The American Journal of Interdisciplinary Innovations and Research
ISSN 2642-7478 Volume 08 - 2026

Explainable AI driven Credit Risk Assessment for Basel III compliant Banking Risk Management

Aritra Pal

FIRST CITIZENS BANK
Charlotte, North Carolina, USA

Received: 07 Aug 2026 | Received Revised Version: 24 Aug 2026 | Accepted: 07 Sep 2026 | Published: 26 Sep 2026

Volume 08 Issue 09 2026 | DOI: 10.37547/tajiir/Volume08Issue09-06

Abstract

The effective identification of potential loan defaults is crucial for banking risk management, enabling informed lending decisions and financial stability. In this paper, a credit risk assessment framework that integrates both predictive modeling and risk interpretability is developed, which leverages Explainable AI (XAI) to enhance banking risk management to meet Basel III requirements. The proposed approach utilizes the probability outputs of the Random Forest (RF) and XGBoost (XGB) models which are combined using a Hybrid Model of Random Forest-XGBoost (RF-XGB). The Kaggle Credit Risk Dataset is preprocessed using duplicate remover, missing value imputer, outlier capping, categorical encoding, stratified data splitting and SMOTE data balancing. The performance of the Hybrid RF-XGBoost model is validated through experimental results with accuracy of 93.14%, precision of 94.19%, recall of 73.13%, F1 of 82.33%, ROC-AUC of 93.28%, and PR-AUC of 88.48%, for the default prediction task. In order to increase transparency of the model, explanations of individual predictions and influential credit-risk factors are provided by means of SHAP and LIME. The predicted Probability of Default (PD) is then translated to five risk categories (Very Low Risk through Very High Risk) and combined with Loss Given Default (LGD), Exposure at Default (EAD), Expected Loss (EL), risk-weighted assets (RWA) and capital requirements. Additionally, a stress testing process is conducted to assess for possible rise in credit losses during adverse conditions. The framework therefore offers a holistic solution which integrates predictive accuracy with explanations, borrower-level risk stratification and Basel III oriented risk assessment. The findings suggest that combining XAI with ensemble credit risk models can boost the transparency and usability of automated banking credit decisions.

Keywords: Explainable Artificial Intelligence (XAI), Credit Risk Assessment, Basel III, Probability of Default, SHAP, LIME, Credit Risk Management.

© 2026 Aritra Pal. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). The authors retain copyright and allow others to share, adapt, or redistribute the work with proper attribution.

Cite This Article: Pal, A. (2026). Explainable AI driven Credit Risk Assessment for Basel III compliant Banking Risk Management. The American Journal of Interdisciplinary Innovations and Research, 8(09), 102–116. https://doi.org/10.37547/tajiir/Volume08Issue09-06

1. Introduction

Financial systems worldwide have been greatly transformed over the past few decades due to rapid technological advancement, digitalization, and access to vast quantities of financial data. These changes have substantially changed the traditional banking processes, altering how financial institutions assess risk, process transactions, detect fraud, and make lending decisions in ever more data-driven environments [1]. The banking industry relies heavily on

credit risk assessments, which comprise determining the probability of a borrower's failure to repay a loan. Typically, this is done by humans who look at a number of financial metrics, including the debt-to-income ratio, credit history, income stability, and other metrics that are specific to the loan [2]. In most cases, these evaluations take a long time, are easy for humans to make mistakes with, and cannot be scaled up.

Artificial intelligence (AI), particularly machine learning (ML), has hastened this transition by allowing financial institutions to analyse complex borrower behaviour patterns and provide highly accurate predictive models for credit risk assessment, fraud detection, customer analytics, and financial forecasting [3].

In the context of the modeling system and decision-making, a ML model's classification accuracy is no guaranty of its performance when it comes to credit risk prediction. However, accuracy does not guarantee trust or reliability in decision-making [4]. However, many prior studies focus on prediction accuracy and explanation and do not treat calibration and uncertainty quantification as complementary analytical activities. Over recent years, several studies have focused on SHAP and LIME for explainability, calibration techniques for probability calibration, and uncertainty quantification methods for prediction interval estimation[5][6]. However, few have investigated how predictive modelling, explainable artificial intelligence (XAI), calibration analysis, and uncertainty quantification can work together along with other methods in a unified approach to open, uncertainty-aware credit risk assessment. This fragmentation reduces their usefulness in practice when dealing with high-risk lending contexts, and stakeholders want to receive not only predicted outcomes but also understandable, trustworthy explanations and confidence estimates.

A. Motivation and Contribution of Paper

The credit risk assessment is a crucial component of a bank's operations since an incorrect estimate of the borrower's default risk might result in financial losses for the bank, inefficient capital deployment, and increased sensitivity to credit risk. The statistical models and set criteria that underpin common credit scoring methods might not be sufficient to capture the intricate relationships between borrower characteristics. Although ML models have advantages in terms of predictive performance, their adoption in regulated financial environments may be hampered by their complexity and opaqueness. Furthermore, the credit risk assessment should include a category of credit risk that is simply comprehensible and metrics that can be used to real-world credit risk management rather than only being a binary forecast of default. This means that an integrated framework is required that integrates a ML default prediction model with explainability and Basel III risk assessment. This proposed work aims to achieve this goal by combining the ensemble learning, XAI techniques, Probability of Default (PD) estimation and risk-based financial measures into a single credit risk assessment framework. This study's main contributions are as follows:

- **Hybrid Credit Risk Prediction:** A Hybrid RF-XGBoost model is developed by combining RF and XGBoost probability outputs to improve automated loan-default prediction.
- **Explainable AI Integration:** The integration of SHAP and LIME allows for both global and local explanations, which in turn helps to pinpoint the most important elements impacting credit-risk forecasts and enhances the model's transparency.
- **Imbalance-Aware Modeling:** SMOTE is applied to the training data to address class imbalance between default and non-default borrowers while preserving an independent test set.
- **PD-Based Risk Stratification:** The PD is classified into five risk levels according on its predicted likelihood of default. To back up differential credit determinations, and have Very Low, Low, Moderate, High, and Very High.
- **Basel III-Oriented Risk Assessment:** The framework integrates PD with LGD, EAD, Expected Loss, stress loss, risk-weighted assets, and capital requirements to provide a practical risk-management perspective for banking applications.

B. Novelty and Justification

Credit risk assessment frameworks that are easy to understand and use are necessary due to the growing complexity of borrower profiles and the demand for open, data-driven lending choices. This helps with both predicted accuracy and practical risk management in the banking industry. Very little effort has gone into understanding specific forecasts or connecting the model's output to other risk-management measures; conventional ML approaches only address the default categorization problem. Borrowers' risk assessment based on Basel III principles, including LGD, EAD, Expected Loss, stress loss, Risk-weighted Assets, and

capital requirements, is combined with the innovative Hybrid RF-XGBoost prediction model with explainability utilizing SHAP and LIME. This unique approach is what makes the research original. This integration represents a step beyond the typical default prediction toward an interpretable and risk-based decision support system, applicable in the banking sector.

C. Structure of the Paper

The remainder of this paper is organized as follows. Section II presents the related work and identifies the research gap in existing credit risk assessment approaches. Section III describes the proposed methodology, including data preprocessing, feature engineering, model development, and explainable AI techniques. Section IV presents the experimental results and performance comparison of the developed models. Finally, Section V concludes the study and presents its limitations and potential directions for future research.

2. Literature Review

The modelling of credit risk the exposure of financial consequences to a borrower's defaulting has progressed over the last few decades. Credit risk, at least initially, is scored primarily by human judgment and rule-based systems. With increasing data availability and progress in computational speed, statistical methods have emerged.

S. Malave et al. (2026) utilized Optical Character Recognition through Easy OCR, PyPDF2, and Pflaumer to extract financial and identity data from applicant documents, followed by data cleaning, normalization, and structured feature engineering. SHAP and LIME are applied to generate global and local explanations, enhancing interpretability. Experimental evaluation shows that XGBoost achieves 92.5% accuracy, outperforming Light GBM (90.8%), RF (87%), Logistic Regression (91%), and LSTM (81%) [7].

M. Indira et al. (2026) attention mechanism literally enhances minority-sensitive characteristics in the latent representation learning stage, whereas attention-balanced loss, which consists of reconstruction, KL-divergence regularization, attention regularization, and class-weighted classification words, avoid the dominance of majority-classes through the training process. The European Credit Card Fraud (ECC) and IEEE-CIS datasets experiments show that the Recall of AB-VAE is equal to 0.812, AUC-PR equals 0.842, and MCC equals 0.846, and it remains higher than that of the imbalance-handling, generative, attention-based, and hybrid DL baselines [8].

S. Kakkar (2025) examined (ML) techniques for credit scoring, comparatively little attention has been paid to systematically quantifying the trade-off between performance and explainability under supervisory requirements. Regulatory acceptability under Basel III and GDPR can be achieved by SHAP's ability to bridge the gap between ensemble model performance and transparency, as opposed to previous studies that have isolated these two concepts. While Logistic Regression is the most compliant method because of its openness, the results demonstrate that Gradient Boosting delivers the best predictive power (AUC = 0.84) [9].

M. J. Lim et al. (2025) produced a balanced selection of 1, 0 5 0 transactions (5 0 0 genuine, 550 fraud) from a Kaggle dataset containing 307,511 transactions (8.07% fraud) in order to enhance the model's performance. The RF model, with 100 trees and tuned parameters, achieved a precision of 0.6203, a recall of 0.7030, and an F 1-score of 0.6591, with a confusion matrix showing 116 true positives, 71 false positives, 49 false negatives, and 79 true negatives. Threshold optimization improved the F1-score to 0.7064 at a 0.4 threshold [10].

S. Chauhan et al. (2024) detection of fraudulent behaviour and to accurately compare the results of various algorithms, balanced datasets are essential, as this shows. There is a 91% F1 score, 92.35% ROC AUC, and 98.01% cross validation accuracy rate for fraud detection in the logistic regression model. Alternatively, a decision tree classifier's fraud detection cross-validation score is 96.67%, ROC AUC was 91.36%, and F1 was 90% [11].

Z. Xu et al. (2023) A local update screening technique is presented to pick reliable model parameters for aggregation each time, ensuring the quality of the global model. According to the study's findings, when compared to algorithms frequently employed in credit risk research, the model error rate is decreased by 6.22% and the recall rate is enhanced by 11.03%. This greatly enhances the capacity to identify defaulters. Finally, the business operations of commercial banks are used to confirm the potential of the proposed model for real-world implementation [12].

Previous work have investigated ML, DL, fraud detection, and explainable AI for financial risk assessment, but most have concentrated on either a predictive performance or a model interpretability only. There is limited work that combines explainable credit risk prediction models, Probability of Default (PD) and Basel III-based risk management in a single framework. Additionally, the

interpretation of borrower-level data, risk categorization, expected loss, stress measurement, and capital-oriented analysis with the combined use of both SHAP and LIME are poorly addressed. Therefore, the purpose of this study is to address this knowledge vacuum by developing a consistent and understandable model for the evaluation of credit risk in banking.

3. Methodology

In Fig. 1 proposed framework performs duplicate removal, median-based missing-value imputation and IQR-based outlier capping on the credit risk dataset. The categorical features are one-hot encoded and the data are split into an 80:20 train-test split stratified by the categorical features. The training set is made more balanced by means of SMOTE. RF and XGBoost models are trained on the balanced data and the predictions of these models are fused to create the Hybrid RF-XGBoost model. Accuracy, Precision, Recall, F1-Score, ROC-AUC, and PR-AUC are used for the evaluation of the model performance. SHAP and LIME are used to give global and local model interpretability. Lastly, the PD values are predicted for credit-risk categorization and Basel III oriented assessment, such as capital requirements, RWA, expected loss, EAD, and LGD.

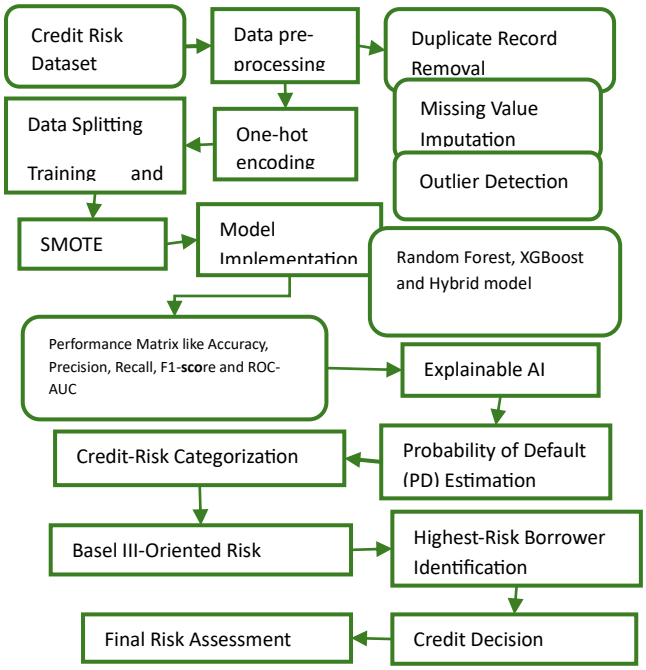


Fig. 1. Flowchart of the Proposed Credit Risk Assessment Framework

This section explains the following phases of the flowchart in a nutshell:

A. Data Analysis and Visualization

The Credit Risk Dataset [13] from Kaggle has 32,581 observations and 12 attributes. Pre-processing involved first identifying and removing duplicate records, leaving 32,416 unique observations. The NAs in person_emp_length and loan_int_rate are filled with the median of their respective variables to keep the data consistent and avoid any possible distributional bias. Subsequently, outliers for numerical attributes such as personage, person income, person_emp_length, loan_amnt, loan_int_rate, loan_percent_income, and cb_person_cred_hist_length are treated by the Tukey's Interquartile Range (IQR) method. The procedure allowed to remove the impact of extreme observation and to maintain the size of the original sample (32 416 × 12) to obtain a cleaned and coherent dataset for subsequent credit risk modeling.

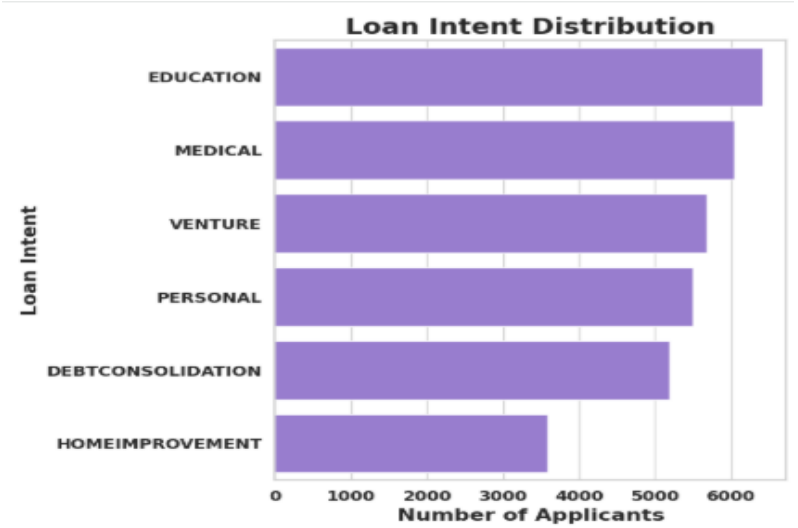


Fig. 2. Distribution of Loan Applicants by Loan Intent

Fig. 2 shows the number of applicants with various loan intents. The leading applicants are Education, followed by Medical, Venture, Personal, Debt Consolidation and Home Improvement. The distribution illustrates the range of lending reasons in the dataset. Fig. 3 displays the distributions of interest rates for loans to borrowers who have not defaulted and those who have. Overall, default situations had higher interest rates, a broader range, and a higher median compared to nondefault instances. Therefore, a key variable—the interest rate on the loans—is associated with the variation in credit risk.

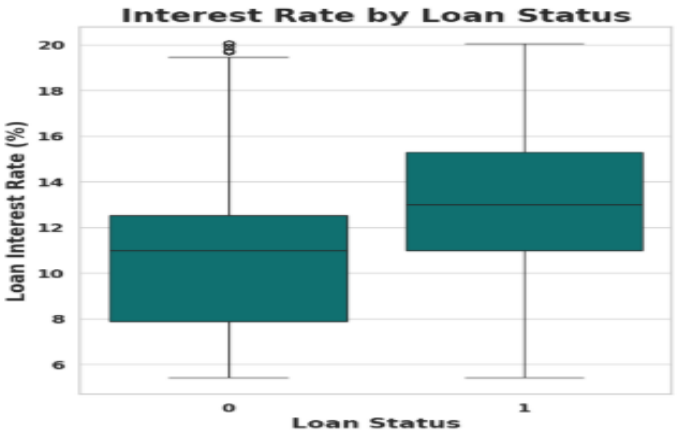


Fig. 3. Distribution of Loan Interest Rates by Loan Status

B. Feature Encoding and Target Definition

The preprocessed data is split between the input features (X) and target variable (y), where loan status is used as the target for prediction. This dataset is composed of 32416 samples, with 11 input features: seven are numerical and four are categorical. Class imbalance is present with the target being 78.13% non-default, 21.87% default. Categorical variables are transformed using one-hot encoding with the first category removed, increasing the feature space from 11 to 22 encoded features for subsequent model development.

C. Training and Testing Data Split

A fixed random state of 42 is used to stratify the encoded data set, which is then split 80/20 between the training and testing sets. The training set has 25,932 samples with 22 features, whereas the testing set contains 6,484 examples with the same number of features. Both subsets are stratified to maintain the original distribution of cases by default and nondefault. This makes sure that the evaluation set is representative, and mitigates potential bias from the imbalance among the target class.

D. Data Balancing using SMOTE

The training set shows a class imbalance with 5,671 default samples (21.87% of the total) and 20,261 nondefault samples (78.13%). To make things more equitable, SMOTE is solely used to enhance training data in order to generate synthetic minority class samples. To ensure that both groups have an equal representation in the sample, utilize SMOTE to raise the minority class's size to 20,261 samples (or 50% of the total), while keeping the independent test set same to assess the model's objectivity. In Fig.4 illustrate the class distribution of before and after SMOTE.

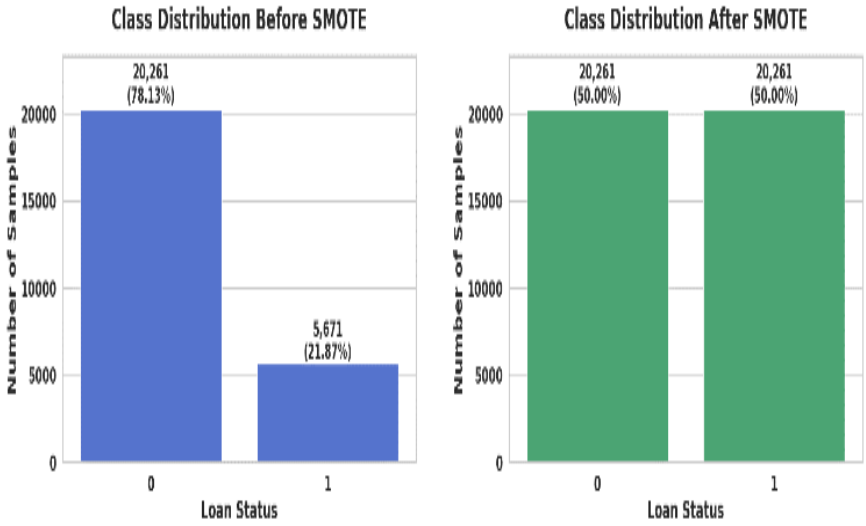


Fig. 4. Class distribution of before and after SMOTE

E. Machine Learning Model Development

Three predictive methods: RF, XGBoost and a Hybrid RF-XGBoost model are created for the classification of credit risk. The hybrid approach consists of equal-weighted average of the probability outputs of RF and XGBoost, which strengthens the robustness of default-risk prediction.

1) Random Forest Model

The RF classifier consisted of 500 decision trees (estimators = 500) to achieve the stability of ensemble predictions. The adept is not specified (None) and min_samples_split=2, min_samples_leaf=1 are chosen to allow detailed growth of the trees. With jobs=-1 and random state=42, the model takes full advantage of all available processor cores and guaranties repeatability. To evaluate classes and generate default probabilities, the trained model is utilized.

2) XGBoost Model

The XGBoost classifier is tuned for 200 boosting estimators (estimators=200) and a maximum tree depth of 6 (max depth=6). The contribution of each boosting round is controlled by a learning rate of 0.05, with subsample=0.8 and colsample_bytree=0.8 during the training of the model. The model has been trained using random state=42, volumetric='log loss', and n_jobs=-1. A trained XGBoost model generated class predictions and default probabilities for the credit risk assessment.

3) Hybrid RF-XGBoost Model

The proposed Hybrid RF-XGBoost model is obtained by taking the average (equal weightage 0.50) of the predicted probabilities of the RF and XGBoost models. The combined probability of default is computed from Class 1 predictions of both models. The probability-fusion approach is based on the output of the two ensemble learning models to get a more robust estimation of borrower default risk and, as a result of that, it is used for the credit risk classification and assessment.

F. Performance Metrics

They assessed the performance of the credit risk classification models by the following measures: Accuracy, Precision, Recall, F1-Score, ROC-AUC, and PR-AUC. The evaluation of a learner's model take into account these outcomes: True Positive (TP) correctly predicted default cases, True Negative (TN) correctly predicted non-default cases, False Positive (FP) non-default cases incorrectly predicted as default, and False Negative (FN) default cases incorrectly predicted as non-default:

Accuracy: The percentage of borrowers that were correctly categorized out of all the borrowers that were analyzed. It is determined by taking the entire number of observations and dividing it by the sum of true positives and true negatives Equation (1), as shown.

$$Accuracy = \frac{TP+TN}{TP+Fp+TN+FN} \quad (1)$$

Precision: Measures the proportion of borrowers predicted as default who are actually default cases, as expressed in Equation (2).

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

F1-Score: Represents the harmonic mean of precision and recall, providing a balanced measure of classification performance. It is calculated by combining precision and recall as presented in Equation (3).

$$F1 - Score = 2\frac{(Precision*Recall)}{Precision + Reall} \quad (3)$$

Recall: The percentage of real default instances that the model accurately identifies is assessed. Equation (4) shows how to compute it as a percentage of the total number of real default instances divided by the number of successfully detected default cases.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

AUC-ROC (Area Under the Receiver Operating Characteristic Curve): Measures the model's ability to distinguish between default and non-default borrowers across different classification thresholds. A higher AUC indicates better discriminative capability.

G. Explainability using SHAP and LIME

The term "explainable AI" (XAI) describes methods that make ML algorithm predictions understandable and trustworthy to humans. Learn how AI models make predictions or take actions with the help of XAI approaches.

- **SHAP:** SHAP is a XAI method or algorithm that provides context for the ML model's prediction and show the relative importance of each factor in forecasting the final result. It is compatible with any ML algorithm, meaning it is not model-specific. Visualization of SHAP results is accomplished by features bar plots and summary (bee swarm)plots. The SHAP value is a way to measure the contribution of each feature in model prediction; it is allocated to each feature.

- **LIME:** One such model-agnostic approach to explaining the model's output for a single data point is LIME and quickly understand how a certain prediction is made with LIME, which is intended to describe the outcome of a given instance or set of data. As a result, quickly understand the reasoning behind a single forecast. one. The LIME feature importance plot is a great tool for seeing which features are bolstering the model's forecast. Their performance is assessed by calculating the weights given to each characteristic.

4. Result Analysis and Discussion

The experimental setup is executed on a state-of-the-art HPC system that includes a 3.70 GHz Intel Core i7-8700K CPU, an NVIDIA GeForce RTX 2080 Ti GPU for DL computation acceleration, and 32 GB of RAM for efficient handling of huge datasets. The software configuration is based on the Windows 10 operating system, with Python serving as the primary programming language for implementing and executing the experiments. Table I shows comparative performance of proposed Hybrid RF-XGBoost model and RF model and XGBoost model on the credit risk assessment problem. The highest overall performance model is the Hybrid RF-XGBoost model, which has an accuracy of 93.14%, a recall value of 73.13%, an F1 score of 82.33%, a ROC AUC of 93.28%, and a PR AUC of 88.48%. While the value of precision for XGBoost is slightly higher at 94.35%, the hybrid model achieves a better balance among precision, recall, and overall discrimination ability. RF performs comparatively worse with accuracy of 92.97%, F1 score of 81.96%, ROC-AUC score of 92.86% and PR-AUC score of 87.90%. These results demonstrate that the suggested framework for explainable credit risk assessment benefits from combining the predictions of RF and XGBoost, and that credit default categorization is generally more robust as a consequence.

TABLE I. PERFORMANCE RESULTS OF MACHINE LEARNING MODELS FOR EXPLAINABLE CREDIT RISK ASSESSMENT

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	PR-AUC
Hybrid RF-XGBoost	0.9314	0.9419	0.7313	0.8233	0.9328	0.8848

XGBoost	0.9314	0.9435	0.7299	0.8231	0.9309	0.8835
Random Forest	0.9297	0.9333	0.7306	0.8196	0.9286	0.8790

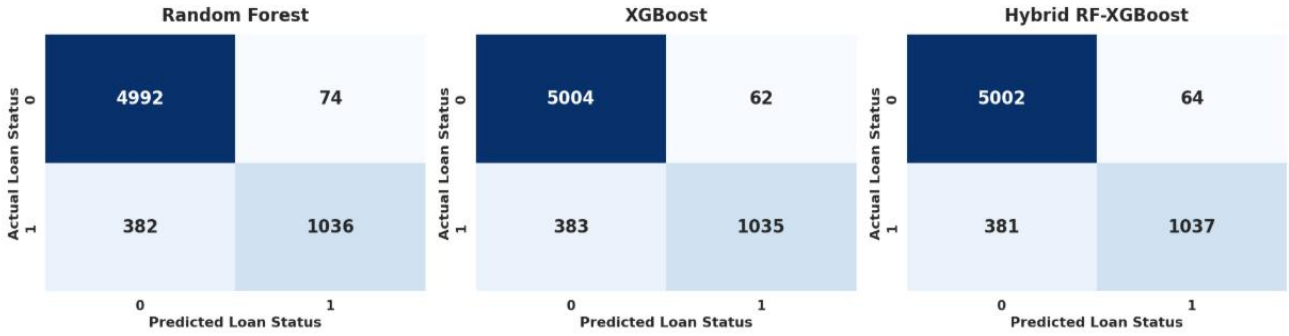


Fig. 5. Confusion Matrices of ML Models for Credit Risk Classification

The confusion matrices for non-default (0) and default (1) classifications of the loan application data, using the RF, XGBoost and proposed Hybrid RF-XGBoost models, are shown in Fig. 5. RF correctly classifies 4992 non-default cases and 1036 default cases, while XGBoost is able to correctly identify 5004 non-default cases and 1035 default cases. The Hybrid RF-XGBoost model correctly classifies 5,002 non-default and 1,037 default cases and makes 381 false negatives and 64 false positives. Specifically, the hybrid model has the best performance in correctly identifying those that default, and the lowest number of false negatives, which means that it can detect potential credit defaults better.

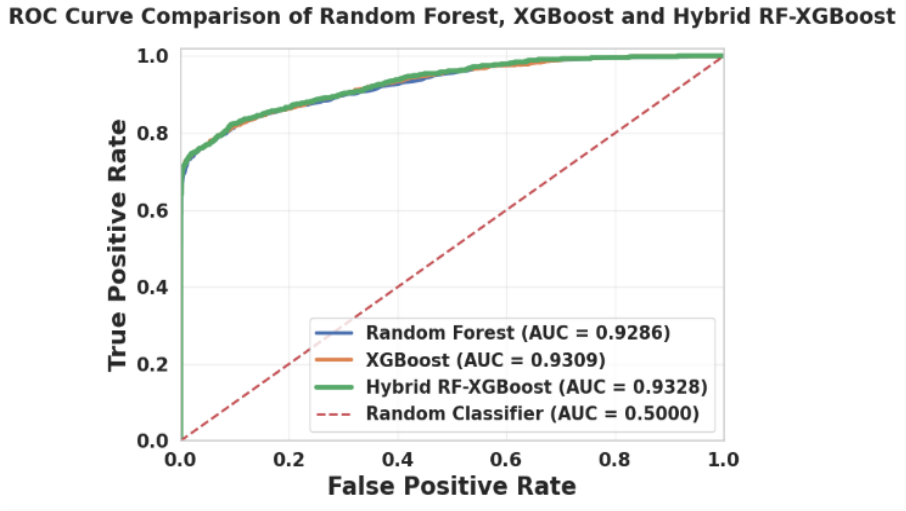


Fig. 6. ROC Curve Comparison of Models for Credit Risk

The ROC curves of the RF model, XGBoost model and Hybrid RF-XGBoost model trained on the credit risk classification task are shown in Fig. 6. The AUC values for the three cases are 0.9286, 0.9309 and 0.9328, respectively, while the baseline value for the random-classifier is 0.5000. The ROC curves of all three models are well above the diagonal reference line demonstrating good classification performance to discriminate between default and nondefault loan cases.

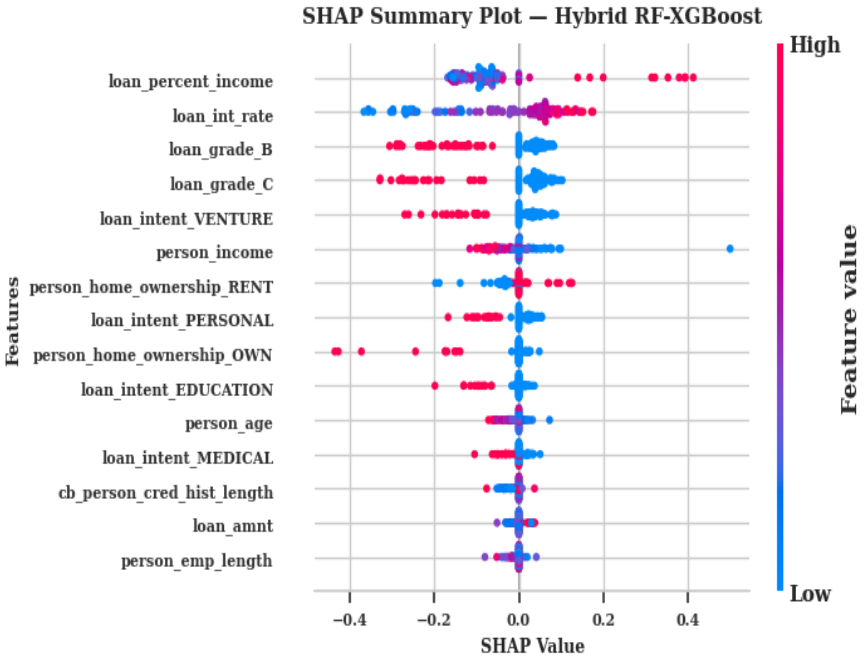


Fig. 7. SHAP Summary Plot of the Hybrid RF-XGBoost Model

The SHAP values from the Hybrid RF-XGBoost model are used to create the global feature importance plot shown in Fig. 7. The results indicate that loan_percent_income and loan_int_rate are the most important features, followed by loan_grade_B, loan_grade_C, and loan_intent_VENTURE. Credit risk assessment is made more understandable by the SHAP values, which reveal the direction and size of each feature's effect on the model's forecast.

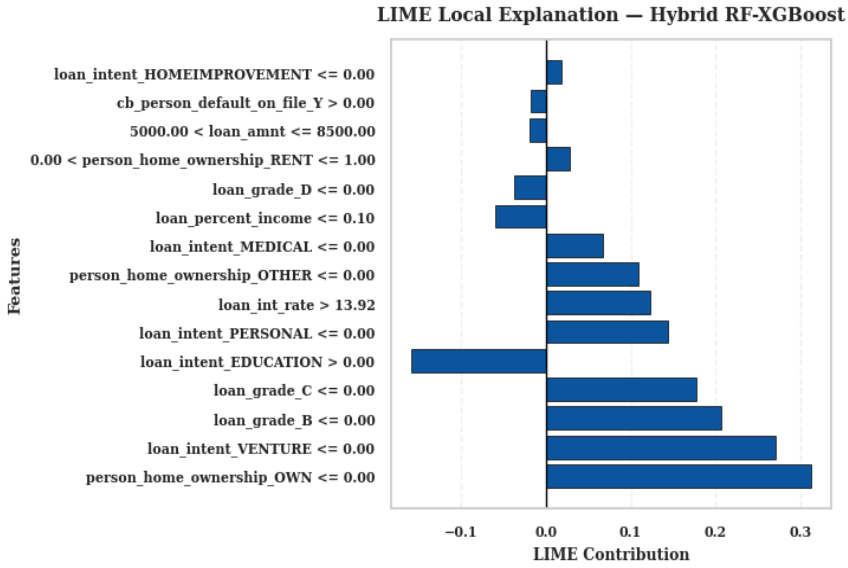


Fig. 8. LIME Local Explanation of the Hybrid RF-XGBoost Model for Credit Risk Assessment

Here, the local explanation of the Hybrid RF-XGBoost model for an individual credit risk prediction is shown in Fig. 8 using LIME. The results are that person_home_ownership_OWN, loan_intent_VENTURE, loan_grade_B, and loan_grade_C are the most significant positive contributors and loan_intent_EDUCATION has the highest negative contribution. This illustrates how the model's prediction varies with the individual borrower characteristics.

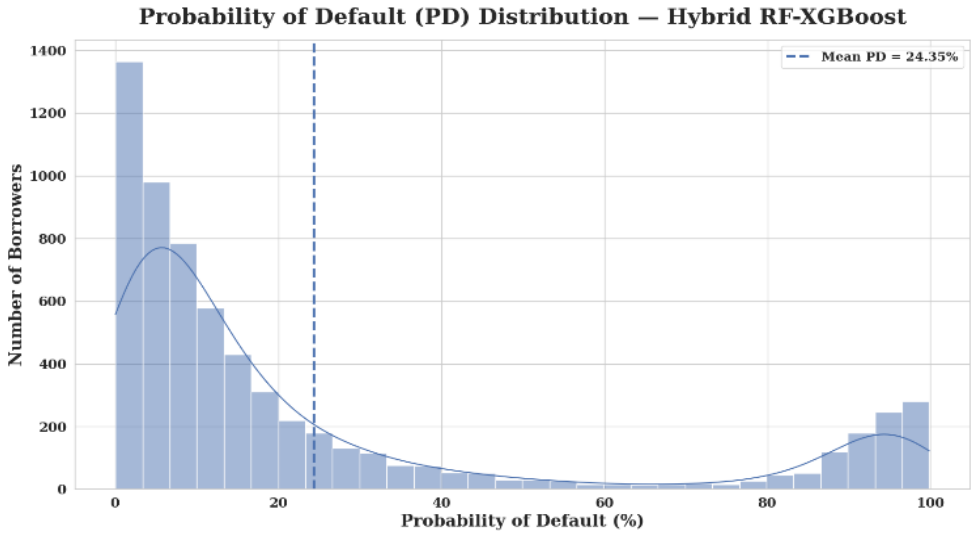


Fig. 9. Probability of Default (PD) Distribution of the Hybrid RF-XGBoost Model

The distribution of borrowers by the predicted Probability of Default (PD) from the Hybrid RF-XGBoost model is shown in Fig. 9. The minimum PD is 0.03% and the maximum PD is 99.84%. The distribution of the mean PD is right-skewed, with the majority of borrowers having relatively low PDs, and a minority having very high PDs, with the median PD at 10.61% and the mean PD at 24.35%. This distribution is used as the basis for differentiation by credit risk estimates.

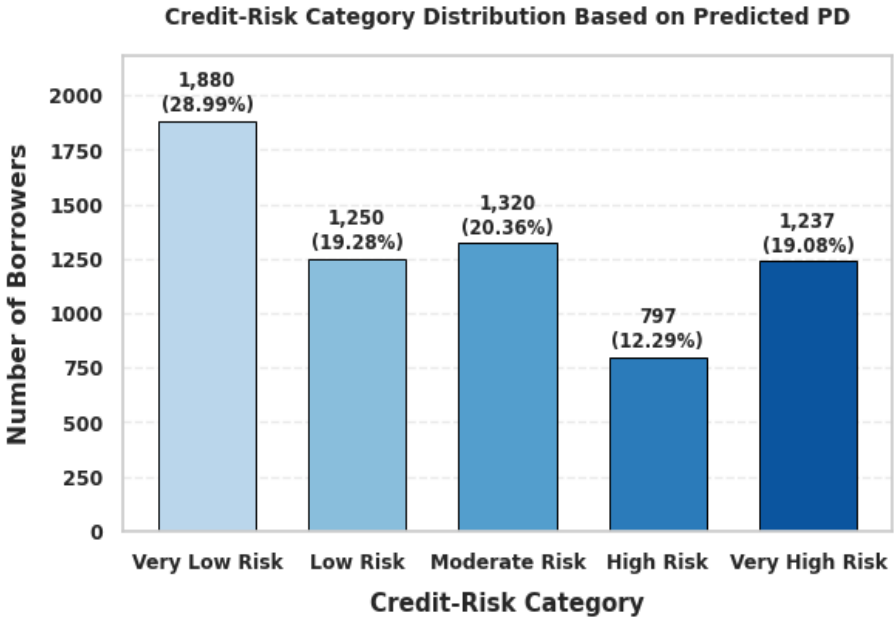


Fig. 10. Credit-Risk Category Distribution Based on Predicted Probability of Default

The distribution of borrowers in the defined credit-risk categories, according to their predicted Probability of Default (PD) is shown in Fig. 10. The results show that 1,880 (28.99%) borrowers are classified as Very Low Risk, 1,250 (19.28%) as Low Risk, 1,320 (20.36%) as Moderate Risk, 797 (12.29%) as High Risk, and 1,237 (19.08%) as Very High Risk. This classification is used to make a meaningful distinction between borrowers based on their perceived credit risk.

The Basel III-based credit risk assessment of borrowers is reported in Table II according to the risk categories assigned to them based on their expected Probability of Default (PD). The results presented indicate that the credit risk of the loan

portfolio is growing from Very Low Risk to Very High Risk, with the average PD rising from 2.32% to 83.60%, and the average LGD rising from 30% to 65%. The "Very High Risk" category has the highest credit-risk exposure, with a total expected loss of 7,453,733.96 and total capital requirement of 1,088,826.00. In general, the outcome of the tests shows that PD, LGD, EAD, expected loss, and risk-weight measures can be combined to facilitate risk-sensitive banking risk management.

TABLE II. BASEL III-ORIENTED CREDIT RISK ASSESSMENT ACROSS DIFFERENT RISK CATEGORIES

Risk Category	Basel Risk Assessment	Borrowers	Avg. PD	Avg. LGD	Total EAD	Avg. EAD	Total Expected Loss	Avg. Expected Loss	Total Stress Loss	Avg. Risk Weight	Total RWA	Capital Requirement
Very Low Risk	Low Credit Risk	1,880	2.32%	30.00%	16,975,650.00	9,029.60	119,033.16	63.32	198,388.60	20.00%	3,395,130.00	271,610.40
Low Risk	Acceptable Credit Risk	1,250	7.34%	35.00%	11,704,425.00	9,363.54	299,398.39	239.52	481,175.98	35.00%	4,096,548.75	327,723.90
Moderate Risk	Moderate Credit Risk	1,320	14.26%	45.00%	11,162,325.00	8,456.31	712,368.17	539.67	1,088,340.26	50.00%	5,581,162.50	446,493.00
High Risk	Elevated Credit Risk	797	27.72%	55.00%	7,470,850.00	9,373.71	1,142,467.32	1,433.46	1,687,735.81	75.00%	5,603,137.50	448,251.00
Very High Risk	Critical Credit Risk	1,237	83.60%	65.00%	13,610,325.00	11,002.69	7,453,733.96	6,025.65	9,430,706.24	100.00%	13,610,325.00	1,088,826.00

Table III presents the highest-risk borrowers identified by the proposed Hybrid RF-XGBoost model based on their predicted Probability of Default (PD). The borrowers listed below are considered Very High Risk with CR assessment between 99.61% and 99.84%, and are listed with extremely high PD values. These borrowers also have very high expected and stress expected losses, both total 100%. Accordingly, the framework recommends strict review or possible rejection for these high-risk borrowers.

TABLE III. TOP 5 HIGHEST-RISK BORROWERS BASED ON PREDICTED PROBABILITY OF DEFAULT

Borrower ID	PD	Risk Category	Basel Risk Assessment	LGD (%)	EAD	Expected Loss	Stress Expected Loss	Risk Weight (%)	RWA	Capital Requirement	Risk Control
6996	0.9984	Very High Risk	Critical Credit Risk	65.00%	6,200.00	4,023.70	4,650.00	100.00%	6,200.00	496.00	Strict review / possible rejection
21987	0.9974	Very High Risk	Critical Credit Risk	65.00%	6,250.00	4,052.06	4,687.50	100.00%	6,250.00	500.00	Strict review / possible rejection
12175	0.9963	Very High Risk	Critical Credit Risk	65.00%	12,000.00	7,771.49	9,000.00	100.00%	12,000.00	960.00	Strict review / possible rejection
23192	0.9961	Very High Risk	Critical Credit Risk	65.00%	5,000.00	3,237.32	3,750.00	100.00%	5,000.00	400.00	Strict review / possible rejection
26738	0.9961	Very High Risk	Critical Credit Risk	65.00%	15,000.00	9,711.63	11,250.00	100.00%	15,000.00	1,200.00	Strict review / possible rejection

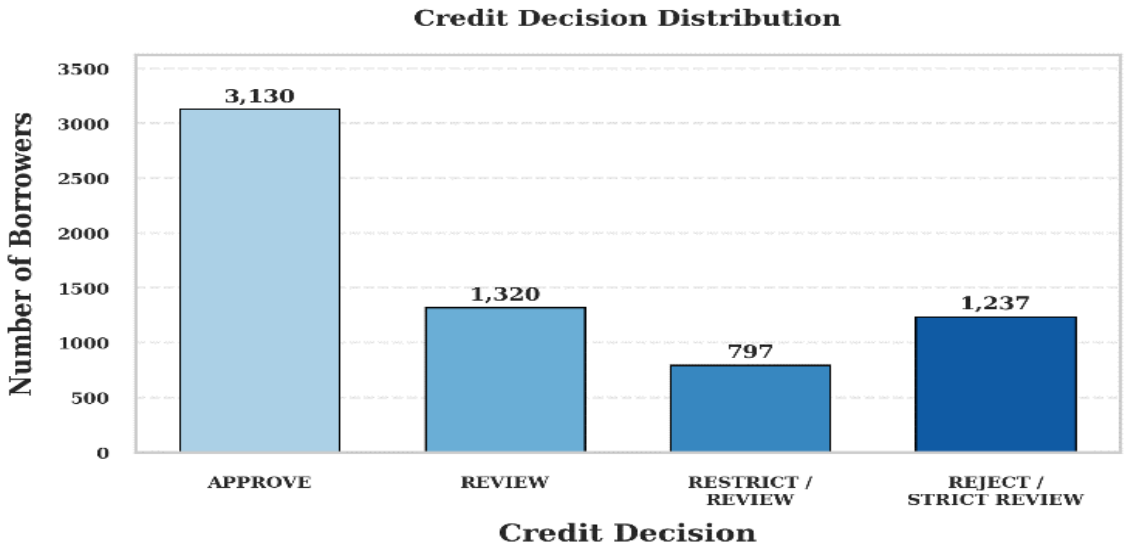


Fig. 11. Credit Decision Distribution Based on Predicted Credit Risk

Fig. 11 shows the distribution of borrowers within the credit decision categories that are created by the proposed framework. In total, 3,130 borrowers are in the APPROVE category, 1,320 are in REVIEW, 797 are in RESTRICT/REVIEW, and 1,237 are in REJECT/STRICT REVIEW. The distribution illustrates how the credit risk is manifested as credit action, or as actionable outcomes in lending decisions.

A. Comparative Analysis and Discussion

Table IV compares the results of the prior ML approaches with those of the suggested Hybrid RF-XGBoost regression model, XGBoost, and RF. The Hybrid RF-XGBoost model achieves an accuracy of 0.9314, precision of 0.9419, recall of 0.7313, F1-score of 0.8233, and ROC-AUC of 0.9328. The standalone XGBoost model achieves the same accuracy of 0.9314 with slightly higher precision (0.9435) and the RF model achieves an accuracy of 0.9297 and 0.9286 ROC-AUC. Compared with these, the LightGBM and XGBoost models in get accuracy of 0.9007 and 0.8874 respectively, while the SVM and Logistic Regression models get accuracy of 0.8800 and 0.8200 respectively. A model has an accuracy of 0.8525 by using the RF model. In general, the proposed models exhibit good performance in the considered classification metrics.

TABLE IV. PERFORMANCE COMPARISON OF THE PROPOSED MODELS WITH EXISTING METHODS

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Hybrid RF-XGBoost	0.9314	0.9419	0.7313	0.8233	0.9328
XGBoost	0.9314	0.9435	0.7299	0.8231	0.9309
Random Forest	0.9297	0.9333	0.7306	0.8196	0.9286
LightGBM[14]	90.07	0.2757	0.1434	-	0.7203
XGBoost[14]	0.8874	0.2199	0.1566	-	0.7033
SVM[15]	0.88	0.87	0.86	0.86	0.89
Logistic Regression[15]	0.82	0.80	0.78	0.79	0.84
Random Forest[12]	0.8525	0.7213	0.7020	0.7114	-

From the comparative analysis it is observed that the proposed ensemble approach gives a more robust solution for the classification of credit-risk as compared to the individual and the conventional ML models. RF coupled with XGBoost is able to learn complementary complex relationships between borrower characteristics. The results also highlight the applicability of ensemble learning in addressing the credit-default prediction challenges. Despite this, the model still struggles with some of the default cases, suggesting more work is required to improve the model. The results indicate that the proposed approach is applicable for the reliable credit risk assessment and interpretable.

5. Conclusion, Limitations and Future Direction

The capacity to analyze credit risk is critical to making prudent loan decisions and maintaining financial stability. Using Explainable AI, this study presented a methodology for credit risk assessment that predicts default using a Hybrid RF-XGBoost, RF, and XGBoost. The proposed hybrid model exhibited good prediction capability with accuracy of 93.14%, ROC-AUC of

93.28%, and PR-AUC of 88.48%. Two other tools, SHAP and LIME, are developed to enhance interpretability by revealing the most important factors in the credit-risk predictions. The Probability of Default is then combined with risk categorization and Basel III related measures, which allowed for an orderly analysis of borrower risk, expected losses and capital needs. In general, the framework offers a data-driven, transparent way to inform banking credit decisions for risks that are sensitive.

A. Limitations

The proposed framework has some drawbacks, but it is effective. The analysis is conducted on only one credit risk data set and the results may not be applicable to other banking populations and to other lending environments. The Basel III-related calculation of the LGD, risk-weight assignments and credit-risk threshold apply to the research-defined values and should not be used as a direct measure of the Basel III calculations. Also, the model is hybrid (RF + XGBoost) with fixed and equal weights for both algorithms without much optimization of the weights scheme. The explainability analysis also relies on SHAP and LIME approximations and might not be a complete image of the inner mechanisms of complex ensemble models.

B. Future Work

The framework can be extended in the future by testing it with other credit data sets from different countries and regions to gain better generalizability. Finally, exploration of advanced ensemble and DL algorithms may be considered to enhance the default prediction further without compromising on interpretability. The hybrid architecture may also include adaptive model weighting and automated hyperparameter optimization. Future research can also combine real-time borrower and transactional data, create more sophisticated stress-testing scenarios and match the risk-weight and capital calculation to institution-specific capital regulatory requirements. Additionally, future studies could focus on continuous model monitoring, assessing fairness, and implementing more sophisticated explainable AI methods to assist in the deployment of reliable and responsible AI in banking contexts.

References

1. P. E. de Lange, B. Melsom, C. B. Vennerød, and S. Westgaard, “Explainable AI for Credit Assessment in Banks,” *J. Risk Financ. Manag.*, vol. 15, no. 12, p. 556, Nov. 2022, doi: 10.3390/jrfm15120556.

2. [V. P. Rambabu, J. Sreerama, and J. T. S. Singh, “AI-Driven Data Integration: Enhancing Risk Assessment in the Insurance Industry,” *Aust. J. Mach. Learn. Res. Appl.*, vol. 2, no. 2, pp. 130–179, July, 2022.

3. D. O. Oyeyemi, O. O. Okosieme, O. Idowu-Kunlere, E. A. Julius, and I. P. Nwinyi, “Explainable AI For Credit Risk Assessment: Integrating Machine Learning With Business Analytics,” *IOSR J. Econ. Financ.*, vol. 16, no. 5, pp. 63–72, 2025, doi: 10.9790/5933-1605016372.

4. M. Rambauli, T. Ravele, and C. Sigauke, “Application of Explainable AI and Uncertainty Quantification in Credit Risk Assessment,” *Risks*, vol. 14, no. 7, p. 152, Jul. 2026, doi: 10.3390/risks14070152.

5. N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, “Explainable Machine Learning in Credit Risk Management,” *Comput. Econ.*, vol. 57, no. 1, pp. 203–216, Jan. 2021, doi: 10.1007/s10614-020-10042-0.

6. P. Chawande, “Model Risk Governance for AI-Based Compliance Systems in Investment Banking,” *Int. J. Multidiscip. Res. Growth Eval.*, vol. 6, no. 3, pp. 2027–2035, 2025, doi: 10.54660/.IJMRGE.2025.6.3.2027-2035.

7. S. Malave, B. Khemani, H. Patil, S. Nandurkar, O. Nandurkar, and A. Nayak, “Transforming financial documents into credit decisions using explainable artificial intelligence and optical character recognition,” *MethodsX*, vol. 16, p. 103962, Jun. 2026, doi: 10.1016/j.mex.2026.103962.

8. M. Indira, R. M. Basha, P. Sivamma, M. S. Babu, and S. H. N. Ketan, “Attention-Balanced Variational Autoencoder Framework for Credit Card Fraud Detection under Extreme Class Imbalance,” in *2026 7th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, IEEE, May 2026, pp. 1060–1067. doi: 10.1109/ICICV68925.2026.11554839.

9. S. Kakkar, “Explainable AI Models for Credit Risk Scoring in Banking: Balancing Accuracy and Regulatory Transparency,” *Int. J. Financ. DATA Sci.*, vol. 3, no. 2, pp. 1–6, Aug. 2025, doi: 10.34218/IJFDS_03_02_001.

10. M. J. Lim, P. W. Alexander, V. M. Karyanto, T. C. Christie, and T. Oktavia, “Detecting Credit Card Fraud using Random Forest and CRISP-DM: A

Machine Learning Approach," in *2025 International Conference on ICT for Smart Society (ICISS)*, IEEE, Sep. 2025, pp. 1–5. doi: 10.1109/ICISS66954.2025.11389400.

11. S. Chauhan, K. S. Gill, R. Chauhan, and H. S. Pokhariya, "Smart Solutions for Utilizing Advanced Machine Learning for Robust Credit Card Fraud Detection," in *2024 First International Conference on Electronics, Communication and Signal Processing (ICECSP)*, IEEE, Aug. 2024, pp. 1–6. doi: 10.1109/ICECSP61809.2024.10698053.

12. Z. Xu, J. Cheng, L. Cheng, X. Xu, and M. Bilal, "MSEs Credit Risk Assessment Model Based on Federated Learning and Feature Selection," *Comput. Mater. Contin.*, vol. 75, no. 3, pp. 5573–5595, 2023, doi: 10.32604/cmc.2023.037287.

13. L. Tse, "Credit Risk Dataset," Kaggle.

14. Shreya and H. Pathak, "Explainable Artificial Intelligence Credit Risk Assessment using Machine Learning," *arXiv Comput. Sci.*, Jun. 2025, doi: 10.48550/arXiv.2506.19383.

15. K. N. Islam, "An Explainable Machine Learning Model for Credit Risk Prediction: Evidence from Commercial Banks in Bangladesh," *Am. J. Ind. Bus. Manag.*, vol. 16, no. 07, pp. 647–673, 2026, doi: 10.4236/ajibm.2026.167034.