A Hybrid Framework Combining Machine Learning and Deep Learning for Intelligent Loan Approval Prediction

***Abstract*— Due to the growing need for automated and smart financial decision-making, predicting whether a loan will be accepted has become an important job for modern banking systems. When working with large, high-dimensional data, traditional methods, which are often manual and based on rules, are likely to make mistakes and take too long. This study presents a hybrid ensemble architecture that integrates the predictive capabilities of both machine learning and deep learning models to enhance the precision, scalability, and interpretability of loan approval systems. The model is based on a structured dataset of 5,000 customer records that includes information about their finances, demographics, and behaviors. A thorough preprocessing pipeline was used to make the data better. This included class balancing with SMOTE, log transformation, outlier removal, and normalization. After testing fifteen models in three groups, XGBoost was the best method, beating both modern deep learning methods and traditional machine learning with an accuracy of 99.20%. Ensemble classifiers and neural networks like MLP and PLSTM also did very well, which shows that the hybrid approach is strong. SHAP and LIME also made it clear how features affected decisions and how the model could be understood. The proposed methodology is suitable for practical application in modern financial systems owing to its high accuracy, scalability, and clarity.**

Keywords—Loan Approval, Machine Learning, Deep Learning, Ensemble Learning, XGBoost, SMOTE, Explainable AI, SHAP, LIME

I. INTRODUCTION

Banks make a lot of money from personal loans, and the financial sector is very important for keeping a country's economy stable. Getting a personal loan has always depended on banking experts looking over the application, which is a long, error-prone, and often inconsistent process. Data-driven decision-making has become very important for making loan approvals more accurate and efficient. This is because data is becoming more and more available and people want automation more and more [1]. Recent advancements in machine learning (ML) and deep learning (DL) have enabled the creation of predictive models that assess a borrower's loan eligibility by analyzing financial and demographic data. These strategies not only speed up operations, but they also reduce human bias and the risks that come with bad loans. Talukder et al. [2] demonstrated the potential of DL models for applications like loan prediction by demonstrating their capacity to recognize complex patterns in high-dimensional financial datasets. Traditional linear models are no longer sufficient for the complexity of financial data, so researchers are turning to ensemble and deep learning techniques to better capture non-linear correlations. Wang et al. [3] significantly improved the performance of their stacking-based ensemble technique for bank loan risk prediction by utilizing DL and optimization. In this study, we combine deep neural networks, ensemble approaches, and traditional machine learning models to estimate a consumer's likelihood of accepting a personal loan.

The dataset has undergone extensive preprocessing to guarantee quality and balance, and it comprises 5,000 records with 14 variables that reflect consumer profiles [1]. Superb performance was shown by models such as XGBoost, CatBoost, LightGBM, and MLP; XGBoost achieved an accuracy of 99.20%. This study provides a comprehensive and comparative examination across several learning paradigms, in contrast to previous research that concentrated on a small number of models or lacked balanced datasets. The efficiency of Random Forest over Decision Tree in loan approval was emphasized by Bhargav and Sashirekha [4]. Similarly, Abdullah et al. [5] found that tree-based models outperformed traditional statistical methods when studying loan behavior in emerging economies. Through balanced data and interpretability for practical applications, our research prioritizes not only accuracy but also fairness. This research advances the development of automated and more dependable loan prediction systems that facilitate quicker and more intelligent decision-making in banking settings.

Key Contributions of This Study:

- To predict loan approval, a hybrid framework that combines ensemble, deep learning, and machine learning techniques is suggested.
- SMOTE is used to clean, normalize, and balance a structured dataset of 5,000 customer profiles.
- After evaluating fifteen models, XGBoost achieved the highest accuracy of 99.20%.
- To determine important features in decision-making, SHAP and LIME are used for model explainability.

II. LITERATURE REVIEW

Because of the growing demand for automation and precise risk assessment, loan prediction has emerged as a major area of study in the financial industry. Istia et al.'s earlier work [1]

A Kaggle dataset was used to create a thorough framework for categorizing loan approval results. Nine deep learning models, such as LSTM and MLP, as well as machine learning models, such as Random Forest (RF), KNN, and SVM, were assessed. When class imbalance was addressed using SMOTE, RF performed best, achieving 94.12% accuracy and a 94.10% F1-score. Additionally, their work improved interpretability through comparative visualizations using ROC curves and confusion matrices, as well as leaf node analysis. Building on this framework, a number of scholars have put forth more sophisticated methods. Bhargav and Sashirekha [4] compared RF and Decision Tree (DT) for loan approval, observing RF's higher precision (79.44%) and reliability. Wang et al. [3] introduced a stacking-based deep learning model with CNN-based feature extraction and counterfactual data augmentation, improving joint loan approval accuracy by 6%. Dasari et al. [6] proposed a voting and bagging ensemble technique, increasing accuracy from 80% to 94%. Abdullah et al. [5] used RF and Extremely Randomized Trees to predict nonperforming loans in 322 banks across 15 emerging economies, emphasizing institutional over macroeconomic variables. Alsaleem and Hasoon [7] evaluated MLP, Naïve Bayes, and BayesNet on 1,000 loan samples, where MLP outperformed traditional classifiers. Ghatasheh [8] found that RF outperforms DT in both precision and robustness on complex banking datasets. Shoumo et al. [9] assessed SVM, Logistic Regression, and RF for credit scoring and noted SVM’s superior performance in identifying eligible applicants, challenging tree-based dominance. Jency et al. [10] performed exploratory analysis, showing income, credit history, and education as key features in predicting loan outcomes. Anand et al. [11] used a dataset of 850 entries to compare CatBoost, LightGBM, and Extra Trees, confirming tree-based ensembles outperform simpler models in high-dimensional settings. Blessie and Rekha [12] applied LR, SVM, NB, and DT for loan approval, where Naïve Bayes surprisingly achieved the highest accuracy under certain data conditions. Kumar et al. [13] applied a hybrid AdaBoost ensemble over RF, KNN, and SVM, showing that boosting weak classifiers improves real-time banking performance. Dosalwar et al. [14] used a Kaggle dataset and reported that Logistic Regression consistently performed well in binary classification, despite the rise of more complex models. While all these studies contribute important insights, many rely on limited models or lack integration of ensemble and deep learning techniques. Our present work builds directly on [1], extending it by using a larger, more balanced dataset and a diverse set of advanced models to further optimize loan approval prediction.

III. Research Methedology

Figure 1 shows the overall structure of the proposed loan approval prediction model, which consists of multiple stages including data preprocessing, balancing, splitting, model training, and evaluation. The workflow integrates ML, DL, and ensemble methods to build an interpretable and high-performing classifier. Techniques such as SMOTE for balancing, grid search for optimization, and SHAP & LIME for explainability are incorporated into the workflow. The visual pipeline provides a step-by-step overview of how raw data is transformed into final model predictions for binary loan decisions.

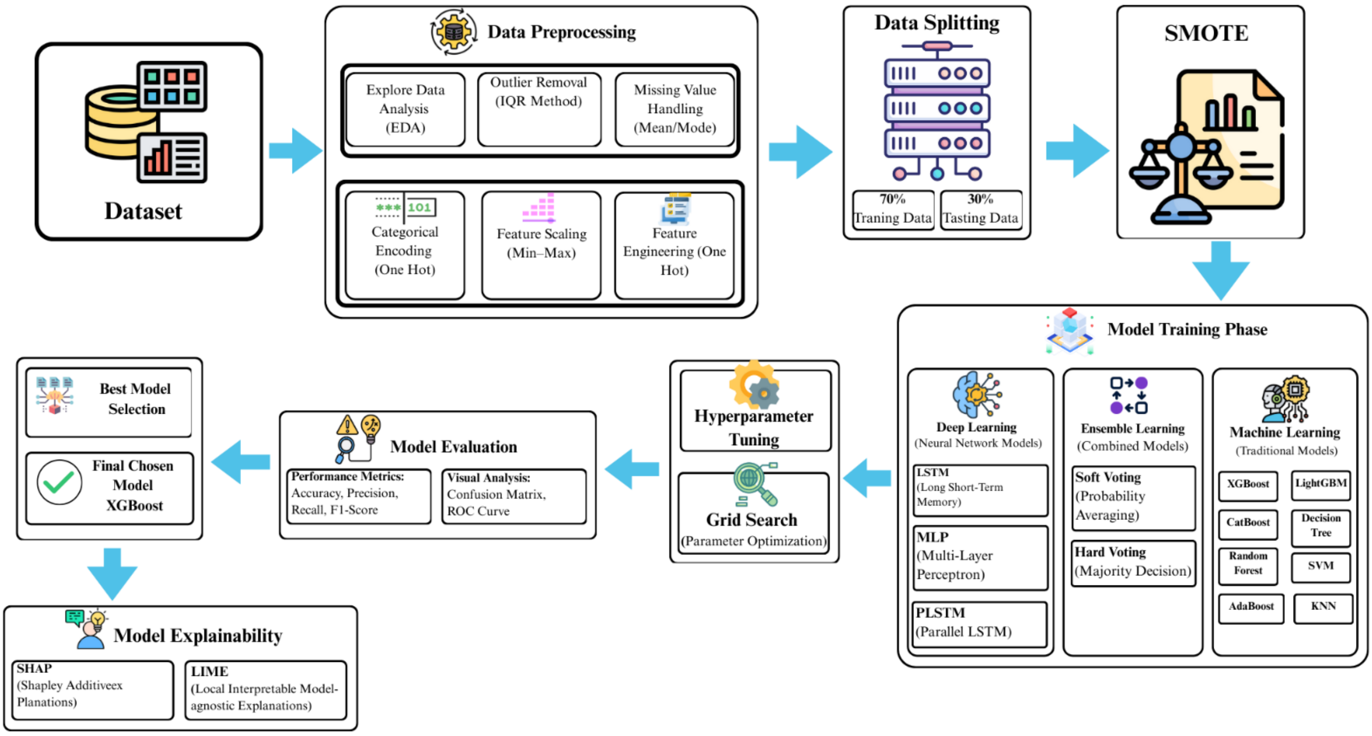


Fig 1. The operational procedure of the proposed technique

A. *Dataset Collection*

Table I provides an overview of the dataset used in this study. The dataset was obtained from Repositories and consists of 5,000 rows, each of which represents the profile of a consumer. The features include personal, financial, and behavioral information. The target attribute, Personal_Loan, is a binary variable that shows whether or not a customer accepted the loan offer. There were no missing values in the dataset, and preliminary analysis showed a combination of numerical and categorical variables appropriate for supervised learning. Due to their large cardinality and irrelevance, two features—ID and ZIP_Code—were not included in the modeling process. For additional processing and model training, the remaining 13 attributes were kept.

TABLE I: DESCRIPTION OF THE DATASET ATTRIBUTES

Attribute Name	Details of Attribute	Data Type
Age	Customer's age (yearly)	Integer
Experience	Professional experience (yearly)	Integer
Income	Annual income	Float
ZIP_Code	5-digit residential code (excluded due to high cardinality)	Integer
Family	Number of family members	Integer
CCAvg	Average monthly credit card spending	Float
Education	Professional, Graduate and Undergraduate	Integer
Mortgage	Home lending amount (in thousands)	Integer
Securities_Account	1 if customer holds a securities account	Binary
CD_Account	1 if customer holds a certificate of deposit	Binary
Online	1 if online banking used	Binary
CreditCard	1 if customer owns a credit card	Binary
Personal_Loan	Loan status (Target variable)	Binary

B. *Preprocessing Dataset and Extracting Features*

A multi-step preparation pipeline was used to guarantee data quality and maximize model performance. This involved removing unnecessary fields, adjusting the distributional skew, managing outliers, dividing the sample for testing and training, balancing class labels, and encoding categorical characteristics. Below is a summary of the general steps:

Initial Cleaning: Irrelevant features such as ID and ZIP_Code were removed A check for missing values verified the dataset's completeness.

Outlier Treatment: Outliers were eliminated from continuous variables like income, experience, and CCAvg using the Interquartile Range (IQR) approach.

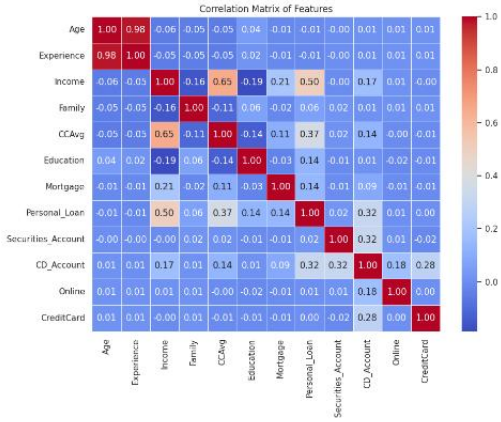
Feature Transformation: Skewed variables were normalized using log transformation. Min-Max scaling was used to standardize numerical data, and one-hot encoding was used to encode categorical variables like education.

Data Balancing: Class imbalance in the Personal_Loan variable was addressed using SMOTE, which synthetically generated samples for the minority class.

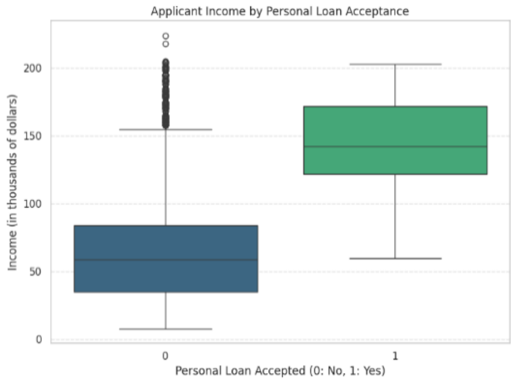
Data Splitting: The cleaned and balanced dataset was split into 70% training and 30% testing sets to ensure unbiased model evaluation.

Model Preparation: Following preprocessing, a total of fifteen models—including machine learning (e.g., XGBoost, RF, SVM), deep learning (e.g., LSTM, MLP, PLSTM), and ensemble classifiers (e.g., Hard and Soft Voting)—were trained and evaluated.

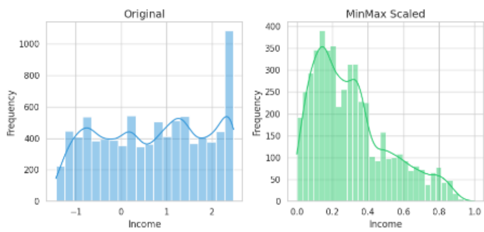
The effects of the preprocessing steps are visualized in **Figure 2**, which includes violin plots, feature correlations, and class distribution charts. These steps ensured a robust and ready dataset for training.



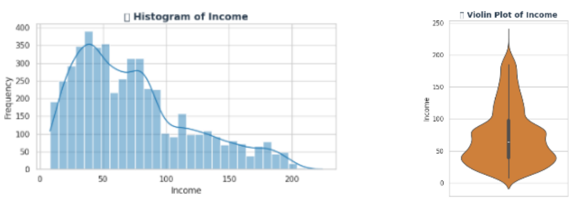
(a) Correlation Matrix of Features



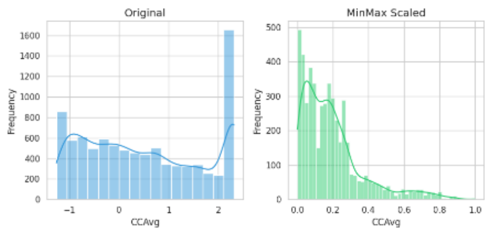
(b) Applicant Income by Personal Loan Acceptance



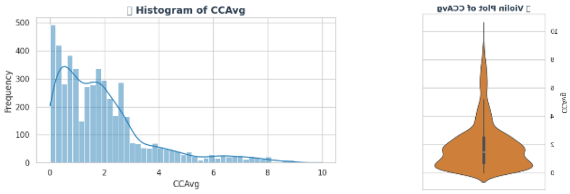
(c) Income Scaling Comparison



(d) Income Distribution Visualization



(e) CCAvg Scaling Comparison



(f) CCAvg Distribution Visualization

Fig 2. Some Visualization of dataset in preprocessing part

IV. Experimental Results and Discussion

Table II presents the performance comparison of all models categorized into three groups: ML, ensemble learning, and ML. Every model was assessed using four standard classification metrics—Accuracy, Precision, Recall, and F1-Score—to ensure a comprehensive assessment of prediction performance. Among the machine learning models, XGBoost achieved the best performance with an accuracy of 99.30% and an F1-Score of 96.62%, followed closely by LightGBM and CatBoost, both exceeding over 99% accuracy. These results demonstrate the strong predictive capabilities of gradient boosting and tree-based models for structured tabular data. Traditional models like SVM, KNN, and Logistic Regression also performed consistently well, while Naïve Bayes showed comparatively lower accuracy. In the ensemble learning category, both Soft Voting and Soft Voting Classifiers produced high performance metrics, achieving over 99% accuracy. These results validate the effectiveness of combining multiple models to improve stability and generalization. For DL models, Multilayer Perceptron (MLP) achieved performance comparable to the top-performing ML models, with an accuracy of 99.30%. PLSTM and LSTM also showed strong results, though LSTM had a slightly lower F1-Score due to a drop in recall.

The detailed evaluation of all models is summarized below:

TABLE II: Performance Comparison of All Models

Machine Learning Models				
Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
XGBoost	99.30%	98.04%	95.24%	96.62%
CatBoost	99.00%	99.00%	99.88%	99.44%
Decision Tree	98.90%	98.89%	98.90%	98.89%
Random Forest	98.80%	98.70%	98.70%	98.67%
LightGBM	99.10%	98.98%	92.38%	95.57%
SVM	97.70%	97.70%	97.70%	97.60%
AdaBoost	97.80%	97.80%	97.80%	97.71%
Logistic Regression	95.40%	95.16%	95.40%	95.16%
KNN	96.00%	95.88%	96.00%	95.72%
Naïve Bayes	89.40%	90.50%	89.40%	89.87%
Ensemble ML Models				
Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Soft Voting Classifier	99.00%	99.00%	99.00%	98.98%
Hard Voting Classifier	98.40%	98.43%	98.40%	98.33%
Deep Learning Models				
Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
MLP	98.80%	98.79%	98.80%	98.79%
PLSTM	98.50%	98.48%	98.50%	98.48%
LSTM	98.30%	93.14%	89.48%	91.79%

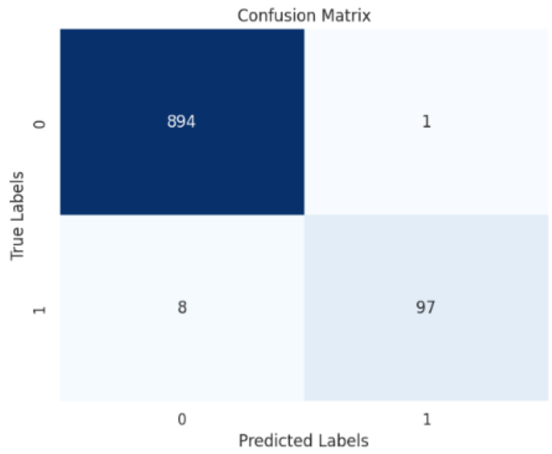


Fig 3. Confusion Matrix (XGBoost)

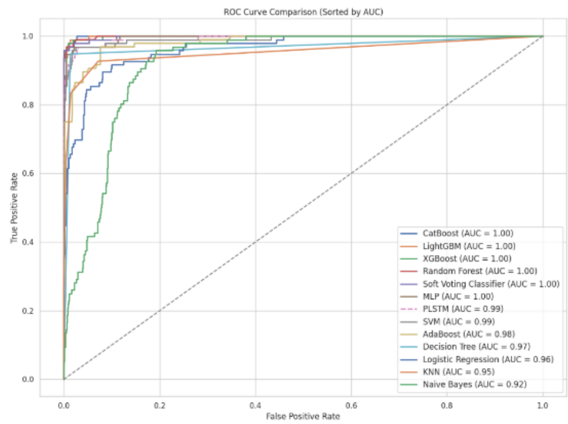


Fig 4. ROC Curve for 13 classifier

XGBoost clearly stands out as the top-performing model, maintaining a near-perfect balance across all evaluation metrics. The confusion matrix of XGBoost, shown in **Figure 3**, confirms its strong classification precision, with almost all predictions falling on the diagonal and only a few misclassifications observed. This demonstrates the effectiveness of the preprocessing pipeline, particularly SMOTE balancing, in ensuring fair representation of both classes.

Figure 4 illustrates the ROC curves of all tested classifiers, where XGBoost again dominates with the highest AUC value, approaching 1.0. The curve highlights its superior ability to distinguish between approved and non-approved loan applicants across varying thresholds, making it more reliable than other baseline models.

Furthermore, **Figure 5** presents the relationship between the number of leaf nodes and XGBoost accuracy. The results show a rapid increase in performance as leaf complexity grows, eventually reaching a peak at the optimal level before stabilizing. This confirms that the model not only achieves high accuracy but also maintains stability, avoiding overfitting at higher complexities.

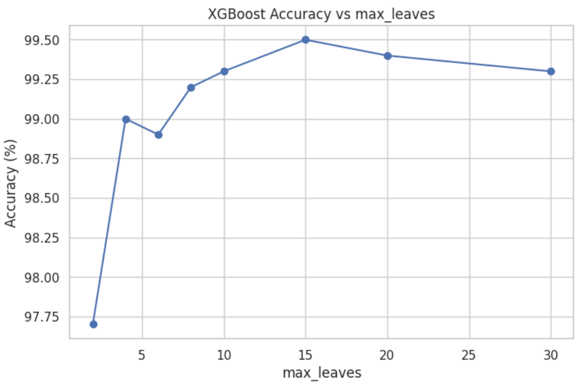


Fig 5. Accuracy vs Max Leaf Nodes

To enhance model interpretability, explainable AI techniques were applied to the best-performing XGBoost model. SHAP (SHapley Additive exPlanations) was used to analyze the global importance of features, while LIME (Local Interpretable Model-agnostic Explanations) provided local, instance-specific explanations.

As shown in **Figure 6**, SHAP results highlight that Income, CD_Account, and CCAvg are the most influential features in loan approval decisions, aligning with financial intuition. Higher income and the presence of a CD account positively impacted approvals, while low average credit card spending and weak mortgage history reduced approval likelihood.

In contrast, **Figure 7** demonstrates LIME's capability to explain an individual prediction. For a specific customer, Income was the strongest positive factor, whereas low education level and absence of a CD account contributed negatively. This instance-level interpretability ensures transparency in decisions and provides explanations that can be directly communicated to applicants.

By combining SHAP and LIME, the framework offers both global interpretability (fairness and consistency across the dataset) and local interpretability (clarity in individual decisions). This dual-layer explanation strengthens the trustworthiness of the model, making it not only accurate but also suitable for deployment in real-world banking systems.

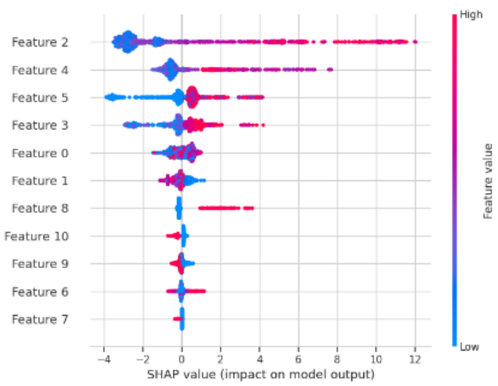


Fig 6. SHAP summary plot showing global feature importance for the XGBoost model.

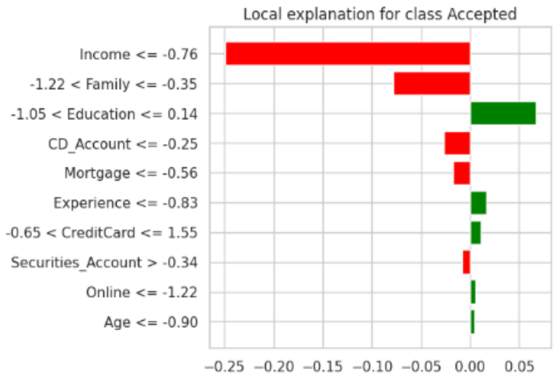


Fig7. LIME visualization explaining individual prediction for a customer

These findings establish that well-tuned ensemble and DL models, when trained on clean and balanced data, can deliver highly reliable results for automated loan approval systems.

A comprehensive comparison with existing studies is presented in **Table III**. Earlier works such as Bhargav and Sashirekha [4] and Abdullah et al. [5] used RF and Extremely Randomized Trees on loan datasets, achieving accuracies of 79.44% and 76.10%, respectively. Although Dasari et al. [6] improved performance using ensemble voting techniques and reached 94%, their dataset size and depth of preprocessing were limited. Alsaleem and Hasoon [7] evaluated MLP and Naïve Bayes on a smaller dataset of 1,000 instances and reported 89.70% accuracy. Similarly, Shoumo et al. [9] applied SVM and Logistic Regression for credit scoring and achieved 90.50% accuracy.

TABLE III: COMPARATIVE ANALYSIS OF PROPOSED MODEL WITH EXIXTING STUDIES

Author(s)	Model Used	Dataset	Best Accuracy
Bhargav and Sashirekha [4]	RF	Kaggle	79.44%
Abdullah et al. [5]	RF, ET	Multi-country banks	76.10%
Dasari et al. [6]	Voting Classifier	Custom Bank Data	94.00%
Alsaleem and Hasoon [7]	MLP, Naïve Bayes	Bank Loan (1000 rows)	89.70%
Shoumo et al. [9]	SVM, Logistic Regression	Credit Scoring	90.50%
Istia et al. [1]	RF	Kaggle	94.12%
Proposed Model	**XGBoost**	Repositories(5000 rows)	**99.30%**

Our own previous study [1], using SMOTE and RF on a Kaggle dataset, achieved 94.12%. However, it lacked integration of DL and ensemble diversity. In contrast, the proposed model in this study combines high-quality data (5,000 balanced rows), robust preprocessing (including outlier removal, normalization, and SMOTE), and a wide range of models including classical ML, DL, and ensemble classifiers. With an accuracy of 99.30%, the XGBoost model developed in this study performed better than earlier research in terms of both performance and generalization. Furthermore, explainability methods like SHAP and LIME were applied to enhance transparency, which was often absent from earlier models. This study's remarkable accuracy, balanced data, model diversity, and interpretability make it a comprehensive and innovative approach to personal loan prediction.

V. Conclusion And Future Work

This study introduces a hybrid ensemble machine learning and deep learning architecture to generate intelligent predictions regarding the acceptance of personal loans. The study used a structured dataset of 5,000 customer records and employed a number of preprocessing techniques, including outlier removal, normalization, feature modification, and class balance using SMOTE. A variety of models from three categories—machine learning, deep learning, and ensemble learning—were trained and evaluated using a single pipeline. With an astounding accuracy rate of 99.30% and exceptional precision, recall, and F1-score values, XGBoost performed better than any other tested model. Other models, such as CatBoost, DT, MLP, and PLSTM, also produced competitive results, indicating the effectiveness of the recommended approach. The use of explainability tools such as SHAP and LIME further enhanced the forecasts' interpretability and transparency, which is essential for real-world application in the financial sector. The proposed hybrid approach successfully addresses important problems in loan approval prediction, including unbalanced data, model generalizability, and trustworthiness. Because of its performance and scalability, it is suitable for modern financial environments where automated, accurate, and explicable decision-making is crucial. In future research, the system can be extended to incorporate real-time transactional and behavioral data in order to further improve prediction accuracy.

Additionally, by adding a feedback loop with user behavior analytics and altering the framework for multi-class risk categorization, financial institutions can obtain greater flexibility and insight. Finally, if the model were integrated into a real-time decision support platform with interactive dashboards, analysts and end users could use its predictions in actual banking situations.

References

[1] U. A. M. Istia, Al-Amain, K. M. M. Uddin, M. T. Ul Islam and M. A. Based, "An Integrated Approach Using Ensemble Machine Learning and Deep Learning for Loan Approval Prediction," 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE), Chittagong, Bangladesh, 2025, pp. 1-6, doi: 10.1109/ECCE64574.2025.11013889.

[2] Nazim Uddin, Md. Khabir Uddin Ahamed, Md Ashraf Uddin, Md. Manwarul Islam, Md. Alamin Talukder, Sunil Aryal,An ensemble machine learning based bank loan approval predictions system with a smart application, International Journal of Cognitive Computing in Engineering, Volume 4,2023,Pages 327-339,ISSN 2666-3074

[3] Y. Wang, M. Wang, Y. Pan, J. Chen. Joint loan risk prediction based on deep learning-optimized stacking model. Eng. Rep. (2023), Article e12748

[4] P. Bhargav, K. Sashirekha. A machine learning method for predicting loan approval by comparing the random forest and decision tree algorithms. J. Surv. Fish. Sci., 10 (1S) (2023), pp. 1803-1813

[5] M. Abdullah, M.A.F. Chowdhury, A. Uddin, S. Moudud-Ul-Huq.Forecasting nonperforming loans using machine learning. J. Forecast.

[6] Y. Dasari, K. Rishitha, O. Gandhi. Prediction of bank loan status using machine learning algorithms. Int. J. Comput. Digit. Syst., 14 (1) (2023), p. 1

[7] M.Y. Alsaleem, S.O. Hasoon. Predicting bank loan risks using machine learning algorithms. AL-Rafidain J. Comput. Sci. Math., 14 (1) (2020), pp. 149-158.

[8] N. Ghatasheh. Business analytics using random forest trees for credit risk prediction: A comparison study. Int. J. Adv. Sci. Technol., 72 (2014), pp. 19-30

[9] S.Z.H. Shoumo, M.I.M. Dhruba, S. Hossain, N.H. Ghani, H. Arif, S. Islam. Application of machine learning in credit risk assessment: A prelude to smart banking. TENCON 2019-2019 IEEE region 10 conference (TENCON), IEEE (2019), pp. 2023-2028

[10] X.F. Jency, V. Sumathi, J.S. Sri. An exploratory data analysis for loan prediction based on nature of the clients. Int. J. Recent Technol. Eng., 7 (4) (2018)

[11] M. Anand, A. Velu, P. Whig. Prediction of loan behaviour with machine learning models for secure banking. J. Comput. Sci. Eng., 3 (1) (2022), pp. 1-13.

[12] E.C. Blessie, R. Rekha. Exploring the machine learning algorithm for prediction the loan sanctioning process.Int. J. Innov. Technol. Explor. Eng., 9 (1) (2019), pp. 2714-2719

[13] Kumar et al., 2022 C.N. Kumar, D. Keerthana, M. Kavitha, M. Kalyani Customer loan eligibility prediction using machine learning algorithms in banking sector 2022 7th international conference on communication and electronics systems (ICCES), IEEE (2022), pp. 1007-1012

[14] Dosalwar et al., 2021 S. Dosalwar, K. Kinkar, R. Sannat, D. Pise Analysis of loan availability using machine learning techniques Int. J. Adv. Res. Sci., Commun. Technol. (September 2021), pp. 15-20

[15] D. Chatterjee. Loan prediction problem dataset https://www.kaggle.com/altruistdelhite04/loan-prediction-problem-dataset (2021)

[16] A. Khairi, B. Bahri, B. Artha. A literature review of non-performing loan. J. Bus. Manag. Rev., 2 (5) (2021), pp. 366-373

[17] M. Musdholifah, U. Hartono, Y. Wulandari. Banking crisis prediction: Emerging crisis determinants in Indonesian banks. Int. J. Econ. Financ. Issue., 10 (2) (2020), p. 124