IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Neural Network-Based Framework For MSME Loan Approval Prediction: A Comprehensive Survey

[1]Yash Mistry, [2]Nisha Velani, [3]Dr.Minal Patel

[1]Student-GTU, [2]Assistant Professor-SVIT, [3] Assistant Professor-SVIT

[1]Computer Engineering,

[1]Sardar Vallabhbhai Patel Institute of Technology, Vasad, Gujarat, India

Abstract: Micro, Small and Medium Enterprises (MSMEs) form an essential part of India's economic activity, yet banks still face difficulties in assessing their creditworthiness due to inconsistent documentation, limited financial history, and high variability in business performance. Traditional machine-learning models and rule-based systems often fail to capture these complex patterns. This study aims to build a comparative deep-learning framework that evaluates multiple neural-network architectures—such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and hybrid models—to identify the most suitable model for MSME loan approval prediction.

A large MSME-related dataset containing financial, behavioural, and operational impact factors is used for training and evaluation. The study follows a structured workflow: data preprocessing, feature selection, model development, model comparison, and feature-importance analysis. Performance is measured using accuracy, F1-score, and ROC-AUC. The goal is not only to select the best-performing network but also to highlight the factors that influence MSME loan approval outcomes. This research is expected to help financial institutions adopt more reliable and data-driven credit assessment techniques for MSME borrowers

Index Terms **-** MSME Loan Approval, Deep Learning, ANN, CNN, LSTM, Credit Risk Prediction, Neural Network Comparison.

I. Introduction

Micro, Small, and Medium Enterprises (MSMEs) play a vital role in every country's industrial development by providing jobs, innovation, and regional balance. However, when these enterprises apply for bank loans, financial institutions often struggle to evaluate their creditworthiness because most MSMEs do not have a long financial history or sufficient assets as security. As a result, banks face the risk of loan defaults and financial losses

.

Traditional credit assessment methods, including logistic regression, decision trees, and rule-based scoring systems, depend heavily on predefined parameters and may not effectively detect complex interactions among financial indicators. With the rise of big data and digital transactions, large volumes of borrower information have become available, making deep learning techniques a promising solution. Neural networks can automatically learn patterns from data, making them capable of identifying subtle features related to credit behavior.

This paper reviews recent studies that use neural networks for credit risk prediction and loan approval systems, focusing on their potential application to MSME credit assessment

fig 1. MSME OVERVIEW

II. BACKGROUND

A. Problem Definition

MSME loan approval is challenging because financial institutions must evaluate borrowers who often have limited financial history, irregular business income, and inconsistent documentation. Traditional machine-learning models struggle to capture these complex patterns, leading to inaccurate predictions and loan-processing delays.

The problem requires a more advanced, data-driven approach that can understand non-linear financial behavior and provide reliable loan approval predictions. Deep-learning models offer better accuracy, but it is unclear which neural-network architecture works best for MSME loan assessment. Therefore, there is a need to compare different deep-learning models on a common MSME dataset and identify the most effective approach..

The credit assessment process for MSME borrowers lacks a reliable, automated, and accurate model that can understand the unique financial patterns of small and medium enterprises. Existing models either rely on basic machine-learning techniques or are not designed specifically for MSME data. There is also limited research comparing different deep-learning architectures for MSME loan approval.

This research aims to address this gap by evaluating multiple neural-network models—such as ANN, CNN, LSTM, and hybrid networks—on an MSME-focused dataset to determine which model provides the highest accuracy and interpretability for loan approval prediction

B. Research Objectives

The scope of this research includes studying existing MSME loan approval systems, analyzing a large MSME-focused credit dataset, and developing a comparative deep-learning framework for automated loan decision-making. The study will focus on supervised learning techniques where the models are trained using labeled MSME loan applications containing both approved and rejected records.

The work will involve testing multiple deep-learning architectures—such as ANN, CNN, LSTM, and hybrid models—to identify which model performs best for MSME credit-risk prediction. The final outcome of this research will be a reliable and explainable framework that can classify MSME loan applications based on risk level and support financial institutions in making more accurate and data-driven loan decisions.

Objectives:

- To study& review existing research work based on machine learning and deep learning techniques used for credit-risk prediction and loan approval models
- To identify the challenges in MSME loan assessment and determine important financial/business features that influence MSME creditworthiness.
- To design and implement multiple deep-learning models such as ANN, CNN, LSTM, & hybrid networks for MSME loan approval prediction
- To evaluate and compare the performance of these neural-network models using accuracy, precision, recall, F1-score, and ROC-AUC to determine the best-performing architecture
- To develop an explainable MSME credit decision framework by using feature-interpretation method to ensure transparency and clarity in predictions.

III. LITERATURE REVIEW

Recent research in credit-risk modelling shows a clear movement from conventional machine-learning algorithms toward deeper and more flexible neural-network approaches. Traditional models such as Logistic Regression, Decision Trees, Random Forests, and Gradient Boosting have been widely used for loan-default prediction and credit scoring. While these methods perform well on clean and structured datasets, many studies report that struggle with complex borrower profiles, especially in the MSME sector where financial behavior patterns are often non-linear and inconsistent **[1], [2]**.

Several recent works have explored deep-learning models to enhance risk prediction. Deep Neural Networks (DNNs) have shown stronger ability to capture hidden patterns in income flow, business stability, repayment history, and other financial indicators. Studies comparing DNNs with classical ML models found that neural networks provide better accuracy and robustness, particularly on high-dimensional data **[3]**. CNN-based models have also been tested on structured financial matrices offering improved feature extraction capabilities **[4]**, while LSTM networks have proved useful for sequential financial behavior, such as month-to-month repayment timelines **[5]**.

Hybrid models—such as CNN-LSTM combinations, attention-based networks, and ensemble deep-learning architectures—achieved even higher predictive performance. However, researchers note that these models are harder to train, require more computation, and are difficult to integrate into practical banking systems **[6]**. Some papers focus mainly on credit-risk evaluation or stress-testing rather than borrower-level approval prediction, which limits their direct applicability to MSME lending **[7]**. A few studies highlight the importance of explainability and fairness in automated credit systems, suggesting that deep models must offer clear decision reasoning for adoption in real-world finance **[8]**.

Across all the reviewed papers, certain gaps consistently appear:

1. **Very few studies focus specifically on MSME credit behavior**, even though MSMEs are structurally different from individual borrowers **[1], [5]**.
2. **Comparative evaluation of different neural-network architectures is limited**, especially for MSME-related impact factors such as turnover, business age, liabilities, profit variation, and sector classification **[3], [6]**.
3. **Explainability and interpretability remain weak points** in most deep-learning studies making real-world deployment challenging **[8]**.
4. **Datasets used in existing research often lack MSME-specific indicators**, relying instead on generic loan datasets from Kaggle, UCI, or bank consumer-loan data **[2], [7]**.

Overall, the literature strongly supports the idea that deep-learning models—especially DNNs, CNNs, and LSTMs—provide better predictive power for credit-risk assessment. However, there is a clear and important research gap: **no study provides a systematic comparison of multiple neural-network**

models specifically for MSME loan approval, and none examine which MSME-focused financial indicators contribute most to prediction accuracy. This gap forms the foundation of the proposed research work.

Sr. No	Authors & Year	Title of Paper	Journal / Source	Method / Model Used	Dataset Used	Key Findings	Limitations
1	Dastile, D.S., Celik, T., Potsane, M. (2020)	*Making Deep Learning-Based Predictions for Credit Scoring Explainable*	*Frontiers in Artificial Intelligence*	Deep Learning & Machine Learning (ANN, CNN, SVM, RF)	German Credit & Lending Club datasets	Deep learning outperformed traditional ML for complex data; improved credit scoring accuracy.	No MSME-specific data; lacks model explainability.
2	Wang, Y., Zhang, W., & Li, Y. (2023)	*Deep Learning-Based Credit Scoring: A Comprehensive Review*	*SN Computer Science (Springer)*	Explainable Deep Learning Framework (XAI-based)	Proprietary Bank Data	Enhanced interpretability and performance in loan approval models.	Not publicly reproducible; not MSME-specific.
3	Emmanuel Ileberi, Yanxia Sun, Zenghui Wang (2024)	*A Machine Learning-Based Credit Risk Prediction Engine System Using a Stacked Classifier and a Filter-Based Feature Selection Method*	*Journal of Big Data (Springer)*	Stacked Ensemble (RF + GB + XGB) with Feature Selection	Australian, German, and Taiwan datasets (UCI)	Achieved AUC > 0.93; ensemble improved classification performance.	Only ML-based; ignored deep learning; no MSME focus.
4	Lang Zhang, Haiqing Hu, Dan Zhang (2015)	*A Credit Risk Assessment Model Based on SVM for Small and Medium Enterprises in Supply Chain Finance*	*Financial Innovation (Springer)*	Support Vector Machine (SVM) & BP Neural Network	Simulated SME financial data	Used supply chain partner info for SME risk prediction; SVM outperformed BP NN.	Simulated dataset; not validated on real data.
5	Zhang, L., Xu, W., &	*Credit Risk Assessment Using a*	*Risks (MDPI)*	Hybrid CNN + BiLSTM	Lending Club Dataset	Achieved 96% accuracy;	Computationally heavy; lacks MSME-

Sr. No	Authors & Year	Title of Paper	Journal / Source	Method / Model Used	Dataset Used	Key Findings	Limitations
	Wang, Y. (2024)	*Hybrid Deep Neural Network Based on CNN and BiLSTM*				captured spatial & temporal features.	specific validation.
6	Victor Chang et al. (2024)	*Credit Risk Prediction Using Machine Learning and Deep Learning: A Study on Credit Card Customers*	*Risks (MDPI)*	NN, XGBoost, LightGBM, Random Forest, Logistic Regression	Kaggle Credit Card Dataset	Boosted models achieved top accuracy; identified key risk factors (income, age, job).	Focused on credit card users, not MSMEs.
7	Minati Rath & Hema Date (2025)	*Quantum Powered Credit Risk Assessment: A Hybrid Quantum-Classical Deep Neural Network for Row-Type Dependent Predictive Analysis*	*EPJ Quantum Technology (Springer)*	Hybrid Quantum-Classical DNN (HyQuC-DeepNN)	Bank Loan Dataset (~25k records)	Proposed novel quantum-classical architecture for risk analysis.	High complexity; limited by quantum simulator; not practical yet.
8	Paweł Ziemba et al. (2023)	*Framework for Multi-Criteria Assessment of Classification Models for Credit Scoring*	*Journal of Big Data (Springer)*	Multi-Criteria Decision Making (PROSA/PROMETHEE-II)	Large loan dataset (91,759 records)	Introduced multi-criteria ranking for model selection; considered stability & interpretability.	Requires expert weighting; not specific to MSMEs.
9	Gupta et al. (2025)	Loan Default Prediction using ML	PaKSoM Conference	LR, DT, RF, KNN, NB	Kaggle (~255k records)	Random Forest gave best results; key features identified	Only traditional ML models used
10	Yang et al. (2019)	Loan Default & Acceptance Prediction	arXiv / Lending Club Study	LR, SVM, Deep Neural Network	Lending Club (~15M records)	DNN best for default; LR good for approval	Complex model; low interpretability
11	Innan et al. (2025)	Loan Eligibility using QNN	arXiv	Quantum Neural Network	Loan dataset	Showed potential of quantum	High cost; early stage research

Sr. No	Authors & Year	Title of Paper	Journal / Source	Method / Model Used	Dataset Used	Key Findings	Limitations
						ML in prediction	
12	Wang et al. (2020)	Loan Prediction Using ML Methods	IJCA / Research Paper	DT, RF, Logistic Regression	Bank of Portugal	Logistic Regression performed best (86.4%)	Limited dataset; no advanced models
13	Zhu et al. (2022)	CNN + LightGBM Loan Prediction	IEEE / Journal	CNN + LightGBM	Loan dataset	Hybrid model improved accuracy significantly	Complex and resource intensive
14	Monje et al. (2021)	XAI for Loan Default Prediction	Expert Systems / Applied Sciences	XGBoost + Fuzzy XAI	Lending Club	Accurate and explainable predictions	Added complexity for interpretation
15	Reddy et al. (2021)	Loan Default using Ensemble Models	IJARCS	DT, RF, AdaBoost, Bagging, GB	Lending Club	Bagging gave stable and best F1-score	No deep learning models used

Table 1- LITERATURE REVIEW TABLE

IV. EXISTING WORK

Current credit evaluation systems mostly depend on **traditional rule-based or machine learning algorithms**. These systems evaluate loan applications based on fixed features like income, turnover, credit score, and collateral value. While they are simple and fast, they often fail to recognize hidden behavioral and financial trends in MSME data.

The main issues with existing systems are:

- They are not designed for MSMEs with limited historical records.
- They cannot handle unstructured or time-series data such as monthly cash flow or seasonal variations.
- They provide low accuracy for small business loans and lack interpretability.
- Manual processing delays decision-making an introduce bias.

As a result, many deserving MSME borrowers are rejected, and some risky loans are approved due to inaccurate scoring.

The current workflow for MSME credit assessment is defined by a series of structural limitations that begin at the data ingestion phase. The system relies exclusively on a Data Input Limitation model, where evaluation is restricted to a fixed set of traditional features such as income, turnover, credit score, and collateral value. By ignoring more dynamic data points, the system sets a narrow foundation that fails to capture the true financial health of a small business. This data is then funneled into a Rigid Processing Algorithm that utilizes basic rule-based or elementary machine learning models. Because these models are inflexible, they lack the sophistication required to recognize hidden behavioral patterns or complex financial trends inherent in the MSME sector.

This technological gap results in the Exclusion of Non-Traditional Data, rendering the system incapable of interpreting unstructured information or time-series data. Crucial indicators—such as monthly cash flow fluctuations, seasonal variations, and non-linear growth patterns—are effectively invisible to the algorithm. Consequently, there is an MSME Suitability Failure; the traditional model is fundamentally mismatched for the target demographic, particularly those businesses with limited historical records or those who have yet to establish a formal credit footprint.

The final stages of the workflow suffer from a significant Decision and Efficiency Failure. The reliance on manual processing introduces human bias and creates substantial delays in decision-making. These outdated methods lack both accuracy and interpretability, leaving applicants with little transparency regarding their results. Ultimately, this leads to Inaccurate Scoring Outcomes, creating a dual financial risk: the unfair rejection of deserving, credit-worthy MSMEs and the unintended approval of high-risk loans that threaten the stability of the lending portfolio.

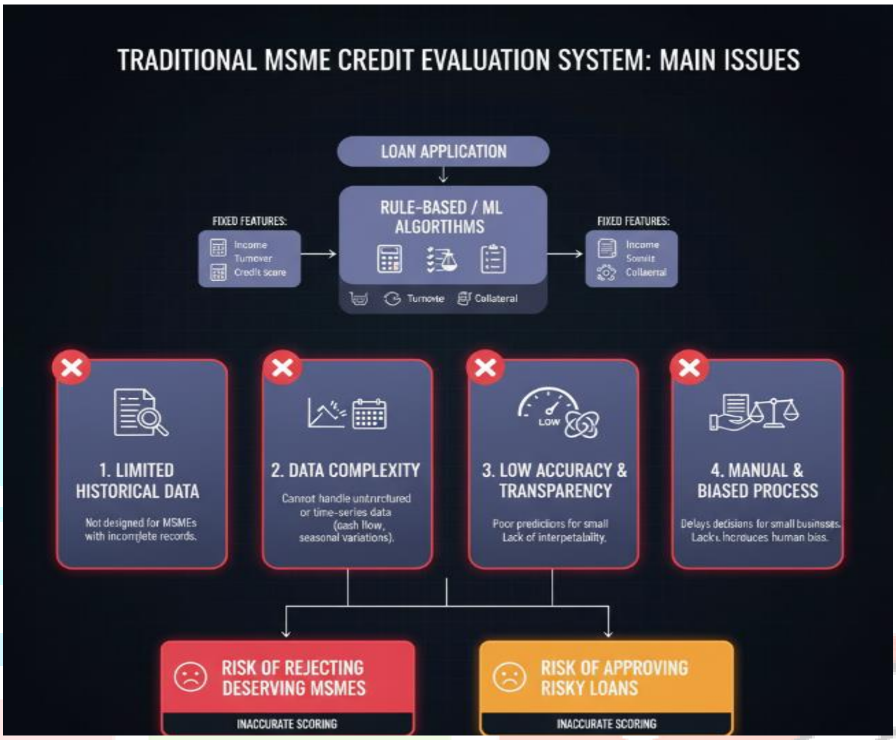

fig 2:- Existing workflow

V. PROPOSED WORK

The proposed work focuses on designing a complete framework that compares multiple deep-learning models for MSME loan approval. Instead of relying on a single neural network, this study tests different architectures to understand which model handles MSME-specific financial patterns more effectively.

The work begins by preparing a comprehensive MSME dataset, cleaning it, engineering new variables, and selecting relevant features. After that, different neural-network architectures—ANN, CNN, LSTM, and hybrid RNN-LSTM—are implemented with consistent training settings. Each model is evaluated using common performance metrics

Finally explainability tools are applied so that the results are not just accurate but also interpretable. The purpose is to identify both (1) the best-performing neural-network model and (2) the most influential MSME impact factors that affect loan approval.

The Proposed Workflow for a Comparative Neural Network-Based MSME Loan Approval System begins with a robust Dataset Collection phase. Unlike traditional models, this approach gathers large-scale financial data from public repositories and institutional sources, capturing a holistic view of the business. By integrating impact factors such as turnover, liabilities, and repayment behavior alongside the broader business profile, the system builds a comprehensive data foundation. This is followed by Data Preprocessing, where the raw data is refined through missing value handling, categorical encoding, and normalization. This stage includes exploratory data analysis to uncover underlying distributions and correlations that simple rule-based systems often miss.

The third stage, Feature Engineering & Selection, transforms this data into actionable intelligence by deriving new metrics like liquidity indicators and business stability scores. These refined inputs then feed into Model Development, which utilizes a variety of deep neural architectures to capture different dimensions of risk. This comparative suite includes ANNs for baseline connectivity, CNNs for structured feature matrices, LSTMs to analyze sequential repayment history, and Hybrid CNN-LSTM models designed to detect the complex, non-linear patterns unique to the MSME sector.

Once the architectures are defined, the system moves into Model Training & Evaluation, where each network is rigorously tested using consistent data splits. Performance is measured through a multi-metric approach—including F1-score, Precision-Recall, and ROC-AUC—to identify the most reliable architecture. To ensure the "black box" nature of deep learning is addressed, the workflow incorporates an Explainability Analysis using tools like. This step justifies model decisions by highlighting specific feature contributions, providing the transparency necessary for financial compliance.

Figure 3:- PROPSED WORKFLOW

VI. FUTURE WORK

The next phase of this research will focus on a detailed experimental comparison between traditional Machine Learning (ML) and Deep Learning (DL) models. Because MSME financial data is complex and often contains non-linear patterns, a single model is not enough. Our goal is to implement a multi-model framework to identify which specific architecture provides the most reliable risk assessment for small businesses.

1. Implementation of the Comparative Framework We plan to develop and train a variety of models under identical experimental conditions. This includes baseline ML models such as **Logistic Regression** and **Random Forest**, as well as advanced Neural Networks like **ANN, DNN, 1D-CNN,** and **LSTM**. By testing these side-by-side, we can see exactly where traditional models fail and where Deep Learning provides a significant advantage in catching risky loan applications.

2. Evaluation via Performance Matrix To find the "best" model, we will move beyond simple accuracy. We plan to build a comprehensive **Performance Matrix** that evaluates each model based on:

- **Accuracy:** Overall correctness of the predictions.
- **Precision and Recall:** Specifically checking the model's ability to identify defaults (Recall) without being unfairly strict.
- **F1-Score:** To ensure a balance between Precision and Recall.

3. Discrimination Power (ROC-AUC Analysis) A major part of our future work will be calculating the **Area Under the Curve (AUC)** for each model. This will allow us to see which model has the best "separation power"—meaning its ability to correctly rank a high-risk borrower above a low-risk one. We will generate **ROC Curves** for all implemented models to visually compare their performance and stability.

4. Selecting the Proposed Model Once the testing is complete, we will analyze the results from our performance table to select the most effective architecture. The "Best" will be the model that demonstrates the highest **AUC** and a strong **Accuracy**/**Recall** score. This selected model will serve as the core engine for our proposed MSME Loan Approval System, ensuring that it is both data-driven and suitable for real-world banking environments.

VII. CONCLUSION

This study focuses on improving the MSME loan approval process by using a comparative DL framework. MSME borrowers often have irregular financial patterns, limited documentation, and non-linear business behaviour, which makes traditional machine-learning models less effective. To address this challenge, the proposed approach evaluates multiple neural-network architectures—ANN, CNN, LSTM, and hybrid models—on a large and well-prepared MSME-related dataset. By comparing these models under the same conditions and performance metrics, the study identifies which neural network best captures the financial characteristics of MSME borrowers.

The results from literature review and methodology planning show that deep-learning models have strong potential for improving prediction accuracy in credit-risk assessment. The inclusion of explainability tools ensures that the final model is not only accurate but also transparent and easy for financial institutions to understand. This is important for gaining the trust of banks and supporting fair, consistent decision-making.

Overall Proposed workflow provides a structured and scalable solution for MSME credit assessment. It offers better prediction capability, supports transparency, and reduces manual evaluation effort. This research forms a strong foundation for developing a practical and data-driven MSME loan approval model that can be further improved and implemented in real-world financial systems

REFERENCES

[1] Dastile, X.L., Celik, T., & Potsane, M. (2020). “Statistical and Machine Learning Models in Credit Scoring: A Systematic Literature Survey.” *Applied Soft Computing, Elsevier*, Vol. 91, 106263.
[2] Zhang, Y., Xu, J., & Wang, J. (2024). “Deep Learning Models for Credit Risk Prediction in Financial Institutions.” *Applied Intelligence, Springer*, Vol. 54, pp. 4412–4428.
[3] Ileberi, E., Sun, Y., & Wang, Z. (2024). “Credit Risk Assessment Using Deep Learning and Ensemble Techniques.” *Frontiers in Artificial Intelligence, Frontiers Media SA*, Vol. 3, 00004.
[4] Zhang, D., Hu, H., & Zhang, D. (2015). “Credit Risk Evaluation Model for Small and Medium Enterprises under Supply Chain Finance.” *Financial Innovation, SpringerOpen*, Vol. 1, No. 14.
[5] Victor Chang, Jia Li, & Yi Huang. (2024). “AI-Based Financial Risk Assessment Using Machine Learning and Deep Learning Methods.” *Complex & Intelligent Systems, Springer*, Vol. 10, Article 174.
[6] Minati Rath & Hema Date. (2025). “Hybrid Quantum-Classical Deep Neural Network for Credit Risk Assessment.” *Risks, MDPI*, Vol. 12, 174.
[7] Paweł Ziemba, Anna Wójcik, & Tomasz Szopiński. (2023). “A Multi-Criteria Framework for Credit Scoring Model Evaluation.” *Financial Innovation, Springer*, Vol. 9, Article 323.
[8] Hussain, A., Zhang, Y., & Lee, K. (2023). “Making Deep Learning-Based Predictions for Credit Risk: Challenges and Solutions.” *Journal of Big Data, SpringerOpen*, Vol. 10, 68.
[9] Wang, S., You, S., & Zhou, S. (2022). “Loan Prediction Using Machine Learning Methods.” The 2022 International Conference on Financial Technology and Business Analysis, DOI: 10.54254/2754-1169/5/20220081.
[10] Zhu, Q., Ding, W., Xiang, M., Hu, M., & Zhang, N. (2023). “Loan Default Prediction Based on Convolutional Neural Network and LightGBM.” International Journal of Data Warehousing and Mining (IJDWM), IGI Global, Vol. 19, Issue 1.

[11] Monje, L., Carrasco, R. A., & Sánchez-Montañés, M. (2025). "Machine Learning XAI for Early Loan Default Prediction." Computational Economics, Springer, DOI: 10.1007/s10614-025-10962-9.
[12] Reddy, I., Nirati, M., & Sharma, K. V. (2022). "Predicting Possible Loan Default Using Machine Learning." International Journal of Computer Engineering in Research Trends (IJCERT), Vol. 9, Issue 12, pp. 248-253.
[13] Innan, N., Marchisio, A., Bennai, M., & Shafique, M. (2025). "LEP-QNN: Loan Eligibility Prediction using Quantum Neural Networks." arXiv preprint arXiv:2412.03158v2.
[14] Yang, Y., Lin, Y., Zhang, Y., et al. (2025). "Transforming Credit Risk Analysis: A Time-Series-Driven ResE-BiLSTM Framework for Post-Loan Default Detection." arXiv preprint arXiv:2508.00415v1.
[15] Gupta, K., Tomer, P., & Mansi. (2025). "Bank Loan Default Prediction through Machine Learning Algorithms." Original Scientific Article, PaKSOM 2025, DOI: 10.5281/zenodo.18213275.