ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

MACHINE LEARNING-BASED PREDICTIVE ANALYTICS FOR CREDIT RISK ASSESSMENT: A COMPREHENSIVE REVIEW

[1*]Muhammad Afnan Abid,
Department of computer science superior university Lahore.
Email: afnanabid157@gmail.com
[2]Majid Hussain Bukhari
Department of Information technology Islamic University of Pakistan.
Email: bukharimajidhussain@gmail.com
[3]Sana Afzal
Department of Business and Economics National College of Business Administration and Economics Lahore.
Email: sanaafzal_12@hotmail.com
*Corresponding Author: Muhammad Afnan Abid Email: afnanabid157@gmail.com

Abstract
Credit risk assessment is one of the most consequential analytical tasks in modern finance, underpinning lending decisions, regulatory capital calculations, and the overall stability of banking systems. For decades, financial institutions relied on statistical scoring methods such as logistic regression and discriminant analysis, but the growing volume, velocity, and variety of financial data have accelerated a shift toward machine learning (ML) and deep learning (DL) approaches. This paper presents a comprehensive review of the literature on ML-based predictive analytics for credit risk assessment, synthesizing evidence from twenty-three peer-reviewed studies, review articles, and book chapters published between 2019 and 2025. The review traces the evolution of credit risk modeling from classical statistical techniques to ensemble learning, deep neural architectures, and hybrid frameworks; examines their application across retail credit scoring, corporate financial distress prediction, credit card fraud detection, rural and microfinance lending, banking supervision, and sector-specific contexts such as shipping finance; and consolidates commonly used datasets and evaluation metrics. The synthesis indicates that ensemble methods (particularly gradient-boosted trees such as XGBoost) and deep learning architectures generally outperform traditional statistical baselines, especially for nonlinear and high-dimensional data, while hybrid and ensemble combinations tend to yield the most consistent gains across benchmark datasets. Persistent challenges include class imbalance, limited model interpretability, data privacy constraints, regulatory compliance, and inadequate data infrastructure in emerging markets. The paper concludes by outlining research directions centered on explainable AI, federated and privacy-preserving learning, and the development of standardized, representative benchmark datasets for credit risk research.

Keywords: *Credit risk assessment; Machine learning; Deep learning; Credit scoring; Predictive analytics; Financial risk management; Ensemble learning.*

1. Introduction

Credit risk, broadly defined as the possibility that a borrower will fail to meet contractual debt obligations, sits at the core of banking and financial intermediation. Its accurate assessment governs decisions ranging from individual loan approvals to portfolio-level capital allocation and macroprudential regulation. Mispriced or mismanaged credit risk was a central contributor to the 2007-2009 global financial crisis, and it remains one of the most heavily scrutinized areas of financial risk management by regulators, rating agencies, and institutional investors alike (Aziz & Dowling, 2019; Leo, Sharma, & Maddulety, 2019).

Traditional credit risk models, including logistic regression, linear discriminant analysis, and expert-based scorecards, have been the industry standard for much of the last half-century because of their transparency, regulatory acceptance, and relative simplicity. However, these linear and parametric methods often struggle to capture the nonlinear interactions and high-dimensional patterns that characterize modern financial data, which increasingly includes transactional records, behavioral indicators, unstructured text, and

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

alternative data sources (Bhattacharya, Biswas, & Mandal, 2023; Shi, Tse, Luo, D'Addona, & Pau, 2022).

The proliferation of machine learning and deep learning techniques over the past decade has opened new possibilities for credit risk analytics. Algorithms such as random forests, gradient boosting machines, support vector machines, and artificial neural networks have demonstrated the capacity to model complex, nonlinear relationships between borrower characteristics and default outcomes, frequently outperforming classical statistical benchmarks in predictive accuracy (Munkhdalai, Munkhdalai, Namsrai, Lee, & Ryu, 2019; Suhadolnik, Ueyama, & Da Silva, 2023). At the same time, the adoption of these techniques introduces new concerns around interpretability, fairness, data privacy, and regulatory compliance that were less salient under simpler statistical frameworks (Guerra & Castelli, 2021; Mashrur, Luo, Zaidi, & Robles-Kelly, 2020).

This paper synthesizes findings from twenty-three studies spanning review articles, empirical papers, and a book chapter published between 2019 and 2025 to provide a structured, comprehensive account of how machine learning has been applied to credit risk assessment. The review is organized to address four guiding questions: (1) how has credit risk modeling evolved from statistical to machine learning-based paradigms; (2) what categories of machine learning techniques have been applied, and with what comparative performance; (3) across which application domains and institutional contexts has this research been concentrated; and (4) what challenges and future research directions does the literature identify. By consolidating this evidence, the review aims to serve both as an orientation for researchers entering the field and as a practical reference for practitioners evaluating machine learning adoption in credit risk functions.

2. Review Methodology

This review follows a narrative synthesis approach informed by systematic review principles, consistent with the methodology adopted in several of the reviewed sources themselves (Kowsar, 2022; Noriega, Rivera, & Herrera, 2023; Shi et al., 2022). The corpus consists of twenty-three sources: peer-reviewed journal articles (in outlets such as Risks, Sustainability, Mathematics, Data, Journal of Risk and Financial Management, Neural Computing and Applications, and Multimedia Tools and Applications), one book chapter (Aziz & Dowling, 2019), and conference proceedings, published between 2019 and 2025. Several of the sources are themselves systematic or narrative literature reviews that collectively synthesize several hundred underlying primary studies (for example, Guerra and Castelli (2021) reviewed 41 papers and two book chapters; Shi et al. (2022) reviewed 76 papers; Kowsar (2022) reviewed 98 peer-reviewed studies), which substantially broadens the effective evidence base underlying this synthesis.

Sources were grouped thematically according to (a) the machine learning technique(s) examined, (b) the application domain or institutional context, and (c) the datasets and evaluation metrics employed. Findings were then synthesized narratively across these groupings to identify convergent conclusions, points of disagreement, and gaps in the literature. Two documents that appeared in the source material were excluded from the evidentiary base: one addressed an unrelated biomedical imaging application, and the other had been formally retracted by its publisher following an investigation into data integrity, and is therefore not treated as reliable evidence in this review.

3. Evolution of Credit Risk Assessment: From Statistical to Machine Learning Models

Credit scoring emerged in the mid-twentieth century as a means of standardizing lending decisions, with discriminant analysis and, later, logistic regression becoming the dominant statistical tools due to their interpretability and ease of validation under regulatory frameworks such as Basel II and III (Bhattacharya et al., 2023; Rudra Kumar & Gunjan, 2020).

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

These approaches model default probability as a linear or log-linear function of borrower characteristics, offering coefficients that are directly interpretable by risk officers and auditors.

As computing power and data availability increased, researchers began exploring nonparametric and machine learning alternatives. Rudra Kumar and Gunjan (2020) trace this progression from statistical scorecards to decision trees, support vector machines, and neural network-based models, noting that each successive generation of techniques has generally improved predictive discrimination at some cost to model transparency. Bhattacharya et al. (2023) similarly compare sixteen statistical, optimization-based (heuristic), and machine learning approaches, concluding that machine learning methods handle nonlinear default patterns more effectively than classical statistical techniques, while hybrid and ensemble algorithms outperform standalone classifiers in the majority of cases they examined across forty-six datasets.

More recently, deep learning architectures, including deep feedforward networks, convolutional neural networks, and long short-term memory (LSTM) networks, have been applied to capture temporal and unstructured patterns in financial data (Khunger et al., 2025; Mosavi et al., 2020; Nosratabadi et al., 2020). This evolution reflects a broader trend across finance in which data-driven, algorithmic approaches are increasingly complementing or replacing rule-based and parametric methods, a trend also documented in adjacent domains such as banking supervision (Guerra & Castelli, 2021) and broader financial risk management (Aziz & Dowling, 2019; Mashrur et al., 2020).

4. Machine Learning Techniques for Credit Risk Assessment

The reviewed literature applies a wide spectrum of machine learning techniques to credit risk problems. This section organizes these techniques into four broad families: classical/traditional machine learning models, ensemble and hybrid models, deep learning architectures, and emerging paradigms such as federated learning.

4.1 Classical Machine Learning Models

Classical machine learning models, including logistic regression, k-nearest neighbors (KNN), naive Bayes, decision trees, and support vector machines (SVM), remain widely used baselines throughout the literature. Munkhdalai et al. (2019) benchmark several such models against a human expert-based scoring system (FICO) using real consumer finance survey data, finding that carefully tuned machine learning models can match or exceed expert-based benchmarks once appropriate feature selection is applied. Murugan and Kala (2023) similarly evaluate cluster-based KNN and logistic regression models for large-scale loan default prediction, reporting that these relatively simple models remain competitive when paired with effective data preprocessing and clustering strategies.

Decision-tree-based models feature prominently across the corpus both as standalone classifiers and as the building blocks of ensemble methods, reflecting their intuitive rule-based structure and relatively strong baseline performance on structured tabular credit data (Suhadolnik et al., 2023; Rudra Kumar & Gunjan, 2020).

4.2 Ensemble and Hybrid Models

Ensemble techniques, which combine multiple base learners to improve predictive robustness, are consistently identified as top performers across the reviewed studies. Suhadolnik et al. (2023) assess ten machine learning algorithms on a dataset of over 2.5 million credit observations and find that ensemble methods, particularly XGBoost, outperform traditional algorithms such as logistic regression. Bhattacharya et al. (2023) reach a comparable conclusion across forty-six datasets, reporting that ensemble classifiers significantly outperform standalone models. Murugan and Kala (2023) also report strong results for cluster-based XGBoost relative to simpler baselines in large-scale financial risk prediction.

Hybrid models, which combine machine learning with optimization techniques, clustering, or rule-based systems, appear as a recurring theme, particularly in shipping and corporate financial distress prediction. Clintworth, Lyridis, and Boulougouris (2021) develop a holistic methodology integrating random forest-based feature evaluation, multiple imputation of missing accounting data, and generalized additive modeling to predict financial distress among shipping companies, demonstrating that combining predictor evaluation with robust missing-data handling materially improves both individual-firm distress prediction and portfolio-level value-at-risk estimation.

4.3 Deep Learning Architectures

Deep learning techniques, including artificial neural networks (ANN), convolutional neural networks (CNN), and recurrent architectures such as LSTM, have been applied where large volumes of data or temporal/sequential structure can be exploited. Sulaiman, Schetinin, and Sant (2022) review ANN-based approaches to credit card fraud detection and propose a hybrid ANN model operating within a federated learning framework, reporting improved detection accuracy while preserving customer data privacy. Khunger et al. (2025) combine correlation-based CNNs with LSTM networks for financial stress testing, reporting substantial reductions in training and testing loss and marked reductions in estimated credit, liquidity, market, and operational risk indicators relative to baseline models.

At a broader methodological level, Mosavi et al. (2020) review deep reinforcement learning (DRL) applications in economics, noting that DRL's capacity to handle high-dimensional, noisy, and nonlinear economic data allows it to outperform traditional algorithms in tasks involving sequential decision-making under risk and uncertainty, such as dynamic portfolio management, though its application specifically to credit risk assessment remains comparatively less developed than to trading and portfolio problems. Nosratabadi et al. (2020) extend this perspective across the broader field of data science in economics, concluding that hybrid deep learning models increasingly outperform single-algorithm approaches across financial application domains, including corporate banking.

4.4 Emerging Paradigms: Federated and Privacy-Preserving Learning

A distinct methodological thread in the literature addresses the tension between model performance and data privacy. Kawa, Punyani, Nayak, Karkera, and Jyotinagar (2019) propose a federated learning framework that allows multiple financial institutions to collaboratively train a shared credit risk model without directly sharing sensitive customer records, addressing both privacy concerns and the substantial communication costs associated with centralized data pooling. Sulaiman et al. (2022) similarly integrate federated learning into their proposed fraud detection architecture, positioning it as a practical route to achieving both higher detection accuracy and stronger data confidentiality guarantees. These contributions suggest that federated and privacy-preserving architectures represent a promising, though still emerging, direction for institutions seeking to pool data-driven insights across organizational boundaries without compromising regulatory or customer privacy requirements.

5. Application Domains

Beyond methodological classification, the reviewed literature can be organized according to the application domain and institutional context in which credit risk models are deployed.

5.1 Retail and Consumer Credit Scoring

The largest concentration of studies addresses retail and consumer credit scoring. Munkhdalai et al. (2019), Suhadolnik et al. (2023), and Rudra Kumar and Gunjan (2020) each examine machine learning approaches to individual borrower default prediction, generally confirming that ensemble and deep learning methods improve on statistical scorecards, though at some cost to interpretability. Jumaa, Saqib, and Attar (2023) apply deep learning to consumer

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

loan default prediction in the United Arab Emirates banking sector using survey-based data from 1,000 participants across the top eleven local and foreign banks, reporting a meaningful improvement in default prediction accuracy over conventional approaches.

5.2 Corporate Credit Risk and Financial Distress Prediction

A second major application area concerns corporate financial distress and bankruptcy prediction. Clintworth et al. (2021) provide a detailed case study of the international shipping industry, an economically important but comparatively understudied sector, using over 5,000 company year-end financial statements combined with macroeconomic and market indicators. Their holistic machine learning methodology, which addresses the noisy and incomplete nature of shipping company financial statements through robust imputation techniques, is presented as a practical early-warning system for both individual company and bank-portfolio-level distress monitoring.

5.3 Credit Card Fraud Detection

Credit card fraud detection, while conceptually adjacent to credit default risk, constitutes its own substantial stream of research. Sulaiman et al. (2022) provide a comparative review of machine learning techniques for fraud detection, encompassing artificial neural networks, random forests, and support vector machines, and highlight data confidentiality as a central concern alongside detection accuracy, motivating their proposed federated ANN-based hybrid solution.

5.4 Rural Finance, Microfinance, and Financial Inclusion

Several studies focus specifically on underserved borrower populations. Kumar, Sharma, and Mahdavi (2021) review machine learning-based digital credit scoring methods for rural finance, emphasizing that small farm holders, rural youth, and women farmers remain largely excluded from mainstream banking despite advances in digitalization, and identifying gaps between available AI-ML methods and their actual adoption by rural lending institutions. Noriega et al. (2023) conduct a systematic review focused specifically on the microfinance credit industry, identifying fifty-two relevant studies and highlighting recurring challenges such as model opacity, the need for explainable AI, feature selection, multicollinearity, and class imbalance in microfinance datasets. Paul and Jana (2023) similarly examine machine learning-based optimization for credit risk evaluation in the context of financial inclusion, concluding that ML's ability to convert soft, judgment-based borrower information into hard, quantifiable indicators can materially improve credit access for borrowers with limited or poor credit history, while cautioning that this conversion process must be handled carefully to avoid introducing new forms of exclusion or bias.

5.5 Banking Supervision, RegTech, and Broader Risk Management

At the institutional and regulatory level, Guerra and Castelli (2021) review machine learning applications specifically from a banking supervision perspective, synthesizing 41 papers and two book chapters and finding that credit risk assessment and stress testing are among the most frequently studied supervisory applications of machine learning. Leo, Sharma, and Maddulety (2019) provide a broader review of machine learning across banking risk management functions, including credit, market, operational, and liquidity risk, concluding that despite substantial academic interest, the depth of machine learning application across all banking risk domains remains uneven and does not yet match the industry's overall level of focus on both risk management and machine learning adoption. Aziz and Dowling (2019) similarly review machine learning and AI across credit risk, market risk, operational risk, and regulatory compliance ('RegTech'), concluding with cautious optimism about the field's trajectory while highlighting persistent limitations in data management practices, model transparency, and the availability of relevant technical skillsets within financial firms. Mashrur, Luo, Zaidi, and Robles-Kelly (2020) propose a comprehensive taxonomy connecting financial

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

risk management tasks to relevant machine learning methods, and highlight significant publications, major research challenges, and emerging trends across the broader financial risk management field, of which credit risk constitutes one of the largest sub-domains.

5.6 Regional and Emerging-Market Applications

A number of studies examine credit risk analytics in specific regional or emerging-market contexts, where institutional constraints and data infrastructure differ substantially from developed markets. Al-Baity (2023) reviews the broader adoption of AI in Saudi Arabia's digital finance sector, proposing a framework for AI implementation while noting that limited published research and nascent institutional readiness currently constrain the sector's ability to fully leverage AI-driven financial services. Kowsar (2022) conducts a systematic review, following PRISMA 2020 guidelines, of credit risk assessment models specifically within Bangladesh's commercial banking sector, drawing on 98 peer-reviewed studies and finding that while machine learning approaches show promise, institutional constraints, limited data infrastructure, and evolving regulatory frameworks continue to shape which models are practically deployable in that context. Bello (2023) similarly reviews the economic and financial implications of machine learning-based credit risk assessment, emphasizing that traditional credit scoring methods, while historically effective, increasingly fall short of capturing the complexity of modern financial markets.

6. Datasets and Evaluation Metrics

The reviewed literature draws on a wide range of datasets, from small, publicly available benchmark datasets to large proprietary institutional records. Bhattacharya et al. (2023) note that among the forty-six datasets used across their reviewed studies, the German, Australian, and Japanese credit datasets are by far the most frequently used benchmark datasets in the credit scoring literature, despite their relatively small size and dated provenance. In contrast, several more recent studies employ substantially larger, real-world datasets: Suhadolnik et al. (2023) use a dataset exceeding 2.5 million observations from a financial institution, while Clintworth et al. (2021) construct a longitudinal dataset of over 5,000 shipping company year-end financial statements combined with macroeconomic and market data.

In terms of evaluation, classification accuracy, the area under the receiver operating characteristic curve (AUC), and related discrimination metrics dominate the literature. Noriega et al. (2023) identify Area Under Curve (AUC) and Accuracy (ACC) as the most commonly reported evaluation metrics across the microfinance credit risk studies they reviewed. Several studies also report loss-based metrics for deep learning models (Khunger et al., 2025) or domain-specific financial indicators such as estimated liquidity, market, credit, and operational risk scores (Khunger et al., 2025), reflecting a broader trend toward evaluating models not only on classification performance but also on their downstream financial and business relevance.

7. Comparative Summary of Reviewed Studies

Table 1 summarizes the primary focus, core machine learning techniques, and principal findings of the twenty-three studies synthesized in this review.

Table 1. Summary of reviewed studies on machine learning-based credit risk assessment

Study (Author, Year)	Application Focus	Core ML/DL Techniques	Key Finding
Al-Baity (2023)	AI adoption in digital finance (Saudi Arabia)	AI/ML (general review)	Proposes an implementation framework; notes limited published research on regional AI adoption
Aziz & Dowling (2019)	Risk management overview	ML/AI across credit, market, operational risk, RegTech	Optimistic outlook tempered by data governance and skillset limitations

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Study (Author, Year)	Application Focus	Core ML/DL Techniques	Key Finding
Bello (2023)	Economic/financial implications of ML in credit risk	ML (general review)	Traditional methods increasingly insufficient for modern market complexity
Bhattacharya, Biswas, & Mandal (2023)	Comparative study of 16 approaches, 46 datasets	Statistical, heuristic, ML models	Ensemble/hybrid models outperform standalone classifiers
Clintworth, Lyridis, & Boulougouris (2021)	Corporate financial distress (shipping industry)	Random forest, GAM, XGBoost, multiple imputation	Holistic ML methodology improves distress prediction and portfolio VaR estimation
Guerra & Castelli (2021)	Banking supervision	ML (review of 41 papers, 2 book chapters)	Credit risk assessment and stress testing are leading supervisory ML applications
Jumaa, Saqib, & Attar (2023)	Consumer loan default (UAE banks)	Deep learning (Keras/TensorFlow)	Notable improvement in default prediction success rate
Kawa, Punyani, Nayak, Karkera, & Jyotinagar (2019)	Cross-institution credit risk assessment	Federated learning	Enables collaborative model training while preserving data privacy
Khunger et al. (2025)	Financial stress testing	CNN-LSTM (CCNN)	Reduced training/testing loss; lower estimated credit, market, liquidity, operational risk
Kowsar (2022)	Credit risk models in emerging economies (Bangladesh)	Statistical and ML models (98 studies)	Institutional and data-infrastructure constraints shape model adoption
Kumar, Sharma, & Mahdavi (2021)	Rural/digital credit scoring	AI-ML methods (systematic review)	Gaps remain between available ML methods and adoption by rural lenders
Leo, Sharma, & Maddulety (2019)	Banking risk management (credit, market, operational, liquidity)	ML (literature review)	ML application uneven across risk domains relative to industry focus
Mashrur, Luo, Zaidi, & Robles-Kelly (2020)	Financial risk management taxonomy	ML/DL (survey)	Proposes taxonomy linking FRM tasks to ML methods; identifies key challenges/trends
Mosavi et al. (2020)	Deep reinforcement learning in economics	Deep RL	DRL outperforms traditional algorithms in complex, high-dimensional economic tasks
Munkhdalai, Munkhdalai, Namsrai, Lee, & Ryu (2019)	Bank client credit assessment	Logistic regression, SVM, ensembles, deep neural networks	Tuned ML models can match/exceed expert-based (FICO) benchmark
Murugan & Kala (2023)	Large-scale financial risk management	Cluster-based KNN, logistic regression, XGBoost	Cluster-based ensemble methods effective for large-scale loan default prediction
Noriega, Rivera, & Herrera (2023)	Credit risk in microfinance	ML (52 studies reviewed)	Boosted models most researched; AUC and accuracy most common

Study (Author, Year)	Application Focus	Core ML/DL Techniques	Key Finding
			metrics; interpretability a key challenge
Nosratabadi et al. (2020)	Data science in economics	DL, hybrid DL, hybrid ML, ensembles	Hybrid deep learning models increasingly outperform single-algorithm approaches
Paul & Jana (2023)	Credit risk evaluation for financial inclusion	ML-based optimization	ML improves conversion of soft borrower information into reliable hard indicators
Rudra Kumar & Gunjan (2020)	Credit scoring model evolution	Statistical, ML, deep learning models	Traces progression from scorecards to deep learning-based scoring
Shi, Tse, Luo, D'Addona, & Pau (2022)	ML-driven credit risk (76 papers)	Statistical, ML, DL	Deep learning generally outperforms classic ML/statistical methods; ensembles improve accuracy
Suhadolnik, Ueyama, & Da Silva (2023)	Credit risk assessment (2.5M+ observations)	10 ML algorithms incl. XGBoost	Ensemble methods (XGBoost) outperform traditional algorithms such as logistic regression
Sulaiman, Schetinin, & Sant (2022)	Credit card fraud detection	ANN, RF, SVM, federated learning	Federated ANN hybrid improves accuracy while preserving data privacy

8. Challenges and Limitations

Despite consistent evidence of predictive gains from machine learning approaches, the reviewed literature converges on several recurring challenges. First, class imbalance is a pervasive issue in credit risk datasets, where defaulting borrowers typically represent a small minority of observations; this imbalance complicates model training and evaluation and is explicitly identified as a limitation by Noriega et al. (2023) and Shi et al. (2022).

Second, model interpretability remains a central concern, particularly given regulatory requirements for explainable lending decisions. The 'black box' nature of ensemble and deep learning models is highlighted as a barrier to adoption by Noriega et al. (2023), while Guerra and Castelli (2021) and Aziz and Dowling (2019) note that transparency requirements under banking supervision and regulatory frameworks constrain how readily complex models can be deployed in practice, motivating growing interest in explainable AI methods.

Third, data privacy and governance present ongoing challenges, particularly where credit risk modeling would benefit from pooling data across institutions. Kawa et al. (2019) and Sulaiman et al. (2022) frame federated learning as a direct response to this tension, though such architectures introduce their own implementation complexity and communication overhead.

Fourth, data infrastructure and quality vary substantially by institutional and regional context. Studies focused on rural finance (Kumar et al., 2021), microfinance (Noriega et al., 2023), and emerging economies (Kowsar, 2022) each report that limited, inconsistent, or incomplete data restricts the practical applicability of otherwise promising machine learning methods, echoing similar data-quality concerns raised in the corporate context by Clintworth et al. (2021) regarding incomplete shipping-industry financial statements.

Finally, several sources note a persistent gap between the pace of academic research and actual industry adoption. Leo et al. (2019) observe that machine learning application across banking risk domains does not yet match the industry's overall stated focus on both risk management and machine learning, while Al-Baity (2023) similarly finds that AI has not yet become a critical, fully embedded component of financial transaction processing in the market she examines, despite growing institutional and governmental interest.

9. Future Research Directions

Based on the gaps and challenges identified across the reviewed literature, several directions for future research emerge. Explainable AI (XAI) methods that preserve the predictive advantages of ensemble and deep learning models while providing interpretable, auditable outputs represent a priority, particularly for regulatory and supervisory applications (Guerra & Castelli, 2021; Noriega et al., 2023). Continued development of federated and privacy-preserving learning architectures could enable broader, cross-institutional data collaboration without compromising customer privacy or regulatory compliance (Kawa et al., 2019; Sulaiman et al., 2022).

There is also a clear need for larger, more representative, and more regularly updated benchmark datasets, given that much of the credit scoring literature still relies heavily on a small number of aging public datasets such as the German and Australian credit datasets (Bhattacharya et al., 2023). Expanding rigorous empirical research into underserved contexts, including rural finance, microfinance, and emerging-market banking sectors, would help close the gap between methodological advances demonstrated on data-rich developed-market datasets and the practical needs of institutions operating under greater data and infrastructure constraints (Kowsar, 2022; Kumar et al., 2021; Noriega et al., 2023). Finally, further sector-specific research, following the model set by shipping-industry financial distress studies (Clintworth et al., 2021), could help extend holistic, imputation-aware machine learning methodologies to other data-sparse but economically significant industries.

10. Conclusion

This review has synthesized evidence from twenty-three studies on the application of machine learning to credit risk assessment, spanning methodological families from classical statistical and machine learning models to ensemble methods, deep learning architectures, and emerging federated learning paradigms. Across application domains, including retail credit scoring, corporate financial distress prediction, credit card fraud detection, rural and microfinance lending, and banking supervision, the literature consistently reports that ensemble and deep learning methods offer meaningful predictive improvements over traditional statistical baselines, particularly for nonlinear and high-dimensional financial data. At the same time, the literature is equally consistent in identifying unresolved challenges around class imbalance, model interpretability, data privacy, and uneven data infrastructure, especially in emerging-market and underserved lending contexts. Addressing these challenges through explainable AI, privacy-preserving learning architectures, and the development of larger and more representative benchmark datasets represents the most promising path toward realizing the full potential of machine learning-based predictive analytics in credit risk assessment.

References

1. Al-Baity, H. H. (2023). The artificial intelligence revolution in digital finance in Saudi Arabia: A comprehensive review and proposed framework. Sustainability, 15(18), 13725. https://doi.org/10.3390/su151813725
2. Aziz, S., & Dowling, M. (2019). Machine learning and AI for risk management. In T. Lynn, J. G. Mooney, P. Rosati, & M. Cummins (Eds.), Disrupting finance: FinTech and strategy in the 21st century (pp. 33-50). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-02330-0_3
3. Bello, O. A. (2023). Machine learning algorithms for credit risk assessment: An economic and financial analysis. International Journal of Management Technology, 10(1), 109-133. https://doi.org/10.37745/ijmt.2013/vol10n1109133

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

4. Bhattacharya, A., Biswas, S. K., & Mandal, A. (2023). Credit risk evaluation: A comprehensive study. Multimedia Tools and Applications, 82, 18217-18267. https://doi.org/10.1007/s11042-022-13952-3
5. Clintworth, M., Lyridis, D., & Boulougouris, E. (2021). Financial risk assessment in shipping: A holistic machine learning based methodology [Working paper]. University of Strathclyde; National Technical University of Athens.
6. Guerra, P., & Castelli, M. (2021). Machine learning applied to banking supervision: A literature review. Risks, 9(7), 136. https://doi.org/10.3390/risks9070136
7. Jumaa, M., Saqib, M., & Attar, A. (2023). Improving credit risk assessment through deep learning-based consumer loan default prediction model. International Journal of Finance & Banking Studies, 12(1), 85-92. https://doi.org/10.20525/ijfbs.v12i1.2579
8. Kawa, D., Punyani, S., Nayak, P., Karkera, A., & Jyotinagar, V. (2019). Credit risk assessment from combined bank records using federated learning. International Research Journal of Engineering and Technology, 6(4), 1355-1358.
9. Khunger, A., Anand, K., Shukla, C., Jagdale, A. D., Chinnakannan, A., & Dbritto, C. (2025). Deep learning for financial stress testing: A data-driven approach to risk management. International Journal of Innovation Studies, 6(1), 79-90.
10. Kowsar, M. M. (2022). A systematic review of credit risk assessment models in emerging economies: A focus on Bangladesh's commercial banking sector. American Journal of Advanced Technology and Engineering Solutions, 2(1), 1-31. https://doi.org/10.63125/p7ym0327
11. Kumar, A., Sharma, S., & Mahdavi, M. (2021). Machine learning (ML) technologies for digital credit scoring in rural finance: A literature review. Risks, 9(11), 192. https://doi.org/10.3390/risks9110192
12. Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. Risks, 7(1), 29. https://doi.org/10.3390/risks7010029
13. Mashrur, A., Luo, W., Zaidi, N. A., & Robles-Kelly, A. (2020). Machine learning for financial risk management: A survey. IEEE Access, 8, 203203-203223. https://doi.org/10.1109/ACCESS.2020.3036322
14. Mosavi, A., Faghan, Y., Ghamisi, P., Duan, P., Ardabili, S. F., Salwana, E., & Band, S. S. (2020). Comprehensive review of deep reinforcement learning methods and applications in economics. Mathematics, 8(10), 1640. https://doi.org/10.3390/math8101640
15. Munkhdalai, L., Munkhdalai, T., Namsrai, O.-E., Lee, J. Y., & Ryu, K. H. (2019). An empirical comparison of machine-learning methods on bank client credit assessments. Sustainability, 11(3), 699. https://doi.org/10.3390/su11030699
16. Murugan, M. S., & Kala T, S. (2023). Large-scale data-driven financial risk management & analysis using machine learning strategies. Measurement: Sensors, 27, 100756. https://doi.org/10.1016/j.measen.2023.100756
17. Noriega, J. P., Rivera, L. A., & Herrera, J. A. (2023). Machine learning for credit risk prediction: A systematic literature review. Data, 8(11), 169. https://doi.org/10.3390/data8110169
18. Nosratabadi, S., Mosavi, A., Duan, P., Ghamisi, P., Filip, F., Band, S. S., Reuter, U., Gama, J., & Gandomi, A. H. (2020). Data science in economics: Comprehensive review of advanced machine learning and deep learning methods. Mathematics, 8(10), 1799. https://doi.org/10.3390/math8101799
19. Paul, R. K., & Jana, A. K. (2023). Credit risk evaluation for financial inclusion using ML based optimization. In Proceedings of the Foundations and Frontiers in Enterprise Analytics Workshop (FFEA WPI-23), Submission 12.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

20. Rudra Kumar, M., & Kumar Gunjan, V. (2020). Review of machine learning models for credit scoring analysis. Ingenieria Solidaria, 16(1). https://doi.org/10.16925/2357-6014.2020.01.11
21. Shi, S., Tse, R., Luo, W., D'Addona, S., & Pau, G. (2022). Machine learning-driven credit risk: A systemic review. Neural Computing and Applications, 34, 14327-14339. https://doi.org/10.1007/s00521-022-07472-2
22. Suhadolnik, N., Ueyama, J., & Da Silva, S. (2023). Machine learning for enhanced credit risk assessment: An empirical approach. Journal of Risk and Financial Management, 16(12), 496. https://doi.org/10.3390/jrfm16120496
23. Sulaiman, R. B., Schetinin, V., & Sant, P. (2022). Review of machine learning approach on credit card fraud detection. Human-Centric Intelligent Systems, 2, 55-68. https://doi.org/10.1007/s44230-022-00004-0