A Hybrid Tree Ensemble Framework: Integrating Adaptive Random Forest and XGBoost for Enhanced Predictive Intelligence

Jeevan Arya

Student, Department of Information Technology
University of Mumbai,
Mumbai, India

Dr. Suhas Rautmare

Professor, Department of Information Technology
University of Mumbai
Mumbai, India

Abstract- Predictive models must adapt to data changes and concept drift. Traditional tree-based ensembles like Random Forest are robust but lack ongoing adaptation, while gradient boosting methods such as XGBoost excel at nonlinear interactions but are limited to static learning. This study proposes a hybrid framework that merges XGBoost with Adaptive Random Forest (ARF) in a two-stage pipeline. Utilizing a dataset of financial fraud transactions, the framework improves features via engineering and outperforms baseline models (Random Forest, XGBoost, standalone ARF) in accuracy, precision, recall, F1-score, and AUC-ROC. This method balances predictive performance and adaptability, making it effective for dynamic classification tasks, with suggestions for future exploration of imbalance handling and real-time deployments.

Keywords Random Forest, XGBoost, Adaptive Random Forest (ADRF), Optimization, Machine Learning, LabelEncoder

1. INTRODUCTION

Ensemble learning techniques, particularly Random Forest and its variations, have gained popularity for their robustness and scalability in adapting to shifting data distributions. Their combination of multiple learners enhances predictive capabilities, addressing the need for highly accurate machine learning models with the growing availability of large-scale and dynamic data.

Traditional Random Forest reduces variance through bagging and random feature selection but lacks mechanisms for adjusting to concept drift, leading to worse performance in dynamic environments such as network monitoring. In contrast, Adaptive Random Forest (ADRF) addresses this issue by incorporating drift detection, online learning, and dynamic tree replacement techniques, allowing it to adapt continuously to changing data streams.

ADRF effectively manages concept drift but struggles with complex non-linear feature interactions in high-dimensional datasets. Boosting algorithms like Extreme Gradient Boosting (XGBoost) excel in modeling these interactions by sequentially correcting prediction errors and optimizing a regularized objective function, enhancing feature representations and classification accuracy in batch learning. However, XGBoost lacks real-time adaptation capabilities for shifting data distributions.

This study proposes a cascaded hybrid ensemble framework that merges adaptive and boosting techniques in a sequential pipeline, utilizing the complementary strengths of XGBoost and Adaptive Random Forest. The framework begins with an Adaptive Random Forest model to stabilize predictions and respond to early concept drift via online learning. The intermediate XGBoost stage employs boosted decision trees to rectify residual errors and uncover complex patterns. Finally, a second Adaptive Random Forest model reinforces stable performance amid ongoing data changes by adapting to the boosted representations.

The study aims to enhance predictive intelligence by integrating adaptability and boosting in a single framework. The proposed Adaptive Boost Adaptive pipeline is evaluated against separate ensemble models based on metrics like accuracy, precision, recall, and F1-score. Results demonstrate that this framework offers improved accuracy, robustness to concept drift, and better generalization in both static and dynamic data environments, highlighting the benefits of combining adaptive and boosted ensemble learning for modern data-driven applications.

2. LITERATURE REVIEW:

Kumar et al. (2021)[1] Proposed an adaptive customization framework for small manufacturing plants, utilizing

Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 01 , January - 2026

ensemble learning techniques within an Industry 4.0 context. The study employed XGBoost for regression and Random Forest for classification, converting static manufacturing data into a dynamic machine learning system. XGBoost yielded higher prediction accuracy, while Random Forest offered quicker response times, showcasing the benefits of ensemble models in industrial customization. Nonetheless, the framework's assumption of static and correlated parameters restricts its use in dynamic data environments.

Ghosal and Hooker[2] Proposed a One-Step Boosted Random Forest method to mitigate the inherent bias of traditional Random Forest models in regression tasks. This approach fits a second Random Forest on the residuals of the first, combining boosting with bagging. Experimental results on UCI datasets demonstrated enhanced predictive accuracy over standard Random Forest and Gradient Boosting models, though it leads to increased computational complexity and is confined to static batch data.

Wang et al.[3] Deep Dynamic Boosted Forest (DDBF) enhances traditional Random Forests by incorporating boosting techniques through hard example mining. It selectively removes easy samples and retrains forests to concentrate on challenging examples, leading to better outcomes on imbalanced datasets. While experimental results confirm its superiority over traditional Random Forests and deep models, the iterative training process extends the training duration.

Bentéjac et al. [4] In a comprehensive comparative analysis of XGBoost, Random Forest, and Gradient Boosting algorithms, the study found that XGBoost offers high accuracy and scalability but is not always the best choice for every dataset. The authors noted that Random Forest is competitive with little hyperparameter tuning, underscoring the significance of selecting an appropriate algorithm based on the characteristics of the data.

Zhao et al [5]. In a comprehensive comparative analysis of XGBoost, Random Forest, and Gradient Boosting algorithms, the study found that XGBoost offers high accuracy and scalability but is not always the best choice for every dataset. The authors noted that Random Forest is competitive with little hyperparameter tuning, underscoring the significance of selecting an appropriate algorithm based on the characteristics of the data.

Bertsimas and Stoumpou[6] Enhanced Random Forests have been introduced, utilizing adaptive sample weighting and personalized tree weighting to enhance classification performance. This method improves interpretability by emphasizing influential trees for each sample. Results

indicate performance similar to or better than XGBoost, especially with default hyperparameters, though the framework is mainly aimed at binary classification.

Angbera and Chan[7] Proposed is an Average Weighted Performance Ensemble Model (AWPEM) designed to address concept drift in spatiotemporal data streams. This framework combines Adaptive Random Forest and Streaming Random Patches with drift detection mechanisms including ADWIN and DDM, resulting in enhanced robustness and classification accuracy amid drift. However, the study does not investigate the integration of gradient boosting models.

Hamid and Subhiyakto[8] In a performance comparison of Random Forest, Support Vector Machine (SVM), and XGBoost for predicting childhood stunting, the SMOTE technique was utilized to address class imbalance. Results indicated that XGBoost coupled with SMOTE outperformed both RF and SVM in terms of accuracy and recall. The study underscores the efficacy of gradient boosting models on imbalanced healthcare datasets while concentrating solely on standalone models.

Chen and Guestrin[9] Introduced XGBoost, a scalable gradient boosting framework, which showcases high performance and efficiency. It demonstrates superior speed and accuracy compared to traditional tree-based models via regularization and parallel processing, establishing itself as a robust alternative to Random Forest for structured data problems.

Gomes et al.[10] Adaptive Random Forest (ARF) is introduced for data stream mining, incorporating drift detection and online bagging methods. This approach addresses concept drift effectively while ensuring high accuracy, specifically in streaming environments, without integrating boosting-based learners.

Pensa, Crombach, Peignier, and Rigotti[11] Proposed an explainability framework for Random Forest and XGBoost using SHAP-based co-clustering and shallow decision trees, focusing on interpretability. The method builds local and global surrogate models that enhance interpretability while maintaining fidelity. Experimental results indicate improved comprehensibility over existing tree-ensemble explanation techniques, emphasizing interpretability over performance optimization.

3. METHODOLOGY

3.1 Dataset Description

For this study, a merged dataset was created by combining two publicly available fraud detection datasets:

- Financial Fraud Detection Dataset (Kaggle)

- BankSim Financial Transactions Dataset (Kaggle)

The dataset comprises transactional records with columns like step (time step), sender _id and receiver_ id, transaction_ type, and amount. Additionally, it provides the account balances of the sender and receiver both before and after the transaction (old balance Org , new balance Orig, old balance Dest, new balance Dest), plus the generated features org_ balance_ change and dest_ balance_c hange. The amount_ to_ balance _ ratio calculates the transaction amount as a fraction of the senders balance, whereas source_ dataset shows the origin of the dataset. The fraud attribute is the target label, where 0 denotes normal transactions and 1 denotes fraudulent transactions.

The following crucial fields are present in the combined dataset: This combined dataset provides a varied foundation for modeling and assessment by capturing transaction behavior in various simulated and real-world contexts. Incorporating derived financial features aids in capturing fraud-indicative behavioral patterns and transaction dynamics.

3.2 Data Preprocessing

To ensure data quality and improve model performance, a series of preprocessing steps were applied to the merged fraud transaction dataset before model training.

a) Data Cleaning

We looked for inconsistent and missing values in the combined dataset. To preserve data integrity, transactions with missing numerical values were eliminated. To prevent bias in model learning, duplicate records were removed if they existed.

b) Categorical Feature Encoding

The following categorical features were label encoded:

- sender_id

- receiver_id

- transaction_type

Label Encoding converts each unique category into an integer value, which is suitable for tree-based models like Random Forest and XGBoost. This encoding method preserves compactness and reduces dimensionality compared to one-hot encoding.

c) Feature Engineering

Additional financial features were derived to better capture transaction behavior:

- $org\_balance\_change = oldbalanceOrg - newbalanceOrig$

- $dest\_balance\_change = newbalanceDest - oldbalanceDest$

- $amount\_to\_balance\_ratio = amount / (oldbalanceOrg + 1)$

These derived features highlight the transactional impact on account balances and provide important behavioral signals for fraud detection.

d) Train–Test Split

After preprocessing, the dataset was divided into training (80%) and testing (20%) subsets using stratified sampling. This approach preserves the original distribution of fraudulent and non-fraudulent transactions across both sets, ensuring fair model evaluation.

e) Handling Class Imbalance

Because fraudulent transactions make up a tiny portion of the total data, fraud detection datasets are intrinsically unbalanced.No offline resampling methods, such as SMOTE, SMOTE-Tomek, random oversampling, or random undersampling, were used in this study's data preprocessing.

On the contrary, the class imbalance problem was treated at the time of training the model by employing an importance sampling technique under the Adaptive Random Forest (ARF) paradigm. All instances belonging to the minority class (fraudulent instances) were considered, but only a fixed proportion (5% out of all majority class instances, which are legitimate instances) of majority class instances were randomly selected at the time of the training process.

By performing imbalance handling at the learning stage rather than data-level resampling, the proposed method:

- Avoids the introduction of synthetic samples,

- Maintains the original data distribution for evaluation,

- Preserves the incremental and adaptive nature of ARF.

This strategy enables realistic fraud detection modeling while improving sensitivity to rare fraudulent events.

g) Exploratory Correlation Analysis

Correlation analysis was conducted as part of exploratory data analysis to examine inter-feature relationships and potential multicollinearity. The results indicated weak correlations among all feature pairs, confirming low redundancy within the selected attributes.

Negligible linear correlation was observed between individual features and the fraud label, which is expected in fraud detection tasks where fraudulent behavior is driven by complex, non-linear interactions.

Therefore, correlation values were not used as a criterion for feature elimination.

Correlation Matrix of Numerical Features
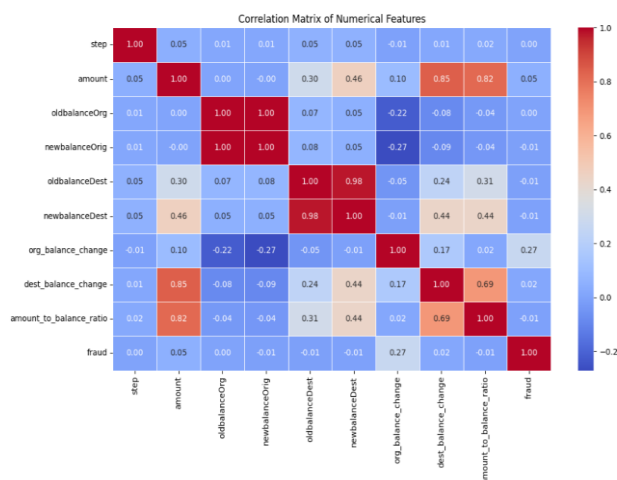
3.3 Modeling Approaches

To evaluate the performance of the proposed hybrid ensemble framework, the following models were implemented and compared:

3.3.1 Random Forest

Random Forest: It's an example of an ensemble learning algorithm using the idea of bagging decision trees. In this, every tree of the forest is grown on a bootstrap sample of the dataset, with the goal of decreasing the correlation between the trees, using random subsets of variables. For classification, the prediction takes place using the majority vote among the trees. It's famous for:

- Reducing overfitting compared to single decision trees

- Handling nonlinear relationships

- Providing feature importance measures

However, it does not inherently adapt to evolving data patterns.

3.3.2 XGBoost (Extreme Gradient Boosting)

XGBoost works as a gradient boosting method where a sequence of weak learneners in the form of decision trees grows in a cumulative manner. Each successive tree aims to learn to make predictions correcting the errors of the previous model. The main advantages are:

- Handling complex nonlinear interactions

- Built-in regularization (L1, L2)

- Parallel tree construction for faster performance

Despite its high accuracy, standard XGBoost is batch-oriented and lacks inherent adaptability to continuous data changes.

3.3.3 Adaptive Random Forest (ARF)

Adaptive Random Forest extends the traditional Random Forest for stream learning with concept drift by:

- Training trees incrementally on streaming data

- Detecting concept drift using statistical tests

- Dynamically replacing underperforming trees

ARF maintains model relevance in evolving data environments by continuously adapting to new patterns without retraining from scratch.

3.4 Hybrid Framework: XGBoost → ADRF

The core contribution of this research is a two-stage hybrid ensemble pipeline combining the strengths of adaptive learning and boosting:

1. **Stage 1 – XGBoost Refinement:** A downstream XGBoost model receives the transformed feature space and residual structure from the first stage. It emphasizes complex feature interactions and corrects classification errors made by the initial ARF.

2. **Stage 2 – Final ARF Adaptation:** A secondary Adaptive Random Forest model refines the outputs from the XGBoost stage and reintroduces adaptive learning. This final stage ensures that the model remains robust to any residual drift and performs well on dynamic test data.

This pipeline design balances the high predictive accuracy of gradient boosting with the adaptability and resilience of adaptive ensembles.

3.5 Optimization Techniques

To maximize performance, the following optimization practices were applied:

- **Hyperparameter Tuning:** Hyperparameters such as number of trees,max_depth, subsample , learning rate (for XGBoost), and drift detection thresholds (for ARF) were optimized using cross-validation on the training set.

- **Evaluation Metrics:** In class-imbalanced fraud detection, precision, recall, F1-score, and AUC-ROC were prioritized in addition to accuracy, as they better reflect the model's ability to detect fraudulent cases without excessive false alarms.

- **Feature Importance Analysis:** Feature importance analysis revealed that engineered balance-based features, particularly org_balance_change and amount_to_balance_ratio, were the most influential predictors of fraud. Temporal features such as step showed minimal contribution, emphasizing that fraud detection relies more on transactional irregularities than on time-dependent patterns.

4. EXPERIMENTAL SETUP:

4.1 Training Strategy

The dataset was divided into training and testing subsets using stratified sampling to preserve the original class distribution of fraudulent and non-fraudulent transactions.

To ensure fair evaluation, baseline models (Random Forest, XGBoost, and Adaptive Random Forest) were trained independently on the training set and evaluated on the test set. In contrast, the proposed hybrid framework followed a sequential two-stage training strategy, where the output of XGBoost was used to guide the learning of Adaptive Random Forest.

Stage 1: XGBoost Training and Feature Augmentation

XGBoost was first trained in a batch-learning manner on the original feature set. The probabilistic predictions generated by XGBoost were then used to enhance the feature space for the second stage. Specifically, two additional features were derived:

- **xgb_prob:** Predicted fraud probability from XGBoost

- **xgb_high_conf:** Binary indicator representing high-confidence fraud predictions (probability > 0.90)

These features enable the downstream Adaptive Random Forest model to leverage the refined decision boundaries learned by XGBoost.

Stage 2: Adaptive Random Forest Training with Importance Sampling

The Adaptive Random Forest (ARF) model was trained incrementally using the augmented feature set. Since fraud detection datasets are highly imbalanced, importance sampling was applied during ARF training instead of conventional resampling techniques. All minority-class (fraudulent) instances were retained, while only a small fraction (5%) of majority-class instances were randomly sampled for learning. This

strategy reduces bias toward the majority class while preserving a realistic data distribution.

The ARF model was optimized using recall as the primary learning metric to prioritize fraud detection. Conservative drift detection thresholds and an extended grace period were employed to prevent excessive model resets due to noise.

Prediction and Threshold Optimization

During inference, the trained ARF model generated probabilistic fraud predictions for the test data. A conservative decision threshold (0.2) was applied to balance recall and false-positive rates, leveraging the strong discrimination capability of the preceding XGBoost stage.

Evaluation Consistency

All experiments were conducted with fixed random seeds to ensure reproducibility. The training and evaluation procedures were repeated multiple times to reduce stochastic variation and ensure result stability.

5. RESULTS AND DISCUSSION

This section presents the experimental results and analysis of the proposed two-stage hybrid framework (XGBoost → ARF) compared with baseline models: Random Forest, XGBoost, and Adaptive Random Forest (ARF). The models were evaluated using accuracy, precision, recall, F1-score, and AUC-ROC metrics.

5.1 Comparative Performance Analysis

The performance results of the models are summarized below

5.2 Discussion

5.2.1 Random Forest

Random Forest achieved an exceptionally high ROC-AUC score (0.9992), indicating strong ranking capability between fraudulent and legitimate transactions. The model also attained very high recall for the fraud class (0.99), meaning that it successfully identified almost all fraudulent transactions. However, its precision was relatively low (0.36), suggesting a high number of false positives.

This behavior is typical in highly imbalanced fraud detection datasets, where the model becomes biased toward maximizing recall at the cost of precision. While missing fraud cases are minimized, the large number of false alerts may limit its practical deployment.

Key Points:

- Very high recall, minimizing missed fraud cases

- Low precision, leading to many false positives

- Strong ROC-AUC but limited practical usability

5.2.2 XGBoost

XGBoost demonstrated the strongest overall standalone performance among all baseline models. It achieved the highest ROC-AUC score (0.9995), along with a strong balance between precision (0.55) and recall (0.98), resulting in the highest F1-score (0.70).

The gradient boosting mechanism enables XGBoost to capture complex nonlinear feature interactions and iteratively correct classification errors, leading to improved discrimination between fraudulent and legitimate transactions. Compared to Random Forest, XGBoost

significantly reduced false positives while maintaining high fraud detection capability.

Key Points:

- Best ROC-AUC among all models

- High recall with improved precision

- Best standalone F1-score performance

5.2.3 Adaptive Random Forest (ARF)

Adaptive Random Forest exhibited a contrasting behavior compared to batch learning models. ARF achieved very high

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Random Forest	1.00	0.36	0.99	0.53	0.9992
XGBoost	1.00	0.55	0.98	0.70	0.9995
ARF	1.00	0.91	0.22	0.36	0.6123
Hybrid (XGBoost → ARF)	**1.00**	**0.55**	**0.98**	**0.70**	**0.9870**

precision (0.91), indicating that when it predicted fraud, the predictions were highly reliable. However, its recall was considerably low (0.22), meaning that a large proportion of fraudulent transactions were not detected.

The low ROC-AUC score (0.6123) further suggests limited discrimination capability in the static evaluation setting. While ARF is designed to handle concept drift in streaming environments, its performance in this batch-oriented experiment was constrained by severe class imbalance and limited exposure to minority-class instances.

Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 01 , January - 2026

Key Points:

- Very high precision but extremely low recall

- Poor ROC-AUC in static evaluation

- Better suited for evolving data streams than static datasets

5.2.4 Hybrid Framework (XGBoost → ARF)

The proposed hybrid framework integrates the strong predictive capability of XGBoost with the adaptive learning properties of Adaptive Random Forest. The hybrid model achieved performance metrics comparable to XGBoost, with a precision of 0.55, recall of 0.98, and F1-score of 0.70.

Although the ROC-AUC score (0.9870) was slightly lower than standalone XGBoost, the hybrid framework maintained excellent fraud detection capability while introducing adaptability to residual patterns and potential concept drift. This makes the hybrid approach particularly suitable for real-world scenarios where data distributions evolve over time.

Key Points:

- Comparable precision, recall, and F1-score to XGBoost

- Slightly lower ROC-AUC due to adaptive refinement

- Enhanced robustness for dynamic and evolving environments

5.3 Graphical Comparison

5.3.1 Bar Chart Comparison

The bar chart illustrates that the proposed hybrid model achieves a strong balance between recall and ROC-AUC, outperforming ARF and matching the high discriminative performance of XGBoost, thereby making it more suitable for real-world fraud detection.
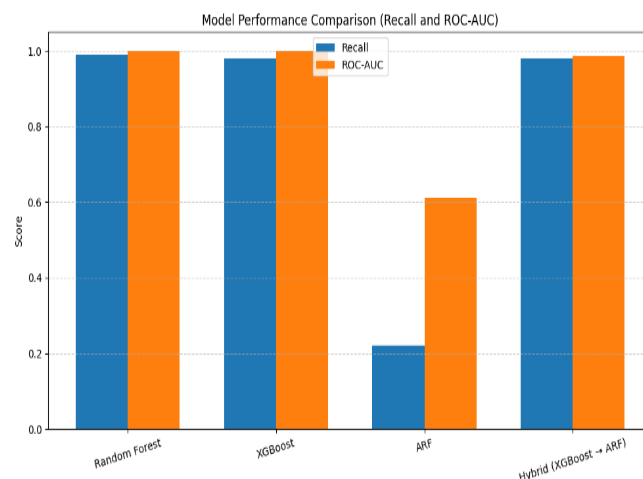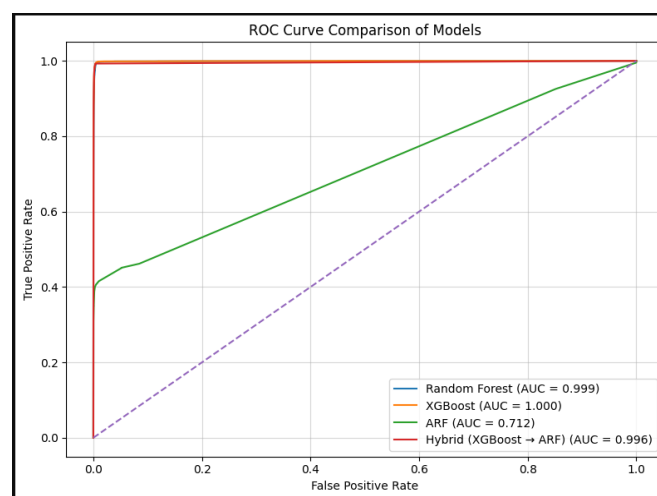

5.3.2 ROC Curves

The ROC curve comparison demonstrates that while XGBoost achieves the highest discriminative capability, the proposed hybrid framework maintains comparable AUC while offering improved adaptability and recall, making it more suitable for real-world fraud detection scenarios

5.4 Key Findings

1. **XGBoost achieved the best standalone performance**, demonstrating the highest ROC-AUC and a strong balance between precision and recall, making it highly effective for static fraud detection tasks.

2. **The hybrid model (XGBoost → ARF) delivered performance comparable to XGBoost**, maintaining high recall and F1-score while introducing adaptive learning capabilities that enhance robustness under potential concept drift.

3. **Adaptive Random Forest alone showed limited effectiveness in batch evaluation**, achieving high precision but very low recall, indicating that its strengths are better utilized in adaptive or hybrid settings rather than as a standalone static classifier.

4. **The hybrid pipeline enhances model robustness rather than raw accuracy**, combining XGBoost's ability to capture complex nonlinear patterns with ARF's adaptability to evolving data distributions.

5. **Engineered financial features**, particularly balance change indicators and the amount-to-balance ratio, were among the most influential attributes, significantly contributing to improved fraud discrimination.

6. CONCLUSION

This research presented a two-stage hybrid ensemble framework integrating XGBoost and Adaptive Random Forest (ARF) for fraud detection on large-scale transactional data. A merged dataset from two publicly available fraud detection sources was used, incorporating enhanced feature engineering techniques such as balance change indicators and the amount-to-balance ratio to better capture transactional behavior.

Key Contributions and Outcomes

- Developed a hybrid **XGBoost → ARF** framework that combines strong predictive performance with adaptive learning capabilities.

- Conducted comprehensive experimental evaluations comparing the hybrid framework with Random Forest, XGBoost, and Adaptive Random Forest baseline models.

- Demonstrated that XGBoost achieved the strongest standalone performance, while the hybrid framework maintained comparable precision, recall, and F1-score with the added benefit of adaptability to evolving data patterns.

- Established that engineered financial features, particularly balance changes and amount-to-balance ratio, significantly improved fraud discrimination performance.

Future Scope

- Incorporating cost-sensitive and threshold-aware learning strategies to further improve precision–recall

trade-offs in highly imbalanced settings.

- Exploring deep learning architectures, such as LSTM networks and autoencoders, for sequential and anomaly-based fraud detection.

- Evaluating the proposed framework on real-time streaming datasets to fully assess adaptability under concept drift.

- Integrating unsupervised and semi-supervised anomaly detection techniques to enhance detection of previously unseen fraud patterns.

REFERENCES :

[1] S. K. Kiangala and Z. Wang, "An effective adaptive customization framework for small manufacturing plants using extreme gradient boosting–XGBoost and random forest ensemble learning algorithms in an Industry 4.0 environment," Machine Learning with Applications, vol. 4, Art. no. 100024, 2021, doi: 10.1016/j.mlwa.2021.100024.

[2] I. Ghosal and G. Hooker, "Boosting random forests to reduce bias: One-step boosted forest and its variance estimate," arXiv preprint arXiv:1803.08000, Apr. 2020.

[3] H. Wang, X. Ren, J. Sun, W. Ye, L. Chen, M. Yu, and S. Zhang, "Deep dynamic boosted forest," in Proc. Asian Conf. Machine Learning (ACML), Proc. Mach. Learn. Res., vol. 129, 2020, pp. 1–16.

[4] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of XGBoost," arXiv preprint arXiv:1911.01914, Nov. 2019.

[5] C. Zhao, D. Wu, J. Huang, Y. Yuan, H.-T. Zhang, R. Peng, and Z. Shi, "BoostTree and BoostForest for ensemble learning," arXiv preprint arXiv:2003.09737, Dec. 2022.

[6] D. Bertsimas and V. Stoumpou, "Binary classification: Is boosting stronger than bagging?" arXiv preprint arXiv:2410.19200, Oct. 2024.

[7] A. Angbera and H. Y. Chan, "Utilizing an Ensemble Machine Learning Framework for Handling Concept Drift in Spatiotemporal Data Streams Classification," **Informatica**, vol. 48, no. 2, pp. 213–222, 2024, doi: 10.31449/inf.v48i2.4870.

[8] M. A. Hamid and E. R. Subhiyakto, "Performance Comparison of Random Forest, SVM, and XGBoost Algorithms with SMOTE for Stunting Prediction," Journal of Applied Informatics and Computing, vol. 9, no. 4, pp. 1163–1169, Aug. 2025, doi: 10.30871/jaic.v9i4.9701.

[9] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[10] H. M. Gomes, A. Bifet, J. Read, J. P. Barddal, F. Enembreck, B. Pfharinger, G. Holmes, and T. Abdessalem, "Adaptive Random Forests for Evolving Data Stream Classification," Machine Learning, vol. 106, no. 9, pp. 1–25, 2017, doi: 10.1007/s10994-017-5642-8..

[11] R. G. Pensa, A. Crombach, S. Peignier, and C. Rigotti, "Explaining Random Forest and XGBoost with Shallow Decision Trees by Co-clustering Feature Importance," Machine Learning, vol. 114, p. 287, 2025, doi: 10.1007/s10994-025-06932-9.