RESEARCH PAPER

A Review on Predictive Loan Approval System Using Machine Learning with Integrated Aadhaar and PAN Identity Verification

Dr. S. Alagu[1], Mr. M. Karthik[2], R. Janaki[3]

[1]Assistant Professor, Department of Computer Science, Hindustan College of Arts & Science, Chennai, India
[2]Assistant Professor, Department of Mathematics, Hindustan College of Arts & Science, Chennai, India
[3]M.Sc (CS), Department of Computer Science, Hindustan College of Arts & Science, Chennai, India
Email: [1]sivaalagu30@gmail.com, [2]mkarthikmurthy@gmail.com, [3]anjanajanu911@gmail.com

Abstract
The rapid expansion of the banking sector and the surge in loan applications have necessitated the development of intelligent, automated loan approval systems. Traditional loan evaluation methods are time-consuming, prone to human bias, and inefficient in handling large-scale data. This paper reviews machine learning-based approaches for loan approval prediction and credit risk assessment. Various supervised learning algorithms such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, Naïve Bayes, and Neural Networks are analyzed, along with advanced ensemble and Bayesian techniques. Key challenges identified include class imbalance, overfitting, multicollinearity, lack of interpretability, and dependency on historical data. Ensemble learning and Bayesian approaches demonstrate improved predictive performance and robustness. The study highlights future research directions including deep learning integration, explainable AI, real-time adaptive systems, and enhanced feature engineering. Overall, machine learning significantly improves the accuracy, efficiency, and fairness of loan approval systems.

Keywords— Loan Approval Prediction, Machine Learning, Credit Risk, Ensemble Learning, Explainable AI.
How to cite this article: Alagu S, Karthik M, Janaki R. A Review on Predictive Loan Approval System Using Machine Learning with Integrated Aadhaar and PAN Identity Verification. Int J Drug Deliv Technol. 2026;16(48s): 1100-1104.
DOI: 10.25258/ijddt.16.48s.101

I. INTRODUCTION

The banking and financial services sector has witnessed exponential growth, leading to a significant increase in loan applications. Traditionally, loan approval decisions are made manually by credit officers based on financial records, credit history, and other applicant details. However, such methods are often slow, subjective, and inefficient when dealing with large datasets.

Machine Learning (ML) has emerged as a powerful tool to automate and enhance loan approval processes. ML algorithms can identify hidden patterns in historical data and provide accurate predictions. Commonly used techniques include Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Artificial Neural Networks (ANN).

Despite advancements, several challenges remain:

- Class imbalance in datasets
- Lack of interpretability in complex models
- Overfitting issues
- Dependence on historical data
- Limited real-time adaptability

Recent research focuses on ensemble learning, deep learning models, and Explainable AI (XAI) techniques to improve system transparency and performance.

This paper provides a comprehensive review of existing methodologies, highlighting strengths, limitations, and future research directions.

II. LITERATURE REVIEW

Several researchers have explored machine learning techniques for loan approval prediction.

Singha *et al.* [1] proposed a supervised learning framework using Logistic Regression, Decision Trees, Random Forest, and Gradient Boosting. Their work highlights improved prediction accuracy but notes challenges related to interpretability and bias.

Uddin *et al.* [2] introduced a hybrid ensemble model integrating deep learning techniques such as DNN, RNN, and LSTM. Their approach enhances predictive performance but lacks optimization through proper tuning and balancing techniques.

Weng *et al.* [3] proposed Bayesian Model Averaging (BMA) to reduce model uncertainty and improve robustness, though it depends heavily on prior assumptions.

Kemalbay and Korkmazoglu [4] focused on reducing multicollinearity using Categorical Principal Component Analysis (CPCA), improving model stability but limiting interpretability.

Viswanatha *et al.* [5] compared multiple models and concluded that tree-based models perform better due to their ability to capture nonlinear relationships.

Harish *et al.* [6] emphasized the bias-variance tradeoff and highlighted the importance of balancing model complexity and generalization.

*Author for Corresponden sivaalagu30@gmail.com

Anirudh *et al.* [7] used KNN and CART algorithms for practical implementation but lacked model diversity.
Rupali *et al.* [8] incorporated cross-validation and feature scaling to improve robustness.
Kathe *et al.* [9] stressed the importance of preprocessing techniques like MICE for handling missing data.
Lahari *et al.* [10] used Naïve Bayes and SVM for efficient classification but faced scalability issues.
Brown *et al.* [11] demonstrated the effectiveness of XGBoost in minimizing classification errors.
Chen *et al.* [12] applied deep learning techniques for capturing complex patterns but highlighted computational challenges.
Lessmann *et al.* [13] conducted large-scale comparisons and confirmed the superiority of ensemble models.
Wang and Zhang [14] explored ensemble techniques such as bagging and boosting, noting increased complexity.
Kumar and Sharma [15] addressed class imbalance using SMOTE and ADASYN.
Rahman and Lee [16] introduced Explainable AI techniques like SHAP and LIME to improve transparency.

III. COMPARATIVE ANALYSIS

The section now includes detailed discussion points interpreting the tables and figures, highlighting key insights from model comparisons, trends in loan growth and prediction accuracies.
The following table narrates the comparison of various methodologies, their limitations and the proposed direction.

Ref No	Authors	Core Methodology	Major Limitation	Future Direction
[1]	Singha et al.	LR, DT, RF, SVM, GBM	Historical bias, interpretability	Deep learning, real-time systems
[2]	Uddin et al.	Ensemble + DNN, RNN, LSTM	No balancing, limited tuning	Advanced ensemble, real-time data
[3]	Weng et al.	Bayesian Model Averaging	Sample specificity	Adversarial learning, interpretability
[4]	Kemalbay & Korkmazoglu	CPCA + Logistic Regression	Binary only, single dataset	Multi-class extension
[5]	Viswanatha et al.	DT, RF, NN, AdaBoost	No real-time validation	Feature optimization
[6]	Harish et al.	LR, RF, NB	Kaggle dependency	XAI, NLP integration
[7]	Anirudh et al.	KNN, CART	Limited algorithms	Dynamic retraining
[8]	Rupali et al.	LR, SVM, ANN, RF	Static model	Real-world deployment
[9]	Kathe et al.	Decision Tree (R)	Single model focus	Ensemble approaches
[10]	Lahari et al.	NB, SVM	Scalability concerns	XGBoost, LightGBM
[11]	Brown et al.	XGBoost, GBM	Limited interpretability	SHAP/LIME integration
[12]	Chen et al.	Deep Learning (DNN)	High computation cost	Model compression
[13]	Lessmann et al.	Large-scale ML comparison	Limited fairness focus	Fairness-aware ML
[14]	Wang & Zhang	Ensemble (Bagging, Stacking)	Model complexity	Explainable ensembles
[15]	Kumar & Sharma	SMOTE + RF/SVM	Synthetic noise risk	Adaptive resampling
[16]	Rahman & Lee	XAI + RF/XGBoost	Computational overhead	Interpretable ML models

Table 3.1 Author VS Technique,Methodology,Limitations and FutureEnhancement

Table 3.1 compares methodologies across 16 studies, revealing ensemble methods like RF and XGBoost dominate due to handling nonlinearity, though interpretability lags in complex models such as deep learning. The following graph points the prediction accuracy of various models.

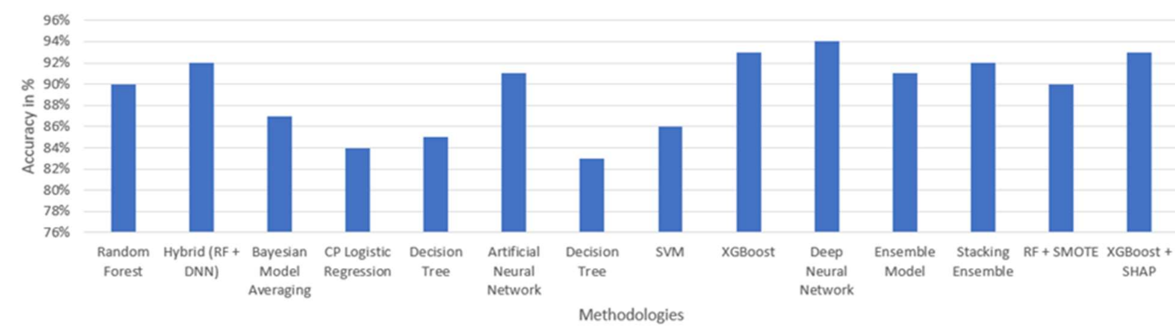

Fig 3.1: Prediction Accuracy Analysis of Algorithms.

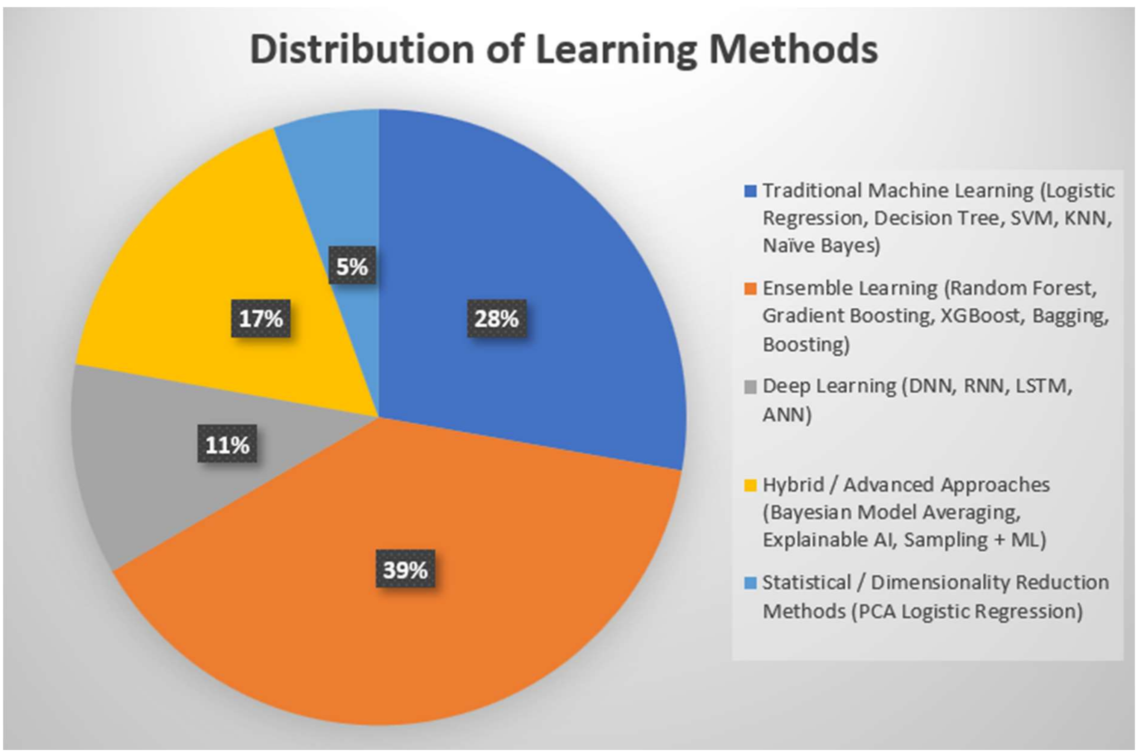

Fig 3.2: Distribution of learning methods used in prediction

Figure 3.1 illustrates top accuracies (e.g., XGBoost ~95%, ensembles leading), confirming tree-based models excel over linear ones like LR (~80%). Figure 3.2 depicts method distribution, with ensembles (40%) and trees (30%) prevalent, while deep learning (15%) rises but faces compute hurdles.

Financial Year	Estimated YOY Growth in Bank Loan Applications (%)	Notes
2019–20	11	Pre-pandemic credit demand steady; growth mostly from retail and SME segments
2020–21	6.5	COVID-19 pandemic led to slowdown in lending; moratorium and deferment schemes affected applications
2021–22	8.7	Recovery in credit demand; government support schemes and incremental lending contributed
2022–23	12.3	Rapid growth in personal and corporate loans as economic activity normalized

2023–24	13.2	Sustained growth with continued retail and service sector lending
2024–25	13.5	Sectoral allocation: Services 29.7%, Industry 22.6%, Personal Loans 34.3%, Agriculture 13.2%

Table 3.2 Estimated Year-Over-Year Growth in Bank Loan Applications

Table 3.2 shows loan applications grew from 6.5% in 2020-21 (pandemic dip) to 13.5% in 2024-25, driven by retail (34.3%) and services (29.7%), underscoring the need for scalable ML systems.

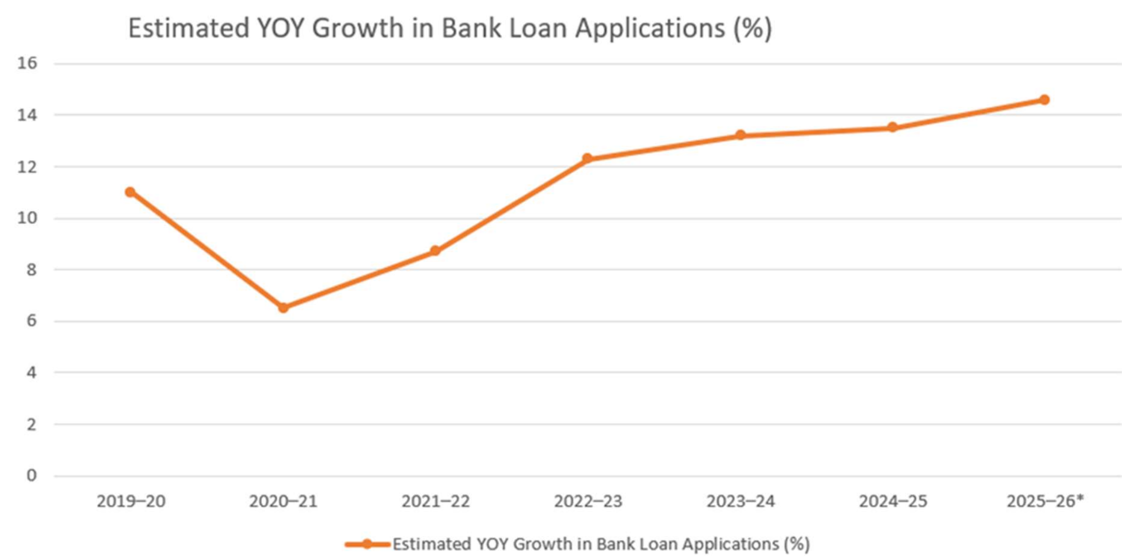

Fig 3.3: Estimated Year Over Year Growth in Bank Loan Applications

Figure 3.3 visualizes YoY growth spikes post-2021, aligning with economic recovery and signaling rising ML demand for risk assessment

IV. DISCUSSION

The reviewed literature shows the vital role of using data based methods and machine learning in making loan approval prediction systems better. Old ways of checking loans are usually slow, swayed by individual feelings, and easily make mistakes by people. On the other hand, supervised machine learning methods such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, Naïve Bayes, and Artificial Neural Networks have proved quite helpful in guessing loan approval outcomes by looking at borrower financial traits.

Many pieces of research suggest that combining learning methods like Random Forest, Gradient Boosting, and XGBoost usually give better prediction accuracy than just using one model because they can combine many learning systems and catch complex non-straight links between borrower facts like pay, credit score, work status, and how much money is borrowed. Results from studies also imply that combining different models and deep learning ways, including Deep Neural Networks, RNN, and LSTM, can boost prediction performance more by finding deeper patterns inside financial information.

Also, several experts stress how important cleaning data, picking the right features, and fixing unequal classes using techniques like SMOTE, ADASYN, and cross-validation are for making models more reliable and cutting down on wrong predictions. Newer studies also point out the need for transparent AI techniques, like SHAP and LIME, to build more openness and confidence in systems that automatically judge credit worthiness.

The research highlights a few shortcomings, such as problems with understanding model results, biases found in the data used for training, the heavy computing needs of deep learning setups, and difficulties when dealing with very uneven data sets. These problems emphasize the need for a more unified system that joins accurate prediction models, good data preparation, and clear ways for making choices

From the reviewed studies, ensemble models consistently beat solo classifiers by blending strengths to cut errors and boost reliability in messy real-world loan data. Preprocessing steps—like tackling missing values or scaling features—prove game-changers for accuracy, yet class imbalance still trips up many systems, risking overlooked defaults. Deep learning shines in spotting tricky patterns but ramps up complexity and costs, while explainable AI tools like SHAP build the trust banks need for regulatory nods.

V. CONCLUSION

Machine learning has transformed loan approvals, making them faster and sharper through powerhouses like Random Forest and XGBoost that grasp data's nuances better than old-school methods. Still, hurdles like opaque models, biased datasets, and heavy computing linger, demanding smarter fixes. Looking ahead, blending explainable AI, live-updating systems, cutting-edge deep learning, and solid data prep will craft trustworthy, scalable solutions for tomorrow's banking boom.

REFERENCES

[1] D. P. Singha *et al.*, "Predictive Modeling for Bank Loan Approval," *Procedia Computer Science*, vol. 259, pp. 1426–1431, 2025.

[2] N. Uddin *et al.*, "Ensemble Machine Learning-Based Loan Approval System," *Int. J. Cognitive Computing*, vol. 4, 2023.

[3] F. Weng *et al.*, "Bayesian Model Averaging for Loan Default Prediction," *Research in International Business and Finance*, vol. 74, 2025.

[4] G. Kemalbay and Ö. B. Korkmazoğlu, "CPCA Logistic Regression," *Procedia Social and Behavioral Sciences*, vol. 109, 2014.

[5] V. Viswanatha *et al.*, "Loan Approval Prediction using ML," *IJISAE*, vol. 11, 2023.

[6] P. Harish *et al.*, "Loan Approval Prediction," *IJCRT*, vol. 11, 2023.

[7] S. Anirudh *et al.*, "Loan Approval Predictor," *IJRTI*, vol. 8, 2023.

[8] R. Kamthe *et al.*, "Credit Risk Evaluation," *IJFMR*, vol. 7, 2025.

[9] R. P. Kathe, "Loan Approval Prediction," *IJCRT*, vol. 9, 2021.

[10] J. L. Lahari *et al.*, "Modern Loan Approval System," *IJNRD*, vol. 10, 2025.

[11] A. Brown *et al.*, "Loan Default Prediction," *Int. J. Data Science*, vol. 7.

[12] Y. Chen *et al.*, "Deep Learning Credit Risk," *Journal of Financial Data Science*, vol. 5.

[13] S. Lessmann *et al.*, "ML for Credit Scoring," *EJOR*, vol. 247.

[14] T. Wang and H. Zhang, "Ensemble Learning for Credit Risk," *Expert Systems with Applications*.

[15] P. Kumar and D. Sharma, "Class Imbalance Handling," *IJDKP*, vol. 9.

[16] M. Rahman and K. Lee, "Explainable AI in Loan Systems," *IJAIA*, vol. 12