11 (1) 2025 54-65 ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
Register: *Jurnal Ilmiah Teknologi Sistem Informasi*

Contents lists available at www.journal.unipdu.ac.id

Register

Journal Page is available to https://journal.unipdu.ac.id/index.php/register/

Research article

Enhancing Bank Financial Performance Assessment: A Literature Review of Deep Learning Applications Using the Kitchenham Method

Mahrus Ali [a], *Rahmat Gernowo* [b], *Budi Warsito* [c] , *Faliha Muthmainah* [d]

[a] *Department of Doctoral Information System Diponegoro University, Jl. Prof. Soedarto No.13, Kec. Tembalang, Kota Semarang 50275, Indonesia.*
[b] *Department of Physics Diponegoro University, Jl. Prof. Soedarto No.13, Kec. Tembalang, Kota Semarang 50275, Indonesia.*
[c] *Department of Statistics Diponegoro University, Jl. Prof. Soedarto No.13, Kec. Tembalang, Kota Semarang 50275, Indonesia.*
[d] *Department of Psychology, State University of Malang, Jl. Cakrawala No.5, Kec. Lowokwaru, Kota Malang 65145, Indonesia.*
email: [a,]mahrusali1606@ub.ac.id, [b]rahmatgernowo@lecturer.undip.ac.id, [c]budiwarsitoundip@gmail.com , [d] faliha.muthmainah.fpsi@um.ac.id*
* Correspondence*

ARTICLE INFO

Article history:
Received December 7th, 2023
Revised February 17th, 2024
Accepted June 27th, 2025
Available online June 30th, 2025

Keywords:
Deep learning;
LSTM;
CNN;
Hybrid model;
Kitchenham.

Please cite this article in IEEE style as:
M. Ali, R. Gernowo, B. Warsito and F. Muthmainah, "Enhancing Bank Financial Performance Assessment: A Literature Review of Deep Learning Applications Using the Kitchenham Method," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 11, no. 1, pp. 54-65, 2025.

ABSTRACT

The assessment of bank financial performance is crucial for ensuring the stability of the banking sector. With advancements in technology, especially deep learning (DL), there is increasing potential to improve the accuracy of risk prediction and financial performance evaluation in banks. However, challenges related to data imbalance and model complexity require more efficient approaches. This study aims to examine the application of DL in assessing bank financial performance, with a focus on credit risk, fraud detection, and bankruptcy prediction. A Systematic Literature Review (SLR) was conducted using the Kitchenham approach, analyzing 697 relevant articles to address nine research questions regarding the application of DL in the banking sector. This study contributes by providing insights into effective DL models that enhance financial performance and risk prediction in banks, while also offering recommendations for the development of more transparent models. The results indicate that models such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) perform well in handling large financial data. Additionally, hybrid models that combine DL with traditional models demonstrate higher accuracy in bankruptcy prediction and fraud detection.

1. Introduction

The banking industry is vital to the global economy, and accurate financial performance assessment is essential. Advances in artificial intelligence and deep learning have enhanced predictive accuracy and improved risk management. Deep learning techniques, such as Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and hybrid models, have demonstrated promising results in areas such as credit risk assessment, fraud detection, and bankruptcy prediction [1][2][3]. These methods have shown a remarkable ability to handle large-scale, high-dimensional datasets, enabling better prediction and classification of default risk, financial fraud, and other critical performance indicators. Despite these advancements, the application of deep learning models in banking remains a topic of growing interest, with many studies focusing on enhancing model transparency [4] and addressing challenges such as data imbalance.

Several gaps remain unaddressed in the current body of research. Many deep learning (DL) models used in banking are treated as "black boxes," raising concerns about transparency, interpretability, and regulatory compliance. Furthermore, issues such as data imbalance, overfitting,

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 54-65

and lack of domain-specific model customization continue to challenge the practical deployment of DL models in real-world financial institutions. Given these challenges, there is a need for a comprehensive synthesis of existing studies to map the evolution, effectiveness, and limitations of deep learning applications in banking. While several previous reviews have explored AI in finance [5], few have conducted a systematic and targeted review focusing specifically on deep learning for credit risk, fraud detection, and bankruptcy prediction–three core components of banking performance evaluation.

This study aims to systematically review the application of deep learning models in banking performance assessment, specifically focusing on credit risk, fraud detection, and bankruptcy prediction. Using a Systematic Literature Review (SLR) employing the Kitchenham method[6], this research synthesizes findings from 798 articles to explore key trends and challenges in this field. The nine research questions identified in this review guide the analysis of how deep learning models can be leveraged to enhance financial performance assessment in banking. The novelty of this study lies in its integrative framework, which combines technical perspectives (e.g., algorithmic effectiveness) with practical considerations (e.g., transparency and applicability in financial systems). The findings are expected to contribute both theoretically and practically by guiding researchers, developers, and banking practitioners toward more effective, explainable, and robust application of deep learning in financial performance evaluation.

2. Materials and Methods

This study employs the Systematic Literature Review (SLR) method using the Kitchenham approach, which follows a rigorous framework consisting of literature search strategies, study selection, classification and quality assessment, data extraction, and data synthesis [6].

2.1. Selection of Studies

The review process began with a comprehensive search of reputable databases, including Scopus, and IEEE Xplore. The search targeted peer-reviewed journal articles published between 2018 and 2024. The inclusion criteria were as follows: (1) studies published within the 2018–2024 period,, (2) focused on the banking sector, (3) applied deep learning methods, and (4) involved empirical research using real-world data and evaluation metrics. The search used the keywords "deep learning" and "bank", resulting in 798 articles,. From this pool. studies were classified into four scopes of banking such as "credit risk prediction" [7], "bankruptcy prediction" [8], "fraud detection" [9], and "financial performance" [10]. Across these four domains of banking, nine types of research were identified, as illustrated in in the inclusion diagram (Fig.1).

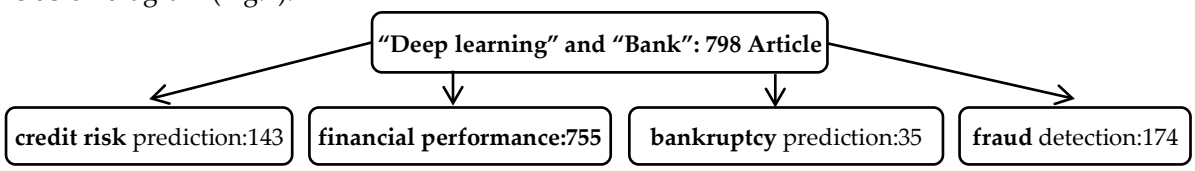

Fig. 1. Inclusion diagram

The subsequent selection step is the exclusion process, based on the following criteria: absence of empirical evaluation, publication not indexed in reputable databases, focus on non-financial sectors, and studies that are duplicates or preprints. As illustrated in Fig. 1, applying these four exclusion criteria resulted in nine selected topics, which are presented in the exclusion diagram in Fig. 2.

2.2. Data Extraction

From the selected articles, data were extracted based on the following elements: (1) Study information: Authors, year of publication, and title. (2) Objective: The main purpose of the study, particularly the problem being addressed, such as credit risk prediction, fraud detection, or bankruptcy forecasting. (3) Deep Learning Models Used: Specific deep learning methods employed (e.g., LSTM, CNN, GAN, etc.). (4) Datasets: Information about the datasets used in the studies, including data sources, types (e.g., financial datasets, credit card transaction data, etc.), and the number of samples. (5) Evaluation Metrics: Performance metrics such as accuracy, precision, recall, F1-score, and G-mean. (6) Contributions: The key contributions of each study in terms of enhancing model performance, addressing challenges, or proposing novel solutions, as summarized in Table 1.

 http://doi.org/10.26594/register.v11i1.4224

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 54-65

Fig. 2. Exclusion Diagram

Table 1. Contribution Table

No	Author and Year	Research Title	DL Model	Dataset	RQ	Evaluation Metrics	Key Results	Contribution
1	Kang et al. (2022)	A CWGAN-GP-based multi-task learning model for consumer credit scoring	CWGAN-GP	Consumer credit data	RQ 1	Accuracy, AUC, F1-Score	Improved credit scoring accuracy for imbalanced datasets	Introduction of CWGAN-GP for imbalanced class handling in credit scoring
…	……	……	……	……	…	……	……	……..
697	Stevenson et al. (2021)	The value of text for small business default prediction: A Deep Learning approach	BERT (Bidirectional Encoder Representations from Transformers)	60,000 textual loan assessments from a lender	RQ 9	AUC, Brier Score	Text alone surprisingly effective for predicting default; combining with traditional data yields no improvement	Demonstrates BERT's robustness for automating mSME lending process and improving loan assessment strategies

2.3. Quality Assessment and Data Synthesis

The quality of included studies was assessed based on the Kitchenham checklist guidelines, which include the following questions:

(Q1) Is the aim of the study clearly stated?
(Q2) Is the methodology described in sufficient detail?
(Q3) Is the data source appropriate and clearly described?
(Q4) Are the evaluation metrics appropriate and adequately applied?
(Q5) Are the results clearly presented and logically interpreted?
(Q6) Does the study discuss its limitations and potential biases?
(Q7) Does the study make a clear contribution to research or practice?

The assessment was conducted using a scale of 0 (no), 0.5 (partial), and 1 (yes/good). The total score was used to classify the quality of the articles as acceptable (very good and good), less acceptable (fair), or excluded (poor) from the synthesis, as shown Table 2.

 http://doi.org/10.26594/register.v11i1.4224

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 54-65

Table 2. Quality Assessment Table

No	Article	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Total Score	Percentage	Quality
1	Sue et al., 2022	1	1	1	1	1	0.5	1	6.5 / 7	92.9%	Very good
……	……..	……...	…… …	……..	…… …	…..	…….	…..	……	….....	……
696	Wang et al., 2023	1	0.5	1	1	0.5	0.5	1	5.5 / 7	78.6%	Good
697	Stevenson et al. (2021)	0.5	0.5	0.5	0.5	0.5	0	0.5	3 / 7	42.9%	Poor

The formulas for calculating the Total Score, Percentage, and Quality columns in the "Quality Assessment" table, as presented, are as follows:

Total Score=$\sum_{i=1}^{7} Qi$ (1)

Percentage=$\left(\frac{Total\ Score}{7}\right) x100\%$ (2)

The categorization of Quality based on the Percentage is as follows: (1) 85% – 100% = Very Good; (2) 70 %– 84.9%= Good; (3) 50 %– 69.9%= Fair, (4) < 50%= Poor.

2.4. Research Questions (RQ) and Analysis

Nine specific research questions were formulated to guide the review process. These questions focused on the following areas:

RQ1. How effective are deep learning models in predicting credit risk in banks?

RQ2. What challenges are associated with applying deep learning to fraud detection in financial institutions?

RQ3. Which deep learning models have shown the best performance in bankruptcy prediction?

RQ4. How do data quality and availability impact the accuracy of deep learning models in financial applications?

RQ5. What are the limitations of current deep learning models used in the banking sector?

RQ6. How do hybrid deep learning models compare to traditional methods in financial risk prediction?

RQ7. What are the key evaluation metrics used to assess deep learning models in banking performance?

RQ8. How do deep learning techniques address the issue of data imbalance in credit risk prediction?

RQ9. What future trends are emerging in the use of deep learning for banking performance assessment?

3. Result And Discussion

Each of the research questions is explored by analyzing relevant articles and summarizing their findings. A comparative analysis is conducted to evaluate the strengths and weaknesses of various deep learning models and their applications in the banking sector. as described below.

RQ1. How effective are deep learning models in predicting credit risk in banks?:

Utilization of More Advanced Deep Learning Models

Deep learning models such as Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) are employed to handle more complex and larger datasets that traditional models, such as logistic regression or decision trees, struggle to process effectively. These models are capable of learning deeper and more intricate patterns from the data, providing more accurate credit score assessments, particularly when dealing with large datasets that contain diverse features [11][12][13].

Enhanced Predictions with More Complex Data

Deep learning facilitates the processing of more complex data, including unstructured data (e.g., text or social media data), alongside structured financial data, enabling a more comprehensive customer evaluation. For example, multi-input deep learning models can integrate financial data with external relevant information to provide more accurate credit scoring predictions [14][15][16].

Improvements in Overfitting and Generalization

When combined with regularization techniques such as dropout and early stopping, Deep Learning can effectively reduce overfitting, which is a common issue in simpler models. These techniques allow the model to generalize better to unseen data, thereby improving its ability to predict credit scores accurately [17][18].

 http://doi.org/10.26594/register.v11i1.4224

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 54-65

Enhanced Feature Engineering

Deep learning also enables automatic feature engineering, allowing models to discover and extract relevant features from input data without manual intervention. This is particularly advantageous in credit scoring, as the model can identify hidden factors that influence creditworthiness that may not be easily detected by humans or traditional models [19].

Hybrid Models

Several studies propose the use of hybrid models that combine deep learning with traditional or other machine learning algorithms, such as Random Forest or Support Vector Machines, to achieve improved accuracy. These hybrid models leverage the strengths of both approaches–deep learning for handling complex data structures and machine learning for providing better interpretability [20)].

Advancements in Model Evaluation

In evaluating model performance, metrics such as Accuracy, Area Under the Curve (AUC), F1-score, and Receiver Operating Characteristic (ROC) Curve are commonly used to demonstrate improvements in predictive accuracy. Studies have shown that deep learning models achieve higher AUC values compared to traditional models, indicating their superior ability to distinguish between creditworthy and non-creditworthy clients [21].

RQ2. What challenges are associated with applying deep learning to fraud detection in financial institutions?:

Based on a review of recent literature, several key challenges have been identified in the application of deep learning (DL) for fraud detection in the financial sector:

Data Quality and Availability

Deep learning models require large volumes of high-quality, labeled data. However, in the context of fraud detection, data is often fragmented across institutions and subject to privacy and regulatory constraints. Labeling fraud-related data also demands domain expertise and is time-consuming, which limits the availability of clean, labeled datasets [22].

Class Imbalance

Fraudulent transactions typically represent a very small proportion of all financial transactions, resulting in highly imbalanced datasets. This imbalance leads to biased models that are less sensitive to fraudulent activity. Techniques such as SMOTE (Synthetic Minority Oversampling Technique) and hybrid oversampling-undersampling approaches have been proposed to address this issue [23].

High Computational Cost

Deep Learning architectures used for fraud detection—such as CNNs, RNNs, or LSTMs—are often computationally intensive. Their training and deployment require significant computing resources (e.g., GPUs or cloud-based infrastructure), which may be a barrier for smaller institutions [24].

Lack of Interpretability and Explainability

One of the primary criticisms of Deep Learning in fraud detection is its "black-box" nature. Financial institutions and regulators demand transparency and interpretability, which deep learning models often lack. Recent efforts in Explainable Artificial Intelligence (XAI) aim to bridge this gap [25].

Adversarial Attacks

Deep learning models are vulnerable to adversarial attacks, in which small, deliberate perturbations in the input data can mislead the model. Techniques such as evasion attacks and data poisoning can compromise the model's robustness and significantly degrade its detection performance [26].

Concept Drift and Evolving Fraud Strategies

Fraudulent behavior constantly evolves, resulting in a phenomenon known as concept drift, in which the statistical properties of the data change over time. Static DL models may struggle to adapt to these changes, reducing their long-term effectiveness. Techniques such as online learning and continual model updates are necessary to mitigate this issue [27].

Limited Access to Labeled Data and Privacy Constraints

Supervised deep learning depends on labeled data, however, fraud cases are both rare and sensitive. Regulatory frameworks such as GDPR further restrict the sharing and centralization of financial data. Emerging approaches such as federated learning and unsupervised methods are being explored to address these challenges without compromising data privacy [28].

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 54-65

RQ3. Which deep learning models have shown the best performance in bankruptcy prediction?:

Long Short-Term Memory (LSTM)

LSTM models are widely recognized for their ability to process sequential and time-series data, making them highly suitable for predicting financial distress and bankruptcy. Studies have demonstrated that LSTM models outperform traditional machine learning algorithms in modeling long-term dependencies in financial ratios and indicators [29].

Bidirectional LSTM (BiLSTM)

BiLSTM networks further enhance prediction by analyzing financial data in both forward and backward directions. This bidirectional architecture has been shown to improve accuracy in bankruptcy prediction compared to unidirectional models [30].

Hybrid Deep Learning Models

Hybrid models that combine LSTM or GRU with optimization algorithms (e.g., Genetic Algorithm, Particle Swarm Optimization) or with traditional models like Random Forest have also been reported to improve predictive performance. For example, WSODL-BPFCA (White Shark Optimizer with Deep Learning) has achieved accuracy above 97% in predicting bankruptcy risk [31]

Deep Neural Networks (DNNs)

DNNs have shown strong performance in modeling complex nonlinear relationships between financial variables and bankruptcy outcomes. DNNs are especially effective when trained on large datasets that include both structured and unstructured features [1].

RQ4. How do data quality and availability impact the accuracy of deep learning models in financial applications?

Impact of Incomplete or Noisy Data

The accuracy of deep learning models in financial tasks is highly dependent on the quality of input data. Incomplete, noisy, or inconsistent financial datasets can lead to poor model generalization and reduced prediction accuracy [32].

Limited Availability of Labeled Financial Data

Supervised deep learning models require large volumes of labeled data for effective training. In the financial sector, such data is often confidential or scarce, especially for events like fraud or bankruptcy, which are relatively rare. This hinders model development and reduces accuracy [33].

Bias in Historical Financial Data

Models trained on historical data may inherit existing biases—such as systemic bias in loan approvals or discrimination in financial access—which can skew predictions and perpetuate unfair outcomes if not corrected [34].

Challenges in Integrating Unstructured Data

While Deep Learning models can leverage unstructured data (e.g., text, images), preprocessing such data (e.g., financial documents, news sentiment) requires significant effort. Poor preprocessing can degrade performance despite the richness of the data [35].

Regulatory and Privacy Constraints

Regulatory frameworks (e.g., GDPR, financial secrecy laws) limit data availability and sharing, reducing access to comprehensive datasets necessary for robust model training. This often leads to fragmented or siloed data, which hampers accuracy [36].

RQ5. What are the limitations of current deep learning models used in the banking sector?

Lack of Interpretability (Black Box Problem)

Deep Learning models, particularly deep neural networks, often lack transparency and interpretability. This "black-box" nature makes it difficult for financial institutions and regulators to understand how predictions are made, which is problematic in risk-sensitive environments like banking [37].

High Data Requirements

These models require large volumes of high-quality, labeled data for effective training. However, in the banking sector, data related to fraud, defaults, or ESG risk events is often rare, imbalanced, or confidential, thereby limiting model performance [38].

 http://doi.org/10.26594/register.v11i1.4224

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 54-65

Overfitting and Limited Generalization

Deep Learning models can overfit the training data, especially in financial domains characterized by high volatility and noise. This reduces their ability to generalize to unseen cases, which is risky when deployed in real-world banking applications [39].

Computational Complexity and Cost

Training Deep Learning models demands substantial computational power, time, and resources, which can be a significant barrier for small- to mid-sized banks.

Regulatory and Ethical Concerns

Due to limited explainability and the potential for bias in training data, there are growing regulatory and ethical concerns regarding the deployment of DL models in credit scoring, fraud detection, and customer profiling.

RQ6. How do hybrid deep learning models compare to traditional methods in financial risk prediction?

Integration of Deep Learning with Traditional Statistical Models

Hybrid models that combine deep learning (e.g., LSTM or CNN) with traditional statistical methods such as ARIMA or logistic regression have demonstrated improved prediction accuracy in financial risk assessments. While traditional models offer greater interpretability, deep learning enhances the model's ability to capture complex, nonlinear patterns.

DL and Machine Learning Ensembles

Some studies propose combining deep learning with ensemble methods such as XGBoost, Random Forest, or Support Vector Machines (SVM). These hybrid models leverage the feature learning capabilities of DL and the interpretability or robustness of classical ML, resulting in improved precision and generalization.

Enhanced Feature Selection and Dimensionality Reduction

Hybrid models often use traditional statistical techniques for preprocessing and dimensionality reduction (e.g., PCA), followed by deep learning for prediction. This combination improves model efficiency and reduces computational burden, especially in high-dimensional financial datasets

Accuracy Comparison with Traditional Models

Empirical evidence indicates that hybrid DL models consistently outperform traditional models, such as logistic regression, linear discriminant analysis (LDA), or ARIMA in financial risk prediction tasks, including bankruptcy prediction, credit scoring, and fraud detection. The improvements are particularly significant in terms of AUC, recall, and F1-score metrics.

Challenges of Hybrid Models

Despite better performance, hybrid models also pose challenges such as increased model complexity, higher computational requirements, and reduced transparency compared to purely statistical approaches.

RQ7. What are the key evaluation metrics used to assess deep learning models in banking performance?.

In assessing the performance of Deep Learning models in banking applications, such as credit scoring, fraud detection, bankruptcy prediction, and stock price forecasting, several key evaluation metrics are commonly used:

Accuracy, Precision, Recall, and F1-Score

These are widely used in classification problems, such as in fraud detection and credit risk prediction.

Accuracy measures the proportion of correct predictions; Precision indicates the proportion of true positives among predicted positives; Recall (Sensitivity) captures the ability to find all relevant positive cases; F1-Score provides a harmonic mean of precision and recall.

Studies such as [40] demonstrate that DL models, e.g. LSTM and CNN, achieved F1-scores above 90% in fraud detection tasks.

Mean Absolute Error (MAE) and Root Mean Square Error (RMSE)

These are typically used in regression tasks, such as stock price or exchange rate forecasting.

MAE quantifies the average magnitude of prediction errors; RMSE gives higher weight to larger errors, making it more sensitive to outliers. For example, LSTM models achieved an RMSE of 0.023 in FOREX prediction [41], significantly outperforming ARIMA and SVM models.

 http://doi.org/10.26594/register.v11i1.4224

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 54-65

Area Under the Curve – Receiver Operating Characteristic (AUC-ROC)

AUC-ROC is used in binary classification tasks (e.g., loan approval or rejection).

Higher AUC values indicate better model discrimination capability.

Shi et al. (2022) reported that CNN-LSTM hybrids achieved an AUC of 0.96 in bank loan classification.

Sharpe Ratio and R-Squared

In portfolio and stock return prediction, performance is often assessed using financial metrics.

Sharpe Ratio evaluates the risk-adjusted return of the predictive model.

R^2 (Coefficient of Determination) indicates how well the model explains the variance in the data. It applied these metrics in DL-based stock price forecasting with R^2 reaching 0.87[42].

Confusion Matrix and Log Loss

Confusion Matrix and Log Loss are used for more detailed performance analysis and error distribution, especially in unbalanced datasets.

RQ8. How do deep learning techniques address the issue of data imbalance in credit risk prediction?

In credit risk prediction, class imbalance is a common challenge, as the number of defaulters (positive class) is usually much smaller than non-defaulters (negative class). Deep learning models have incorporated several strategies to address this issue:

Use of Weighted Loss Functions

Many Deep Learning models apply weighted cross-entropy or focal loss to penalize misclassification of minority classes more heavily. This approach allows the model to learn better representations of the rare (risky) class, improving recall for defaulters.

Oversampling the Minority Class

Techniques such as SMOTE (Synthetic Minority Over-sampling Technique) and ADASYN are commonly applied in preprocessing steps to balance datasets before training DL models such as CNN or LSTM. This synthetic oversampling allows models to receive more representative samples of high-risk customers.

One study compared models trained with and without oversampling and found that those using SMOTE achieved a 7–10% increase in F1-score for the minority class [43].

Ensemble Deep Learning Models

Combining multiple DL models (e.g., ensemble of LSTM and CNN) can help mitigate bias toward the majority class. These ensemble models often aggregate predictions using voting or averaging schemes, which give more attention to the minority class.

Data Augmentation for Financial Features

Some models create synthetic financial behavior patterns using Generative Adversarial Networks (GANs) to enrich training data for the minority class. This approach improves the model's exposure to edge cases related to default behavior.

Dropout and Regularization

Regularization techniques like Dropout are also used to avoid overfitting to the majority class and to improve generalization, especially important in the context of imbalanced dataset.

RQ9. What future trends are emerging in the use of Deep Learning for banking performance assessment?:

Use of Complex and Hybrid Models

Recent studies increasingly utilize hybrid or multi-stage models to enhance prediction performance. For instance, the XGBoost model is used to transform credit features into a high-dimensional feature matrix, which is then processed using graph-based neural networks (forgeNet) to mine the relationships between these features. In this way, Deep Learning can capture more complex patterns in credit data that conventional models often miss.

Use of Graph and Attention Techniques

Graph Attention Networks (GAT) are employed to predict loan default risks, allowing the system to understand non-linear and high-level relationships between borrowers. GAT helps the model identify more complex patterns between historical credit data, borrowers' financial profiles, and transaction histories, leading to improved predictive accuracy for default.

 http://doi.org/10.26594/register.v11i1.4224

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 54-65

Model Optimization with Imbalanced Data

In many cases, the data used has an imbalanced distribution. DL can address this issue by using cost-adjustment algorithms or SMOTE (Synthetic Minority Over-sampling Technique) to place more focus on the minority class (default). This techniques enhance the model's sensitivity in detecting potential default risks even when the data is sparse.

Efficient Feature Extraction and Data Preprocessing

Transfer Light Gradient Boosting Machine (TrLightGBM), which combines transfer learning and feature engineering, addresses data imbalance issues while providing deeper insights into the factors influencing default prediction. With efficient data processing, the model can learn more effectively and make more accurate predictions.

Utilization of Deep Reinforcement Learning (DRL) Models

DRL can be used to predict default risk in a more adaptive manner. By formulating this as a Markov Decision Process (MDP), the model can prioritize high-risk borrowers for default. DRL allows the model to learn from previous decisions and optimize predictions based on potential financial losses.

Model Interpretability and Explanations

Most DL models used for loan default prediction, as demonstrated in studies with Explainable AI (XAI), provide transparency in decision-making processes. By leveraging SHAP (Shapley Additive Explanations), the model can provide insights into which features have the greatest influence on default predictions. This level of interpretability is crucial for banks and financial institutions to communicate risks to decision-makers and mitigate losses with more precise strategies.

Combination of Various Types of Data (Multiview Data)

In some studies, particularly those employing Multi-view GCN, DL combines various types of data, including loan information, credit history, and borrowers' personal information. This approach allows the model to account for more aspects of borrower behavior, providing a more comprehensive view and improving the accuracy of default predictions.

Overall Benefits of Deep Learning in Default Prediction

Overall, the use of deep learning in predicting bank loan defaults provides significant advantages in terms of accuracy, proactivity, and the ability to identify patterns that conventional models often overlook. With the ability to process large, complex data, these models enable banks to respond to potential default risks more quickly and efficiently, thereby reducing financial losses caused by troubled loans.

4. Conclusion

This research has systematically explored the role of Deep Learning in enhancing banking performance assessment across nine thematic research questions. The findings underscore that deep learning models–particularly LSTM, CNN, and hybrid architectures–significantly outperform traditional methods in tasks such as credit scoring, bankruptcy prediction, fraud detection, and financial ratio analysis. These models offer superior predictive accuracy, especially when combined with unstructured data sources like sentiment and ESG reports. Furthermore, deep learning techniques demonstrate a strong capacity to handle data quality issues and class imbalance, which are common challenges in financial datasets. Despite these advantages, notable limitations persist, particularly regarding model interpretability, limited integration of ESG dimensions, and the dependency on high-quality data inputs. Emerging trends, including the adoption of graph neural networks, explainable AI, and multimodal learning, signal a shift toward more transparent, adaptive, and comprehensive analytical frameworks in banking. Overall, deep learning is not only transforming predictive modeling in the financial sector but also enabling more informed and proactive decision-making. However, further research is necessary to address ethical, regulatory, and technical challenges, ensuring the responsible and sustainable adoption of Deep Learning in banking applications.

Based on the gaps and trends identified in the review, future research in this domain should focus on: (1) Explainable AI (XAI) and Model Interpretability: Developing frameworks that make DL predictions in banking more transparent and explainable for decision-makers and regulators; (2) ESG and Sustainable Finance Modeling: Integrating environmental, social, and governance (ESG) indicators more comprehensively with DL to assess long-term financial sustainability and risk; (3) Graph-Based and Attention Mechanisms: Expanding the use of Graph Neural Networks (GNNs), GATs, and

 http://doi.org/10.26594/register.v11i1.4224

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 54-65

attention-based models to capture complex interdependencies in customer and transaction networks; (4) Data Privacy and Ethics in Financial AI: Addressing privacy, fairness, and ethical implications of DL applications in personalized banking and credit evaluation; (5) Benchmarking and Standardization: Establishing standardized benchmarks, datasets, and evaluation protocols to improve reproducibility and cross-study comparison; (6) Real-time and Federated Learning Models: Exploring DL models that can learn from decentralized data sources (e.g., federated learning) and deliver insights in real-time without compromising data privacy.

Author Contribution

M. Ali: Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, and software. R. Gernowo: Supervision, validation, and writing – review & editing. B. Warsito: Supervision, validation, and writing – review & editing. F. Muthmainah: Visualization, writing – original draft, and writing – review & editing.

Acknowledgment

The authors would like to express their sincere gratitude to the academic staff of the Department of Information Systems, Diponegoro University, and the Faculty of Computer Science, State University of Malang, for their guidance and support. We also appreciate the access to digital library resources that greatly facilitated this research. Special thanks are extended to colleagues and reviewers for their constructive feedback during the writing process.

Declaration of Competing Interest

We declare that we have no conflict of interest

References

[1] L. Li and B. M. Muwafak, "Adoption of deep learning Markov model combined with copula function in portfolio risk measurement," *Appl. Math. Nonlinear Sci.*, vol.7, no.1, pp. 901-916, 2021, doi: 10.2478/amns.2021.2.00112.

[2] M. Andronie *et al.*, "Generative artificial intelligence algorithms in Internet of Things blockchain-based fintech management," *Oeconomia Copernicana*, vol. 15, no. 4. pp. 1349–1381, 2024. doi: 10.24136/oc.3283.

[3] F. Liu, "Improve the Bi-LSTM Model of University Financial Information Management Platform Construction," *J. Electr. Syst.*, vol. 20, no. 1, pp. 124–138, 2024, doi: 10.52783/jes.671.

[4] M. Ali, R. Gernowo, B. Warsito, and F. Muthmainah, "Markov Switching Autoregressive in Information Systems for Improving Islamic Banks," *Data Metadata*, vol. 3, pp. 1-10, 2024, doi: 10.56294/dm2024.681.

[5] M. Ali, R. Gernowo, and B. Warsito, "Performance Analysis of Islamic Banks in Indonesia Using Machine Learning," *E3S Web Conf.*, vol. 448, pp. 1-7, 2023, doi: 10.1051/e3sconf/202344802026.

[6] B. Kitchenham *et al.*, "Systematic literature reviews in software engineering-A tertiary study," *Inf. Softw. Technol.*, vol. 52, no. 8, pp. 792–805, 2010, doi: 10.1016/j.infsof.2010.03.006.

[7] E. V Orlova, "Methodology and models for individuals' creditworthiness management using digital footprint data and machine learning methods," *Mathematics*, vol. 9, no. 15, pp. 1-28, 2021, doi: 10.3390/math9151820.

[8] P. K. Viswanathan, S. Srinivasan, and N. Hariharan, "Predicting Financial Health of Banks for Investor Guidance Using Machine Learning Algorithms," *J. Emerg. Mark. Financ.*, vol. 19, no. 2, pp. 226–261, 2020, doi: 10.1177/0972652720913478.

[9] A. M. Ozbayoglu, M. U. Gudelek, and O. B. Sezer, "Deep learning for financial applications: A survey," *Appl. Soft Comput. J.*, vol. 93, pp. 106384, 2020, doi: 10.1016/j.asoc.2020.106384.

[10] N. Majidi, M. Shamsi, and F. Marvasti, "Algorithmic trading using continuous action space deep reinforcement learning[Formula presented]," *Expert Syst. Appl.*, vol. 235, pp. 121245, 2024, doi: 10.1016/j.eswa.2023.121245.

[11] K. L. Sue, C. F. Tsai, and H. M. Tsau, "Missing value imputation and the effect of feature normalisation on financial distress prediction," *J. Exp. Theor. Artif. Intell.*, vol. 36, no. 8, pp. 1467–1483, 2022, doi: 10.1080/0952813X.2022.2153278.

[12] T. Kristóf and M. Virág, "EU-27 bank failure prediction with C5.0 decision trees and deep learning neural networks," *Res. Int. Bus. Financ.*, vol. 61, pp. 101644, 2022, doi: 10.1016/j.ribaf.2022.101644.

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 54-65

[13] L. O. Hjelkrem, P. E. de Lange, and E. Nesset, "The Value of Open Banking Data for Application Credit Scoring: Case Study of a Norwegian Bank," *J. Risk Financ. Manag.*, vol. 15, no. 12, pp. 1-15, 2022, doi: 10.3390/jrfm15120597.

[14] A. Guarino, L. Grilli, D. Santoro, F. Messina, and R. Zaccagnino, "To learn or not to learn? Evaluating autonomous, adaptive, automated traders in cryptocurrencies financial bubbles," *Neural Comput. Appl.*, vol. 34, no. 23, pp. 20715–20756, 2022, doi: 10.1007/s00521-022-07543-4.

[15] J. A. Bastos and S. M. Matos, "Explainable models of credit losses," *Eur. J. Oper. Res.*, vol. 301, no. 1, pp. 386–394, 2022, doi: 10.1016/j.ejor.2021.11.009.

[16] M. Corstjens, M. Bakhshandeh, P. Kahraman, and J. Bosman, "Predicting the daily number of payment transactions in the largest bank in the Netherlands: Application to Banking Data," 2019, pp. 5507–5512. doi: 10.1109/BigData47090.2019.9005538.

[17] E. Saberi, J. Pirgazi, and A. Ghanbari sorkhi, "A machine learning approach for trading in financial markets using dynamic threshold breakout labeling," *J. Supercomput.*, vol. 80, no. 17, pp. 25188–25221, 2024, doi: 10.1007/s11227-024-06403-3.

[18] I. Pratama, P. T. Prasetyaningrum, A. Y. Chandra, and O. Suria, "Measuring Resampling Methods on Imbalanced Educational Dataset's Classification Performance," *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 10, no. 1, pp. 1–11, 2024, doi: 10.26594/register.v10i1.3397.

[19] Y. R. Wang and Y. C. Tsai, "The Protection of Data Sharing for Privacy in Financial Vision," *Appl. Sci.*, vol. 12, no. 15, pp. 1-22, 2022, doi: 10.3390/app12157408.

[20] V. G. Krishnan, M. V. V. Saradhi, T. A. M. Prakash, K. G. Kannan, and A. G. N. Julaiha, "Development of Deep Learning based Intelligent Approach for Credit Card Fraud Detection," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 10, no. 12, pp. 133–139, 2022, doi: 10.17762/ijritcc.v10i12.5894.

[21] G. A. Chandok, V. A. M. Rexy, H. A. Basha, and H. Selvi, "Enhancing Bankruptcy Prediction with White Shark Optimizer and Deep Learning: A Hybrid Approach for Accurate Financial Risk Assessment," *Int. J. Intell. Eng. Syst.*, vol. 17, no. 1, pp. 140–148, 2024, doi: 10.22266/ijies2024.0229.14.

[22] A. Oguntimilehin, M. L. Akukwe, K. A. Olatunji, O. B. Abiola, O. A. Adeyemo, and I. A. Abiodun, "Mobile Banking Transaction Authentication using Deep Learning," in *2022 5th Information Technology for Education and Development (ITED)*, Abuja, Nigeria: IEEE, pp. 1-7, 2022, doi: 10.1109/ITED56637.2022.10051553.

[23] D. Singh and B. K. Gupta, "Closing Price Prediction of Nifty Stock Using LSTM with Dense Network," in *Lecture Notes in Networks and Systems*, vol. 302, pp. 382–392, 2022, doi: 10.1007/978-981-16-4807-6_37.

[24] S. P. Sharma, L. Singh, and R. Tiwari, "Integrated feature engineering based deep learning model for predicting customer's review helpfulness," *J. Intell. Fuzzy Syst.*, vol. 44, no. 6, pp. 8851–8868, 2023, doi: 10.3233/JIFS-223546.

[25] T. Baabdullah, A. Alzahrani, D. B. Rawat, and C. Liu, "Efficiency of Federated Learning and Blockchain in Preserving Privacy and Enhancing the Performance of Credit Card Fraud Detection (CCFD) Systems," *Futur. Internet*, vol. 16, no. 6, pp. 1-22, 2024, doi: 10.3390/fi16060196.

[26] R. Chakraborty, A. Samanta, K. M. Agrawal, and A. Dutta, "Towards smarter grid: Policy and its impact assessment through a case study," *Sustain. Energy, Grids Networks*, vol. 26, pp. 100436, 2021, doi: 10.1016/j.segan.2021.100436.

[27] J. El Fiorenza Caroline, P. Parmar, S. Tiwari, A. Dixit, and A. Gupta, "Accuracy prediction using analysis methods and f-measures," in *Journal of Physics: Conference Series*, 2019, vol. 1362, no. 1. pp. 1-8, 2019, doi: 10.1088/1742-6596/1362/1/012040.

[28] A. Kesa and T. Kerikmäe, "Artificial Intelligence and the GDPR: Inevitable Nemeses," *TalTech J. Eur. Stud.*, vol. 10, no. 3, pp. 68–90, 2020, doi: 10.1515/bjes-2020-0022.

[29] D. Baishya, J. J. Deka, G. Dey, and P. K. Singh, "SAFER: Sentiment Analysis-Based FakE Review Detection in E-Commerce Using Deep Learning," *SN Comput. Sci.*, vol. 2, no. 6, pp. 479, 2021, doi: 10.1007/s42979-021-00918-9.

[30] Y. Zhao, "The Data Analysis of Enterprise Operational Risk Prediction Under Machine Learning: Innovations and Improvements in Corporate Law Risk Management Strategies," *J. Organ. End User Comput.*, vol. 36, no. 1, pp. 1-24, 2024, doi: 10.4018/JOEUC.355709.

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
11 (1) 2025 54-65

[31] A. Maroof, S. Wasi, S. I. Jami, and M. S. Siddiqui, "Aspect-Based Sentiment Analysis for Service Industry," *IEEE Access*, vol. 12, pp. 109702–109713, 2024, doi: 10.1109/ACCESS.2024.3440357.

[32] Q. Li, H. Wu, W. Qian, X. Li, Q. Zhu, and S. Yang, "Portfolio Optimization Based on Quantum HHL Algorithm," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2022, vol. 13339 LNCS, pp. 90–99. doi: 10.1007/978-3-031-06788-4_8.

[33] D. C. Yıldırım, I. H. Toroslu, and U. Fiore, "Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators," *Financ. Innov.*, vol. 7, no. 1, pp. 1-36, 2021, doi: 10.1186/s40854-020-00220-2.

[34] X. Y. Liu *et al.*, "Dynamic datasets and market environments for financial reinforcement learning," *Mach. Learn.*, vol. 113, no. 5, pp. 2795–2839, 2024, doi: 10.1007/s10994-023-06511-w.

[35] S. Pol, M. Hudnurkar, and S. S. Ambekar, "Predicting Credit Ratings using Deep Learning Models – An Analysis of the Indian IT Industry," *Australas. Accounting, Bus. Financ. J.*, vol. 16, no. 5, pp. 38–51, 2022, doi: 10.14453/aabfj.v16i5.04.

[36] E. Politou, E. Alepis, and C. Patsakis, "Profiling tax and financial behaviour with big data under the GDPR," *Comput. Law Secur. Rev.*, vol. 35, no. 3, pp. 306–329, 2019, doi: 10.1016/j.clsr.2019.01.003.

[37] C. Y. Lee, S. K. Koh, M. C. Lee, and W. Y. Pan, "Application of Machine Learning in Credit Risk Scorecard," in *Communications in Computer and Information Science*, 2021, vol. 1489 CCIS, pp. 395–410. doi: 10.1007/978-981-16-7334-4_29.

[38] S. C. Tékouabou Koumétio and H. Toulni, "Improving KNN Model for Direct Marketing Prediction in Smart Cities," *Studies in Computational Intelligence*, vol. 971. Springer Science and Business Media Deutschland GmbH, Faculty of Sciences, Department of Computer Sciences, Chouaib Doukkaly Univercity, B.P. 20, El Jadida, 2400, Morocco, pp. 107–118, 2021. doi: 10.1007/978-3-030-72065-0_7.

[39] Y. Yang, X. Su, and S. Yao, "Nexus between green finance, fintech, and high-quality economic development: Empirical evidence from China," *Resour. Policy*, vol. 74, pp. 102445, 2021, doi: 10.1016/j.resourpol.2021.102445.

[40] E. Parkar, S. Gite, S. Mishra, B. Pradhan, and A. Alamri, "Comparative study of deep learning explainability and causal ai for fraud detection," *Int. J. Smart Sens. Intell. Syst.*, vol. 17, no. 1, pp. 1-24, 2024, doi: 10.2478/ijssis-2024-0023.

[41] Z. Hu, Y. Zhao, and M. Khushi, "A survey of forex and stock price prediction using deep learning," *Appl. Syst. Innov.*, vol. 4, no. 1, pp. 1–30, 2021, doi: 10.3390/ASI4010009.

[42] R. A. Mulla, S. Saini, P. S. Mane, B. W. Balkhande, M. E. Pawar, and K. A. Deshmukh, "A Novel Hybrid Approach for Stock Market Index Forecasting using CNN-LSTM Fusion Model," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 12, pp. 266–279, 2024, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85185306859&partnerID=40&md5=d6cb8fa432122fa4a792e9b2b7c99186

[43] A. Akinjole, O. Shobayo, J. Popoola, O. Okoyeigbo, and B. Ogunleye, "Ensemble-Based Machine Learning Algorithm for Loan Default Risk Prediction," *Mathematics*, vol. 12, no. 21, pp. 1-31, 2024, doi: 10.3390/math12213423.