Mapping Machine Learning Research Networks in Financial Risk Prediction

Cosmin COJOCARU*
*National University of Science and Technology Politehnica Bucharest, Bucharest, Romania *Corresponding author, cojocaru.cosmin@gmail.com*

Sorin IONESCU
National University of Science and Technology Politehnica Bucharest, Bucharest, Romania
sc.ionescu@gmail.com

Abstract. *Financial risk forecasting is a more and more relevant topic in contemporary financial systems, since the correct anticipation of credit defaults, bankruptcy, and systemic vulnerabilities is essential for sound decision-making. Classical statistical models have been the cornerstone of risk prediction for decades; however, they are likely to fail in capturing the complex and nonlinear relationships of financial data. The latest developments in machine learning brought forth new methods with increased predictability and response to massive, disparate data. The existing science literature best illustrates a shift away from risk models towards evidence-based approaches involving artificial intelligence, but the subtle understanding of the intellectual process engaged is still poorly investigated. This work fills this void through the use of a network-oriented bibliometric methodology integrating co-citation analysis, bibliographic coupling, and author keyword co-occurrence analysis. The research explores root works, reveals upcoming thematic clusters, and analyzes shifting paradigms affecting financial risk prediction methods. Major findings indicate a systematic intellectual environment where classical models remain the gold standard, and ensemble learning and deep learning techniques are being applied increasingly to enhance risk estimation. Regulatory requirements as well as model transparency also gain increasing attention. These results hold deep implications for both theoretical and practical applications because they provide fresh insights into the dynamic between old and new methods and thus open doors to scalable and interpretable models of risk evaluation.*

Keywords: financial risk prediction, machine learning, bibliometric analysis, ensemble learning, deep learning, explainability

Introduction

Predictive financial risk allows us to forecast credit defaults, bankruptcy, and systemic risk to enable financial institutions, regulators, and investors to reduce losses. Although the traditional techniques like logistic regression have seen extensive usage, these are bad at identifying subtle, non-linear financial trends. With markets today more dynamic and data-driven, the emergence of digital banks, fintech, and alternative credit assessment technologies has necessitated better predictive models.

As a result, machine learning (ML) methods have much enhanced financial risk forecasting compared to conventional approaches. Ensemble learning methodology like Random Forests and Gradient Boosting has attained prominence in credit risk modeling, while deep learning methods continue to advance automatic financial decision-making. Use of ML in hybrid models, with structured and unstructured data, has further enhanced predictive capability. Still, with such development, there exist problems with interpretability of the models, adherence to regulation, and ethical explainability.

DOI: 10.2478/picbe-2025-0099

As increased financial risk investigation is driven by ML, intellectual and collaborative communities must be understood. Bibliometric network analysis yields a systematic manner to depict them as relations in terms of co-citation, bibliographic coupling, and keyword co-occurrence analysis. In contrast with citation-based analysis that emphasizes metrication, network-based bibliometric methods emphasize that influential papers, authors, and institutions interact and influence one another to determine what research to undertake. Through citation patterns and thematic clusters, researchers can monitor seminal studies, detect emerging trends, and gauge the evolution of financial risk modeling.

This research rigorously maps ML-based financial risk prediction research networks, determining landmark works, theme-bearing clusters, and impactful academic contributions. This systematic mapping captures a snapshot of current knowledge, spotlights areas for future research, and promotes industry-academia convergence. By assessing the manner in which risk models generated by ML meet regulatory requirements and actual financial concerns, this research consolidates theory with practice.

With changing financial markets and regulatory scrutiny increasing, the implementation of ML models for risk analysis continues to be hindered by issues of interpretability, compliance, and validation protocols. Overcoming these challenges calls for matching research breakthroughs with applied financial use cases. This research offers insights for future research approaches, interdisciplinary research, and adoption of ML to ensure emerging financial technologies continue to be innovative while remaining compliant.

Literature review

Background

Prediction of financial risk has been traditionally relying on statistical models that are basically using financial ratios and in general a combination of regression models. The reference papers in this domain are Altman's (1968) Z-score model and Ohlson's (1980) logistic regression. These established the groundwork for modelling bankruptcy risk and evaluation of corporate financial stability. These traditional models usually are used as a benchmark for evaluating a firm's financial health. They use structured financial indicators and ratios, such as liquidity ratios, profitability measures, and leverage indicators. Still, despite their historical significance and reliance as a widespread approach, these models are limited by their reliance on linearity due to the model functions, thus failing to capture non-linear characteristics that are often observed in financial distress scenarios.

Thus, the rise of big data technologies and evolution of computational power have enabled the possible adoption of AI and ML techniques, through analysis of multi-dimensional, structured and unstructured data sources such as transactions, social media, and macroeconomic indicators.

State-of-the-Art in ML for Financial Risk Prediction

Due to the limitations of conventional methods, ML methods have been popular in financial risk modeling. These methods, such as ensemble learning, deep learning, and hybrid AI models, have extensive benefits in processing complex data sets and enhancing predictive accuracy and have been widely applied in structured credit risk measurement owing to their predictability and stability. Ensemble methods have emerged to prominence, with Random Forests from Breiman (2001) providing a formidable challenger that is both accurate and interpretable. Even newer breakthroughs, such as Chen and Guestrin's (2016) development of XGBoost, have gone on to advance the boundaries of predictive capability in credit risk scoring and bankruptcy prediction. These methods exploit boosting processes that improve classification and regression methods through iteratively optimizing weak predictors to make models more robust.

Apart from ensemble techniques, deep learning models such as convolutional neural networks (CNNs) and long short-term memory (LSTM) networks are increasingly being seen as essential instruments to address unstructured and large financial data, as revealed by Li et al. (2023). Deep learning models perform dramatically better in representation learning and feature extraction and therefore can learn complex patterns underlying textual and sequential financial data like financial disclosures, earnings reports, and loan applications. Further, the integration of deep learning with other sources of data, i.e., news sentiment analysis and web search trends, has expanded the range of AI-based applications for financial risk measurement even wider.

But their growing sophistication has also prompted concerns about interpretability and compliance with regulations. Their compliance with regulations is one of the essential factors in the choice of ML techniques for measurement of financial risks. Table 1 presents various ML models as far as compliance, interpretability, and key applications are concerned.

Table 1: Comparison of ML Techniques and Regulatory Compliance in Financial Risk Assessment

ML Technique	Regulatory Compliance	Interpretability	Primary Use Cases	Challenges
Traditional ML (LR, Decision Trees)	High (audit-friendly, simple models)	Easy to interpret	Credit scoring, risk ranking	Lower predictive power on complex datasets
Ensemble Learning (Random Forest, XGBoost)	Moderate (Requires XAI for Compliance)	Moderate (can use SHAP/LIME)	Credit risk, bankruptcy prediction	Computational cost, hyperparameter tuning
Deep Learning (CNN, LSTM)	Low (Lacks Transparency for Regulation)	Difficult to interpret	Fraud detection, alternative credit risk	Requires large datasets, lacks explainability
Explainable AI (SHAP, LIME, Counterfactuals)	High (Meets GDPR & Basel Standards)	Very high	Used as an add-on to other ML models	Computational overhead, limited adoption

Source: Authors' own research.

Regulatory requirements such as Basel III and IFRS 9 focus on transparency and ability to explain how the model predicts, thus decision trees and logistic regression are preferred due to their transparent logic and auditability. Even though the predictive accuracy of ensemble methods is superior, their use of boosting makes them difficult to interpret and requires the use of explainable AI (XAI) tools like SHAP and LIME for audit purposes. Deep learning models, being black boxes, are under severe scrutiny by regulatory regimes, restricting their application to high-risk domains such as credit scoring and stress testing, where transparency is a prime requirement. GDPR requires transparency in automated financial decision-making, with widespread application of explainable AI methods to facilitate ML-based risk assessment. Without such interpretability augmentations, ensemble and deep learning algorithms are not compliant, limiting their adoption in regulated financial environments. As daunting as these barriers are, deep learning lies at the heart of fraud detection, where more weight has to be placed on accuracy than interpretability. Advances in explainable deep learning, such as self-explainable neural networks and hybrid AI models, seek to close this gap, potentially expanding regulatory clearance in credit risk assessment and stress testing, while keeping ML progress in sync with changing compliance needs. Hybrid AI models combining deep learning with explainability-driven approaches are likely where future research is trending, pushing compliance while leveraging the predictability of AI in financial risk assessment.

Although ML models have made significant improvements in prediction performance, their general use still has a lot of challenges that need to be resolved. Some of these challenges include the overfitting risks from noisy financial information, the need to constantly update the models to respond to market evolutions, black swan events, and also compliance challenges faced from financial regulators. Future work will need to use more and more hybrid approaches that leverage the strength of conventional econometric methods and the flexibility of ML approaches, enabling the building of more stable and interpretable models.

Network-Based Bibliometric Analysis

Even though the presence of many research studies have reported high rates of superior performance by ML methods in forecasting financial risk, little is known about the larger intellectual and collaborative environment of this fast-developing field. Conventional bibliometric studies tend to look at publication and citation frequencies but ignore the nuanced interactions between front-runner studies, new research communities, and cooperative networks. Bibliometric network analysis provides a general view that shows such structural connections (Donthu et al., 2021). Through mapping of citation networks and also of thematic clusters, researchers can easily identify core contributions and trace how new knowledge has impacted developments in the field. Co-citation analysis, for instance, reveals how early research on bankruptcy prediction continues to be relevant to current ML research, while bibliographic coupling reveals the intersection of classical financial theories and AI-based innovations. Keyword co-occurrence analysis also identifies new research trends, including the increasing prominence of explainability, alternative credit risk models, and the integration of fintech innovations into conventional banking paradigms.

This research is advantaged by these network-facilitated methods by employing tools such as VOSviewer to systematically chart the development of financial risk forecasting studies. Through the identification of core publications, thematic clusters, and co-occurring keywords, analysis not only closes the divide between conventional financial model and contemporary applications of ML, but it is also a foundation for further investigation and interdisciplinarity. This bibliometric analysis assists in putting the accelerating change that is happening in financial risk forecasting into perspective and provides a systematic understanding of how AI and ML are reshaping the industry.

Methodology

This study uses a mixed-method approach combining qualitative and quantitative network-based bibliometric analysis to chart the intellectual and collaborative frameworks within machine learning (ML)-based financial risk prediction. The research questions are: What are the intellectual pillars and influential publications in this study area? What are the emerging thematic clusters visible by theme? How are the shifting research topics, as substantiated by co-occurrence of author keywords, used in the creation of ML financial risk prediction applications?

The data were accessed from the Scopus database between 2000 and 2024. Focused keywords within a search strategy were applied in machine learning, financial risk forecasting, credit scoring, and bankruptcy prediction. The original data were first screened to exclude duplicates and include records with full bibliographic data and relevance to the study, ultimately yielding a filtered dataset for network analysis.

The methodological basis of the study is bibliometric analysis through networks. Co-citation analysis has been carried out in order to create intellectual evolution map and seminal works of the topic area. Networks in which documents that have been co-cited together are shown as nodes and strength of ties indicates the degree of co-citation. Bibliographic coupling was utilized to identify upcoming trends in research using comparison between the shared references of existing studies and therefore the contemporary clusters portraying the trend of prevailing research streams. Author keyword co-occurrence mapping was also utilized to graphically map thematic clusters and identify changes in research direction over time and therefore comprehend mature and emerging themes in financial risk prediction.

In order to represent and analyze the bibliometric data, VOSviewer was employed, a very tested software to build and study bibliometric networks (Donthu et al., 2021). The co-citation network was built at the document level, considering those studies that have at least 20 cited sources to involve strong sources of financial risk modeling. For relevance, cited reference bibliographic coupling analysis was later done on fractional count, excluding publications receiving fewer than 100 citations to show significant thematic relationships among modern studies. Keyword co-occurrence analysis targeted author keywords as well as fractional counting, with at least 15 occurrences, to reflect the changing research context. These metrics were selected to trade completeness and concentration so that there can be a perceptive representation of research clusters, connection strengths, and citation influence of machine learning techniques in financial risk prediction. In view of this integrated methodological framework, the research presents an overall description of the background and status quo of ML applications for financial risk forecasting and gains new insights to guide upcoming investigations and spur interdisciplinary collaborations.

Results and discussions

Co-citation analysis of studies of financial risk prediction illustrates that there exists a systematic intellectual landscape, in which distinct research traditions and methodological breakthroughs are interconnected.

Four significant clusters emerge from network visualization, each an autonomous research discipline, ranging from traditional financial risk estimation models to innovations using machine learning, deep learning techniques, and explanation techniques.

Table 2: Co-Citation analysis with main cluster statistics

Cluster	Nodes	Avg Links	Avg Total Link Strength	Avg Citations	Max Citations	Min Citations	Std Dev Citations
Blue	44	55	112	30	51	20	9
Red	34	68	179	30	62	21	10
Green	22	59	157	35	106	20	23
Yellow	21	34	58	27	70	20	11

Source: Authors' own research, VoSViewer analysis.

Blue Cluster is where firm bankruptcy prediction serves as the foundation for the subsequent machine learning innovations. It is the largest with 44 studies and further contributes to its image as the powerhouse of financial risk analysis. Groundbreaking contributions such as Altman's (1968) Z-score model and Ohlson's (1980) logistic regression model form the dominant cluster, exemplifying traditional use of financial ratios and regression-type models. Its average of numbers of citations (29.8) and peak value (51) indicate study significance, and 8.93 standard deviation gives a smooth trend of citation. Network connectedness indicates ongoing impact, thus application of traditional ML practice to such highly ingrained models continues to be appropriate.

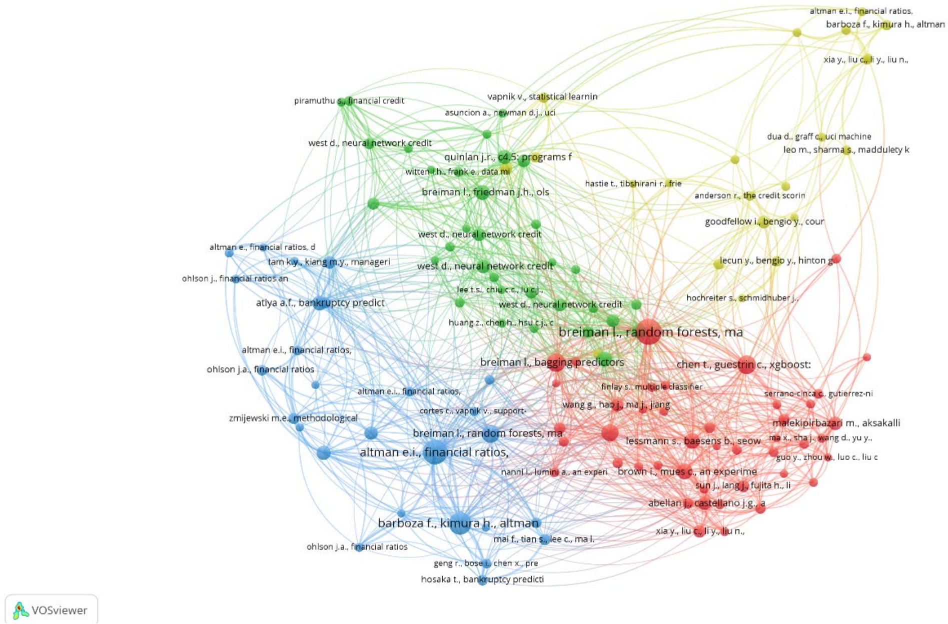

Figure 1. Co-Citation Network of References

Source: www.scopus.com, VoSViewer analysis

Red Cluster indicates migration toward machine learning-based methodologies, and ensemble techniques are the standard approach. This 34-paper Red Cluster is best linked (Avg Links: 68.0, Total Link Strength: 178.8) and turns into the finance risk literature core. Breiman's (2001) Random Forests and Chen & Guestrin's (2016) XGBoost transformed credit risk and bankruptcy forecasting, and Baesens et al. (2003) cemented the superiority of ensemble methods over traditional models. The average citations (29.6), max (62), and range (10.31) indicate extreme but reasonably evenly spread impact. Though predominant, the cluster also shows very strong association with the Blue Cluster, an expression of the residual memory of orthodox finance theory paradigms.

The Green Cluster indicates the emergence of deep learning and hybrid AI models for financial risk forecasting. It has 22 studies and is more cited (Avg: 35.2, Max: 106) with more dispersed citations (Std Dev: 22.95), suggesting a handful of extremely influential papers propelling its influence. Studies in this cluster combine neural networks with conventional financial metrics, with West (2000) leading the way in applying neural networks to credit scoring and risk assessment. Though, its network interconnectivity is lower than that of the Red Cluster, demonstrating that deep learning is not yet adopted on a large scale by regulators due to issues of interpretability, computational intensity, and lack of transparency.

The Yellow Cluster depicts increased share of boosting models, hybrid AI technologies, and credit risk evaluation using ML. It is a moderate size cluster, being the one with the lowest interconnectivity. It also has moderate citation impact. The paper written by Xia et al. (2017) cites the emergence of boosting algorithms like XGBoost and LightGBM. These models have showed improved predictive performance in financial classification problems. Another paper by Anderson

(2007) supports the use of credit scoring models with the integration of conventional methods enhanced by models using AI/ML. This cluster also reveals a shifting approach towards engineered ML pipelines and structured datasets for financial risk modeling.

These clusters of research directly influence ML adoption in financial risk modeling. They show a potential reconciliation between performance metrics, explainability, and compliance. With regulations like Basel III, IFRS 9 or GDPR that require explainability and auditability, the research indicates explainability and compliance currently drive AI uptake in financial decision-making.

Co-citation network gives us important trends defining the industry: traditional models are still industry standards and used as benchmarks. Deploying deep learning poses regulatory challenges, and explainability is an emergent priority. These findings aid financial institutions and policymakers in navigating the evolving risk assessment climate as AI intersects with traditional finance methods.

Bibliographic coupling analysis indicates the recent trends of research done in financial risk forecasting by identifying recent trends and emerging evolutions which are linked or have a separation to already established research.

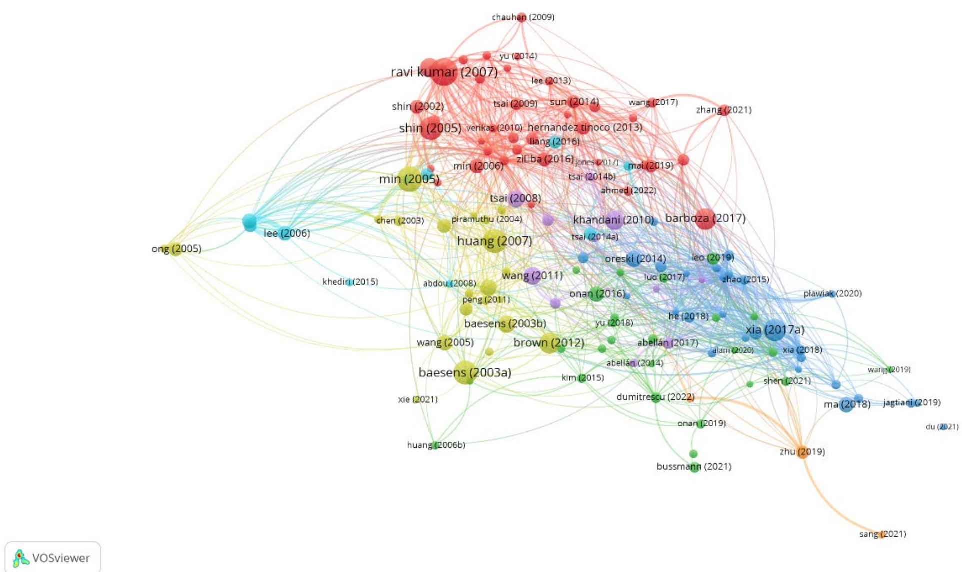

Figure 2. Bibliographic Coupling of Studies

Source: www.scopus.com, VoSViewer analysis

In contrast to the co-citation analysis tracing historical origins, bibliographic coupling reveals the interactive and dynamic character of new literature in the field.

Table 3: Bibliographic coupling analysis with main cluster statistics

Cluster	Nodes	Avg Links	Avg Total Link Strength	Avg Citations	Max Citations	Min Citations	Std Dev Citations
Red	39	79	30	225	819	102	152
Green	28	49	14	158	343	100	59
Blue	28	75	31	177	590	100	101
Yellow	23	73	23	270	704	109	191
Magenta	11	91	28	229	454	109	129
Light Blue	10	67	34	216	309	117	72

Source: Authors' own research, VoSViewer analysis.

The research identifies that machine learning credit risk evaluation and bankruptcy prediction studies are highly linked. Red Cluster with 39 studies, is the most central and largest one. The high average number of citations (224.8) and high network connectivity (Avg Links: 79.4) shows the combination of conventional statistical techniques with contemporary machine learning based techniques. Ravi Kumar and Ravi (2007) show that bankruptcy prediction models have progressed from initial statistical methods to dynamic AI methods that continuously adapt. This shift is justified by the network topology, and it confirms that ensemble learning and support vector machines from Green Cluster are key in the new way credit risk modeling is performed. This cluster is an important hub, showing that ML based methods have a better performance than traditional, classical models – Baesens et al. (2003).

The inclusion of deep learning methods and approaches in the Blue Cluster also indicates the move towards a more sophisticated data-driven decision-making requiring large volumes of data and complex algorithms. This shows that machine learning further improves and refines risk estimation in financial forecasting field. Though ensemble learning is still at the center, recent studies show that the field is working on improving these models with advanced tuning methods and optimization strategies. The study by Xia et al. (2017) captures the effectiveness of boosting-based methodologies through Bayesian hyper-parameter optimization. In contrast to previous studies comparing ML with statistical models, this cluster focuses on adaptive optimization strategies. This promotes increased heterogeneity and prediction accuracy.

One of the key bibliographic coupling trends is increasing focus on interpretability, auditability and regulatory compliance. In the Yellow Cluster we identify increased focus on AI fairness, transparency, and accountability. Onan (2019) proposes consensus clustering-based techniques for under-sampling to mitigate imbalanced learning issues. This cluster has high citation impact and connectivity. This indicates that explainability has become a primary need from a secondary one. This requirement is driven by regulatory frameworks, like Basel III, IFRS 9, and GDPR, that all require increased model interpretability.

Apart from classical and risk models based on machine learning, bibliographic coupling also identifies some new trends in financial risk analysis. The Magenta Cluster indicates the increasing significance of fintech lending and alternative credit risk evaluation. The study by Jagtiani and Lemieux (2019) presents evidence of how alternative data points and emerging ML methods are revolutionizing credit risk evaluation with P2P lending and electronic finance.

Light Blue Cluster reveals other new AI methods in financial risk modeling. Its high network interconnectivity reflects that non-traditional lending systems and credit risk measurement via unstructured data are being more incorporated into ML-based financial models. These methods are not yet predominant. Still, their high degree of bibliographic connectivity with traditional ML techniques ensures that fintech, P2P lending, and AI-based risk models will keep on expanding.

Author keyword co-occurrence analysis further reveals the overarching research interest in machine learning. The Red Cluster has the highest correlations of words. Main ones are artificial intelligence, imbalanced data, fintech, peer-to-peer lending, convolutional neural networks, transfer learning, and financial inclusion. The strength of the link, with 28 average links and a total link strength of 59 indicates that AI has moved from the conventional credit risk evaluation to venturing into alternative lending, fraud prevention, and financial inclusion.

Table 4: Author keyword co-occurrence analysis with main cluster statistics

Cluster	Nodes	Avg Links	Avg Total Link Strength	Occurrences			Std Dev Occurrences	Avg Pub Year	Avg Citations
				Avg	Max	Min			
Red	20	28	59	69	792	16	172	2021	15
Green	20	29	32	37	149	15	35	2022	13
Blue	17	39	71	82	368	16	96	2017	35
Yellow	14	35	60	69	500	17	128	2019	35
Purple	11	37	61	74	207	15	58	2020	17
Light Blue	8	33	50	53	157	15	58	2019	26

Source: Authors' own research, VoSViewer analysis

Another clear focus of bankruptcy prediction is also emerging in the network, led by Blue Cluster. Co-occurring keywords like artificial neural networks, support vector machines, genetic algorithms, and financial crisis validate the application of machine learning in financial distress prediction and systemic risk estimation. High frequency of prediction entails that this stream of research is more and more combining ML methods with traditional models of risk for better prediction of financial crises.

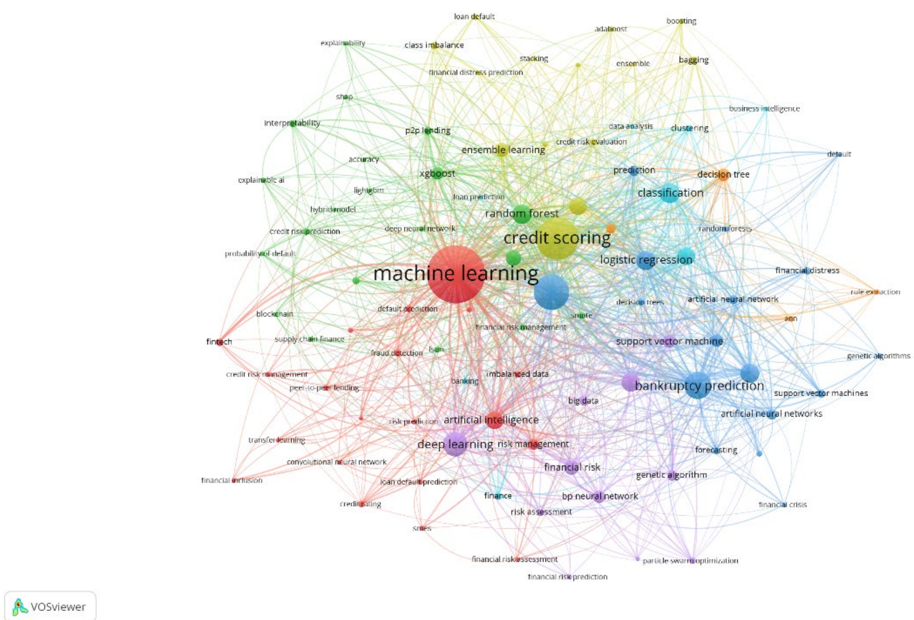

Figure 3. Thematic clusters in keyword networks

Source: www.scopus.com, VoSViewer analysis

A very large thematic cluster containing ensemble learning for credit risk forecasting, can be seen in Green Cluster. Key words like XGBoost, random forest, explainable AI, and interpretability are an emerging intersection of predictive accuracy and regulatory clarity. The focus on accuracy, default probability, and hybrid models in this cluster aligns with previous research in noting the need for balancing machine learning efficiency with financial regulation requirements. Furthermore, the common co-occurrence of clustering and classification techniques such as decision trees, business intelligence, and data analysis indicates their sustained use across structured credit risk estimation and general financial decision-making. The cluster's very high interconnectivity attests to the proof that ensemble approaches are presently in the limelight of financial risk modeling, most especially in structured credit assessments.

Deep learning seems like an independent line of research within Purple Cluster, typically in combination with artificial neural networks, backpropagation networks, big data, measurement tools for financial risk, and risk prediction. Contrary to ensemble learning's dominance in measuring structured credit risk (Green Cluster), deep learning-based models are increasingly being used in applications with high-dimensional and unstructured data. The occurrence of big data in this group also entails the application of deep learning in analyzing large-scale financial data sets, enhancing pattern detection and forecasting in risk modeling.

Machine learning continues to be the prevailing paradigm but with deep learning and ensemble methods being used for various areas of credit risk, bankruptcy prediction, and fintech-based financial products. The increasing emphasis on transparency and explainability casts a light on a new priority: that new AI risk models not just maximize predictive effectiveness but also satisfy regulatory requirements as well as ethical requirements.

Conclusion

This research provides an exhaustive bibliometric analysis of machine learning research in financial risk analysis, tracing prevailing thematic clusters and outlining new directions in predictive modeling, explainability, and compliance. The results depict how the subject progressed from conventional financial risk models to ML models, with ensemble learning, deep learning, and fintech applications emerging as mainstream areas of research.

One of the important observations from the citation and bibliographic coupling analysis is that machine learning models are supplementing, but not substituting, conventional risk assessment models. The prevalence of ensemble techniques is a testament to their capacity to cope with variability in structured credit risk prediction where they offer accuracy vs. interpretability trade-off. But the growing adoption rate of deep learning methods in fraud detection, credit scoring, and financial forecasting suggests there is an even greater shift toward data-driven modeling in high-dimensional spaces. While promising, such a transition entails novel concerns regarding regulatory sanction and interpretability, especially in cases where strong justification is required of the associated risk.

Among the interesting trends in keyword co-occurrence and bibliometric networks is the increased focus on explainable AI (XAI) in modelling and building solutions for financial risk prediction. Regulatory requirements as Basel III, IFRS 9, and GDPR have increased the importance of explainable decision-making and auditability for predictions using ML methods.

Bearing in mind improvements in methodology is not the primary concern, available methods and techniques used in risk modeling are still being overwhelmed by data issues. Newer techniques that handle class imbalance such as under-sampling, oversampling and ensemble learning adaptive strategies are being used currently. Hybrid modeling and feature engineering approaches, where more than one model or advanced techniques are used, are being researched and adopted with increasing interest.

As increasing numbers of financial institutions use ML to perform risk modeling, seizing the successful application along with it is a delicate calibration between predictability and regulation. They will be needed to enhance the transparency and auditability of models developed through ML algorithms. Consolidating these findings into best practices will position them better to leverage machine learning more efficiently in risk assessment.

DOI: 10.2478/picbe-2025-0099, pp1248-1259, ISSN 2558-9652 |
Proceedings of the 19th International Conference on Business Excellence 2025

Regulators, on the other hand, will be in a position to apply the evidence to develop more informed guidelines on credit scoring using the new emerging methods and accurately measure and value the risks created by this innovation with the financial system's overall stability.

Investors, for their part, see fast development of machine learning and other AI risk models, mostly in fintech and creation of alternative lending markets provide new opportunities. But these too pose challenges, although not as they do to systemic organizations. But interpretability and fairness also are one of the most important points also in this context. The future research must continue to improve and assess the application of ML in financial risk so that the models can be generalizable, implementable in realities, and regulate compliant. By creating prediction models that combine the different sources of financial and non-financial data in an optimum manner, either the source structured or unstructured, the models have balance among prediction performance, ethical, as well as regulatory considerations. The same can assist in building durable and stable estimation of financial risks, facilitate innovation and protect the whole financial system.

This research adds to the literature on how ML techniques are transforming financial risk modeling and serves as a platform for subsequent research on AI adoption, regulation, and implementation techniques. The findings confirm that machine learning not only is reshaping the manner in which financial risk analysis is performed—it indeed is reframing its very essence, bridging the automation-explainability-accountability void.

References

Altman, E. I. (1968). Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. *The Journal of Finance, 23*(4), 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x

Anderson, R. (2007). *The credit scoring toolkit: Theory and practice for retail credit risk management and decision automation.* Oxford University Press. https://doi.org/10.1093/oso/9780199226405.001.0001

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society, 54*(6), 627–635. https://doi.org/10.1016/j.ejor.2015.05.030

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research, 133*, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint* arXiv:1702.08608. https://doi.org/10.48550/arXiv.1702.08608

Jagtiani, J., & Lemieux, C. (2019). The roles of alternative data and machine learning in fintech lending: Evidence from the LendingClub consumer platform. *Financial Management, 48*(4), 1009–1029. https://doi.org/10.1111/fima.12295

Li, J., Xu, C., Feng, B., & Zhao, H. (2023). Credit risk prediction model for listed companies based on CNN-LSTM and attention mechanism. *Electronics, 12*(7), 1643. https://doi.org/10.3390/electronics12071643

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research, 18*(1), 109–131. https://doi.org/10.2307/2490395

Onan, A. (2019). Consensus clustering-based undersampling approach to imbalanced learning. *Scientific Programming, 2019*, Article ID 5901087. https://doi.org/10.1155/2019/5901087

Ravi Kumar, P., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques—A review. *European Journal of Operational Research, 180*(1), 1–28. https://doi.org/10.1016/j.ejor.2006.08.043

West, D. (2000). Neural network credit scoring models. *Computers & Operations Research, 27*(11-12), 1131–1152. https://doi.org/10.1016/S0305-0548(99)00149-5

Xia, Y., Liu, Y., Li, Y., & Liu, X. (2017). Boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications, 78*, 225–241. https://doi.org/10.1016/j.eswa.2017.02.017