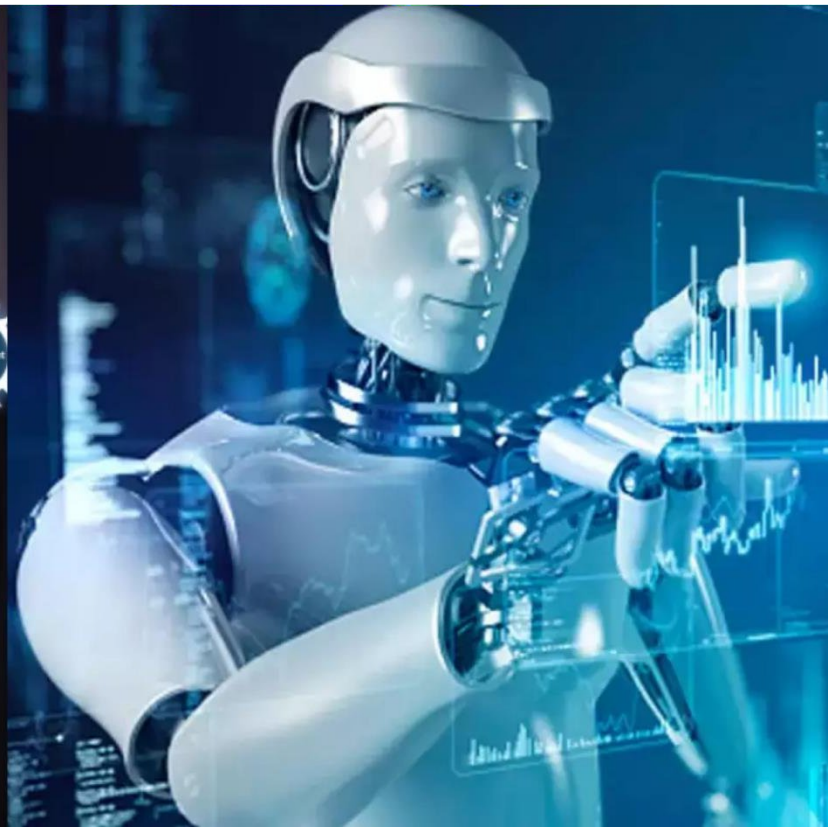
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 4, April 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504211|

Machine Learning and Deep Learning for Loan Prediction in Banking: Exploring Ensemble Methods and Data Balancing

Pooja Ravi Raut, Prof. A. D. Gujar

Department of Computer Engineering, TSSM's Bhivarabai Sawant College of Engineering and Research, Narhe, Pune, India

ABSTRACT: Loan default prediction is a critical task in the banking sector as it directly affects financial stability, profitability, and risk management. Traditional credit evaluation systems rely heavily on manual verification and expert judgment, which are time-consuming and prone to human error. With the rapid growth in loan applications and digital banking services, there is an urgent need for automated, accurate, and intelligent loan prediction systems. This study proposes a comprehensive framework integrating Machine Learning (ML), Deep Learning (DL), ensemble methods, and advanced data balancing techniques to enhance loan prediction accuracy. The dataset undergoes extensive preprocessing including feature scaling, categorical encoding, dimensionality reduction, and class imbalance handling using SMOTE and SMOTE-TOMEK techniques. Multiple ML models such as Logistic Regression, Random Forest, Decision Tree, Gradient Boosting, and Gaussian Naïve Bayes are compared with advanced DL architectures including MLP, CNN, LSTM, GRU, ResNet, and DenseNet. Performance is evaluated using Accuracy, Precision, Recall, F1-Score, AUC-ROC, and MCC. Experimental results demonstrate that ensemble and deep learning approaches, particularly DenseNet and ResNet, significantly outperform traditional models. The proposed system provides a robust, scalable, and intelligent solution for minimizing credit risk and improving banking decision-making processes.

KEYWORDS: Loan Prediction, Credit Risk Assessment, Machine Learning, Deep Learning, Ensemble Methods.

I. INTRODUCTION

The banking and financial sector has witnessed exponential growth in digital transactions and loan applications over the past decade. Loans form a major revenue source for banks, but they also introduce significant financial risk due to potential defaults. Accurately identifying trustworthy borrowers has become increasingly complex due to large volumes of applicants and diverse financial backgrounds. Traditional loan approval systems rely on manual verification processes and credit officers' judgment, which are time-intensive and susceptible to Type I and Type II errors. A Type I error (False Negative) occurs when a good applicant is rejected, resulting in loss of business opportunity. A Type II error (False Positive) occurs when a risky borrower is approved, leading to financial loss.

With advancements in Artificial Intelligence, Machine Learning and Deep Learning techniques have emerged as powerful tools for predictive analytics in finance. ML algorithms can automatically analyze large datasets, identify hidden patterns, and make accurate predictions based on historical data. However, single-model approaches often struggle with complex nonlinear relationships and imbalanced datasets common in loan prediction tasks. Therefore, integrating ensemble learning and deep neural networks has gained research attention for improving prediction robustness and generalization ability.

Furthermore, real-world banking datasets are often imbalanced, where non-default cases significantly outnumber default cases. This imbalance negatively affects model performance, especially in identifying minority class defaults. Advanced preprocessing techniques such as SMOTE and SMOTE-TOMEK help in generating synthetic minority samples and cleaning noisy boundaries. This research focuses on combining advanced preprocessing, ML algorithms, DL architectures, and ensemble methods to develop a highly accurate and reliable loan prediction framework capable of supporting automated banking decisions.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504211|

Motivation

The primary motivation of this research is to reduce financial losses caused by loan defaults while improving approval accuracy. Manual evaluation methods are inefficient and inconsistent when handling large-scale loan applications. Additionally, increasing competition among financial institutions demands faster and more reliable credit scoring systems. By leveraging AI-driven predictive modeling, banks can automate loan approval processes, minimize human bias, enhance customer satisfaction, and ensure better risk management.

II. ALGORITHM

Random Forest Algorithm

Random Forest is an ensemble learning algorithm that constructs multiple decision trees during training and outputs the final prediction based on majority voting (classification).

Steps:

1. Select random samples from the dataset (Bootstrap Sampling).
2. Construct multiple decision trees using randomly selected feature subsets.
3. Each tree makes an independent prediction.
4. Final prediction is determined by majority voting.
5. Evaluate model using Accuracy, Precision, Recall, F1-score.

Random Forest reduces overfitting, improves stability, and provides high predictive accuracy.

III. LITERATURE REVIEW

SefikIlkinSerengil [1], In this paper, we studied on several machine learning algorithms to solve this problem and we propose a comparative study of some of the mostly used non performing loan models on a customer portfolio dataset in a private bank in Turkey. We also deal with a class imbalance problem using class weights. A dataset, composed by 181.276 samples, has been used to perform the analysis considering different performance metrics (i.e. Precision, Recall, F1 Score, Imbalance Accuracy (IAM), Specificity). In addition to these, we evaluated the performance of the algorithms and compared the obtained results.

Jian Chen [2], In this study, we approximate financial stress by agricultural loan delinquency, and predict it by employing a logistic regression and several machine learning methods. The main datasets include the Call Reports and Summary of Deposits from the Federal Deposit Insurance Corporation (FDIC). Our results show that ensemble learning methods have the best performance in prediction accuracy, with improvement of 26 percentage points at most and that the Naïve Bayes classifier is the best method to maintain the lowest cost from false predictions when the failure of identifying potentially high-risk loans is very costly.

DebabrataDansana [3], The purpose of this article is to determine whether to grant loans to specific individuals or organizations. The Random Forest Regressor model has been utilized to mea sure performance and identify suitable customers for loan approval. The model suggests that banks should not only target affluent clients but also consider other customer characteristics that are critical in credit granting and predicting loan default.

NavdeepSood [4], Network controllers have available to them all manner of data from the different network elements and the various network links. Techniques are presented herein that leverage that data to support a new heat map concept. The presented techniques begin with the generation of node-level heat maps and services-level heat maps. Those artifactsmay be employed to study the impact of a given failure on the nodes and the services that are present in a network.

L. UdayaBhanu [5], Customer loan prediction is usually life time issue so; each and every retail bank faces the issue at the minimum lifetime. If done exactly, it can spare a lot's of man hours at the conclusion of a retail bank. If Company wants to semi automate the loan acceptability process (real time) based on customer detail provided while filling online application form.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504211|

MayankAnand [6], Given loan default prediction has such a large impact on earnings, it is one of the most influential factor on credit score that banks and other financial organisations face. There have been several traditional methods for mining information about a loan application and some new machine learning methods of which, most of these methods appear to be failing, as the number of defaults in loans has increased. For loan default prediction, a variety of techniques such as Multiple Logistic Regression, Decision Tree, Random Forests, Gaussian Naive Bayes, Support Vector Machines, and other ensemble methods are presented in this research work.

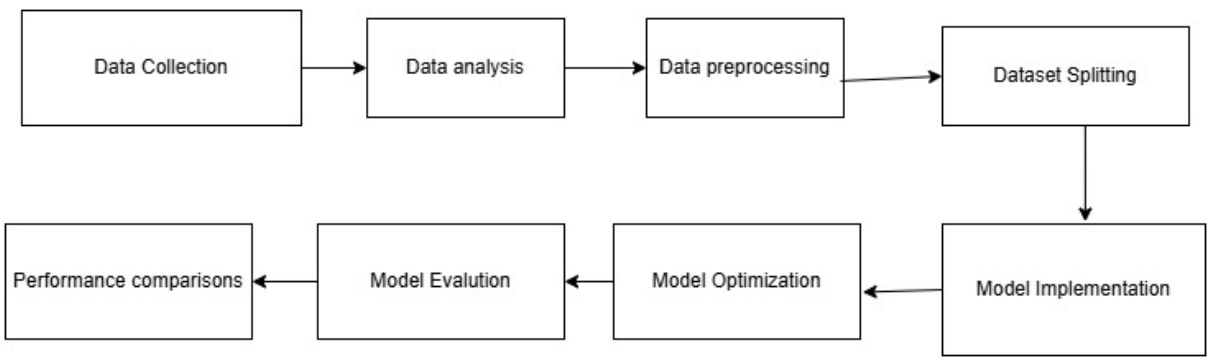
Velvigneswar. J [7], Loan approval is a critical process in the banking and financial sector, requiring accurate and timely decision making to ensure effective risk management for institutions and financial support for applicants. Traditional loan processing methods are often manual, time-consuming, and susceptible to human bias or inconsistency, which can result in delayed or inaccurate decisions. To address these challenges, this project proposes a machine learning-based solution using Random Forest, Support Vector Machine, Logistic Regression and Decision Tree algorithms to predict the likelihood of loan approval. The system is trained on historical loan data, including features such as income, employment status, credit history, education level, marital status, and loan amount, to identify meaningful patterns that distinguish approved from rejected applications.DapsePunamLaxman [8], In our banking system, banks have many products to sell but main source of income of any banks is on its credit line. So they can earn from interest of those loans which they credits. A bank's profit or a loss depends to a large extent on loans i.e. whether the customers are paying back the loan or defaulting. By predicting the loan defaulters, the bank can reduce its Non Performing Assets. This makes the study of this phenomenon very important. Previous research in this era has shown that there are so many methods to study the problem of controlling loan default.

AnantShinde [9], As the needs of people are increasing, the demand for loans in banks is also frequently getting higher every day. Banks typically process an applicant's loan after screening and verifying the applicant's eligibility, which is a difficult and time-consuming process. In some cases, some applicants default and banks lose capital. The machine learning approach is ideal for reducing human effort and effective decision making in the loan approval process by implementing machine learning tools that use classification algorithms to predict eligible loan applicants.Sara Zahi [10], Logistic regression is a widely used machine-learning model for predicting categorical outcomes. It uses a number of explanatory variables to predict the values of the target variable that can be either binomial or multinomial. It is used in a number of fields such as cancer detection problems, risk modeling and any subject that requires computing the probability of an event occurrence. In this paper, we propose to apply this supervised machine-learning model to study prepayment risk and its determinants concerning car loans based on a number of characteristics.

Gap analysis

1. Many previous studies focused only on traditional ML algorithms without integrating deep learning models.
2. Limited research addressed severe class imbalance using advanced balancing techniques.
3. Lack of comprehensive comparison between ML and DL models.
4. Insufficient focus on ensemble approaches for boosting performance.
5. Few studies evaluated models using multiple robust metrics such as MCC and AUC-ROC.

System Architecture



International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504211|

Proposed system

The proposed system is a hybrid intelligent loan prediction framework that combines advanced data preprocessing, machine learning algorithms, deep learning architectures, and ensemble techniques to enhance predictive performance. Initially, raw banking data is collected and preprocessed to remove noise, handle missing values, encode categorical variables, and scale numerical features. Class imbalance is addressed using SMOTE and SMOTE-TOMEK techniques to ensure fair representation of default cases. The processed dataset is split into training and testing sets. Multiple ML and DL models are trained and optimized using hyperparameter tuning methods such as Adam optimizer. Finally, ensemble strategies are applied to improve prediction robustness. The system outputs whether a loan should be approved or rejected, minimizing both false positives and false negatives.

IV. METHODOLOGY

1. Data Collection
- Dataset collected from banking records.
- Contains 252,000 records and 13 features.

2. Data Analysis
- Exploratory Data Analysis (EDA).
- Heatmap correlation analysis.
- Distribution analysis of features.
- Identification of target variable imbalance.

3. Data Preprocessing
- Removal of unnecessary features.
- Handling missing values.
- Encoding categorical variables (Ordinal Encoding, One-Hot Encoding).
- Feature scaling (Standardization, Min-Max Scaling).
- Dimensionality reduction.
- Class imbalance handling using SMOTE and SMOTE-TOMEK.

4. Dataset Splitting
- 80% Training data.
- 20% Testing data.

5. Model Implementation
- Machine Learning Models:
  - Random Forest

6. Model Optimization
- Hyperparameter tuning.
- Adam optimizer for DL models.
- Cross-validation techniques.

7. Model Evaluation
- Accuracy
- Precision
- Recall
- F1-score
- AUC-ROC
- MCC

8. Performance Comparison
- Compare ML vs DL models.
- Identify best performing architecture.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504211|

V. RESULT

Experimental results show that ensemble-based ML models like Random Forest performed strongly among traditional algorithms. However, Deep Learning architectures such as DenseNet and ResNet achieved superior predictive performance. After applying data balancing techniques, model accuracy significantly improved, especially in identifying minority class defaults. DenseNet achieved the highest overall accuracy and AUC-ROC score, demonstrating strong generalization capability. The integration of SMOTE-TOMEK improved recall for default cases, reducing misclassification rates.

VI. CONCLUSION

This study presents a comprehensive AI-based loan prediction framework integrating Machine Learning, Deep Learning, ensemble learning, and advanced data preprocessing techniques. The results confirm that deep learning architectures, particularly DenseNet and ResNet, outperform traditional ML models in predicting loan defaults. The use of data balancing methods significantly enhances minority class detection. The proposed system can assist financial institutions in making faster, more accurate, and data-driven loan approval decisions while minimizing financial risk. Future work may include incorporating explainable AI techniques for improved model transparency and regulatory compliance.

REFERENCES

[1] M. Anand, A. Velu, and P. Whig, ‘‘Prediction of loan behaviour with machinelearningmodelsforsecurebanking,’’J.Comput.Sci.Eng.(JCSE), vol. 3, no. 1, pp. 1–13, Feb. 2022.
[2] L. U. Bhanu and D. S. Narayana, ‘‘Customer loan prediction using super vised learning technique,’’ Int. J. Scientific Res. Publications (IJSRP), vol. 11, no. 6, pp. 403–407, Apr. 2021.
[3] J. Chen, A. L.Katchova, and C. Zhou, ‘‘Agricultural loan delinquency pre diction using machine learning methods,’’ Int. Food Agribusiness Manage. Rev., vol. 24, no. 5, pp. 797–812, Jul. 2021.
[4] S.ZahiandB.Achchab, ‘‘Modeling car loan prepayment using supervised machine learning,’’ Pro. Comput. Sci., vol. 170, pp. 1128–1133, Jan. 2020.
[5] T. Ndayisenga, ‘‘Bank loan approval prediction using machine learning techniques,’’ Master’s dissertation, 2021.
[6] S. I. Serengil, S. Imece, U. G. Tosun, E. B. Buyukbas, and B. Koroglu, ‘‘A comparative study of machine learning approaches for non performing loan prediction with explainability,’’ Int. J. Mach. Learn. Comput., vol. 12, no. 5, pp. 1102–1110, 2022.
[7] A.Shinde,Y.Patil, I. Kotian, A. Shinde, and R. Gulwani, ‘‘Loan prediction system using machine learning,’’ in Proc. ITM Web Conf. Les Ulis, France: EDPSciences, 2022, p. 3019.
[8] M. A. Sheikh, A. K. Goel, and T. Kumar, ‘‘An approach for prediction of loan approval using machine learning algorithm,’’ in Proc. Int. Conf. Electron. Sustain. Commun. Syst. (ICESC), Jul. 2020, pp. 490–494.
[9] D. Dansana, S. G. K. Patro, B. K. Mishra, V. Prasad, A. Razak, and A. W. Wodajo, ‘‘Analyzing the impact of loan features on bank loan prediction using random forest algorithm,’’ Eng. Rep., vol. 6, no. 2, 2023, Art. no. e12707.
[10] P. Kumar and N. Sood, ‘‘Controller-driven event-based network heat map visualisation,’’ Tech. Rep., 2023.
[11] T. Segal. Nonperforming Loan – NPL. (Apr. 2021). [Online]. Available: https://www.investopedia.com/terms/n/nonperformingloan.asp
[12] Banking Regulation and Supervision Agency, Turkish Banking Sector Main Indicators, March 2020.
[13] E. Mortaz, “Imbalance accuracy metric for model selection in multi class imbalance classification problems,” Knowledge-Based Systems, vol. 210, p. 106490, 2020.
[14] S. I. Serengil, S. Imece, U. G. Tosun, E. B. Buyukbas, and B. Koroglu, “A comparative study of machine learning approaches for non performing loan prediction,” in Proc. 2021 6th International Conference on Computer Science and Engineering (UBMK), September 2021, pp. 326-331.
[15] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, “Explainable AI in Fintech risk management,” Frontiers in Artificial Intelligence, vol. 3, no. 26. 2020.
[16] B. H. Misheva, J. Osterrieder, A. Hirsa, O. Kulkarni, and S. F. Lin, “Explainable AI in credit risk management,” arXivpreprint arXiv:2103.00949, 2021.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504211|

[17] Sheikh MA,GoelAK,KumarT.Anapproachforprediction of loan approval using machine learning algorithm. Paper presented at: 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 490–494. IEEE. 2020, July.

[18] MadaanM,KumarA,Keshri C, Jain R, Nagrath P. Loan default prediction using decision trees and random forest: a comparative study. IOP ConfSer Mater Sci Eng. 2021;1022:012042.

[19] Gupta A, Pant V, Kumar S, Bansal PK. Bank loan prediction system using machine learning. Paper presented at: 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART), pp. 423–426, IEEE. 2020, December.

[20] Bhoomi Patel, HarshalPatil, JovitaHembram, Shree Jaswal "Loan default forecasting using data mining" Department of Information Technology, St. Francis Institute of Technology, Mumbai, India (2020)

[21] Aakanksha, Tamara Denning, VivekSrikumar, Sneha Kumar Kesera "secrets in source code: reducing false positives using ML" software engineering (Microsoft) school of computing, USA (2020)

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details