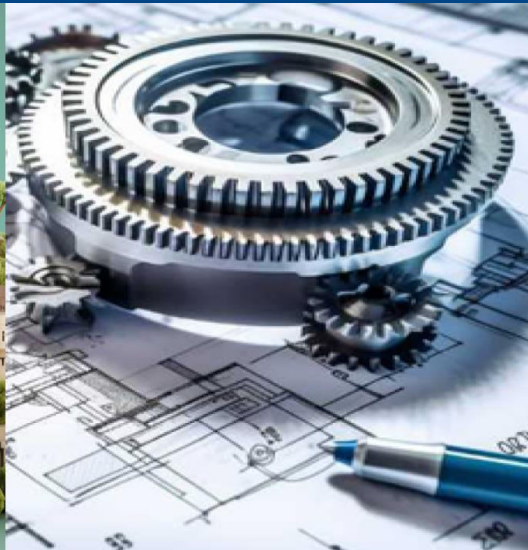
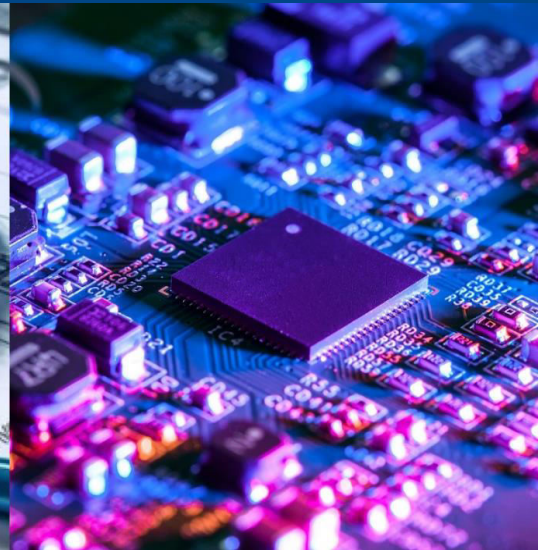
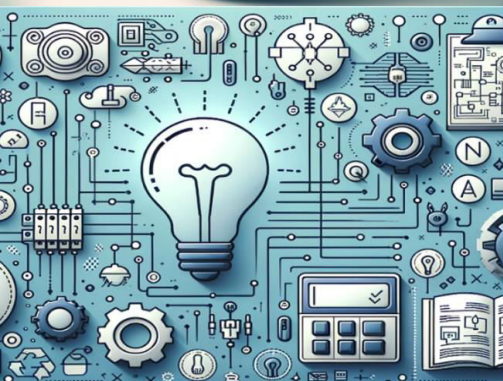




International Journal of Multidisciplinary Research in Science, Engineering and Technology

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



Impact Factor: 9.864

Volume 9, Issue 10, October 2026



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

Knowledge Graph-Enhanced XGBoost Framework for Loan Default Risk Prediction Using Integrated Financial and Relational Features

Neeta Yashwant Raut, Pradnya Ashish Vikhar, Dr. Sidhhant Nanabhau Patil

Research Scholar, Vikrant University, Gwalior, Madhya Pradesh, India

Assistant Professor, Vikrant University, Gwalior, Madhya Pradesh, India

Assistant Professor, Shatabdi Institute of Engineering and Research Agaskhind, Nashik, India

ABSTRACT: Accurate prediction of loan default risk is essential for financial institutions because incorrect credit-risk assessment can increase financial losses, weaken portfolio quality, and affect the reliability of automated lending decisions. Conventional credit-scoring approaches generally rely on structured borrower attributes and may not adequately represent the complex relationships among borrowers, financial institutions, previous credit accounts, and loan activities. This study presents a hybrid machine-learning framework that combines Knowledge Graph Embedding (KGE) with Extreme Gradient Boosting (XGBoost) for loan default prediction. The proposed framework integrates conventional tabular borrower characteristics with relational information extracted through knowledge-graph representations. The Home Credit Default Risk (HCDR) dataset is used as the experimental benchmark. A systematic preprocessing pipeline is applied, including missing-value imputation, numerical normalization, categorical encoding, filter-based feature selection, and stratified train-test partitioning. An 80:20 train-test split is employed to preserve the distribution of default and non-default classes. In the proposed architecture, borrower-, institution-, account-, and loan-related relationships are represented as knowledge-graph triples and transformed into dense embeddings. These embeddings are concatenated with original features to form an enriched representation for XGBoost classification. Reported results show an accuracy of 96.79%, precision of 80.83%, recall of 78.75%, F1-score of 79.00%, and ROC-AUC of 83.60%. The reported confusion matrix contains 55,661 true negatives, 3,876 true positives, 919 false positives, and 1,046 false negatives. The proposed model is also compared with reported baseline results for Random Forest, Neural Network, and Decision Tree models, with respective accuracies of 94.20%, 89.00%, and 73.00%. The findings indicate that combining structured financial information with relational semantic representations provides a useful framework for credit-risk classification. Future work should investigate explainability, cross-dataset validation, calibration, privacy-preserving learning, and real-time deployment.

KEYWORDS: Loan Default Prediction; Credit Risk Assessment; Knowledge Graph Embedding; XGBoost; Machine Learning; Ensemble Learning; Financial Risk Management; Credit Scoring; HCDR Dataset; Feature Engineering.

I. INTRODUCTION

Credit lending represents an important component of modern financial systems because it enables individuals and organizations to obtain capital for consumption, investment, business expansion, and other economic activities. Banks, microfinance institutions, peer-to-peer lending platforms, and digital financial-service providers consequently require reliable mechanisms for evaluating the probability that a borrower will repay a loan. Accurate credit-risk assessment allows lenders to control potential losses while maintaining an appropriate balance between credit availability and financial risk. Conventional credit assessment commonly relies on borrower income, employment information, collateral, previous repayment history, and credit records. Although these variables remain important, manual and rule-based assessment processes can be time-consuming and may have difficulty representing complex interactions among heterogeneous borrower characteristics.

Loan default occurs when a borrower fails to satisfy the repayment obligations associated with a credit agreement. Large-scale defaults can negatively affect lenders through direct financial losses, increased operational costs, and deterioration in portfolio quality. The prediction problem is challenging because default behavior is influenced by multiple interacting factors, including income instability, employment conditions, previous credit behavior, repayment



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

history, loan characteristics, and broader economic circumstances. Therefore, a reliable prediction system must identify nonlinear relationships and interactions among potentially large numbers of borrower attributes.

Machine-learning methods have increasingly been applied to credit-risk prediction because they can identify statistical patterns from historical financial data. Logistic Regression, Decision Trees, Random Forests, Support Vector Machines, neural networks, and gradient-boosting methods have all been investigated for borrower-risk classification. Ensemble learning is particularly useful because multiple predictive mechanisms can be combined to improve robustness and reduce the limitations associated with an individual classifier.

Despite these developments, many conventional ML approaches represent financial information primarily as independent tabular features. Such a representation can lose important relational information. For example, a borrower may have multiple historical loans, relationships with different financial institutions, repeated repayment events, and connections between current and previous applications. These relationships can potentially provide additional information for risk assessment.

Knowledge Graph Embedding provides a mechanism for transforming relational information into numerical vector representations. In a knowledge graph, entities and their relationships can be represented using triples of the form (head, relation, tail). Embedding methods such as TransE, DistMult, and ComplEx transform these entities and relationships into dense numerical representations. These representations can subsequently be combined with conventional tabular features.

This study therefore develops a hybrid XGBoost + KGE framework for loan-default prediction. The central idea is to integrate conventional statistical features with relational semantic representations before classification. XGBoost is selected as the principal prediction model because gradient-boosted decision trees can model nonlinear feature interactions and are well suited to structured tabular data.

The principal contributions of the study are:

- Development of a hybrid credit-risk prediction framework combining conventional borrower features and knowledge-graph representations.
- Application of a preprocessing pipeline incorporating missing-value treatment, normalization, categorical encoding, and filter-based feature selection.
- Construction of an enriched feature representation by concatenating tabular features with KGE vectors.
- Application of XGBoost for binary loan-default classification.
- Evaluation using accuracy, precision, recall, F1-score, ROC-AUC, and confusion-matrix analysis.
- Comparison of the reported proposed-model accuracy with Random Forest, Neural Network, and Decision Tree baselines.
- Identification of future opportunities involving lightweight embeddings, explainable AI, real-time deployment, and broader cross-dataset validation.

II. LITERATURE SURVEY

2.1 Machine Learning for Loan Default Prediction

Machine learning has become an important approach for automated credit-risk assessment. Traditional statistical models provide useful baseline approaches, but ML models can represent nonlinear relationships among borrower attributes. Random Forest, Decision Tree, Logistic Regression, K-Nearest Neighbors, neural networks, and boosting algorithms have therefore been investigated extensively. Akinjole et al. investigated ensemble-based ML methods for loan-default risk prediction, while Franco et al. examined loan-default prediction using LendingClub data [1], [2]. Turiel and Aste investigated peer-to-peer loan acceptance and default prediction using artificial intelligence [7].

2.2 Ensemble Learning

Single classifiers can be affected by overfitting, feature limitations, and sensitivity to particular data distributions. Ensemble learning addresses some of these issues by combining multiple learners or learning stages. Stacking is an ensemble strategy in which predictions or learned representations from one or more models are combined to support a



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

final predictive model. Research has investigated stacking approaches for personal credit-risk assessment and P2P lending [15], [16]. Han et al. further explored multi-layer and multi-view stacking for credit-risk assessment [17].

2.3 Deep Learning and Hybrid Models

Deep-learning approaches have also been explored for loan-default prediction. Lakshmanarao et al. compared conventional ML approaches with a feed-forward neural network and reported 91% accuracy [24]. Gao et al. investigated BP neural networks and KNN for bank-loan default prediction and reported accuracy exceeding 70% for the BP network [23]. Zhu et al. investigated CNN and LightGBM for loan-default prediction [12].

2.4 Privacy-Preserving and Human-in-the-Loop Approaches

Recent research has increasingly considered privacy and human oversight in automated credit-risk prediction. Ramachandra and Vaithyanathan proposed a federated Wasserstein GAN framework with differential privacy for synthetic loan-default data generation and reported 95% accuracy [18]. Khan et al. investigated a human-in-the-loop approach combined with XGBoost and reported 99.4% accuracy on LendingClub data [19].

2.5 Personalized and Knowledge-Driven Models

Zhou proposed stacked models using personalized borrower features. The study used a large dataset containing approximately 15 million applicants and 468 features, with 41 selected features. The reported stacked model achieved an AUC of 0.772 and F1-score of 0.188 [20]. Knowledge-driven approaches provide an alternative means of representing financial relationships by encoding associations among borrowers, loans, institutions, accounts, and repayment events.

2.6 Research Gap

The reviewed literature demonstrates significant progress in machine learning, deep learning, ensemble modeling, privacy-preserving learning, and personalized credit-risk prediction. However, many studies rely primarily on conventional tabular representations. Knowledge-driven feature enrichment through Knowledge Graph Embedding remains comparatively less explored in combination with advanced ensemble classifiers. This gap motivates the integration of KGE with XGBoost so that statistical borrower characteristics and relational financial information can be learned together.

III. METHODOLOGY

3.1 Overall Framework

The proposed methodology consists of six major stages: HCDR dataset acquisition, exploratory data analysis, data preprocessing, feature selection and stratified partitioning, knowledge-graph construction and embedding, and integration of KGE with tabular features followed by XGBoost classification.

Overall workflow: HCDR Dataset → EDA → Preprocessing → Feature Selection → Knowledge Graph Embedding → Feature Fusion → XGBoost → Performance Evaluation

3.2 Dataset

The Home Credit Default Risk (HCDR) dataset is used as the experimental dataset. It contains anonymized information associated with applicants, previous loans, financial histories, repayment behavior, and related credit information. The principal data sources described in the study include application.csv, bureau.csv, bureau_balance.csv, previous_application.csv, POS_CASH_balance.csv, credit_card_balance.csv, and installments_payments.csv.

3.3 Exploratory Data Analysis

Exploratory Data Analysis is used to understand feature distributions, correlations, demographic patterns, occupational characteristics, and potential class imbalance. Correlation analysis is used to identify strongly associated numerical variables and potential multicollinearity. The source study also reports descriptive demographic and occupational patterns, which should be interpreted as dataset-level observations rather than causal relationships.

3.4 Missing-Value Treatment

Mean imputation is applied to numerical variables and mode imputation to categorical variables using a Scikit-learn preprocessing pipeline.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

3.5 Feature Normalization

Numerical features are standardized using StandardScaler:

$$z = (x - \mu) / \sigma$$

where x is the original value, μ is the training-data mean, σ is the training-data standard deviation, and z is the standardized value.

3.6 Categorical Feature Encoding

One-Hot Encoding is applied to categorical variables. Each category is represented using a binary feature, avoiding the artificial ordering that can arise from simple integer encoding.

3.7 Feature Selection

A filter-based feature-selection approach is applied to remove irrelevant or redundant variables. This reduces dimensionality and computational requirements while retaining informative predictors.

3.8 Train-Test Partitioning

The processed dataset is divided using an 80:20 stratified train-test split with a reported random state of 10. Stratification preserves the approximate default/non-default class distribution.

3.9 Knowledge Graph Construction

A knowledge graph represents relationships among borrowers, financial institutions, accounts, loans, and geographic entities. Relationships are represented as triples (h, r, t) , where h is the head entity, r is the relation, and t is the tail entity. KGE methods such as TransE, DistMult, and ComplEx can transform these relationships into dense vectors.

3.10 Feature Fusion

The conventional feature vector is $x_i \in \mathbb{R}^m$ and the entity embedding is $h_i \in \mathbb{R}^d$. They are concatenated as:

$$z_i = [x_i \parallel h_i] \in \mathbb{R}^{m+d}$$

This enriched representation combines statistical borrower information with relational semantic information.

3.11 XGBoost Classification

The enriched representation is supplied to XGBoost. For binary classification, the predicted probability is:

$$P(y_i = 1 | z_i) = 1 / [1 + \exp(-f(z_i))]$$

The binary cross-entropy objective is:

$$L = - \sum [y_i \log(P(y_i)) + (1 - y_i) \log(1 - P(y_i))]$$

3.12 Performance Evaluation

The evaluation uses accuracy, precision, recall, F1-score, ROC-AUC, and the confusion matrix.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{Precision} = TP / (TP + FP)$$

$$\text{Recall} = TP / (TP + FN)$$

$$F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

$$FPR = FP / (FP + TN)$$

IV. RESULTS & DISCUSSION

4.1 Experimental Environment

The reported experiments were implemented in Python using a Jupyter Notebook environment on Google Colab. The reported libraries include Keras, TensorFlow, NumPy, Pandas, Seaborn, and Matplotlib. The reported hardware includes an NVIDIA GTX 1660-class GPU with 8 GB VRAM and 16 GB RAM.

4.2 Proposed Model Performance

Metric	Reported Value (%)
Accuracy	96.79
Precision	80.83



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

Recall	78.75
F1-score	79.00
ROC-AUC	83.60

The proposed XGBoost + KGE model achieves a reported accuracy of 96.79%, precision of 80.83%, recall of 78.75%, F1-score of 79.00%, and ROC-AUC of 83.60%. The difference between accuracy and F1-score highlights the importance of evaluating multiple metrics in an imbalanced classification problem.

4.3 Confusion Matrix Analysis

Actual / Predicted	Non-Default	Default
Non-Default	55,661	919
Default	1,046	3,876

The reported confusion matrix contains TN = 55,661, FP = 919, FN = 1,046, and TP = 3,876. False-negative cases are important because they correspond to default cases predicted as non-default, potentially exposing a lender to additional credit risk.

4.4 ROC-AUC Analysis

The reported ROC-AUC is 0.836 (83.60%). The ROC curve evaluates the relationship between true-positive rate and false-positive rate across classification thresholds.

4.5 Comparative Analysis

Model	Accuracy (%)
Proposed XGBoost + KGE	96.79
Random Forest	94.20
Neural Network	89.00
Decision Tree	73.00

The source reports 96.79% accuracy for XGBoost + KGE, compared with 94.20% for Random Forest, 89.00% for Neural Network, and 73.00% for Decision Tree. These values should be interpreted as reported comparisons because the source does not establish that every baseline was retrained under an identical preprocessing, partitioning, and hyperparameter-optimization protocol.

4.6 Contribution of Knowledge Graph Embedding

The principal methodological contribution is the integration of relational information with conventional tabular features. The fused representation $z_i = [x_i \parallel h_i]$ allows XGBoost to process both borrower attributes and learned relational representations.

4.7 Practical Implications

- Automated credit-risk screening.
- Portfolio monitoring and risk stratification.
- Decision-support for lending systems.
- Financial analytics using graph-derived and tabular representations.

4.8 Limitations

- The reported experiments use the HCDR dataset; cross-dataset generalization requires further validation.
- KGE construction introduces additional computational complexity.
- Additional metrics such as PR-AUC, MCC, calibration error, and cost-sensitive measures would strengthen evaluation.
- The baseline comparison does not establish a fully matched experimental protocol.
- Case-level explainability is not extensively demonstrated in the source study.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

V. CONCLUSION

5.1 Conclusion

This study presented a hybrid machine-learning framework for loan-default prediction based on the integration of Knowledge Graph Embedding and XGBoost. The framework combines conventional borrower and loan characteristics with relational information encoded through knowledge-graph embeddings. The methodology includes exploratory data analysis, missing-value imputation, normalization, categorical encoding, filter-based feature selection, stratified data partitioning, knowledge-graph representation, feature fusion, and XGBoost classification.

The reported results on the HCDR dataset include 96.79% accuracy, 80.83% precision, 78.75% recall, 79.00% F1-score, and 83.60% ROC-AUC. The reported confusion matrix contains 55,661 true negatives, 3,876 true positives, 919 false positives, and 1,046 false negatives. The study also reports baseline accuracies of 94.20% for Random Forest, 89.00% for Neural Network, and 73.00% for Decision Tree. Overall, the results demonstrate the potential of combining graph-based relational representations with gradient-boosting models for credit-risk prediction.

5.2 Future Work

- Investigate lightweight Knowledge Graph Embedding methods to reduce computational overhead.
- Integrate SHAP, LIME, and counterfactual explanations for case-level model transparency.
- Validate the framework on LendingClub, peer-to-peer lending, banking, and other credit-risk datasets.
- Evaluate SMOTE, ADASYN, focal loss, cost-sensitive learning, and threshold optimization for class imbalance.
- Perform probability-calibration analysis using calibration curves, Brier score, and expected calibration error.
- Combine federated learning and differential privacy with graph-enhanced credit-risk prediction.
- Develop temporal and dynamic knowledge graphs to model changes in borrower behavior.
- Investigate cost-sensitive optimization based on the financial consequences of false positives and false negatives.
- Explore real-time deployment within digital lending and credit-risk management systems.

REFERENCES

- [1] A. Akinjole, O. Shobayo, J. Popoola, O. Okoyeigbo, and B. Ogunleye, "Ensemble-Based Machine Learning Algorithm for Loan Default Risk Prediction," *Mathematics*, vol. 12, p. 3423, 2024, doi: 10.3390/math12213423.
- [2] C. Franco, P. Madrazo-Lemarroy, J. Beltrán, J. A. Núñez Mora, and P. Moncayo, "Loan Default Prediction: A Complete Revision of LendingClub," *Revista Mexicana de Economía y Finanzas*, vol. 18, no. 3, 2023, doi: 10.21919/remef.v18i3.886.
- [3] B. Chaudhari, S. Verma, and S. Somu, "A Review of Secure API Gateways with Java Spring for Financial Lending Platforms," vol. 14, pp. 315–326, 2024, doi: 10.56975/ijcsp.v14i4.303090.
- [4] B. Chaudhari, S. C. G. Verma, and S. R. Somu, "Transforming Financial Lending: A Scalable Microservices Approach using AI and Spring Boot," *International Journal of Scientific Research in Modern Technology*, pp. 72–81, 2024, doi: 10.38124/ijrsmt.v3i8.527.
- [5] A. Egwa, B. Habeeb, A. Ajiya, and M. Suleiman Bizi, "Default Prediction for Loan Lenders Using Machine Learning Algorithms," pp. 1–12, 2022.
- [6] X. Zhu, Q. Chu, X. Song, P. Hu, and L. Peng, "Explainable Prediction of Loan Default Based on Machine Learning Models," *Data Science and Management*, 2023, doi: 10.1016/j.dsm.2023.04.003.
- [7] J. D. Turiel and T. Aste, "Peer-to-peer Loan Acceptance and Default Prediction with Artificial Intelligence," *Royal Society Open Science*, 2020, doi: 10.1098/rsos.191649.
- [8] V. Padimi, S. T. Venkata, and D. N. Devarani, "Applying Machine Learning Techniques to Maximize the Performance of Loan Default Prediction," *Journal of Neutrosophic and Fuzzy Systems*, vol. 2, no. 2, pp. 44–56, 2022.
- [9] V. Singh, "Predicting Loan Default Risk in P2P Lending Platforms: A Study of Lending Club Borrowers," *International Journal of Scientific Research*, vol. 12, no. 11, pp. 2255–2260, 2023.
- [10] J. Chen, "Research on Financial Loan Default Prediction Based on Multi-Model Ensemble and Custom Thresholds," *Transactions on Computer Science and Intelligent Systems Research*, vol. 7, pp. 666–674, 2024, doi: 10.62051/7dnjhn18.
- [11] B. Chaudhari, S. Chitraju, and G. Verma, "Synergizing Generative AI and Machine Learning for Financial Credit Risk Forecasting and Code Auditing," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 11, no. 2, pp. 2882–2893, 2025, doi: 10.32628/CSEIT25112761.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

- [12] Q. Zhu, W. Ding, M. Xiang, M. Hu, and N. Zhang, "Loan Default Prediction Based on Convolutional Neural Network and LightGBM," *International Journal of Data Warehousing and Mining*, 2022, doi: 10.4018/IJDWM.315823.
- [13] J. C. Alejandrino, J. P. Bolacoy, and J. V. B. Murcia, "Supervised and Unsupervised Data Mining Approaches in Loan Default Prediction," *International Journal of Electrical and Computer Engineering*, 2023, doi: 10.11591/ijece.v13i2.pp1837-1847.
- [14] V. Verma, "Security Compliance and Risk Management in AI-Driven Financial Transactions," *International Journal of Engineering Science and Mathematics*, vol. 12, no. 7, pp. 1–15, 2023.
- [15] W. Yin, B. Kirkulak-Uludag, D. Zhu, and Z. Zhou, "Stacking Ensemble Method for Personal Credit Risk Assessment in Peer-to-Peer Lending," *Applied Soft Computing*, 2023, doi: 10.1016/j.asoc.2023.110302.
- [16] Z. Kun, F. Weibing, and W. Jianlin, "Default Identification of P2P Lending Based on Stacking Ensemble Learning," *Proceedings of the 2020 2nd International Conference on Economic Management and Model Engineering*, 2020, doi: 10.1109/ICEMME51517.2020.00203.
- [17] W. Han, X. Gu, and L. Jian, "A Multi-Layer Multi-View Stacking Model for Credit Risk Assessment," *Intelligent Data Analysis*, 2023, doi: 10.3233/IDA-220403.
- [18] P. Ramachandra and S. Vaithyanathan, "Fed-DPSDG-WGAN: Differentially Private Synthetic Data Generation for Loan Default Prediction via Federated Wasserstein GAN," *IEEE Access*, vol. 13, pp. 52069–52084, 2025, doi: 10.1109/ACCESS.2025.3552487.
- [19] S. A. Khan, R. Salman, A.-H. M. Majid, M. Mythili, M. B. Alazzam, et al., "Enhancing Loan Default Prediction with Human-in-the-Loop and XGBoost Ensemble Learning," *2025 Fifth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies*, 2025, pp. 1–6.
- [20] L. Zhou, "Loan Defaults Prediction Based on Stacked Models Trained by Personalized Features," *Highlights in Business, Economics and Management*, vol. 40, pp. 422–428, 2024, doi: 10.54097/rd657111.
- [21] Y. Kumar Jain, P. K. Mannepalli, K. Kaur, A. Maheshwari, J. Singh, and M. Ranka, "Effective Machine Learning-Based Predictive Analytics for Loan Default Prediction in Banking Sector," *2024 International Conference on Communication, Control, and Intelligent Systems*, 2024, pp. 1–6, doi: 10.1109/CCIS63231.2024.10931843.
- [22] P. Pathak, A. Jain, M. Bansal, and P. S. Rana, "SentiNet: Empowering Robust Loan Default Prediction through Ensemble Modeling," *2023 IEEE International Conference on Computer Vision and Machine Intelligence*, 2023, pp. 1–6, doi: 10.1109/CVMI59935.2023.10464518.
- [23] B. Gao, X. Yang, and Z. Wang, "An Empirical Study of BP Neural Network and KNN for Bank Loan Default Prediction," *IEEE Xplore*, 2023, doi: 10.1109/ICEMI58056.2023.10471076.
- [24] A. Lakshmanarao, C. Gupta, C. S. Koppireddy, U. V. Ramesh, and D. R. Dev, "Loan Default Prediction Using Machine Learning Techniques and Deep Learning ANN Model," *2023 Annual International Conference on Emerging Research Areas: International Conference on Intelligent Systems*, 2023, pp. 1–5, doi: 10.1109/AICERA/ICIS59538.2023.10420221.
- [25] A. M. Abidemi, M. D. Ajegbile, Y. B. Ajegbile, J. Adedeji, and C. Dada, "A Deep Learning Prediction Model for Loan Default," *Global Scientific Journal*, vol. 11, no. 7, 2023.
- [26] M. A. Kheneifar and B. Amiri, "A Novel Hybrid Model for Loan Default Prediction in Maritime Finance Based on Topological Data Analysis and Machine Learning," *IEEE Access*, vol. 13, pp. 81474–81493, 2025, doi: 10.1109/ACCESS.2025.3566066.
- [27] J. Olusegun, "The Impact of Data Preprocessing and Attribute Scaling on Neural Network Accuracy for Loan Default Prediction," 2025.
- [28] B. John, "Feature Selection Techniques for Enhancing Neural Network Performance in Loan Default Prediction," 2025.
- [29] D. A. Agustina Pertiwi, K. Ahmad, T. L. Nikmah, Alamsyah, B. Prasetyo, and M. A. Muslim, "Combination of Stacking with Genetic Algorithm Feature Selection to Improve Default Prediction in P2P Lending," *2023 5th International Conference on Cybernetics and Intelligent Systems*, 2023, doi: 10.1109/ICORIS60118.2023.10352271.
- [30] I. R. Berrada, F. Barramou, and O. B. Alami, "Towards a Machine Learning-Based Model for Corporate Loan Default Prediction," vol. 15, no. 3, pp. 565–573, 2024.
- [31] E. S. Jayaram, "Machine Learning-Based Loan Default Prediction: Models, Insights, and Performance Evaluation in Peer-to-Peer Lending Platforms," *Educational Administration: Theory and Practice*, vol. 30, no. 5, pp. 12975–12989, 2024, doi: 10.53555/kuey.v30i5.5637.
- [32] M. Madaan, A. Kumar, C. Keshri, R. Jain, and P. Nagrath, "Loan Default Prediction Using Decision Trees and Random Forest: A Comparative Study," *IOP Conference Series: Materials Science and Engineering*, vol. 1022, no. 1, p. 012042, 2021, doi: 10.1088/1757-899X/1022/1/012042.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

- [33] Rathod, N.J., Chopra, M.K., Chaurasiya, P.K. *et al.* Optimization on the Turning Process Parameters of SS 304 Using Taguchi and TOPSIS. *Ann. Data. Sci.* **10**, 1405–1419 (2023). <https://doi.org/10.1007/s40745-021-00369-2>
- [34] Rathod, N.J., Chopra, M.K., Shelke, S.N. *et al.* Investigations on hard turning using SS304 sheet metal component grey based Taguchi and regression methodology. *Int J Interact Des Manuf* **18**, 2653–2664 (2024). <https://doi.org/10.1007/s12008-023-01244-5>
- [35] N.J. Rathod, M.K. Chopra, U.S. Vidhate, N.B. Gurule, U.V. Saindane, Investigation on the turning process parameters for tool life and production time using Taguchi analysis, *Materials Today: Proceedings*, Volume 47, Part 17, 2021, Pages 5830-5835, ISSN 2214-7853, <https://doi.org/10.1016/j.matpr.2021.04.199>.
- [36] Rathod, N.J., Chopra, M.K., Chaurasiya, P.K. *et al.* Design and optimization of process parameters for hard turning of AISI 304 stainless steel using Taguchi-GRA-PCA. *Int J Interact Des Manuf* **17**, 2403–2414 (2023). <https://doi.org/10.1007/s12008-022-01021-w>
- [37] Rathod, N.J., Chopra, M.K., Chaurasiya, P.K. *et al.* Optimization of Tool Life, Surface Roughness and Production Time in CNC Turning Process Using Taguchi Method and ANOVA. *Ann. Data. Sci.* **10**, 1179–1197 (2023). <https://doi.org/10.1007/s40745-022-00423-7>
- [38] T. Dhande, Mahesh, Sanjaykumar Tiwari, and Nikhil J. Rathod. 2025. "Design of an Efficient Malware Prediction Model Using Auto Encoded & Attention-Based Recurrent Graph Relationship Analysis". *International Research Journal of Multidisciplinary Technovation* 7 (1):71-87. <https://doi.org/10.54392/irjmt2515>.
- [39] Kalangi, C., Rathod, N.J., Madhuri, K.S. *et al.* Performance optimization of ethanol blends in diesel model using Taguchi and grey relational approach. *Sci Rep* **15**, 36048 (2025). <https://doi.org/10.1038/s41598-025-20009-6>
- [40] N.J. Rathod, M.K. Chopra, U.S. Vidhate, U.V. Saindane, Multi objective optimization in turning operation of SS304 sheet metal component, *Materials Today: Proceedings*, Volume 47, Part 17, 2021, Pages 5806-5811, ISSN 2214-7853, <https://doi.org/10.1016/j.matpr.2021.04.143>.
- [41] P. K. Chaurasiya, N. J. Rathod, P. K. Jain, V. Pandey, Shashikant and K. Lala, "Material Selection for Optimal Design Using Multi-Criteria Decision Making," *2023 3rd International Conference on Advancement in Electronics & Communication Engineering (AECE)*, GHAZIABAD, India, 2023, pp. 206-210, doi: 10.1109/AECE59614.2023.10428303.
- [42] Rathod, N.J., Bonde, P. & Nehete, H.R. Parametric optimization of WEDM of SS 304 stainless steel for material removal rate and surface roughness using Taguchi and Response Surface Methodology. *Interactions* **246**, 60 (2025). <https://doi.org/10.1007/s10751-025-02273-0>
- [43] Mahesh T. Dhande, et al. (2023). HMCMA: Design of an Efficient Model with Hybrid Machine Learning in Cyber security for Enhanced Detection of Malicious Activities. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(11s), 721–734. <https://doi.org/10.17762/ijritcc.v11i11s.9729>
- [44] Janardan Rathod N, B. M P, Gitay M, N. Patil S, Patel MT. Optimization Of Multiple Objectives in The Machining Process of SS304 Sheet Metal Components. *J Neonatal Surg* [Internet]. 2025 Apr. 20 [cited 2026 Oct. 21];14(14S):801-9. Available from: <https://jneonatalurg.com/index.php/jns/article/view/4132>
- [45] Nikhil Janardan Rathod, Markala Karthik, Rajendra Sopan Narkhede, Prem Kumar Chaurasiya, Experimental data integration with fuzzy multi-criteria decision-making framework for matrix material selection of 2XXX series aluminium alloys, *Next Materials*, Volume 13, 2026, 103554, ISSN 2949-8228, <https://doi.org/10.1016/j.nxmate.2026.103554>
- [46] Nikhil J. Rathod, Rajendra S. Narkhede, Markala Karthik, Mayur J. Gitay, Automotive brake material selection using various MCDM methods and their comparisons, *Next Materials*, Volume 13, 2026, 102998, ISSN 2949-8228, <https://doi.org/10.1016/j.nxmate.2026.102998>.
- [47] Rathod, Nikhil Janardan; Gitay, Mayur Jayant; Karthik, Markala; Ghongade, Harshwardhan; Rajak, Upendra; and Narkhede, Rajendra Sopan (2026) "Optimization of Mechanical Properties of Copper Nanofiller Reinforced Banana Fiber Polyester Composites Using Grey–Taguchi Method," *Mansoura Engineering Journal*: Vol. 51 : Iss. 5 , Article 20. Available at: <https://doi.org/10.58491/2735-4202.3439>
- [48] Rathod, Nikhil Janardan; Karthik, Markala; Ghongade, Harshvardhan P.; Shaikh, Javed Sikandar; and Deshmukh, Rahul Soma (2026) "The Multiwalled Carbon Nanotube-Incorporated CFRP Laminates Interlaminar Fracture Toughness," *Mansoura Engineering Journal*: Vol. 51 : Iss. 5 , Article 18. Available at: <https://doi.org/10.58491/2735-4202.3451>
- [49] Rathod, N.J. *et al.* (2026). Wire Electrical Discharge Machining Process Parameter Optimization via the Taguchi Method. In: Al-Ramahi, N., Musleh Al-Sartawi, A.M.A., Kanan, M. (eds) *Artificial Intelligence in the Digital Era. Studies in Systems, Decision and Control*, vol 594. Springer, Cham. https://doi.org/10.1007/978-3-031-89771-9_6



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

- [50] Rathod, N.J. *et al.* (2026). Aluminum Machining Process Parameter Optimization in WEDM with the GRA Approach. In: Al-Ramahi, N., Musleh Al-Sartawi, A.M.A., Kanan, M. (eds) Artificial Intelligence in the Digital Era. Studies in Systems, Decision and Control, vol 594. Springer, Cham. https://doi.org/10.1007/978-3-031-89771-9_16
- [51] Rathod, N.J. *et al.* (2026). Optimizing Wire Electric Discharge Machining Process Parameters for AISI 304 Stainless Steel via Taguchi Design of Experiments. In: Al-Ramahi, N., Musleh Al-Sartawi, A.M.A., Kanan, M. (eds) Artificial Intelligence in the Digital Era. Studies in Systems, Decision and Control, vol 594. Springer, Cham. https://doi.org/10.1007/978-3-031-89771-9_8



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com