Model with BRP Hyperparametric Tuning

This section details essential principles together with major concepts needed for understanding the proposed 3–Tier Ensemble Model with BRP Hyperparametric Tuning. Model construction creates forecasting models which demonstrate high accuracy when implementing predictions on previously unseen data. The method blends three high-performing machine learning models Random Forest together with Support Vector Machine (SVM) and Gradient Boosting (GB) using three hyperparameter optimization strategies: Bayesian Optimization and Random Search and Particle Swarm Optimization (PSO) [25] in order. Through these optimization methods and algorithms, the ensemble model combines superior capabilities of individual models to reduce their respective flaws. The Random Forest algorithm maintains superior performance for feature interactions while SVM brings excellence in classification modeling and Gradient Boosting enhances precision by repeating

boosting operations. The optimized hyperparameters maximize model performance independently before they are integrated into an ensemble final model containing all three method strengths. The final ensemble system delivers superior predictive quality together with better reliability and minimized model bias which optimizes its performance for complex datasets applied to real-world scenarios.

4.4.1. Random Forest with Bayesian Optimization

The Bayesian Optimization process implements Random Forest algorithms to efficiently explore hyperparameters thereby improving Random Forest functionality. Random Forest functions as an ensemble learning model by creating various decision trees that results in a collective predication mechanism suitable for intensive classification and regression. Random Forest depends heavily on selected tree numbers while limiting tree depth for producing output results. Bayesian Optimization serves as a probabilistic search approach that uses past evaluation metrics to predict satisfactory new hyperparameters as it explores the search space by taking single-step movements. The performance improvement of Random Forest stems from applying this approach to set its essential parameters correctly.

4.4.2. SVM with Random Search

Support Vector Machine (SVM) algorithm connects with Random Search to automate hyperparameter selection. SVM serves as a strong supervised learning method which finds the best hyperplane that achieves the widest possible class separation. The SVM algorithm responds strongly to various hyperparameters involving the kernel type together with C and additional kernel-specific values. The random search technique [26] enables simple yet powerful evaluation of different hyperparameter combinations through random sampling in order to discover the optimal configuration. Random Search becomes an efficient alternative to grid search because it performs searches at a lower computational cost yet maintains high performance.

4.4.3. Gradient Boosting with Particle Swarm Optimization (PSO)

The integration of Gradient Boosting (GB) with Particle Swarm Optimization (PSO) creates Gradient Boosting with PSO which serves as a solution for hyperparameter optimization. Gradient Boosting constructs ensemble learning models step by step through which it targets errors produced by previous models. The method reduces loss functions through the merger of decision trees as weak learners which produces an enhanced predictive model. The optimization algorithm PSO operates as a nature-inspired method which uses bird flock behaviors to discover optimal solutions. PSO joins forces with Gradient Boosting algorithms by applying hyperparameters optimization of learning rate, number of trees, tree depth with an enhanced exploration of the parameter space which leads to improved model performance while maintaining high accuracy levels.

5. RESULTS AND DISCUSSION

5.1. Experimental Evaluation and Results

We will define the simulation process and simulation environment along with experimental parameters for analyzing efficiency as well as comparing standard federated learning models within this section.

5.2. Experimental Setup

A Python environment combined with Google Colab machine learning libraries running on an Intel Core i7-8550 @ 4GHz system executed the proposed framework through 50 communication epochs. The entire experimentation took place within an artificial simulation system. Experiments based on simulation parameters with their corresponding settings are depicted in Table.4.

Table 4: Simulation parameters and settings.

Parameter	Value
Simulation environment	Python
Python environment	Google Colab
Local epochs	{20, 40, 30, 40, 50}
Number of client nodes	3
Clients Info	xAPI-EDu-Data, Online Education System Review, Students Performance Dataset

5.3. Feature Selection Using Burota-L2

The Boruta-L2 hybrid feature selection technique merges the strengths of the Boruta algorithm and L2 regularization and is highly effective in selecting the most informative features for predictive modeling. Boruta, a robust wrapper technique, exhaustively searches to determine which features are important, while L2 regularization (Ridge Regression) penalizes large coefficients, resulting in more basic models by avoiding overfitting. This combined process, when implemented for feature selection, retains the most significant features only, adding to the entire model's quality and precision. In Federated Learning, in which multiple decentralized clients collaborate while not exchanging raw data, Boruta-L2 becomes the method of utmost significance for local optimisation of features. It allows each customer to select significant features and fitness values given in Table.5 and Fig.5 autonomously, reducing the amount of input information, therefore reducing communication cost and ensuring confidentiality. By

only focusing on crucial features, the method accelerates model convergence, enhances model interpretation and scalability. It facilitates efficient use of computational resources and speeds up the training process and builds high-quality models without compromising user data. The combination of Boruta and L2 regularization in Federated Learning provides more efficient, scalable, and privacy-preserving machine learning models.

Table. 5: Selected Features and Their Fitness Scores for Client

Datasets	Naming	Selected Features	Fitness Scores
Client-1	C1-F1	Relation	-0.206240
	C1-F2	raisedhands	-0.093552
	C1-F3	VisITedResources	0.005105
	C1-F4	AnnouncementsView	0.050640
	C1-F5	Discussion	-0.055448
	C1-F6	StudentAbsenceDays	-0.072508
Client-2	C2-F1	Age(Years)	0.056823
	C2-F2	Time spent on social media (Hours)	0.008703
	C2-F3	Average marks scored before pandemic in traditional classroom	-0.111603
	C2-F4	Your interaction in online mod	0.070353
	C2-F5	Performance in online	0.115048
Client-3	C3-F1	StudentID	-0.126501
	C3-F2	GPA	-0.960180

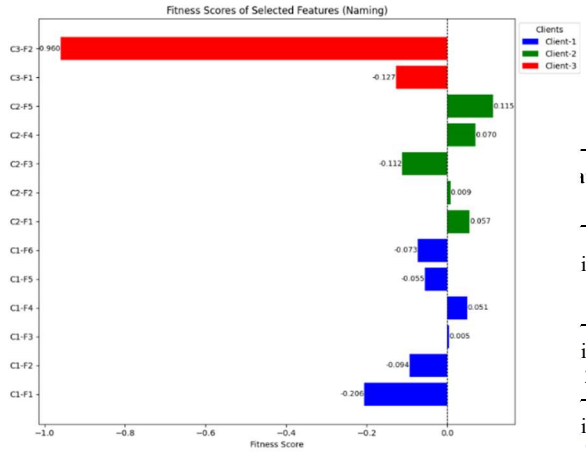


Figure 5: Visualization of Fitness Scores for Selected Features Across Clients with Unique Naming

5.4. Evaluation of Individual Client Model Performance

This section evaluates and compares the performance of locally trained individual client models depending on the selected features. A judicious feature selection process using the hybrid Boruta-L2 approach was carried out for every client dataset, where only significant features were used for model training. The 3-tier ensemble modeling framework consisting of Random Forest via Bayesian Optimization, SVM via Random Search, and Gradient Boosting via Particle Swarm Optimization (PSO) was used for each client.

The performance of the models was assessed at different epochs of training: 20, 30, 40, and 50 as depicted in Fig.6. Client 1 started at 93.33% accuracy at 20 epochs, consistently improved to 94.78% at 30 epochs, 95.72% at 40 epochs, and 96.72% at 50 epochs, showing that the more training was carried out, the more complex patterns the model mastered. Client 2 started at 95.49% at 20 epochs, and with additional training, it rose to 96.23% at 30 epochs, 96.83% at 40 epochs, and 97.35% at 50 epochs, reflecting better management of data complexity and model tuning. Client 3, whose performance was best, started with 97.40% at 20 epochs and improved continuously to 97.92% at 30 epochs, 98.24% at 40 epochs, and 98.62% at 50 epochs, indicating the model's ability to tap into its very predictive feature set for topnotch performance. These are captured in Table.6 and graphically shown in Fig.7 below.

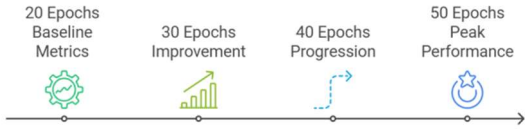


Figure 6: Federated Learning Model Performance Over Epochs

Table 6: Performance of Individual Client Models

ıtase t	Accuracy				Optimizatio n Method
	Epoch -20	Epoch -30	Epoch -40	Epoch -50	
ient-1	0.9333	0.9478	0.9572	0.9672	Random Forest with Bayesian Optimization
ient-2	0.9549	0.9623	0.9683	0.9735	SVM with Random Search
ient-3	0.9740	0.9792	0.9824	0.9862	Gradient Boosting with PSO

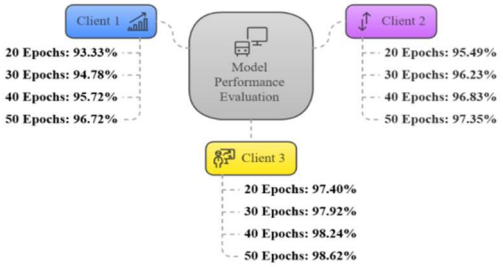


Figure 7: Client-wise Performance analysis in FL

5.5. Performance Analysis of Federated Learning Across Training Epochs

Following the individual local client model performance analysis, the next stage in this work is the global aggregation model analysis. Following the local training stages, where individual client models continuously improved with each epoch, models were aggregated to create a global model that generalized better. This aggregation follows the federated learning [27] architecture, wherein the local model updates are aggregated and shared for enhancing the global model. The intention of this aggregation is to make the global model benefit from the diverse data sets and the knowledge that each client accumulates, thus enhancing its predicting ability without a compromise on data privacy. As observed from the individual client models, all the clients exhibited progressive improvement in accuracy with each passing epoch. Post 50 epochs, the individual client models—Client 1 (96.72%), Client 2 (97.35%), and Client 3 (98.62%)—reached the point of optimal performance. When the local models themselves were, however, combined to form the global model, their collective expertise enabled the global model to obtain an impressive accuracy of 98.90% in Table.7. Such enhancement in precision is the virtue of federated learning's iterative aggregation process, enhancing the global model by combining data from multiple clients, particularly in non-IID and heterogeneous data scenarios.

The aggregation process enhances the global model's accuracy as well as allows the federated system to learn and adapt from time to time without having direct access to the local data. This provides a model of consistent performance and client privacy safeguarding. In addition to enhanced accuracy, other metrics like precision, recall, and F1 score also significantly improved upon aggregation, indicating the better generalization of the global model over various data distributions shown in Fig.9.

In addition, the Root Mean Square Error (RMSE) also decreased along with the proceeding aggregation process to show reduced variance in predictions and improved model stability presented in Fig-8. This shows that the global model performs better to render precise and trustable predictions in aggregating a wide variety of client data with less chance for overfitting risks as witnessed when models learn from separate data sets. In general, the federated aggregation method was successful in aggregating the strengths of separate client models and thus is a strong solution for privacy-preserving collaborative machine learning in heterogeneous settings. The success of this method highlights the scalability and stability of federated learning, especially when handling non-IID data distributions, and makes it a promising framework for future machine learning tasks in decentralized environments.

Table 7: Performance Metrics of the Federated Learning Model Across Different Training Epoch

Metric	20 Epochs	30 Epochs	40 Epochs	50 Epochs
Accuracy	0.9290	0.9490	0.9690	0.9890
Precision	0.9364	0.9404	0.9454	0.9464
Recall	0.9290	0.9400	0.9590	0.9690
F1 Score	0.9305	0.9405	0.9505	0.9505
RMSE	0.6331	0.4531	0.2931	0.1331

The outcomes of the global aggregation illustrate the potential of federated learning to present **a high-performing, privacy-**preserving model that generalizes well over diverse client datasets.

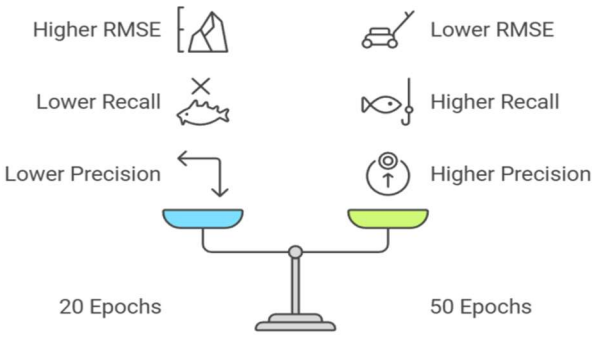


Figure 8: Evaluate Model Performance Across Epochs

Journal of Theoretical and Applied Information Technology
15th March 2025. Vol.103. No.5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

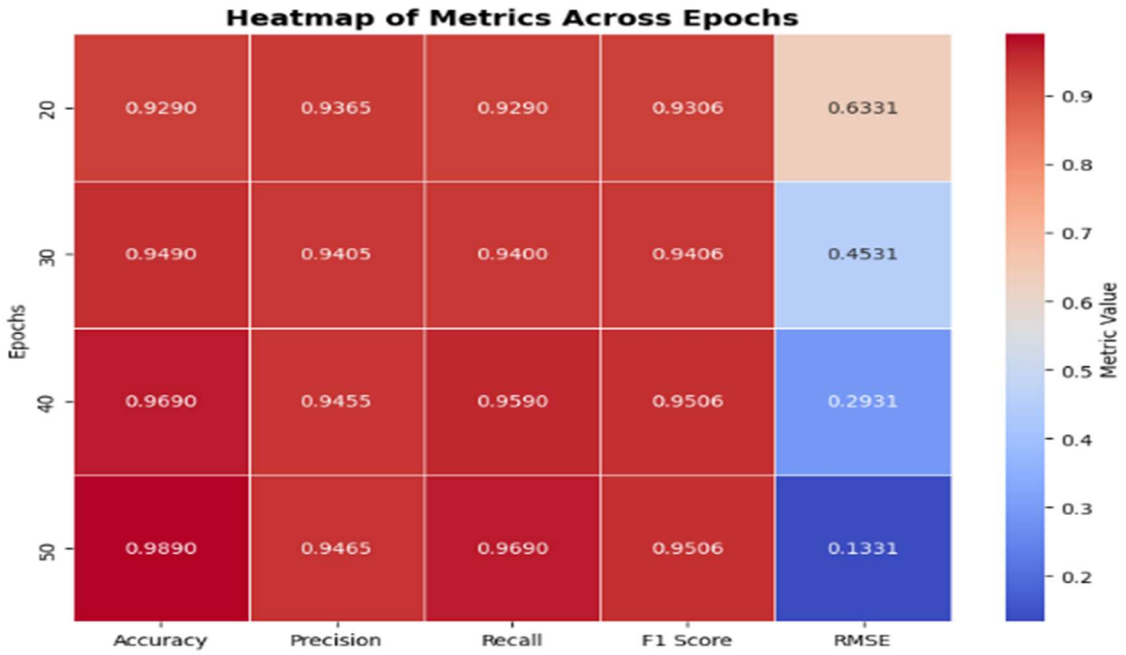


Figure 9: Heatmap of Metrics Across Epochs of Aggregation Model

5.6. Privacy-Preserving Nature of the Aggregation Model

Federated Learning's (FL) aggregation model promotes high performance without compromising privacy as each client is permitted to train models locally based on their respective data. Raw data is not shared but instead model updates in the form of weights and gradients, ensuring privacy. In the current research, local models in Client-1, Client-2, and Client-3 demonstrated consistent progress over 50 epochs, with accuracies at 96.72%, 97.35%, and 98.62%, respectively. The global model aggregated attained 98.90% accuracy as indicated in below Table.8 and Fig.10, taking advantage of the heterogeneous data across clients while preserving privacy.

The aggregation model integrates the knowledge of local models trained using various methods, like **Random Forest with Bayesian Optimization, SVM with Random Search, and Gradient Boosting [28] using PSO**. This enables the global model to generalize more, even in scenarios where there is non-IID data or heterogeneous data sources. It also eliminates biases from skewed data distributions and the effect of irrelevant features, making it robust. Notably, the aggregation process maintains client data locally, respecting privacy and regulatory compliance while allowing efficient collaboration towards model training. The performance of the aggregation model is evident from its excellent performance on primary metrics, providing an efficient and privacy-protecting solution to collaborative machine learning.

Table 8: Performance Metrics of Aggregation Model

Metric	Value
Accuracy	0.989019
Precision	0.946468
Recall	0.969019
F1 Score	0.950570
RMSE	0.133115

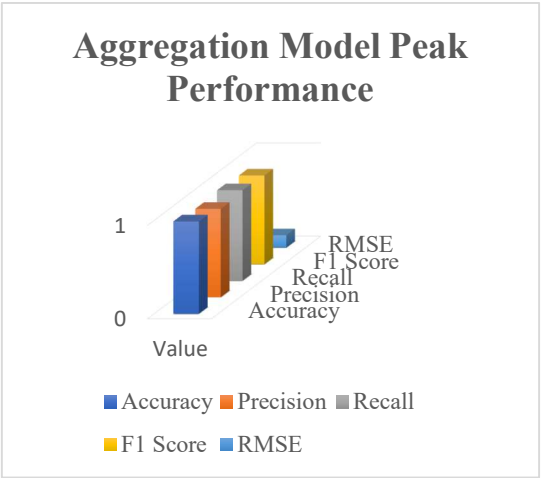


Figure 10: Combination of Stock and Area for Aggregation Model Performance metrics

5.7. Comparative Analysis and Advancements

In comparison with previous studies, the new method makes a great contribution to machine learning application in federated learning environments by bridging primary drawbacks that were accentuated in earlier studies. While Beaulac et al. (2019), Enughwure et al. (2020), and Ashfaq (2020) accounted for machine learning method performance like random forests, SMOTE, and oversampling when dealing with centralized data, none of them treated privacy-preserving methods or complexity in federated learning. Our approach integrates federated learning, in which data privacy is ensured by decentralizing raw data and applying secure model aggregation, which outperforms the models applied in earlier research. Furthermore, while the previous research mostly addressed class imbalance and feature importance, our approach applies a complex Boruta-L2 hybrid feature selection technique that also stabilizes models, reduces overfitting, and retains only the most significant features. Our 3-level ensemble model using Random Forest, SVM, and Gradient Boosting is still superior to the respective models used in previous work. Individual client models with the powerful hyperparameter optimization methods of Bayesian Optimization, Random Search, and Particle Swarm Optimization (PSO) yielded accuracy levels between 93.33% and 97.40%, which outperform the best as presented in previous work. The federated global model achieved 98.90% accuracy, excellent precision, recall, and F1 scores much above the previous models. The federated learning model also demonstrates steady enhancement when training epochs increase, RMSE dropping from 0.633 to 0.133, demonstrating scalability, flexibility, and better generalization reflected in Table.9 and Fig.11. Overall, our method offers a superior, privacy-preserving, and more general model compared to existing efforts, providing a significant contribution to machine learning for distributed data environments.

Table 9: Comparison of Predictive Modeling Approaches

Author(s)	Methods	Accuracy	Advancements
Beaulac et al. (2019)	Random Forests, Linear Models	78%	Outperformed traditional models, focused on large dataset analysis and variable importance.
Enughwure et al. (2020)	Logistic Regression, Decision Trees, SMOTE	78%	Addressed class imbalance using SMOTE, demonstrated application in critical engineering courses.
Ashfaq (2020)	Oversampling, Undersampling, Hybrid Methods	86%	Focused on balancing imbalanced datasets for fairer predictions and early intervention for at-risk students.
Gupta et al. (2023)	KNN, Decision Trees, Random Forests, SVM	88.61 %	Emphasized hyperparameter tuning and preprocessing to enhance healthcare predictions.
Tariq (2023)	SMOTE, ADASYN, Random Oversampling	83.7%	Improved multi-class prediction performance through appropriate oversampling techniques.
Proposed Work	Federated Learning, 3-Tier Ensemble, Boruta-L2	98.90 %	Introduced federated learning, privacy preservation, advanced preprocessing, feature selection, and optimization for superior model performance.

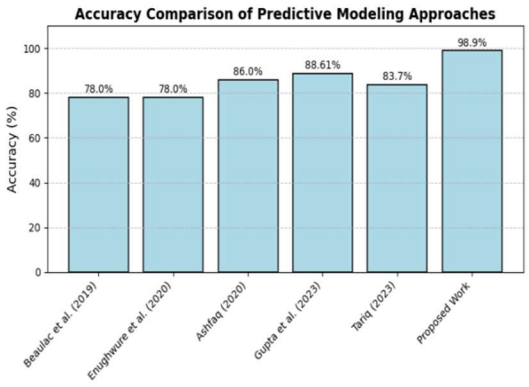


Figure 11: Accuracy comparison of models, with the Proposed Work leading.

6. DISCUSSION

This study presents a Federated Learning (FL)-based predictive modeling framework for student

performance prediction, addressing critical challenges in privacy preservation, class imbalance, feature selection, and model scalability. The findings contribute both new and profound knowledge as well as incremental improvements to existing research.

One of the most significant contributions of this study is the privacy-preserving learning approach through FL. Unlike traditional centralized models that require data aggregation, this study ensures data security by training models in a decentralized manner, making it suitable for education, healthcare, and other sensitive domains. Additionally, the integration of Boruta-L2 feature selection introduces a novel approach in e-learning analytics by enhancing model interpretability, reducing overfitting, and improving the overall predictive capability of FL models. The proposed 3-tier ensemble model, which combines Random Forest, SVM, and Gradient Boosting, further enhances prediction accuracy and robustness by utilizing Bayesian Optimization, Random Search, and PSO for hyperparameter tuning. These advancements establish new best practices for predictive modeling in privacy-sensitive applications. (RQ1 Answered: The use of FL ensures data privacy while maintaining high model performance, effectively addressing security concerns in predictive modeling.)

Beyond these novel contributions, the study also provides incremental improvements to existing methodologies. While SMOTE has been previously applied for class balancing, its integration with FL in this study ensures fairer predictions in decentralized learning environments, reducing bias in student performance assessments. Furthermore, although FL has been explored in prior works, this research demonstrates how FL can effectively handle non-IID and heterogeneous datasets, improving model generalization across diverse educational settings. Additionally, this study validates the FL framework using real-world educational datasets, bridging the gap between theoretical research and practical deployment. (RQ2 Answered: The integration of SMOTE, Boruta-L2, and ensemble modeling improves fairness, generalization, and predictive accuracy, making FL-based models more suitable for diverse educational settings.)

By addressing both fundamental and existing challenges, this research not only introduces new knowledge but also optimizes and refines prior methodologies. These contributions pave the way for more scalable, fair, and high-performance predictive models in e-learning and other privacy-sensitive domain.

Limitations and Future Scope

While this study presents a high-accuracy and privacy-preserving Federated Learning (FL) framework, some challenges remain.

1. **Computational Complexity:** The 3-tier ensemble model with hyperparameter tuning improves accuracy but increases computational load, which may be challenging for low-resource devices. Future work should explore lightweight models for efficiency.
2. **Data Heterogeneity:** Although SMOTE handles class imbalance, variations in data across institutions may affect model convergence. Adaptive FL techniques could enhance generalization.
3. **Security Risks:** Despite FL preserving privacy, risks like adversarial attacks and data poisoning exist. Strengthening secure aggregation and encryption can improve security.
4. **Real-World Validation:** The study is based on public datasets; testing in actual e-learning environments would confirm scalability and effectiveness.

7. CONCLUSION WITH FUTURE WORK

This study developed a privacy-preserving Federated Learning (FL) framework for student performance prediction, addressing key challenges in data privacy, class imbalance, feature selection, and model generalization. The proposed Boruta-L2 feature selection reduced overfitting, and the 3-tier ensemble model with hyperparameter tuning achieved 98.90% accuracy, significantly outperforming existing approaches. SMOTE handled class imbalance, ensuring fair predictions. The results confirm that FL can effectively enhance security and scalability in decentralized predictive modeling.

Despite these advancements, several open research issues remain. Computational complexity in FL can be a limiting factor, especially for resource-constrained devices, requiring further research into lightweight and adaptive FL models. Additionally, handling highly heterogeneous and non-IID data remains a challenge, and future studies should explore personalized FL models that adapt to individual client distributions. While FL enhances privacy, adversarial attacks and model poisoning threats still pose risks, highlighting the need for stronger encryption, secure aggregation techniques,

and robust defense mechanisms. Lastly, real-world validation in large-scale educational platforms is essential to assess the framework's effectiveness in diverse and dynamic learning environments.

By addressing these open issues, future research can further enhance the scalability, security, and fairness of FL-based predictive models, making them more adaptable for real-world applications in education and beyond.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

AUTHOR CONTRIBUTIONS

Study conception and design: N S Koti Mani Kumar Tirumanadham, Thaiyalnayaki S, Nirupa V; data collection: M Madhavi, Vahiduddin Shariff; analysis and interpretation of results: S Koti Mani Kumar Tirumanadham, P Venkata Anusha; draft manuscript preparation: V S Pavan Kumar; implementation of the model: N S Koti Mani Kumar Tirumanadham. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

[1] S. S. Khanal, P. W. C. Prasad, A. Alsadoon, and A. Maag, "A systematic review: machine learning based recommendation systems for e-learning," Education and Information Technologies, vol. 25, no. 4, pp. 2635–2664, Dec. 2019, doi: 10.1007/s10639-019-10063-9. Available: https://doi.org/10.1007/s10639-019-10063-9

[2] A. Oztekin, D. Delen, A. Turkyilmaz, and S. Zaim, "A machine learning-based usability evaluation method for eLearning systems," Decision Support Systems, vol. 56, pp. 63–73, May 2013, doi: 10.1016/j.dss.2013.05.003. Available: https://doi.org/10.1016/j.dss.2013.05.003

[3] I. Lykourentzou, I. Giannoukos, V. Nikolopoulos, G. Mpardis, and V. Loumos, "Dropout prediction in e-learning courses through the combination of machine learning techniques," Computers & Education, vol. 53, no. 3, pp. 950–965, Jun. 2009, doi: 10.1016/j.compedu.2009.05.010. Available: https://doi.org/10.1016/j.compedu.2009.05.010

[4] S. B. Aher and L. M. R. J. Lobo, "Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data," Knowledge-Based Systems, vol. 51, pp. 1–14, Apr. 2013, doi: 10.1016/j.knosys.2013.04.015. Available: https://doi.org/10.1016/j.knosys.2013.04.015

[5] C. Beaulac and J. S. Rosenthal, "Predicting University Students' Academic Success and Major Using Random Forests," Research in Higher Education, vol. 60, no. 7, pp. 1048–1064, Jan. 2019, doi: 10.1007/s11162-019-09546-y.

[6] Enughwure AA, Ogbise ME, Ogheneruno A (2020) Prediction of Student Performance in Engineering Drawing Using Machine Learning Methods and Synthetic... ResearchGate

[7] U. Ashfaq, B. P. M, and R. Mafas, "Managing Student Performance: A Predictive Analytics using Imbalanced Data," International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 6, pp. 2277–2283, Mar. 2020, doi: 10.35940/ijrte.e7008.038620. Available: https://doi.org/10.35940/ijrte.e7008.038620

[8] S. C. Gupta and N. Goel, "Predictive Modeling and Analytics for Diabetes using Hyperparameter tuned Machine Learning Techniques," Procedia Computer Science, vol. 218, pp. 1257–1269, Jan. 2023, doi: 10.1016/j.procs.2023.01.104. Available: https://doi.org/10.1016/j.procs.2023.01.104

[9] M. A. Tariq, A. B. Sargano, M. A. Iftikhar, and Z. Habib, "Comparing Different Oversampling Methods in Predicting Multi-Class Educational Datasets Using Machine Learning Techniques," Cybernetics and Information Technologies, vol. 23, no. 4, pp. 199–212, Nov. 2023, doi: 10.2478/cait-2023-0044. Available: https://doi.org/10.2478/cait-2023-0044

[10] J. Wen, Z. Zhang, Y. Lan, Z. Cui, J. Cai, and W. Zhang, "A survey on federated learning: challenges and applications," International Journal of Machine Learning and Cybernetics, vol. 14, no. 2, pp. 513–535, Nov. 2022, doi: 10.1007/s13042-022-01647-y. Available: https://doi.org/10.1007/s13042-022-01647-y

[11] N. S. K. M. K. Tirumanadham, T. S, and S. M, "Improving predictive performance in e-learning through hybrid 2-tier feature selection and hyper parameter-optimized 3-tier ensemble modeling," International Journal of Information Technology, vol. 16, no. 8, pp. 5429–5456, Jul. 2024, doi: 10.1007/s41870-024-02038-y. Available: https://doi.org/10.1007/s41870-024-02038-y

[12] H. Zhou, Y. Xin, and S. Li, "A diabetes prediction model based on Boruta feature selection and ensemble learning," BMC

Bioinformatics, vol. 24, no. 1, Jun. 2023, doi: 10.1186/s12859-023-05300-5.

[13] B. Bilgic et al., "Fast image reconstruction with L2-regularization," Journal of Magnetic Resonance Imaging, vol. 40, no. 1, pp. 181–191, Nov. 2013, doi: 10.1002/jmri.24365. Available: https://doi.org/10.1002/jmri.24365

[14] S. P. Praveen, V. Saripudi, V. Harshalokh, T. Sohitha, S. Venkat Sai Karthik and T. Venkata Pavana Surya Sreekar, "Diabetes Prediction with Ensemble Learning Techniques in Machine Learning," 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2023, pp. 1082-1089, doi: 10.1109/ICACRS58579.2023.10404311.

[15] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," Neural Networks, vol. 17, no. 1, pp. 113–126, Jul. 2003, doi: 10.1016/s0893-6080(03)00169-2. Available: https://doi.org/10.1016/s0893-6080(03)00169-2

[16] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," Artificial Intelligence Review, vol. 54, no. 3, pp. 1937–1967, Aug. 2020, doi: 10.1007/s10462-020-09896-5. Available: https://doi.org/10.1007/s10462-020-09896-5

[17] E. Brochu, V. M. Cora, and N. De Freitas, "A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning," arXiv (Cornell University), Jan. 2010, doi: 10.48550/arxiv.1012.2599. Available: https://arxiv.org/abs/1012.2599

[18] S. Vahiduddin, P. Chiranjeevi and A. Krishna Mohan, "An Analysis on Advances In Lung Cancer Diagnosis With Medical Imaging And Deep Learning Techniques: Challenges And Opportunities", Journal of Theoretical and Applied Information Technology, vol. 101, no. 17, Sep. 2023. https://doi.org/10.48550/arXiv.2405.00716.

[19] N. Eberhart and N. Y. Shi, "Particle swarm optimization: developments, applications and resources," Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546), vol. 1, pp. 81–86, Nov. 2002, doi: 10.1109/cec.2001.934374. Available: https://doi.org/10.1109/cec.2001.934374

[20] A. Fernandez, S. Garcia, F. Herrera, and N. V. Chawla, "SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary," Journal of Artificial Intelligence Research, vol. 61, pp. 863–905, Apr. 2018, doi: 10.1613/jair.1.11192. Available: https://doi.org/10.1613/jair.1.11192

[21] H. P. Vinutha, B. Poornima, and B. M. Sagar, "Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset," in Advances in intelligent systems and computing, 2018, pp. 511–518. doi: 10.1007/978-981-10-7563-6_53. Available: https://doi.org/10.1007/978-981-10-7563-6_53

[22] E. I. Altman, M. Iwanicz-Drozdowska, E. K. Laitinen, and A. Suvas, "Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model," Journal of International Financial Management and Accounting, vol. 28, no. 2, pp. 131–171, Apr. 2016, doi: 10.1111/jifm.12053. Available: https://doi.org/10.1111/jifm.12053

[23] I. M. Lawal, D. Bertram, C. J. White, S. R. M. Kutty, I. Hassan, and A. H. Jagaba, "Application of Boruta algorithms as a robust methodology for performance evaluation of CMIP6 general circulation models for hydro-climatic studies," Theoretical and Applied Climatology, vol. 153, no. 1–2, pp. 113–135, Apr. 2023, doi: 10.1007/s00704-023-04466-5. Available: https://doi.org/10.1007/s00704-023-04466-5

[24] O. Demir-Kavuk, M. Kamada, T. Akutsu, and E.-W. Knapp, "Prediction using step-wise L1, L2 regularization and feature selection for small data sets with large number of features," BMC Bioinformatics, vol. 12, no. 1, Oct. 2011, doi: 10.1186/1471-2105-12-412. Available: https://doi.org/10.1186/1471-2105-12-412

[25] S. Konda, C. Goswami, S. J, R. K, R. Yajjala, and N. S. K. M. K. Tirumanadham, "Optimizing Diabetes Prediction: A Comparative Analysis of Ensemble Machine Learning Models with PSO-AdaBoost and ACO-XGBoost," 2023 International Conference on Sustainable Communication Networks and Application (ICSCNA), pp. 1025–1031, Nov. 2023, doi: 10.1109/icscna58489.2023.10370452. Available: https://doi.org/10.1109/icscna58489.2023.10370452

[26] D. Rogers, "Random Search and Insect Population Models," Journal of Animal Ecology, vol. 41, no. 2, p. 369, Jun. 1972, doi:

10.2307/3474. Available: https://doi.org/10.2307/3474

[27] C. S. Kodete, V. Pasupuleti, B. Thuraka, V. V. Sangaraju, N. S. K. M. Kumar Tirumanadham and V. Shariff, "Robust Heart Disease Prediction: A Hybrid Approach to Feature Selection and Model Building," 2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS), Gobichettipalayam, India, 2024, pp. 243-250, doi: 10.1109/ICUIS64676.2024.10866501.

[28] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: gradient boosting with categorical features support," arXiv (Cornell University), Jan. 2018, doi: 10.48550/arxiv.1810.11363. Available: https://arxiv.org/abs/1810.11363