

A Hybrid Review of Supervised Machine Learning for Banking Transaction Fraud Detection: A Bibliometric and PRISMA-Based Analysis

Nguyen Thi HANG
Thu Dau Mot University, Viet Nam
nguyenthhang@tdmu.edu.vn

This study provides a systematic review of research on the application of supervised machine learning to banking transaction fraud detection, aiming to identify gaps in the existing literature. The study integrates bibliometric analysis using the bibliometrix package in R and VOSviewer with the PRISMA screening method. The dataset comprises 292 Scopus-indexed publications for bibliometric analysis and 171 articles meeting PRISMA criteria for qualitative synthesis. Results reveal significant growth in the field, dominated by decision tree-based models and ensemble techniques. However, four key gaps are identified: (i) research bias toward credit card fraud; (ii) over-reliance on binary classification; (iii) predominant use of traditional SMOTE to address data imbalance; and (iv) limited feature interpretability, particularly regarding the direction of feature effects and behavioral attributes on predicted risk. The study recommends expanding research to debit and mobile banking fraud, developing risk-score models, and evaluating SMOTE variants with concept drift handling to enhance transparency.

Keywords: Banking Transaction Fraud, Bibliometric Analysis, PRISMA, Supervised Machine Learning

DOI: 10.24818/issn14531305/30.3.2026.03

1 Introduction

The rapid expansion of digital banking and electronic payment systems has substantially increased the volume of online transactions in recent years, while also enabling more sophisticated and harder-to-detect fraudulent activities. Transaction fraud in banks not only results in direct financial losses but also undermines risk management effectiveness, regulatory compliance, and the long-term reputation of financial institutions [1].

Parallel to technological advancements, the rise of high-tech crime has transformed the nature and methods of fraud in the banking sector. At the same time, given transaction data characterized by severe class imbalance, where fraud cases represent only a small fraction, combined with constantly evolving user behavior and the need for near real-time processing, traditional rule-based and manual fraud detection methods increasingly show limitations in flexibility and scalability [2]. These approaches struggle to keep pace with rapidly changing fraud patterns [3].

In this context, machine learning, particularly supervised machine learning algorithms, has become a crucial tool for enhancing fraud

detection efficiency. Supervised machine learning techniques enable models to learn from labeled historical data to classify legitimate and fraudulent transactions, contributing to higher prediction accuracy, lower false alarm rates, and greater adaptability to emerging fraud patterns. Numerous studies indicate that algorithms such as Decision Trees, Random Forest, Support Vector Machine, and boosting models are widely applied in financial and banking fraud detection [4] [5].

Although the application of machine learning in financial fraud detection has attracted significant attention, with numerous comprehensive literature reviews [6] [7] [8], most of these studies approach machine learning broadly, without in-depth analysis of specific algorithm groups. In practical bank transaction fraud detection, supervised machine learning models are highly valued for their ability to learn from labeled data, suitability for risk scoring mechanisms, and relative ease of integration into financial institutions' internal control systems. Furthermore, user behavior continuously evolves over time, and fraudulent behavior adapts to evade control systems, causing shifts in data distribution and

predicted risk [9] [10]. However, most previous review studies have primarily focused on evaluating model performance, without fully considering the impact of data dynamics and behavioral changes on model stability and adaptability in real-world contexts [11]. This highlights a research gap concerning the consideration of supervised learning from a dynamic perspective and its relation to an ever-changing risk environment.

Addressing this research gap, this study adopts an integrated approach, combining the PRISMA document screening process and bibliometric analysis to provide a systematic overview of supervised machine learning in bank transaction fraud detection. Bibliometric analysis is employed to identify publication trends, academic networks, and the knowledge structure of the field, while the PRISMA process ensures transparency and rigor in document selection. Building on this, the study reviews supervised machine learning cohorts and synthesizes the role of input features in risk prediction. Through this approach, it aims to clarify development trends and future research directions in bank transaction fraud detection based on supervised machine learning.

2 Literature review

2.1. Supervised Machine Learning in Fraud Detection

Supervised machine learning is a branch of artificial intelligence in which models are trained on labeled historical data to predict or classify new events based on input features. In bank transaction fraud detection, supervised machine learning is essential for classifying transactions as legitimate or fraudulent according to known behavioral patterns. This approach employs algorithms to process large volumes of highly dynamic data and supports real-time analysis, overcoming the limitations of traditional rule-based methods, which are inflexible and easily circumvented by cybercriminals.

Popular supervised machine learning algorithms in fraud detection include Logistic Regression, often used as a baseline model for its simplicity and interpretability; Decision Tree

and Random Forest, valued for their ability to handle nonlinear data and provide relatively high explanatory power; Support Vector Machine, effective in complex feature spaces and capable of managing imbalanced data through kernel techniques; and K-Nearest Neighbors and Naïve Bayes, applied in early-stage empirical studies. Additionally, Ensemble and Boosting methods such as AdaBoost, Gradient Boosting, XGBoost, and LightGBM are increasingly employed to enhance predictive performance by combining multiple weak models into a single strong model. Neural Networks, including deep learning models, further enable the detection of complex fraud patterns through multi-layered hidden structures, particularly when processing large and highly dynamic datasets.

Previous studies have shown that supervised machine learning generally outperforms unsupervised methods when labeled data are available, with metrics such as accuracy, precision, recall, F1-score, and AUC significantly improved. This is particularly relevant for banking transaction data, where low fraud rates and class imbalance are common challenges. Supervised machine learning is employed in fraud prediction due to its ability to adapt dynamically to changes in fraudulent behavior, reduce false alarms, and support financial risk management. Numerous studies have demonstrated that these models not only enhance predictive accuracy but also integrate effectively with digital banking systems, contributing to financial security worldwide [12] [13]. However, challenges such as the need for high-quality data and model interpretability require the use of comprehensive analytical frameworks to properly assess effectiveness [2].

2.2 Systematic Review and Bibliometric Methods

Within the literature review framework, a PRISMA-based approach provides a transparent and systematic method for identifying, screening, and selecting research literature. PRISMA was developed as a reporting standard to improve the transparency and reproducibility of systematic reviews. According to

[14], applying PRISMA ensures that literature search and selection are conducted consistently, in a controlled and verifiable manner, thereby reducing selection bias and enhancing the reliability of review findings.

The PRISMA framework, including flowcharts and reporting criteria, enables readers to trace the entire process from identifying initial data sources to duplicate removal and screening according to research criteria. In addition to standardizing reporting, PRISMA enhances the academic quality of review studies by requiring clear descriptions of search strategies, inclusion and exclusion criteria, and methods for assessing study quality. This is particularly important in rapidly evolving fields such as machine learning and financial fraud detection, where publication volume is growing quickly, and approaches are diverse [15]. Using a systematic framework ensures that selected studies are truly relevant to the research question, rather than being chosen based on convenience or popularity.

To enhance transparency and update methodological standards, the 2020 version of PRISMA was published with several significant adjustments. [16] emphasize that this updated version expands and clarifies reporting criteria, particularly in the context of big data and complex analytical methods. PRISMA 2020 requires more detailed descriptions of data sources, duplicate handling, data extraction procedures, and bias assessment, thereby improving comparability across review studies. More recently, PRISMA-trAIce has been proposed as an extension specifically for studies in artificial intelligence and machine learning. The study [17] argues that AI studies often involve complex data, algorithms, and model training processes, necessitating more detailed reporting on data sources, preprocessing, training-testing splits, and performance evaluation. PRISMA-trAIce provides specific guidance to ensure that AI-focused review studies are transparent not only about literature selection but also about methodological context and technical characteristics of the analyzed models.

As research output grows, quantitative literature reviews offer greater efficiency and

convenience compared to traditional manual methods. In this context, bibliometric analysis serves as a key methodological framework for systematizing and structuring published scientific knowledge. Unlike traditional reviews, which tend to be descriptive, bibliometrics builds on the foundations of scientometrics and informetrics to assess the structure, impact, and dissemination of knowledge through bibliographic data. This approach enables the identification of influential authors, leading institutions, and key countries within the global research network.

In addition to measuring publication productivity and citation impact, bibliometrics analyzes the structure of knowledge networks using techniques such as co-citation, bibliographic coupling, and co-word analysis. Examining co-citations and keyword co-occurrences helps identify research clusters and emerging trends. Regarding tools, software such as VOSviewer and the bibliometrix package in RStudio play a crucial role in facilitating large-scale bibliometric analyses. VOSviewer is designed to optimize network visualization using the Visualization of Similarities algorithm, highlighting the connections between research entities. Bibliometrix provides a comprehensive analytical ecosystem, enabling transparent data processing, index construction, and reproducible result generation within the R environment.

Using quantitative indicators such as publication counts, total citations, network node centrality, and keyword co-occurrence frequency, bibliometric methods help reconstruct a comprehensive picture of the structure and evolution of a research field. This approach not only identifies key topics and influential research groups but also highlights academic gaps and unexplored approaches.

Study [18] emphasizes that integrating bibliometrics with a PRISMA-guided systematic review, as in the Bibliometric Systematic Literature Review framework, provides a robust synthesis method, offering both a comprehensive quantitative view of the knowledge structure and ensuring that literature screening and selection are conducted transparently and systematically.

3 Methodology

3.1 Bibliometric method

This study employs two bibliometric tools, VOSviewer and the bibliometrix package in RStudio, to extract and analyze data. Bibliometric analysis typically relies on a single raw dataset from one platform to ensure data integrity and compatibility across analytical tools. Accordingly, bibliometric data on supervised machine learning in bank transaction fraud detection were retrieved from Scopus using the keywords “supervised machine learning” AND “transaction fraud detection” in the “article title, abstract, and keywords” fields, yielding 292 articles published between 2007 and 2026. Scopus was selected for its extensive coverage, comprehensiveness, and high academic credibility, which aligns with the objectives of bibliometric analysis. Combining data from multiple platforms may introduce normalization challenges and hinder network visualization; therefore, only Scopus data were used for the bibliometric analysis. Articles from Web of Science and Google Scholar will be collected separately and screened using a PRISMA-based approach for content analysis and systematic review.

VOSviewer is employed to visualize collaborative networks among authors, institutions, and countries, and to identify prominent research themes through keyword co-citation and co-occurrence analysis. Its similarity-based node placement algorithm enables the grouping of related elements into thematic clusters, producing a visual map of the knowledge structure in supervised machine learning for bank transaction fraud detection (Van Eck & Waltman, 2010). This tool is especially effective for illustrating international collaboration patterns and comparing research activity across geographic regions.

The RStudio environment with the bibliometrix package is employed to conduct comprehensive analyses of knowledge structure and research performance. Bibliometrix enables the assessment of publication counts, citations, influential authors, and organizations, while constructing conceptual, social, and intellectual maps of the field [19]. This approach not only depicts the current knowledge

structure but also provides insights into the evolution of research streams, supporting predictions of future trends in supervised machine learning for bank transaction fraud detection. An additional strength is the application of Bradford's Law, which identifies core journals with significant contributions, highlighting key knowledge sources and ensuring analytical rigor.

The combined use of VOSviewer and bibliometrix allows for a comprehensive perspective, integrating network visualization and knowledge structure analysis to capture both the current state and the developmental trajectory of supervised machine learning in bank transaction fraud detection.

3.2 PRISMA-Guided Screening Process

Phase 1. Identification

In addition to the 292 records retrieved from the Scopus database for bibliometric analysis, additional studies were identified from Web of Science and Google Scholar to support the PRISMA-based systematic review. The search strategy employed keywords such as “supervised machine learning,” “transaction fraud detection,” and “digital banking,” combined with specific algorithm-related terms including Random Forest, Support Vector Machine, XGBoost, Neural Networks, Logistic Regression, and Decision Tree. This approach was designed to ensure comprehensive coverage of supervised learning techniques applied to detecting digital banking transaction fraud. The initial search yielded 292 records from Scopus and 177 records from Web of Science and Google Scholar, resulting in a total of 469 identified records before duplicate removal.

Phase 2. Screening

All 469 records were imported into EndNote for duplicate removal, using a function based on title, author, publication year, and DOI. This process eliminated 116 duplicates, leaving 353 records for title and abstract screening. Applying the inclusion and exclusion criteria, 134 records were excluded - 86 for not being scientific articles and 48 for not being

published in English, resulting in 219 records advancing to the full-text review stage.

Phase 3. Eligibility

The full texts of the 219 retained records were assessed in detail to determine their eligibility and alignment with the research objectives. During this eligibility phase, 48 studies were excluded for the following reasons: 8 were review articles, 24 mentioned supervised machine learning but did not specifically address transaction fraud in the banking context, and 16 were qualitative in nature, lacked empirical data, or did not report model performance

metrics. After applying these criteria, 171 studies met all inclusion requirements and were included in the final qualitative synthesis.

Phase 4. Included

A total of 171 studies were included in the final systematic synthesis and subjected to qualitative content analysis. This analysis aimed to examine methodological trends, the distribution of supervised learning algorithms, and the characteristics of datasets employed in digital banking transaction fraud detection.

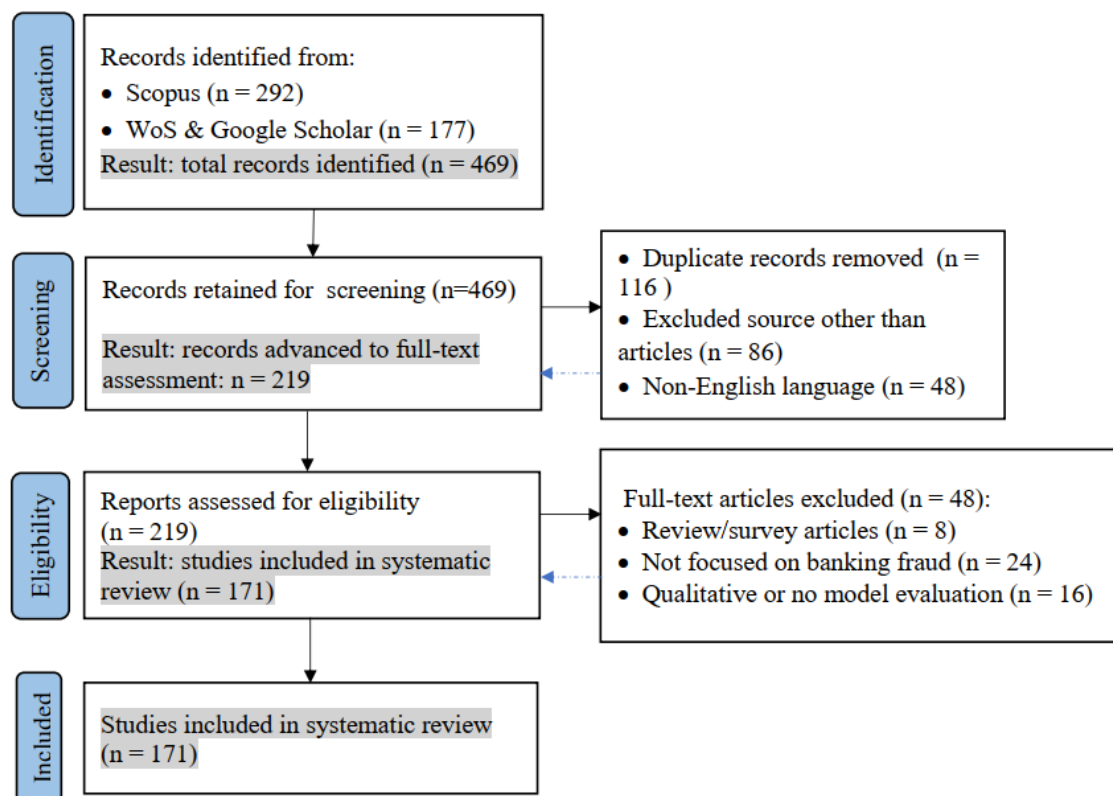


Fig. 1. PRISMA Flow Diagram of Study Selection Process

4 Findings and Discussion

4.1. Bibliometric analysis

The bibliometric dataset comprises 292 peer-reviewed publications focusing on the application of supervised machine learning techniques in bank transaction fraud detection over the period 2007-2026. General information is summarized in Table 1.

The 216 publications analyzed reveal a clear interdisciplinary link between finance, computer science, and information systems. The

average annual growth rate of 10.78% indicates stable and increasing research output, coinciding with the rise of digital banking and online transaction systems. The literature's average age of 3.35 years reflects its recency, aligned with the rapid evolution of supervised learning algorithms such as Random Forest, SVM, and XGBoost. With an average of 13.72 citations per article and 2,064 references, the field demonstrates a solid academic foundation and notable influence.

Table 1. Data Description

Description	Results
MAIN INFORMATION ABOUT DATA	
Timespan	2007:2026
Sources (Journals, Books, etc)	216
Documents	292
Annual Growth Rate %	10.78
Document Average Age	3.35
Average citations per doc	13.72
References	2064
DOCUMENT CONTENTS	
Keywords Plus (ID)	1120
Author's Keywords (DE)	647
AUTHORS	
Authors	1007
Authors of single-authored docs	16
AUTHORS COLLABORATION	
Single-authored docs	16
Co-Authors per Doc	3.69
International co-authorships %	17.81
DOCUMENT TYPES	
article	72
book chapter	16
conference paper	190
conference review	6
erratum	1
review	7

The dataset includes 1,120 Keywords Plus and 647 keywords, illustrating diverse methodological approaches, from traditional classification to imbalanced data handling and anomaly detection. A total of 1,007 authors contributed, averaging 3.69 co-authors per paper, with 16 single-author articles; international collaboration accounts for 17.81%. Conference papers dominate (190/292) over journal articles (72) and other formats, reflecting the technological focus and rapid publication cycle. These descriptive indicators highlight a fast-growing, highly current, and interdisciplinary field, providing the basis for subsequent analyses knowledge structure and research networks.

The annual publication distribution shows three distinct development phases. During

2007–2016, research activity was limited and sporadic, with a maximum of 3 articles in 2014 and most years having 0 - 1 publication, reflecting an emerging field relying on traditional statistical methods or basic machine learning algorithms such as SVM and Decision Trees, in the context of large-scale banking transaction data and early-stage AI adoption.

From 2017, the field entered an acceleration phase with 8 articles, followed by 11 and 15 publications in 2018 - 2019, representing an annual growth of approximately 36 - 37%. A slight decline to 12 articles occurred in 2020, likely due to the COVID-19 pandemic affecting experimental research and data collection.

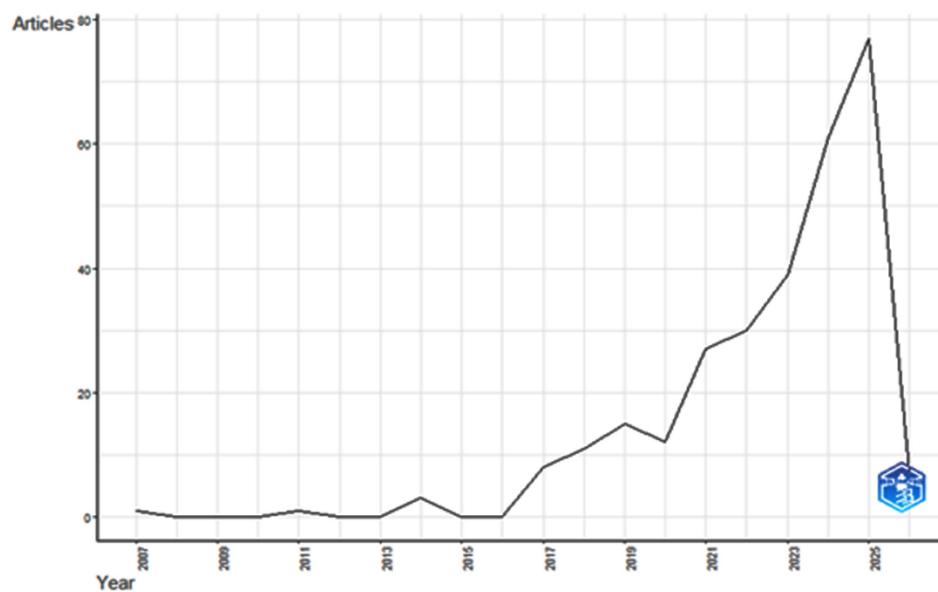


Fig. 2. Chart of Annual Scientific Production

Between 2021 and 2025, publications surged, reaching 27 in 2021 (125% growth), 30 in 2022, 39 in 2023, 61 in 2024, and peaking at 77 in 2025. This phase corresponds with the widespread adoption of ensemble methods such as Random Forest and XGBoost, deep learning models, and imbalanced data

handling techniques for fraud detection, reflecting both technological advances and rising practical demand amid the expansion of digital banking and online fraud. The 2026 count of 7 articles (up to February) does not yet capture the full-year trend.

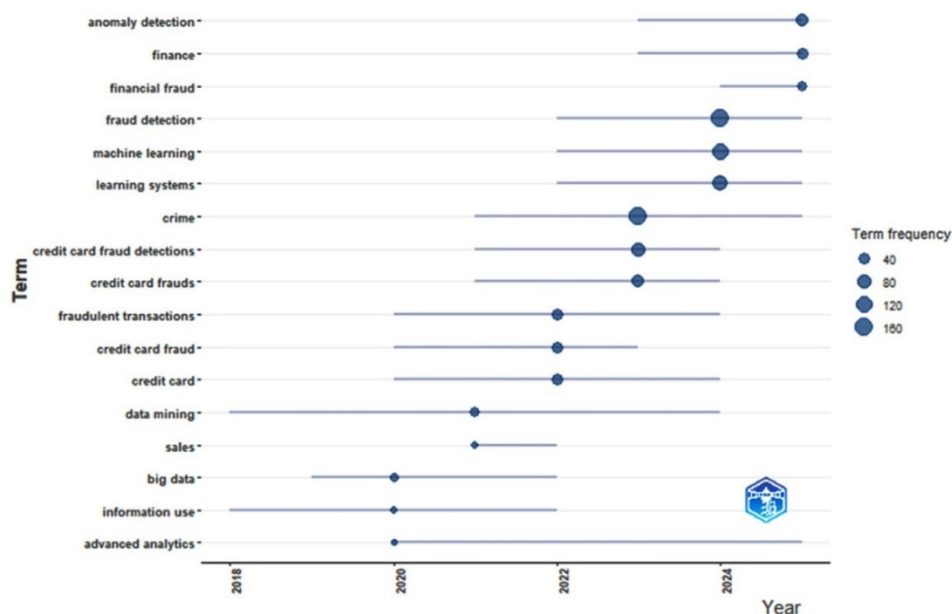


Fig. 3. Most Relevant Sources

Although the overall average growth rate is 10.78%, the 2021–2025 period exhibits exceptional annual increases of 30–50%, confirming that supervised machine learning

research in bank transaction fraud detection is in a strong growth and maturing phase. This trend also opens new research directions, including model optimization, imbalanced data

handling, and model interpretability in real-world applications.

Figure 3 shows that conference proceedings and book series dominate, reflecting the AI/ML field's preference for rapid dissemination through international conferences on artificial intelligence, data science, and system security rather than the lengthy journal peer-review process. Lecture Notes in Networks and Systems leads with 23 papers, highlighting the role of the Springer series in spreading research on intelligent systems and AI applications in finance. Other sources, including Communications in Computer and Information Science, Advances in Intelligent Systems and Computing, Lecture Notes in

Computer Science, and IEEE and ACM proceedings, confirm the focus on computer science and engineering venues. Prestigious journals like IEEE Access and Expert Systems with Applications appear less frequently, indicating that while the field is gradually maturing academically, most research prioritizes conferences for rapid technological updates and applied relevance. Many conferences scheduled for 2024–2025 feature 1–2 papers per source, aligning with the recent surge in publications and reflecting the urgent need to share findings on ensemble learning, deep learning, and imbalanced data handling amid rising digital transaction fraud.

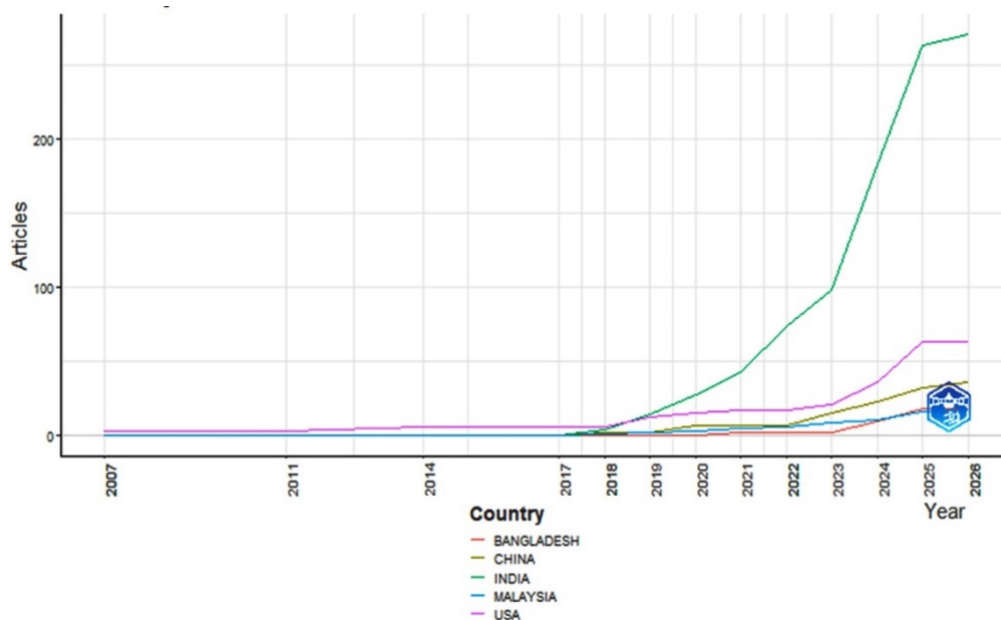


Fig. 4. Chart of Country Production Over Time

Figure 4 presents the annual publication output by country, while Figure 5 illustrates the structure of international research collaboration in studies on supervised machine learning applications for detecting bank transaction fraud. The findings reveal an uneven distribution of research output, with a clear concentration in a limited number of leading countries over the past decade. India emerges as the most productive country, demonstrating strong growth since 2018 and particularly rapid expansion during the period from 2021 to 2025. The United States maintains a

prominent position, with stable and steadily increasing output over time. China entered the field later but has shown notable growth in recent years. Several other countries contribute a smaller volume of publications, yet they also display an upward trend beginning around 2020.

Overall, the publication structure reflects the rapid expansion of the field, particularly in countries that place strong emphasis on artificial intelligence and financial technology development.



Fig. 5. Co-author Country Network

The international collaboration network indicates that India represents the most prominent central node, maintaining numerous connections with other countries. The United States and China also occupy important positions within the overall network structure. Nevertheless, the overall rate of international collaboration remains relatively modest at 17.81%, suggesting that a large proportion of research activity is still conducted at the domestic or

regional level. Collaborative clusters are concentrated around a few major hubs, whereas many countries appear on the periphery with limited cross-national linkages. This pattern reflects rapid expansion in research productivity but limited depth of international integration. At the same time, it highlights considerable potential for strengthening transnational research networks in the future.

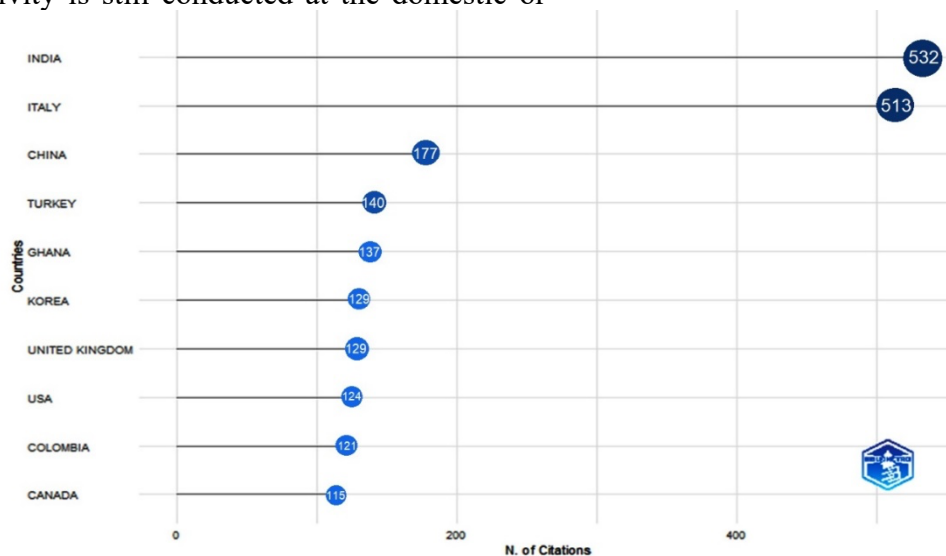


Fig. 6. Most Cited Countries

Figure 6 illustrates the concentration of academic influence by country in applying supervised machine learning to bank transaction fraud detection. India leads with 532 citations, followed by Italy with 513, forming a central, high-impact group. China ranks third with 177 citations, while Turkey, Ghana, South Korea, the United Kingdom, the United States, Colombia, and Canada show markedly lower citations. This right-skewed distribution reflects

a core-periphery structure consistent with Bradford's law. The prominence of India and Italy over traditional research powers such as the United States and China indicates that in niche domains, academic influence depends not only on national research scale but also on practical relevance, engagement with real-world applications, and publication quality.

Learning, and Crime have large nodes and high connectivity, confirming their core role. The green cluster focuses on credit card fraud and electronic transactions, with keywords like Credit Card Fraud, Fraudulent Transactions, Decision Trees, Random Forests, and Supervised Machine Learning, indicating the continued use of traditional supervised models in practical classification tasks. The blue cluster covers technical and algorithmic research, including Logistic Regression, Machine Learning Techniques, SMOTE, and Classification, reflecting efforts to improve

model performance with rare fraud data. The red cluster highlights interdisciplinary and advanced approaches, with keywords such as Anomaly Detection, Artificial Intelligence, Network Security, and Financial Fraud, showing the increasingly blurred boundaries between supervised learning and other methods. To conclude, the keyword map indicates that the field is evolving toward integrating classical supervised learning models with modern approaches to address the growing complexity of the digital financial environment.

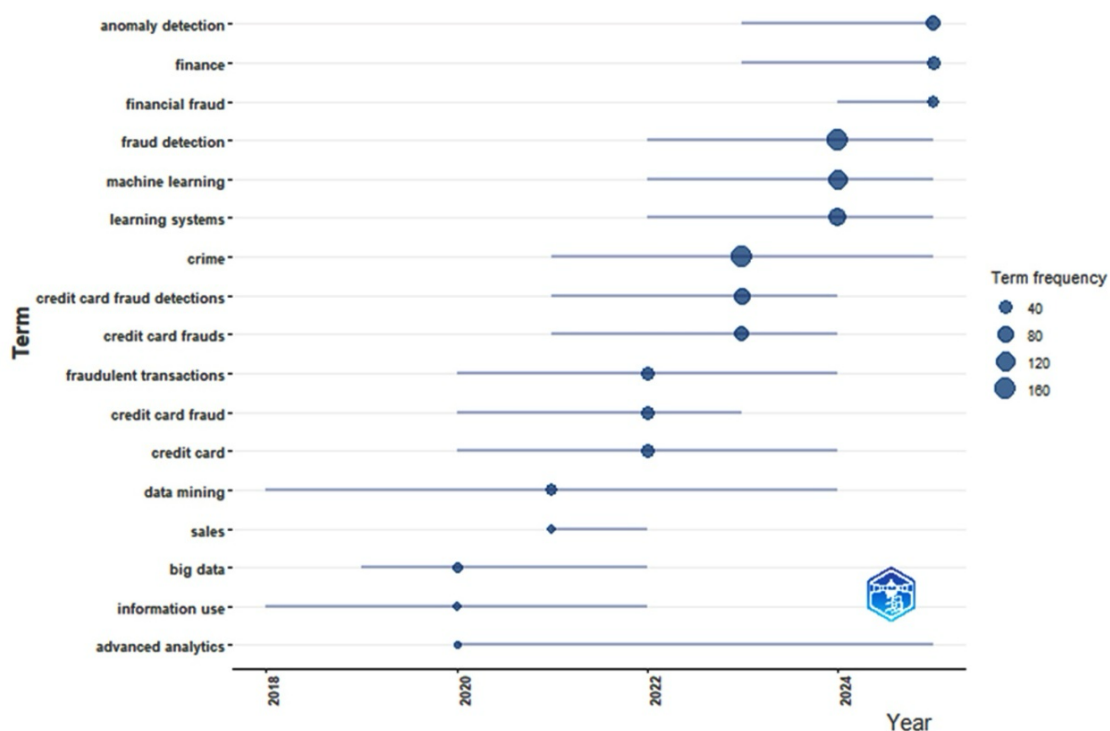


Fig. 9. Evolving Research Trends

Figure 9 shows a structural shift in the field from Big Data and Data Mining toward Supervised Machine Learning (SML) as the central approach during 2018–2024. In 2018–2020, research focused on large-scale data collection and analysis, reflecting infrastructure and data processing development. From 2020 onward, topics directly related to Machine Learning and Fraud Detection increased significantly and persisted through 2024, establishing SML as the dominant methodology. Keywords such as Fraud Detection, Credit Card Fraud, and Fraudulent Transactions are

closely linked to supervised classification models, confirming SML's central role in building transaction fraud detection systems. The emergence and persistence of Anomaly Detection and Learning Systems further reflect integration trends and model optimization toward adaptive, intelligent systems. Overall, the Trend Topics chart illustrates the evolution of the field, positioning Supervised Machine Learning not only as a technical tool but also as a methodological pillar in modern bank fraud detection research and applications.

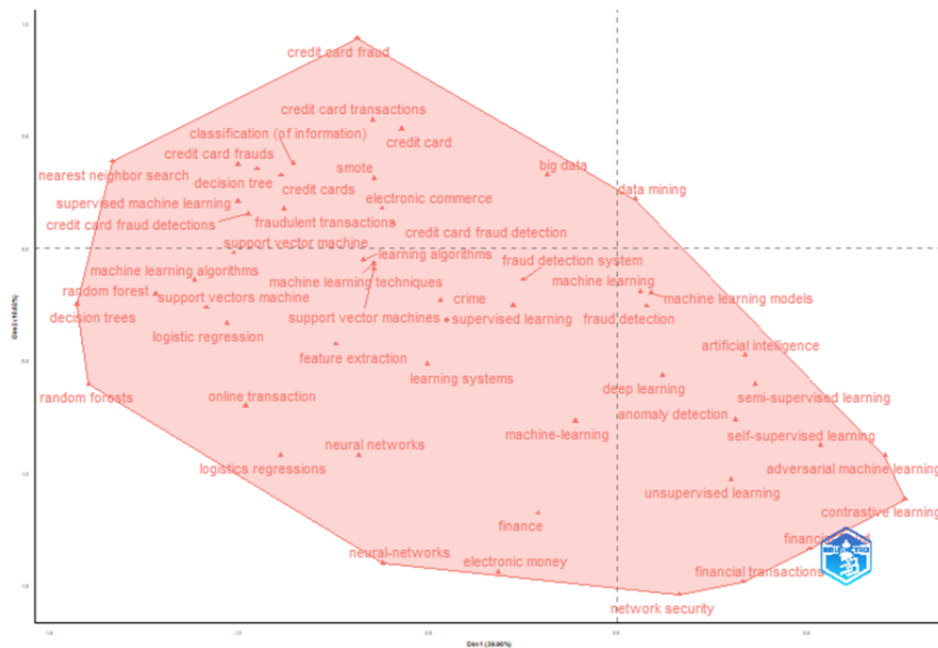


Fig. 10. Conceptual Structure Map – Method MCA

From Figure 10, the field of supervised machine learning in bank transaction fraud detection is organized around a clear and stable methodological core. Central and upper areas of the map cluster concepts such as Credit Card Fraud, Credit Card Fraud Detection, Fraud Detection, Supervised Machine Learning, Machine Learning, Random Forest, Decision Tree, Support Vector Machine, Logistic Regression, and SMOTE, reflecting the dominance and standardization of supervised classification models in credit card fraud detection.

Linking terms such as Fraudulent Transactions, Credit Card Transactions, Online Transaction, Electronic Commerce, Feature Extraction, Learning Algorithms, and Machine Learning Models connect methodology with application context, particularly addressing data imbalance via SMOTE. Middle and lower regions show Neural Networks, Deep Learning, Anomaly Detection, Unsupervised Learning, Semi-Supervised Learning, and Adversarial Machine Learning. These topics are dispersed and distant from the supervised learning core, representing emerging or expanding directions with potential but not yet

forming a central cluster. In the lower-right area, Big Data, Data Mining, Artificial Intelligence, Finance, Financial Transactions, and Network Security reflect technological foundations and broader application contexts, supporting but not forming the main methodological axis.

In conclusion, the MCA map indicates a knowledge structure centered on traditional supervised learning, with Random Forest, Support Vector Machine, Decision Tree, and Logistic Regression at the core, while Deep Learning and semi-/unsupervised methods appear as potential growth branches without replacing the supervised learning core.

4.2 Content Analysis of Selected Studies

Following the PRISMA screening, 171 studies were selected for content analysis. Literature was systematized using five core criteria: fraud type, machine learning model, performance evaluation, data imbalance handling, and feature analysis. These criteria capture methodological, data-related, and evaluative developments in the field and highlight prevailing research trends. Results are summarized in Table 2.

Table 2. Key Trends and Findings from the Selected Studies

No.	Criteria	Research Focus	Key Findings
1	Type of Fraud	Credit card fraud; Debit card fraud; Mobile banking fraud	Credit card fraud dominates; debit card and mobile banking fraud are underrepresented; fraud contexts are rarely analyzed separately
2	Supervised Machine Learning Models	Random Forest; Decision Tree; XGBoost; Logistic Regression; SVM; Neural Networks	Tree-based and ensemble models are core; classical models are benchmarks; neural networks are limited; advanced ensembles are increasingly applied
3	Model Performance Evaluation	AUC-ROC; F1-score; Recall; Accuracy	Metrics suitable for imbalanced data prioritized; minority-class detection emphasized; accuracy alone is insufficient
4	Imbalanced Data Handling	SMOTE; Cost-sensitive learning; Ensemble approaches	Oversampling improves model learning; cost-sensitive applied selectively; imbalance is recognized as a structural challenge
5	Feature Analysis	Tree-based feature importance; SHAP; LIME	Key features: transaction amount, timing, frequency, merchant info; systematic impact analysis and explainable AI limited

Regarding fraud type, most studies focus on credit card fraud, with some representative studies in 2021 - 2025 such as [22] [23] [24] [25], due to the availability and popularity of credit card datasets. Debit card fraud appears limited and is usually merged in general studies, while online money transfer fraud from mobile banking remains rare [1] [3] [26]. This creates bias in the experimental base, implying that detection models may need adjustment for specific risk behaviors of each fraud type.

Regarding supervised machine learning models, decision tree-based and ensemble techniques such as Random Forest, Gradient Boosting, and XGBoost occupy the central position because of their ability to handle non-linear relationships, complex interactions among variables, and stability under noisy and imbalanced data [12] [23] [27] [28]. Traditional models such as Logistic Regression and Support Vector Machines are often used as baselines for comparison, while Neural Networks appear more frequently in recent

studies, reflecting deeper exploration of data structures [29] [30]. Most studies still focus on binary classification, while risk-score or continuous probability approaches are underexplored.

For model performance evaluation, Area Under the Receiver Operating Characteristic curve, F1-score, and Recall are most used, emphasizing attention to the minority class and prioritizing fraud detection ability rather than overall accuracy [31] [32] [33] [34]. Accuracy is less used alone because it can be dominated by the non-fraud class, giving a misleading evaluation of real detection ability [35] [36]. Some studies discuss probability calibration and imbalance handling to improve risk estimation [9].

Data imbalance handling is a core methodological challenge [37]. Synthetic Minority Over-sampling Technique is widely applied to increase minority class observations and improve model learning [33] [38]. Some studies combine cost-sensitive learning or ensemble methods like boosting to adjust class weights

and reduce prediction bias [35]. These approaches improve performance on static data, but reliance on synthetic data raises representativeness and generalization concerns in real-time environments, where data distribution may change continuously [36].

For feature analysis, most studies use feature importance from tree-based models like Random Forest, Gradient Boosting, or XGBoost to rank input variable impact, such as transaction amount, transaction time, transaction frequency, and merchant information [39]. Recent studies propose frameworks integrating supervised learning and explanation methods like SHAP or LIME to evaluate each feature's impact on individual transactions [40] [41]. However, most studies still mainly identify relative importance, without systematically analyzing the effect direction on risk. SHAP provides a consistent explanation at global and local levels [42], but full integration in real banking environments remains limited.

Conclude, Table 2 shows that supervised machine learning in bank transaction fraud detection has achieved significant growth in algorithms and performance evaluation. However, high focus on credit card fraud, binary classification tendency, and limited feature impact analysis and risk-score modeling still create important gaps, forming the basis for deeper research directions in the future.

4.3 Research Gaps and Future Research Directions

Based on the bibliometric analysis and content analysis of 171 studies selected through the PRISMA process, it can be confirmed that the field of bank transaction fraud detection using supervised machine learning has reached a considerable level of maturity. However, the synthesized results also reveal several research gaps in the existing body of literature, including:

Bias in fraud types: Previous studies mainly focus on credit card fraud, mostly due to the availability of public datasets. In contrast, debit card fraud and mobile banking transaction fraud appear less frequently and are often merged in general analyses, while debit card and mobile banking fraud have distinct

characteristics that have not been systematically studied.

Limitations of the binary classification approach: Most prior studies treat the problem as a binary classification task, while in banking practice, risk management usually relies on risk ranking or continuous fraud probability. The lack of studies on risk-score modeling and calibration indicates a methodological gap in the current literature.

Dependence on traditional SMOTE: In published studies, SMOTE and cost-sensitive learning are the most common approaches for handling class imbalance, but advanced variants have not been systematically compared. Moreover, concept drift in real-time operational environments has not been fully integrated into current imbalance handling studies.

Limitations in feature analysis: Most studies only identify relative importance through feature importance from tree-based models and rarely analyze the effect of each feature on fraud probability. Current feature sets focus on traditional factors such as transaction amount, time, and frequency, and do not fully exploit multi-channel behavior such as device changes, authentication methods, or geographic location.

Based on these identified gaps, future research should shift from pure algorithm optimization to a more systematic approach, closely linked to real-world digital banking operations:

First, expand fraud types: Future studies should include various types of fraud beyond credit card, to improve coverage and generalizability. Diversifying data contexts will enhance practical value and more accurately reflect multi-channel risk structures in modern banking.

Second, shift to continuous risk modeling: Move from binary classification to continuous risk probability modeling, focusing on risk stratification and evaluating risk-score stability over time. Integrating risk-based decision frameworks will help bridge the gap between academic research and practical implementation.

Third, advanced imbalance handling: Future research should systematically test advanced

variants like Borderline-SMOTE, ADASYN, or combined oversampling and under sampling methods. Imbalance handling should be applied in dynamic data contexts, where concept drift can change fraud behavior distribution over time, and integrate drift detection and adaptation in the model pipeline.

Finally, systematic feature analysis: Future studies should expand feature sets to include multi-channel and multi-device behaviors and systematically analyze the effect of each feature on fraud probability. Combining advanced model explanation techniques with a risk evaluation framework can improve transparency, interpretability, and regulatory compliance in banking systems.

5 Conclusion

This study provides a comprehensive overview of supervised machine learning in bank transaction fraud detection using a two-tier approach: bibliometric analysis and systematic content analysis. Specifically, 292 documents extracted from the Scopus database were analyzed using R and VOSviewer to identify publication trends, academic collaboration structures, co-citation networks, and main keyword clusters. At the same time, 171 studies selected through the PRISMA process were analyzed in-depth to summarize methods, data processing techniques, and modeling approaches. Bibliometric results show a strong increase in publications from 2021, with research clusters focused on supervised learning, ensemble models, class imbalance, and explainable AI. Keyword and co-citation networks indicate a shift from traditional models to ensemble and deep learning, with increased attention to explainability and practical applications. Tree-based and ensemble models such as Random Forest, Gradient Boosting, and XGBoost dominate due to their ability to handle non-linear and imbalanced data. SMOTE and cost-sensitive learning are widely used to address class imbalance. However, most studies still focus on credit card fraud, binary classification, and optimization of statistical metrics without integrating economic objective functions. Feature analysis mainly identifies relative importance, while

the effect direction and feature design by transaction type are not fully explored. By combining bibliometric and content analysis, this study not only systematizes the knowledge structure of the field but also highlights methodological and empirical gaps. These findings provide a foundation for future research directions, including continuous risk-scoring models, systematic testing of advanced imbalance handling techniques, expansion of multi-channel behavioral feature sets, and integration of economic cost-based evaluation in real-time operational contexts. In conclusion, supervised machine learning has become a pillar in bank transaction fraud detection research. However, to improve adaptability, transparency, and practical implementation value, future studies need to focus on building comprehensive, dynamic fraud detection systems closely aligned with modern digital banking risk management.

References

- [1] George, M. Z. H., Alam, M. K., & Hasan, M. T. (2023). Machine learning for fraud detection in digital banking: A systematic literature review. *ASRC Procedia: Global Perspectives in Science and Scholarship*, 3(1), 37–61.
- [2] Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: A review of anomaly detection techniques and recent advances. *Expert Systems with Applications*, 193, 116429.
- [3] Ali, A., Razak, S. A., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial fraud detection based on machine learning: A systematic literature review. *Applied Sciences*, 12(19), 9637.
- [4] Yanto, Y., Lisah, L., & Tandra, R. (2024). The best machine learning model for fraud detection in banking sector: A systematic literature review. *ECo-Buss*, 7(2), 1361–1384.
- [5] Compagnino, A. A., Maruccia, Y., Cavuoti, S., Riccio, G., Tutone, A., Crupi, R., & Pagliaro, A. (2025). An introduction to machine learning methods for fraud detection. *Applied Sciences*, 15(21), 11787.

- [6] Hernandez Aros, L., Bustamante Molano, L. X., Gutierrez-Portela, F., Moreno Hernandez, J. J., & Rodríguez Barrero, M. S. (2024). Financial fraud detection through the application of machine learning techniques: A literature review. *Humanities and Social Sciences Communications*, 11, Article 1130.
- [7] Thakkar, H., Datta, S., Bhadra, P., Barot, H., & Jadav, J. (2025). Artificial intelligence and machine learning in fraud detection: A comprehensive bibliometric mapping of research trends and directions. *Annals of Library and Information Studies*, 72(2), 138–150.
- [8] Naqvi, S. H. (2026). Fraud detection using machine learning in financial transactions: A systematic review. *Veredas do Direito*, 23(2), e234187.
- [9] Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2015). Credit card fraud detection and concept-drift adaptation with delayed supervised information. In *2015 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). IEEE.
- [10] Soemers, D., Brys, T., Driessens, K., Winands, M., & Nowé, A. (2018). Adapting to concept drift in credit card transaction data streams using contextual bandits and decision trees. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 32, No. 1).
- [11] Suárez-Cetrulo, A. L., Quintana, D., & Cervantes, A. (2023). A survey on machine learning for recurring concept drifting data streams. *Expert Systems with Applications*, 213, 118934.
- [12] Afriyie, J. K., Tawiah, K., Pels, W. A., Addai-Henne, S., Dwamena, H. A., Owiredo, E. O., ... & Eshun, J. (2023). A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. *Decision Analytics Journal*, 6, 100163.
- [13] Gupta, S., & Mehta, S. K. (2024). Data mining-based financial statement fraud detection: Systematic literature review and meta-analysis to estimate data sample mapping of fraudulent companies against non-fraudulent companies. *Global Business Review*, 25(5), 1290–1313.
- [14] Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery*, 8(5), 336–341.
- [15] Yuhertiana, I., & Amin, A. H. (2024). Artificial intelligence driven approaches for financial fraud detection: A systematic literature review. *KnE Social Sciences*, 448–468.
- [16] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372.
- [17] Holst, D., Moenck, K., Koch, J., Schmedemann, O., & Schüppstuhl, T. (2025). Transparent reporting of AI in systematic literature reviews: Development of the PRISMA-trAIce checklist. *JMIR AI*, 4, e80247.
- [18] Marzi, G., Balzano, M., Caputo, A., & Pellegrini, M. M. (2025). Guidelines for bibliometric-systematic literature reviews: 10 steps to combine analysis, synthesis and theory development. *International Journal of Management Reviews*, 27(1), 81–103.
- [19] Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975.
- [20] Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences*, 479, 448–455.
- [21] Trivedi, N. K., Simaiya, S., Lilhore, U. K., & Sharma, S. K. (2020). An efficient credit card fraud detection model based on machine learning methods. *International Journal of Advanced Science and Technology*, 29(5), 3414–3424.
- [22] More, R., Awati, C., Shirgave, S., Deshmukh, R., & Patil, S. (2021). Credit

- card fraud detection using supervised learning approach. *International Journal of Scientific & Technology Research*, 9(10), 216–219.
- [23] Sopiyan, M., Fauziah, F., & Wijaya, Y. F. (2022). Fraud detection using random forest classifier, logistic regression, and gradient boosting classifier algorithms on credit cards. *JUITA: Jurnal Informatika*, 10(1), 77–87.
- [24] Talukder, M. A., Hossen, R., Uddin, M. A., Uddin, M. N., & Acharjee, U. K. (2024). Securing transactions: A hybrid dependable ensemble machine learning model using iht-lr and grid search. *Cybersecurity*, 7(1), 32.
- [25] Sun, J. (2025). Decision tree-based credit card fraud detection system: Design and optimization. *Economics & Management Information*, 4(4), 1–5. <https://doi.org/10.62836/emi.v4i4.508>
- [26] Pratama, S. F., & Wahid, A. M. A. (2025). Fraudulent transaction detection in online systems using random forest and gradient boosting. *Journal of Cyber Law*, 1(1), 88–115.
- [27] Nijwala, D. S., Maurya, S., Thapliyal, M. P., & Verma, R. (2023). Extreme gradient boost classifier based credit card fraud detection model. In *2023 International Conference on Device Intelligence, Computing and Communication Technologies (DICCT)* (pp. 500–504). IEEE.
- [28] Yu, S., Chang, V., Huynh, G. L., Jesus, V., & Luo, J. (2025). Advanced supervised machine learning algorithms in credit card fraud detection. In *Proceedings of the 7th International Conference on Finance, Economics, Management and IT Business (FEMIB 2025)* (pp. 126–138). SciTePress.
- [29] Reddy, S. T. S., & Sriramy, P. (2023). Comparison of the support vector classifier algorithm with the decision tree algorithm for credit card fraud detection with the goal of improving accuracy. *J. Surv. Fish. Sci*, 10, 2304–2313.
- [30] Wang, Y. (2025). A data balancing and ensemble learning approach for credit card fraud detection. In *2025 4th International Symposium on Computer Applications and Information Technology (ISCAIT)* (pp. 386–390). IEEE.
- [31] Dhankhad, S., Mohammed, E., & Far, B. (2018). Supervised machine learning algorithms for credit card fraudulent transaction detection: A comparative study. In *2018 IEEE International Conference on Information Reuse and Integration (IRI)* (pp. 122–125). IEEE.
- [32] Sailusha, R., Gnaneswar, V., Ramesh, R., & Rao, G. R. (2020). Credit card fraud detection using machine learning. In *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1264–1270). IEEE.
- [33] Sinap, V. (2024). Comparative analysis of machine learning techniques for credit card fraud detection: Dealing with imbalanced datasets. *Turkish Journal of Engineering*, 8(2), 196–208.
- [34] Btoush, E. A. L. M., Zhou, X., Gururajan, R., Chan, K. C., & Alsodi, O. (2025). Enhancing credit card fraud detection with a stacking-based hybrid machine learning approach. *PeerJ Computer Science*, 11, e3007.
- [35] Randhawa, K., Loo, C. K., Seera, M., Lim, C. P., & Nandi, A. K. (2018). Credit card fraud detection using AdaBoost and majority voting. *IEEE Access*, 6, 14277–14284.
- [36] Baisholan, N., Dietz, J. E., Gnatyuk, S., Turdalyuly, M., Matson, E. T., & Baisholanova, K. (2025). FraudX AI: An interpretable machine learning framework for credit card fraud detection on imbalanced datasets. *Computers*, 14(4), 120.
- [37] He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284.
- [38] Albalawi, T., & Dardouri, S. (2025). Enhancing credit card fraud detection using traditional and deep learning models with class imbalance mitigation. *Frontiers in Artificial Intelligence*, 8, Article 1643292.
- [39] Carcillo, F., Le Borgne, Y. A., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2021). Combining unsupervised and

- supervised learning in credit card fraud detection. *Information Sciences*, 557, 317–331.
- [40] Keerthana, C. S., Nalluri, S. C., Muskaan, S., & Sadagopan, P. (2025, February). Explainable AI in Credit Card Fraud Detection: SHAP and LIME for Machine Learning Models. In *2025 10th International Conference on Signal Processing and Communication (ICSC)* (pp. 387-392). IEEE. DOI: 10.1109/ICSC64553.2025.10968935
- [41] Ojo, I. P. (2025). Explainable AI for credit card fraud detection: Bridging the gap between accuracy and interpretability. *World Journal of Advanced Research and Reviews*, 25(2), 1246–1256. <https://doi.org/10.30574/wjarr.2025.25.2.0492>
- [42] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30. Curran Associates, Inc.



Nguyen Thi HANG graduated with a master's degree in finance and Banking in 2022 and a master's degree in English Linguistics in 2025. With over 10 years of experience in the education sector, she is currently a full-time lecturer at Thu Dau Mot University, Vietnam. In addition, she is pursuing a PhD in Finance and Banking. Her research interests focus on the application of Artificial Intelligence and Machine Learning in economics, finance, and banking, with an emphasis on data-driven decision-making and intelligent financial systems.

tems.