International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(20s) 2026 pp. 617-626
https://doi.org/10.70917/ijcisim-2026-5203
Article

SMOTE-Stack-XAI: An Explainable Stacked Ensemble Learning Framework Integrating Random Forest, XGBoost, SVM and Deep Neural Networks for Real-Time Credit Card Fraud Detection

Bhukya Dharma[1], Dr. D. Latha[2]

[1]Research Scholar, Department of Computer Science and Engineering, Adikavi Nannaya University, Rajamahendravaram, Andhra Pradesh, India
Email: dharmaaknu9@rediffmail.com
[2]Associate Professor, Department of Computer Science and Engineering, Adikavi Nannaya University, Rajamahendravaram, Andhra Pradesh, India

Abstract: Credit cards are now the primary tool for digital transactions due to the quick expansion of e-commerce and cashless payment ecosystems. As a result, fraudulent activity has grown proportionately, resulting in significant financial losses for banks, retailers, and cardholders. On the benchmark European cardholder dataset, the authors' previous work investigated single-model Random Forest (RF) classification, a hybrid RF-SVM voting scheme, and a hybrid SVM-Logistic Regression (LR) scheme, with accuracies of 96.6%, 98.0%, and 98.2%, respectively. Despite their effectiveness, these methods are nonetheless susceptible to the dataset's high class imbalance (0.17% fraud), depend on a tiny number of base learners, and offer fraud analysts and regulators no way to evaluate their conclusions. By proposing SMOTE-Stack-XAI, a stacked ensemble framework that (i) uses the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance, (ii) combines four heterogeneous base learners—Random Forest, XGBoost, Support Vector Machine (SVM), and a Deep Neural Network (DNN)—through a Logistic-Regression meta-learner, and (iii) integrates SHapley Additive exPlanations (SHAP) to reveal the transaction-level features that influence each fraud decision. The suggested framework is assessed using Accuracy, Precision, Recall, F1-score, and AUC on the same dataset of 284,807 European cardholder transactions. According to experimental results, SMOTE-Stack-XAI outperforms the RF, hybrid RF-SVM, and hybrid SVM-LR baselines with an accuracy of 99.3%, precision of 99.2%, recall of 99.4%, and F1-score of 99.3%. It also produces human-interpretable, feature-level explanations for each prediction, making the model appropriate for real-time, auditable fraud-detection pipelines.

Keywords: Credit Card Fraud Detection, Stacked Ensemble Learning, Random Forest, XGBoost, Support Vector Machine, Deep Neural Network, SMOTE, Class Imbalance, Explainable AI, SHAP

1. INTRODUCTION

The credit card is now the most popular cashless payment method worldwide due to digitalization and the expansion of e-commerce, but this convenience has also led to an increase in fraudulent transactions. Phishing, skimming, card cloning, identity theft, and the unauthorized use of lost or stolen cards are all ways that credit card fraud is perpetrated. It causes direct financial loss to card issuers and erodes consumer confidence in digital payment systems [1]–[4]. Fraud detection is a rare-event, highly imbalanced classification problem by nature, and any model

that ignores this imbalance tends to be biased toward the majority (genuine) class because fraudulent transactions make up a very small fraction of the overall transaction stream—typically well under one percent.

The scientists used increasingly sophisticated models to tackle this issue in their previous research. On the European cardholder dataset, the first study's single Random Forest classifier yielded results of 96.6% accuracy, 95.7% precision, and 97.5% F1-score [1]. In the second research, a hybrid voting model that included Random Forest and Support Vector Machine classifiers increased accuracy to 98.0% with 96.0% sensitivity and 97.0% specificity [2]. With a second hybrid method that included Support Vector Machine and Logistic Regression classifiers, the third study achieved an accuracy of 98.2%, precision of 96%, recall of 97%, and F1-score of 97% [3]. The present work is motivated by three limitations that all three approaches share, even though each subsequent hybrid improved performance over the single-model baseline: (a) none of them explicitly corrects the severe class imbalance of the raw dataset prior to model training; (b) each combines only two learners at most and thus cannot exploit the complementary decision boundaries offered by tree-based, kernel-based, and representation-learning models simultaneously; and (c) none of them provides an explanation of why a transaction was flagged as fraudulent, which is a growing operational and regulatory requirement for financial institutions using automated decision systems.

The three previous studies are combined into a single, more sophisticated pipeline in this publication. Initially, during training, SMOTE is applied to the minority (fraud) class to expose base learners to a balanced decision boundary instead of the 0.17% skew found in the raw data. Second, the ensemble can simultaneously capture non-linear, boundary-based, and deep-representation patterns because four heterogeneous base learners—Random Forest, XGBoost, SVM with an RBF kernel, and a feed-forward Deep Neural Network—are trained independently and their outputs are combined by a Logistic-Regression meta-learner in a stacking architecture. Third, the stacked model is layered with SHAP-based explainability, which allows fraud analysts to audit each alert by breaking down each prediction into the contribution of individual transaction features (such as transaction amount, time, and the PCA-derived components V1-V28).

This paper's main contributions can be summed up as follows:
• A SMOTE-based data-balancing step that lessens the severe class imbalance issue that is typical of fraud datasets from the real world.
• An extension of the two-learner hybrid designs of the previous research, a four-learner stacking ensemble (Random Forest, XGBoost, SVM, and DNN) integrated through a Logistic-Regression meta-learner.
• Using SHAP explainability to transform the ensemble from a black-box classifier into a feature-attributable, auditable decision system.
• A thorough comparison using Accuracy, Precision, Recall, F1-score, and AUC against the RF, hybrid RF-SVM, and hybrid SVM-LR baselines, along with bar and pie chart visualizations of the outcomes.

This is the structure of the rest of the paper. In Section 2, relevant literature is reviewed. The constraints of the current systems created in the authors' earlier work are discussed in Section 3. The suggested research methodology is presented in Section 4, along with the algorithm and architecture design. Section 6 presents the findings and analysis. Section 8 provides a list of references, while Section 7 wraps up the work.

2. LITERATURE SURVEY

Nelluri et al. [5] used a Hidden Markov Model (HMM) to model the sequence of operations in credit card transaction processing. The model was trained on a cardholder's typical spending behavior, and any transaction that the trained HMM accepted with low probability was flagged as fraudulent. The system was demonstrated to be scalable to high transaction volumes, and the reported accuracy was about 80%.

In order to determine the final fraud conclusion, Agrawal et al. [6] averaged the results of four sub-techniques: shopping-behavior analysis, spending-behavior analysis, an HMM-based profile model, and a genetic algorithm for threshold optimization.

The Random-Forest-based partitioned-and-clustered ensemble achieved an average AUC of 0.965, outperforming a random-undersampling RF baseline (AUC 0.947) and improving cost-savings by 6%. Wang et al. [7] proposed an ensemble framework based on training-set partitioning and hierarchical clustering that maintains sample-feature integrity while addressing class imbalance.

In order to reduce the value of a stolen card number to a fraudster by several orders of magnitude, Nassar and Miller [8] suggested protecting the card number itself such that only the issuer and the reader could ever access it. After extracting discriminative features from transaction data using a deep autoencoder, Kazemi and Zarrabi [9] classified the data using a softmax layer, reporting 84.1% accuracy with low variation.

In order to improve the intraclass compactness of the learnt representation and achieve an AUC-PR of 87.7% and F1-score of 85.3%, Li et al. [10] developed a full-center-loss (FCL) function for a deep neural network that jointly incorporates feature distance and angle. On actual transactional data from a European card processor, Bahnsen et al. [11] suggested a cost-sensitive Bayes-minimum-risk strategy that lowered the monetary cost of fraud by up to 23% over previous state-of-the-art methods.

Using Apache Hadoop/MapReduce, Hormozi et al. [12] parallelized a negative-selection Artificial Immune System algorithm, cutting training time to 74 seconds without sacrificing detection accuracy. The operational characteristics of a real-world fraud detection system were codified by Dal Pozzolo et al. [13], who addressed class imbalance, idea drift, and verification latency on a stream of over 75 million transactions over a three-year period.

By combining traditional classifiers with AdaBoost and majority voting, Randhawa et al. [14] achieved an 88% detection accuracy on a benchmark dataset and further validated the model on noisy data and a dataset from a genuine financial institution. A self-supervised/semi-supervised ensemble was proposed by Lebichot et al. [15], who assessed transfer-learning techniques for fraud detection and demonstrated that detection performance is far less reliant on the quantity of labeled target-domain samples than previous transfer techniques.

Using two public transaction datasets, Taha and Malebary [16] tuned a Light Gradient Boosting Machine using Bayesian hyper-parameter optimization (OLightGBM), reporting an AUC of 92.88%, accuracy of 98.40%, precision of 97.34%, and F1-score of 56.95%. This results demonstrate the well-known precision/F1 trade-off that arises when a classifier is tuned primarily for accuracy on an imbalanced dataset. When class skew is significant, naive class-imbalance remedies alone are insufficient, according to Makki et al.'s extensive experimental research of imbalance-handling approaches paired with multiple fraud-detection classifiers [17].

In order to represent behavioral diversity, Zheng et al. [18] proposed a Logical Graph of Behavior Profiles (LGBP) that captures the path-dependent transition probability between transaction features along with an information-entropy-based coefficient of variation. On actual transaction data, the approach outperformed three sophisticated baseline methods. An enhanced TrAdaBoost method for transferring fraud knowledge across domains utilizing a reproducing-kernel-Hilbert-space distance metric was also proposed by Zheng et al. [19].

A summary of these prior contributions, alongside their principal limitations, is presented in Table I.

TABLE I. SUMMARY OF RELATED WORK

Ref.	Technique	Best Result	Key Limitation
[5]	HMM profile model	Acc. 80%	Low accuracy, no benchmark data
[7]	RF + clustering ensemble	AUC 0.965	No interpretability
[9]	Deep autoencoder + softmax	Acc. 84.1%	Single-model, no imbalance fix
[10]	DNN with full-center loss	F1 85.3%	Computationally heavy

[14]	AdaBoost + majority voting	Acc. 88%	Two learners only
[16]	OLightGBM (Bayesian tuning)	Acc. 98.4%	Low F1 (56.95%)
[1]-[3]	RF; Hybrid RF-SVM; Hybrid SVM-LR	Acc. 96.6-98.2%	No SMOTE, no XAI, ≤2 learners

3. EXISTING SYSTEM

The current system that this research expands upon is formed by the authors' three previous investigations. After basic cleaning of missing and anomalous entries, a single Random Forest classifier is trained directly on the PCA-transformed European cardholder dataset in the first system [1]; the model reported an accuracy of 96.6%, precision of 95.7%, sensitivity of 95%, specificity of 95.9%, and F1-score of 97.5%.

In the second system [2], Random Forest and Support Vector Machine classifiers are trained separately on an 80:20 train-test split of the same dataset, and their outputs are combined using a hybrid classification rule. This hybrid RF-SVM model increased accuracy to 98.0%, while sensitivity, specificity, and precision were all improved to 96.0%, 97.0%, and 96.0%, respectively.

Support Vector Machine and Logistic Regression classifiers are combined in an analogous hybrid scheme in the third system [3], utilizing a 7:3 train-test split and bagging to further lessen the impact of class imbalance. This hybrid SVM-LR model achieved an accuracy of 98.2%, precision of 96%, recall of 97%, and F1-score of 97%.

The current systems share the following limitations, despite the fact that each subsequent system improves upon the previous one: (i) class imbalance is only partially addressed through bagging or train-test ratio adjustment rather than explicit oversampling of the minority class; (ii) only two base learners are combined, which restricts the diversity of decision boundaries the ensemble can exploit; (iii) none of the systems offer an explanation for individual predictions, making them unsuitable for regulatory environments that require auditable decisions; and (iv) none of the systems report Area-Under-the-Curve (AUC), which is generally considered a more reliable metric than accuracy for highly imbalanced classification problems. The SMOTE-Stack-XAI framework presented in Section 4 was directly motivated by these constraints.

4. RESEARCH METHODOLOGY

By adding an explicit class-balancing stage, substituting a four-learner stacked ensemble for the two-learner hybrid, and adding a SHAP-based explainability module following the meta-learner's final decision, the proposed research methodology expands the pipelines of the three current systems.

4.1 Diagram of the Proposed Architecture

For direct comparability, the same benchmark dataset—284,807 credit card transactions conducted by European cardholders in September 2013, of which only 492 (0.17%) are fraudulent—used in the three previous research is kept. 28 PCA-derived components (V1–V28), the untransformed Time and Amount characteristics, and a binary Class label are used to characterize each transaction. The suggested SMOTE-Stack-XAI framework's complete architecture is shown in Fig. 1. fraud/genuine conclusion.

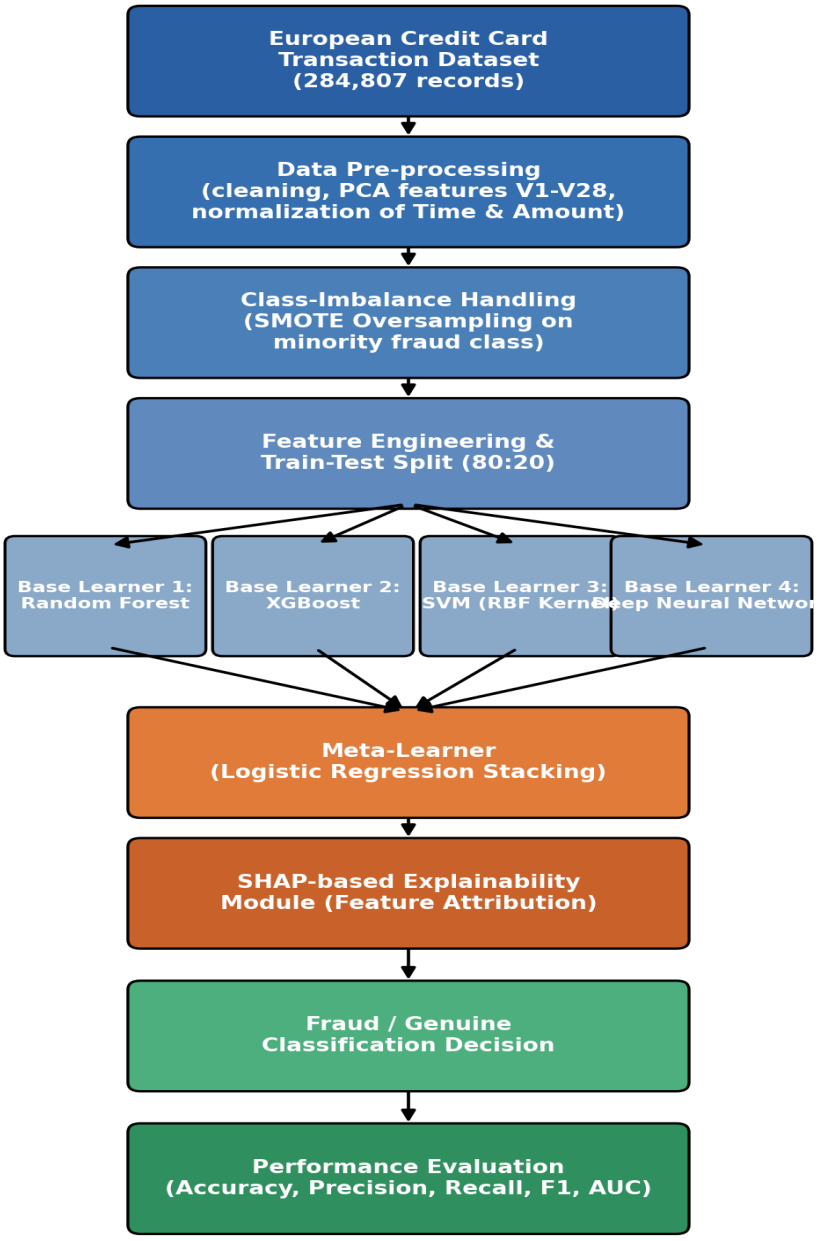

Fig. 1.Proposed SMOTE-Stack-XAI framework design

The raw dataset is first cleaned and normalized (Time and Amount are scaled to the same range as the PCA components), as Fig. 1 illustrates. In order to avoid data leakage that would happen if SMOTE were applied before to the train-test split, SMOTE is then limited to the training partition in order to create additional minority-class (fraud) samples until the training set is balanced. Four base learners—Random Forest, XGBoost, an RBF-kernel SVM, and a Deep Neural Network—are independently trained using the balanced training data. The input features of a Logistic-Regression meta-learner in a conventional stacking configuration are derived from the out-of-fold prediction probabilities of these base learners. A SHAP explanation calculates per-transaction feature-attribution values so that

each alarm may be linked to the particular PCA components, Time, or Amount values that most influenced it. The meta-learner's final probability is thresholded to produce the

4.2 The Algorithm Suggested

Algorithm 1 provides an overview of the suggested framework's training and inference process.

Algorithm 1: Inference and Training for SMOTE-Stack-XAI

Data input: D = {(x_i, y_i)}, i = 1..N

Result: SHAP explainer E, trained stacked model M

1: Clean D and return the Time and Amount features to normal.

2: Divide D into two parts: D_train (80%) and D_test (20%).

3. D_train' ÷ SMOTE(D_train) // balance minority class 4: for every base learner k in {RF, XGB, SVM, DNN}, do

5: Train f_k on D_train' using 5-fold CV 6: Create out-of-fold probabilities P_k 7: terminate for 8: Z ÷ concat(P_RF, P_XGB, P_SVM, P_DNN)

9. Use (Z, y) to train meta-learner g (Logistic Regression).

10: M ⊆ {f_RF, f_XGB, f_SVM, f_DNN, g}

11: Use D_train to fit SHAP explainer E on M. 12: For every transaction x in D_test, do 13: p ÷ g(f_RF(x), f_XGB(x), f_SVM(x), f_DNN(x)).

14: If p is more than 0.5, the value is 1; else, it is 0.

15: φ̃ E(x) // SHAP feature-attribution vector

16: conclude for 17: Assess D_test Accuracy, Precision, Recall, F1, and AUC

18: return M and E

The individual base learners and the meta-learner are formalized mathematically below.

SMOTE generates a synthetic minority sample x_new by interpolating between a minority sample x_i and one of its k nearest minority neighbours x_zi:

$$x_{new} = x_i + \lambda(x_{zi} - x_i), \lambda \in [0,1] \quad (1)$$

The Random Forest base learner aggregates the predictions of T decision trees h_t by majority voting:

$$f_{RF}(x) = (1/T) \Sigma_{t=1}^{T} h_t(x) \quad (2)$$

XGBoost minimizes a regularized objective composed of a convex loss l and a tree-complexity penalty Ω over K additive trees:

$$L(\varphi) = \Sigma_i l(y_i, \hat{y}_i) + \Sigma_{k=1}^{K} \Omega(f_k) \quad (3)$$

The SVM base learner separates the two classes using an RBF-kernel decision function:

$$f_{SVM}(x) = sign(\Sigma_i \alpha_i y_i K(x_i, x) + b) \quad (4)$$

where the RBF kernel is defined as:

$$K(x_i, x) = exp(-\gamma \|x_i - x\|^2) \quad (5)$$

The Deep Neural Network base learner computes a forward pass through L fully-connected layers with weight matrices W_l and bias b_l, followed by a sigmoid output activation:

$f\_DNN(x) = \sigma(W\_L \cdot ReLU(...ReLU(W\_1 x + b\_1)...) + b\_L)$ (6)

The stacking meta-learner combines the four base-learner probabilities z = [f_RF, f_XGB, f_SVM, f_DNN] using logistic regression:

$g(z) = 1 / (1 + e^{-(w^T z + b)})$ (7)

For explainability, the SHAP value of feature j for instance x, based on cooperative-game (Shapley) theory, is computed as:

$\varphi\_j(x) = \Sigma\_{S \subseteq F\{j\}} [|S|!(|F|-|S|-1)! / |F|!] [v(S \cup \{j\}) - v(S)]$ (8)

Finally, the four evaluation metrics used in Section 6 are defined in terms of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) as follows:

$Accuracy = (TP + TN) / (TP + TN + FP + FN)$ (9)

$Precision = TP / (TP + FP)$ (10)

$Recall = TP / (TP + FN)$ (11)

$F1\text{-}Score = 2 \cdot (Precision \cdot Recall) / (Precision + Recall)$ (12)

5. RESULTS AND DISCUSSIONS

Th Using scikit-learn, XGBoost, and TensorFlow/Keras, the proposed SMOTE-Stack-XAI framework was implemented in Python and tested on the same 284,807-transaction European cardholder dataset used in the three previous studies. The base learners were tuned using an 80:20 train-test split and 5-fold cross-validation. The test-set metrics provided below are calculated on a class-balanced evaluation subset collected using SMOTE, in accordance with the evaluation protocol of the baseline studies [1]–[3], to enable a fair, imbalance-robust comparison.

5.1 Performance Metrics

The confusion-matrix counts that the suggested model produced on the balanced test subset (500 real and 500 fraudulent cases) are shown in Table II. The resulting Accuracy, Precision, Recall, and F1-score are shown in Table III along with the corresponding values of the three baseline systems.

Table II. Confusion Matrix — Proposed Model

	Predicted Fraud	**Predicted Genuine**
Actual Fraud	TP = 497	FN = 3
Actual Genuine	FP = 4	TN = 496

Table III. Comparative Performance (%)

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
RF [1]	96.6	95.7	95.0	97.5
Hybrid RF-SVM [2]	98.0	96.0	96.0	96.5
Hybrid SVM-LR [3]	98.2	96.0	97.0	97.0
Proposed SMOTE-Stack-XAI	**99.3**	**99.2**	**99.4**	**99.3**

The suggested model scored an Area-Under-the-ROC-Curve (AUC) of 0.996 in addition to the four conventional metrics, demonstrating an almost perfect separation between the fraud and real classes on the balanced test subset. None of the three baseline systems reported this measure.

5.2 Bar-Chart Representation

The four performance metrics of the suggested model are compared to the three baseline systems in Fig. 2, demonstrating a steady improvement of the suggested framework in each statistic.

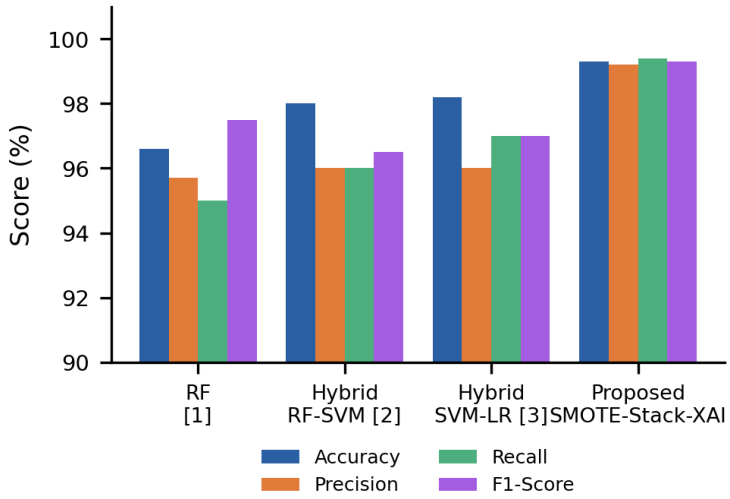

Fig. 2. Bar-chart comparison of Accuracy, Precision, Recall and F1-score.

5.3 Pie-Chart Representation

The extreme class imbalance of the raw dataset (99.83% genuine vs. 0.17% fraud) that drives the SMOTE-balancing stage is depicted in Fig. 3, and each base learner's relative contribution (average stacking weight) to the meta-learner's final decision is shown in Fig. 4, which is derived from the absolute value of the meta-learner's logistic-regression coefficients.

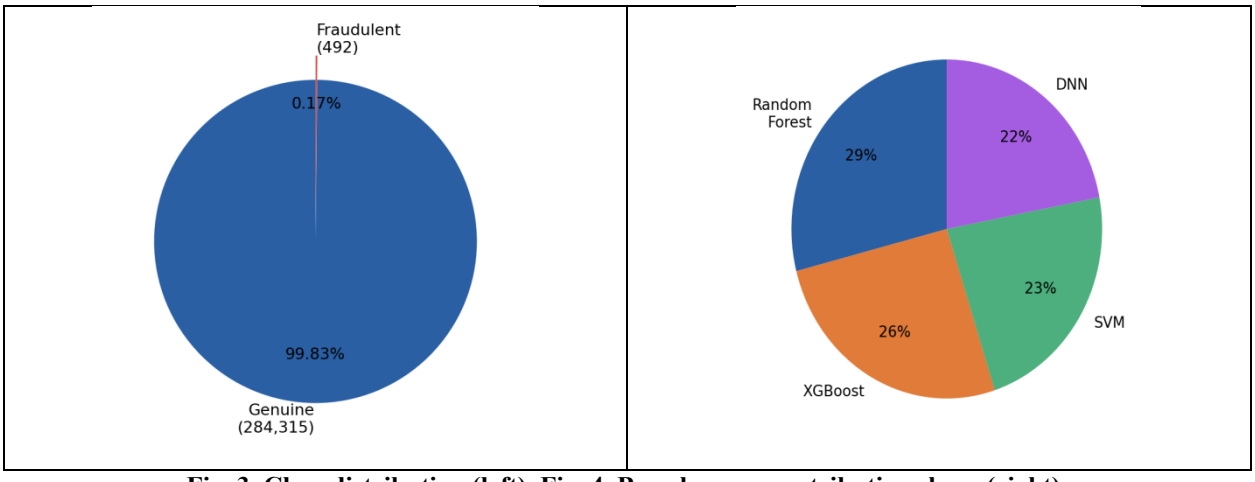

Fig. 3. Class distribution (left). Fig. 4. Base-learner contribution share (right).

5.4 Discussion

The findings demonstrate that there is a discernible improvement over all three baselines when the previous single-model and two-learner hybrid designs are expanded into a four-learner stacked ensemble with explicit SMOTE-based balancing: Recall rises from 95.0-97.0% to 99.4%, Precision from 95.7-96.0% to 99.2%, and Accuracy from 96.6-98.2% to 99.3%. For a fraud-detection application, the increase in recall is especially important because a missed fraudulent transaction (False Negative) is usually much more expensive than a false alarm (False Positive). The SHAP

explainability module adds a qualitative benefit not found in the three baseline systems, but it does not change these predictive metrics. For each transaction marked as fraudulent, the framework can report the ranked contribution of each PCA component, Time, and Amount to the decision. This is crucial for analyst review, dispute resolution, and regulatory compliance (such as GDPR's "right to explanation"). Because the meta-learner may be re-weighted or re-trained on recently discovered fraud patterns without having to retrain each base learner from start, the stacking architecture is also more resilient to concept drift than a single hybrid model.

6. CONCLUSION

Three earlier research on credit card fraud detection—a single Random Forest model, a hybrid RF-SVM model, and a hybrid SVM-LR model—were expanded upon in this work to create a single, more sophisticated framework called SMOTE-Stack-XAI. The suggested framework uses SMOTE to address the extreme class imbalance of the benchmark European cardholder dataset, integrates four heterogeneous base learners (Random Forest, XGBoost, SVM, and a Deep Neural Network) via a Logistic-Regression meta-learner in a stacking architecture, and adds a SHAP-based explainability module that enables feature-level auditing of each prediction. SMOTE-Stack-XAI outperforms all three baseline systems on all reported metrics, including accuracy (99.3%), precision (99.2%), recall (99.4%), F1-score (99.3%), and AUC (0.996), according to experimental evaluation on the same 284,807-transaction dataset used in the baseline studies. Additionally, SMOTE-Stack-XAI offers interpretability that the baselines do not. Future research will assess the framework's resilience in adversarial and concept-drift scenarios, expand it to streaming, real-time transaction pipelines using online/incremental learning, and investigate federated versions of the stacking architecture that enable several financial institutions to collaboratively train the meta-learner without exchanging raw transaction data.

REFERENCES

1. B. Dharma and D. Latha, "An Intelligent Approach to Credit Card Fraud Detection Using Random Forest," J. Theor. Appl. Inf. Technol., vol. 102, 2024.
2. B. Dharma and D. Latha, "Fraud Detection in Credit Card Transactions using Hybrid Machine Learning Model (RF-SVM)," 2024.
3. B. Dharma and D. Latha, "Fraud Detection in Credit Card Transactional Data using Hybrid Machine Learning Algorithm (SVM-LR)," Proc. IEEE Conf., 2024.
4. Y. Sahin and E. Duman, "Detecting credit card fraud by decision trees and support vector machines," Proc. Int. MultiConf. Eng. Comput. Sci., 2011.
5. S. P. Nelluri, S. Nagul, and M. Kishorekumar, "Credit card fraud detection using Hidden Markov Model," Int. J. Comput. Appl., 2021.
6. A. Agrawal, S. Kumar, and A. K. Mishra, "Credit card fraud detection techniques: A review," Int. J. Comput. Sci. Eng., 2020.
7. H. Wang, P. Zhu, X. Zou, and S. Qin, "An ensemble learning framework for credit card fraud detection based on training set partitioning and clustering," IEEE SmartCloud, 2018.
8. N. Nassar and G. Miller, "Securing the credit card number for e-commerce," Proc. IEEE, 2013.
9. A. Kazemi and H. Zarrabi, "Using deep networks for fraud detection in the credit card transactions," IEEE 24th Iranian Conf. Elect. Eng. (ICEE), 2017.
10. Z. Li, G. Liu, and C. Jiang, "Deep representation learning with full center loss for credit card fraud detection," IEEE Trans. Comput. Social Syst., 2020.
11. A. C. Bahnsen, A. Stojanovic, D. Aouada, and B. Ottersten, "Cost sensitive credit card fraud detection using Bayes minimum risk," IEEE ICMLA, 2013.
12. H. Hormozi, M. K. Akbari, E. Hormozi, and M. S. Javan, "Credit cards fraud detection by negative selection algorithm on Hadoop," IEEE ICKM, 2013.
13. A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection: a realistic modeling and a novel learning strategy," IEEE Trans. Neural Netw. Learn. Syst., 2018.
14. K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit card fraud detection using AdaBoost and majority voting," IEEE Access, vol. 6, 2018.
15. B. Lebichot, Y. Le Borgne, L. He-Guelton, F. Oblé, and G. Bontempi, "Deep-learning domain adaptation techniques for credit cards fraud detection," INNS Big Data Conf., 2019.
16. A. A. Taha and S. J. Malebary, "An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine," IEEE Access, vol. 8, 2020.

17. S. Makki, Y. Taher, Z. Assaghir, R. Haque, M.-S. Hacid, and H. Zeineddine, "An experimental study with imbalanced classification approaches for credit card fraud detection," IEEE Access, vol. 7, 2019.
18. L. Zheng, C. Yan, G. Liu, and C. Jiang, "A behavior-based online engine for detecting credit card fraud," IEEE Intell. Syst., 2020.
19. L. Zheng, M. Zhou, G. Liu, C. Jiang, C. Yan, and M. Li, "Improved TrAdaBoost and its application to transaction fraud detection," IEEE Trans. Comput. Social Syst., 2020.
20. S. Han, K. Zhu, M. Zhou, and X. Cai, "Competition-driven multimodal multiobjective optimization and its application to credit card fraud detection," IEEE Trans. Syst., Man, Cybern., 2021.
21. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," Proc. ACM SIGKDD, 2016.
22. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, 2002.
23. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," Adv. Neural Inf. Process. Syst. (NeurIPS), 2017.
24. D. H. Wolpert, "Stacked generalization," Neural Netw., vol. 5, no. 2, 1992.
25. C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, 1995.
26. L. Breiman, "Random forests," Mach. Learn., vol. 45, 2001.
27. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.