International Research Journal of Multidisciplinary Scope (IRJMS), 2024; 5(3):457-474

Review Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2024.v05i03.0779

Machine Learning in Financial Distress: A Scoping Review

Alay Peralungal*, Natchimuthu Natchimuthu

School of Commerce, Finance and Accountancy, CHRIST (Deemed to Be University), Bengaluru, Karnataka, India. *Corresponding Author's Email: alay.p@res.christuniversity.in

Abstract

Predicting financial distress is crucial for stakeholders, policymakers, governments, and management in decision-making processes. Researchers have developed various prediction models encompassing both traditional and machine-learning approaches. Notably, recent attention has shifted towards employing machine learning models to address the limitations of traditional methods. This study seeks to offer insights into current trends, identify gaps, and suggest future research directions using machine learning models for financial distress prediction, employing the PRISMA Extension for Scoping Reviews methodology. To achieve this, a comprehensive search was conducted across three databases—Science Direct, EBSCO, and ProQuest—spanning from 2020 to 2023, identifying 34 relevant articles for analysis. The findings underscore the prevalent use of Support Vector Machine in financial distress prediction, followed by the Random Forest Classifier and Artificial Neural Network, with little attention paid to other models. Furthermore, the study underscores the necessity for more research in developing countries, noting the predominance of studies from developed nations. While machine learning models hold promise for enhancing the accuracy and efficiency of financial distress prediction, additional research is imperative to evaluate their effectiveness and applicability across diverse contexts. This scoping review aims to furnish researchers, policymakers, and institutions with valuable insights and policy recommendations, shedding light on underexplored machine-learning techniques.

Keywords: Artificial Neural Networks, Financial Distress Prediction, Machine Learning, Support Vector Machine.

Introduction

The trend of the past several decades has revealed that financial crises have been one of the most significant issues the world has been facing (1). Researchers have predominantly focused on this domain during financial crises like those that occurred from 2008 to 2010 and the crisis brought on by the Coronavirus pandemic that resulted in several corporate failures (2). A crisis may affect the economy internally and externally (3). It can be interpreted in several ways, including financial market collapse, a high unemployment rate, low oil prices, notable swings in credit volume, asset prices, economic fragility, and recession (4). A business can declare bankruptcy for many reasons, such as immediate environmental changes, inappropriate management decisions, low profitability, underutilization of assets, and inefficient working capital management. Before declaring bankruptcy, the company passes through several stages, including profit decline, mild illiquidity, distress stage, severe illiquidity, and bankruptcy (5). One such stage is financial distress (FD). FD occurs when an individual or a company cannot satisfy its financial obligations due to poor economic performance, as evidenced by revenue loss, low profitability, asset underutilization, and poor working capital management (6). Numerous studies have emphasized the significance of predicting bankruptcy, insolvency, and FD as, in economic decision-making; the anticipation of business collapse is of great importance. Moreover, corporate failures might impact the overall economy of a country (7, 8). Thus, several researchers have turned their attention to foreseeing FD. Over the past five decades, predicting FD has been a prominent research issue that continues to grow. This topic remains relevant and trending (9). Predicting FD has become crucial, especially in a competitive world. As a result, it has become a hot topic in finance and is essential for giving decision-makers early warning signals about possible FD (10).

Researchers in the early 1960s developed various models to predict FD. Beaver introduced the first statistical model in 1966 using a univariate analysis, which examines a single variable to find patterns and trends (11). Shortly after, in 1968, Altman enhanced Beaver's research by developing

(Received 17th March 2024; Accepted 18th July 2024; Published 30th July 2024)

a multiple discriminant analysis and presented the first multivariate study. Altman developed a Z-score model using a discriminant function to categorize the corporates into financially healthy and distressed. His work laid the ground for developing bankruptcy prediction models, which paved the way for various bankruptcy prediction models (12). Over time, there have been many articles on Altman's Z-score model. In 1980, Ohlson developed the O-score model using a conditional logit model to address the issues associated with multivariate discriminant analysis (13). Several other researchers presented various alternative models, which encompassed univariate analysis, multiple discriminant statistical methods, techniques involving discriminant analysis, Conditional Logit Analysis, as well as models like Logit, Probit, and logistic regression (14) to achieve higher accuracy and to predict FD early. However, this field of study is still expanding, and each model has drawbacks and limitations. These models are mainly designed to provide predictions based on available data but do not comprehend the underlying causes of FD. In the real world, assumptions of the linear relationship between variables may not hold, and overfitting the data can result in poor generalization of new data. Statistical models require a large sample size to be accurate. However, FD events are relatively rare, and obtaining a sample size that is large enough is challenging. Financial data can be subject to errors and biases, such as data entry errors or selective reporting. These issues can affect the accuracy of statistical models. Therefore, Machine Learning (ML) Models have emerged as a breakthrough solution to address traditional models' limitations and enhance prediction accuracy in FD. Their evolution in this domain has been driven by their ability to uncover intricate patterns within complex financial data, adapt to the ever-changing dynamics of the market, and deliver superior precision when contrasted with traditional techniques (15). Through autonomous learning from historical data and adept handling of nonlinear relationships, these models have introduced a data-centric and resilient approach to identifying enterprises vulnerable to financial instability, facilitating efficacious risk management strategies.

Over the past several decades, the development of ML methods has significantly impacted the prediction and management of financial crises, with each major financial crisis spurring advancements in ML techniques and their application to FD prediction. During the 1980s Savings and Loan Crisis in the United States, initial explorations into ML methods began amidst the dominance of traditional statistical models. Although ML methods were still in their infancy and did not significantly influence crisis management at that time, this period marked the beginning of substantial breakthroughs in ML applications that emerged later in the decade (16). The global financial crisis of 2008 was a pivotal moment for ML applications in predicting market trends and crises, as researchers increasingly utilized extensive datasets and sophisticated algorithms, including decision trees (DT), support vector machines (SVM), and ensemble methods like random forests (RF), to anticipate market failures and fluctuations (15). More recently, the COVID-19 pandemic further accelerated the adoption of advanced ML techniques to predict FD in real time. The profound disruptions to global trade and business activity precipitated economic decline, heightened unemployment, and increased poverty, prompting financial experts and institutions to turn to advanced ML models such as deep learning (DL), recurrent neural networks (RNN), and ensemble methods to navigate the rapidly evolving market dynamics. These models leveraged large datasets to develop early warning systems, underscoring the critical integration of ML with domain expertise for accurate and interpretable predictions (17).

Artificial intelligence (AI) and ML have become important in the 4th Industrial Revolution (18). Integrating AI expedites the model development process for FD prediction. ML techniques yield models capable of making predictions without explicit programming, deriving insights from sample data commonly known as "training data." The inception of the term "machine learning" is attributed to Arthur Samuel in 1959. As early as 1985, ML models were used to forecast FD during its nascent stages (19). Later, in 1990, Marcus D. Odom and Ramesh Sharda developed a neural network model to predict bankruptcy and compared the model with statistical models. The neural network model proved more robust than the discriminant analysis method on reduced sample sizes (20). ML models have gained widespread prominence in predicting FD in recent years. Recently, extensive research on bankruptcy prediction strategies has primarily centered on traditional statistical methods, neglecting the exploration of ML techniques (21).

Nevertheless, existing reviews have approached this topic from diverse perspectives. Some studies meticulously analyzed influential articles and their disseminating journals, emphasizing co-authorship patterns and prevalent methods (22). Other research delved into various aspects, such as definitions, modelling techniques, and sampling methodologies, and focused on private firms in advanced economies (23). Conversely, specific reviews concentrated on individual countries, offering localized insights. However, most of this research is outdated, offering limited insight into recent techniques and complicating the identification of future research trends. Additionally, prior reviews predominantly focused on statistical methods or core techniques like neural networks, expert systems, and hybrid intelligence systems (24). In recent years, ML techniques have gained prominence in predicting FD due to technological advancements and the availability of vast datasets (25). Studies also highlighted the need for diverse AI methods in various finance areas (26). The current scoping review aims to answer the question: What are the current trends and gaps in modelling FD using ML techniques? With the growing emphasis on ML models for predicting FD, this study comprehensively reviews recent advancements in the field. To the author's knowledge, no previous scoping review has analyzed the recent developments in AI techniques and their application in predicting FD in academic research. The authors intend to fill this research gap by examining the latest ML models used, the variables considered, the countries studied, and sector-specific analyses. Additionally, the study seeks to analyze the volume of articles in this field and assess the journals publishing them based on citation metrics and ranking. Consequently, it is imperative to evaluate the advancements in this field and outline the prevailing research trends through this scoping review. The study encompasses data from key databases - Science Direct, EBSCO, and ProQuest- from 2020 to 2023.

Study Design

The current study adopts the scoping review technique, as a scoping study determines the worth of a thorough systematic review, summarizes and distributes research findings, and discovers gaps in current literature by examining the scope, range, and character of research activity (27). Numerous academics have emphasized the need to conduct a scoping review. The scoping review identifies and maps available evidence (28). Researchers have improved scoping appraisals' methodological and reporting quality over the years (29). Hence, the current study uses the PRISMA Extension for Scoping Reviews (PRISMA-scr, 30).

Search Strategy

The current study predominantly used three databases - Science Direct, EBSCO, and ProQuest as they provide extensive content coverage and size (31). The researcher ensured consistent use of similar search terms across all databases, aiming for comprehensive coverage and minimizing the chance of missing relevant papers. The search incorporated BOOLEAN operators like AND and OR in conjunction with specific phrases. The study utilized the following search strings: (Machine Learning) AND (Financial Distress) OR (Bankruptcy) AND (Prediction). This analysis included academic papers published in English from 2020 to 2023, focusing on these four years to provide an updated overview of research and its current state, including new studies published since the previous review. This can aid in identifying recent trends, insights, and developments in ML models for predicting FD. Grey literature, such as books, book chapters, conference papers, trade publications, editorials, periodicals, and news articles, were excluded. Following a comprehensive literature search, 927 articles were identified from Science Direct (n=718), EBSCO Host (n=139), and ProQuest (n=70) (Figure 1).

Evaluation and Selection Procedure

Initial duplicate checking, abstract, and title screening were done using Rayyan software, a web and mobile tool created to streamline the screening process for researchers performing systematic reviews, scoping reviews, and other types of literature review projects (32). After eliminating three duplicate articles using Rayyan, 924 articles remained. These articles underwent initial screening based on their abstract and title, excluding 882 articles that did not match the key terms and context (33), such as studies not directly related to ML applications in FD or Bankruptcy Prediction. The criteria for initial screening included relevance to the topic, presence of empirical data, and publication in peer-reviewed journals. The remaining 42 articles underwent a full screening based on a detailed review of the full text, focusing on the study objectives, methods, and outcomes. During this process, the researchers found eight articles irrelevant to the study due to factors such as insufficient data or non-empirical nature. Finally, 34 articles were reviewed, and the extracted data were recorded in Microsoft Excel for further analysis.

Data Extraction and Synthesis

The extracted data, encompassing article details like title, authors, year, country, objectives, study period, sample size, sectors, sample design, models used, variables, journal, publisher, journal ranking, number of citations, study method, and relevant outcome, were recorded in a Microsoft Excel file. The use of Microsoft Excel facilitated organized data management and streamlined the synthesis process. The results section discusses the review of selected articles, and the search strategy for the scoping review is illustrated using the PRISMA 2020 flow diagram in Figure 1.

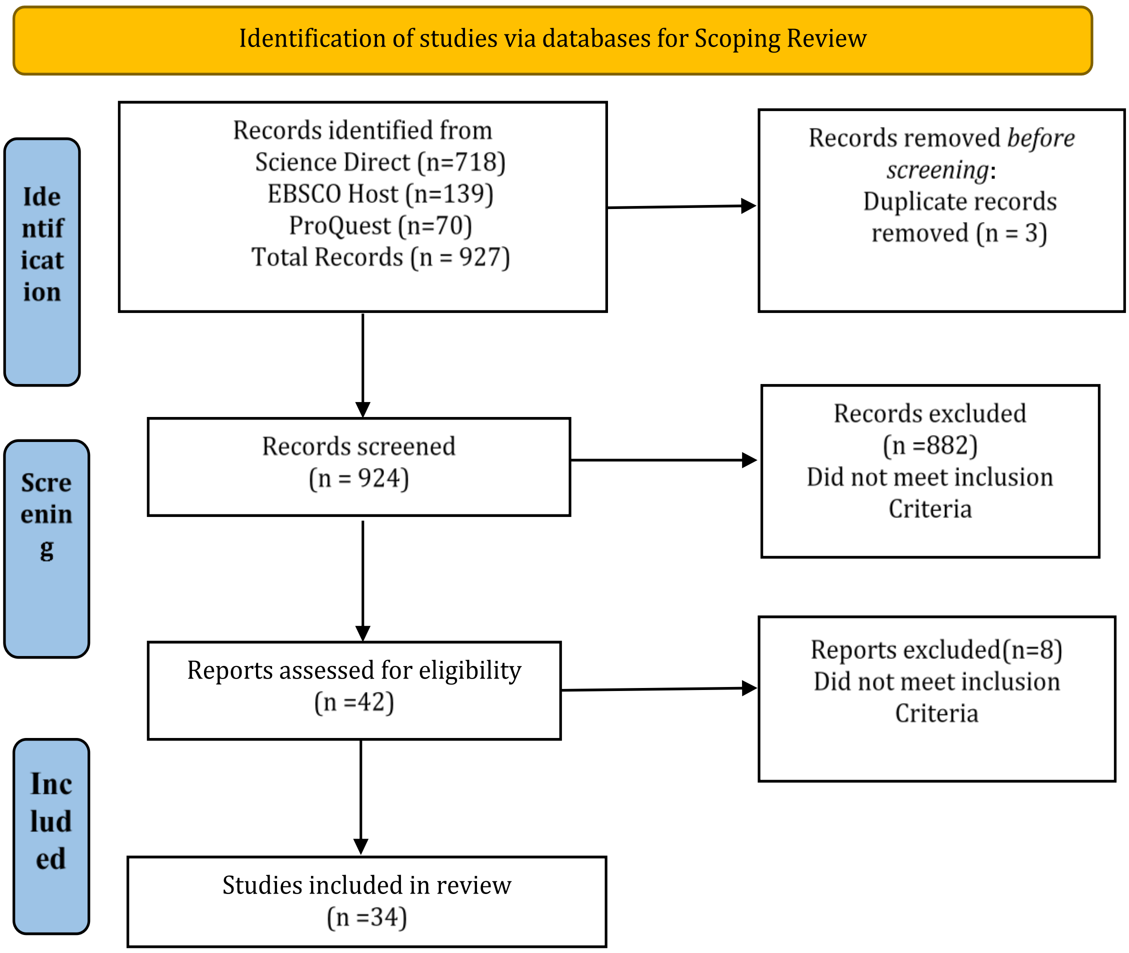

Figure 1: PRISMA 2020 Flow Diagram for New Scoping Reviews

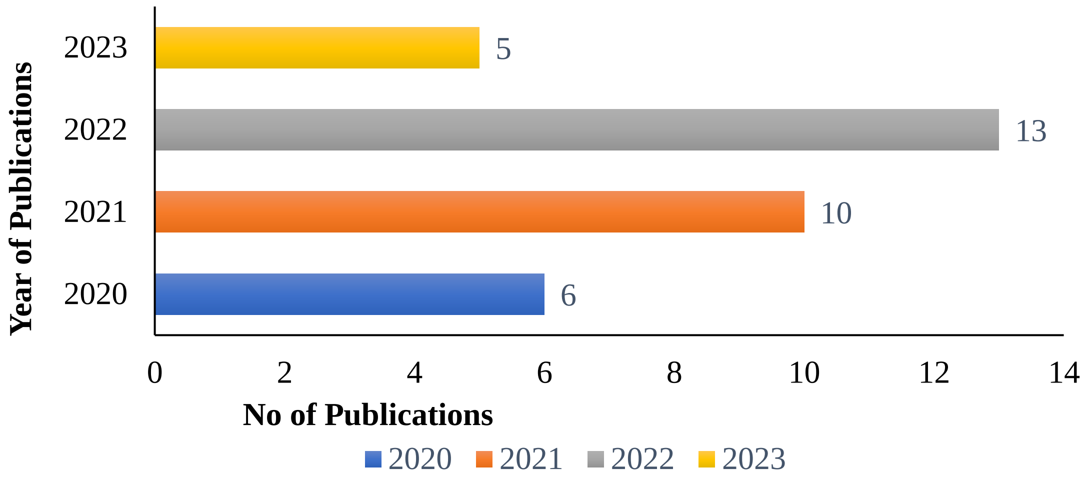

Figure 2: Yearly Trend in FD Prediction Research

Results

Yearly Trend in FD Prediction Research

The number of articles published on this topic has doubled from 6 in 2020 to 13 in 2022, as shown in Figure 2. This indicates a growing interest among researchers in using ML models for FD prediction.

Publication Metrics and Rankings

Table 1 presents a comprehensive overview of the 34 papers analyzed in this study, including the publishers, journals, number of articles, and the respective journals' rankings. The rankings were determined according to the Australian Business Deans Council (ABDC) Journal Quality List as of March 15, 2023, and the SCImago Journal Rank (SJR) based on Quartile Ranking. The selected journals, which are multidisciplinary in scope, were assessed for their quality and impact. Elsevier contributed 11 articles among the major publishers, MDPI published seven articles, and Springer International Publishing released two. The ABDC rankings revealed one article as A*, four articles categorized as A, seven as B, and two as C, while 20 articles were not assigned any ranking. Additionally, based on the SCImago Journal Rank (SJR) Quartile Ranking, 13 articles were in Q1, six in Q2, one in Q3, and two in Q4, with 12 articles not receiving a ranking in the current study.

Table 1: Publication Metrics and Rankings

Journals	Publisher	No of Articles	ABDC Journal Ranking	SCImago Journal Rank (SJR) - Quartile Ranking
Ekonomicky casopis	Institute of Economic Research Slovak Academy of Science	1	B	Q4
Information Systems Frontiers	Springer International Publishing	1	A	Q1
Computational Economics	Springer International Publishing	1	B	Q2
Journal of Risk and Financial Management	MDPI	2	B	-
Risks	MDPI	1	B	Q2
Sustainability	MDPI	1	-	Q1
Mathematics	MDPI	1	-	-
Turkish Journal of Computer and Mathematics Education	Karadeniz Technical University	2	-	-

Iranian Journal of Management Studies	University of Tehran, College of Farabi	1	-	Q4
PLoS ONE	Public Library of Science	1	-	Q1
International Journal of Electrical and Computer Engineering	Institute of Advanced Engineering and Science (IAES)	1	-	Q2
Montenegrin Journal of Economics	Economic Laboratory for Transition Research	1	-	Q2
Journal of Innovation and Knowledge	Elsevier	1	-	Q1
Engineering, Construction, and Architectural Management	Emerald Group Publishing	1	A	Q1
Expert Systems with Applications	Elsevier	2	C	Q1
The Journal of Prediction Markets	University of Buckingham Press	1	B	-
Intelligent Systems with Applications	Elsevier	1	-	Q1
International Journal of Management Economics and Business	Zonguldak Bulent Ecevit University	1	-	-
Borsa Istanbul Review	Borsa Istanbul Anonim Sirketi	1	-	Q2
Machine Learning with Applications	Elsevier	2	-	-
IAES International Journal of Artificial Intelligence	Institute of Advanced Engineering and Science (IAES)	1	-	Q3
Buildings	MDPI	1	-	Q1
Knowledge Engineering Review	Cambridge University Press	1	-	Q2
Data	MDPI	1	-	-
Sustainability Analytics and Modeling	Elsevier	1	-	-
Research in International Business and Finance	Elsevier	1	B	Q1
International Review of Financial Analysis	Elsevier	1	A	Q1
European Journal of Operational Research	Elsevier	1	A*	Q1
Ege Academic Review	Ege University	1	-	-
Finance Research Letters	Elsevier	1	A	Q1

Citation Metrics and Research Influence

Citation analysis evaluates the impact and influence of research articles or academic works by examining the number of times other researchers have cited an article. It provides insights into the reach and significance of the work. This analysis helps researchers understand the importance of their work, identify research trends, evaluate the quality of research, and even find potential collaborators within the scholarly community. In the current study, the researcher determined the citation count of various articles using Google citations. Citation analysis reflects the citation as of 01-Nov-2023. Figure 3 presents this study's top 10 most cited articles. Figure 3 highlights that Horak (3) received the highest citation for the article, with 76 citations, published in the Journal of Risk and Financial Management. The majority of the reviewed studies had more than ten citations.

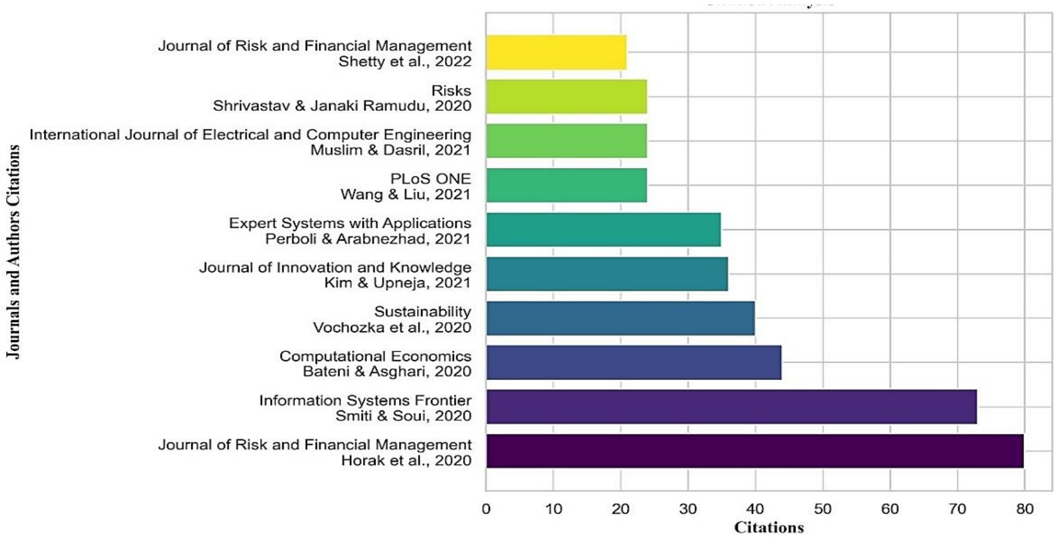

Figure 3: Citation Analysis

Regional Representation in FD Prediction Research

Figure 4 displays that studies utilizing ML models for predicting FD were performed in Europe, with 13 articles published, followed by the United States with 9 studies, China and Iran, followed closely with 3 studies. India published 2 studies, and other countries, such as Indonesia, Africa, Vietnam, and Korea, published 1 article each in this area of research. It is worth noting that the predominant focus of these studies has been on developed countries, emphasizing the imperative for further research in developing nations.

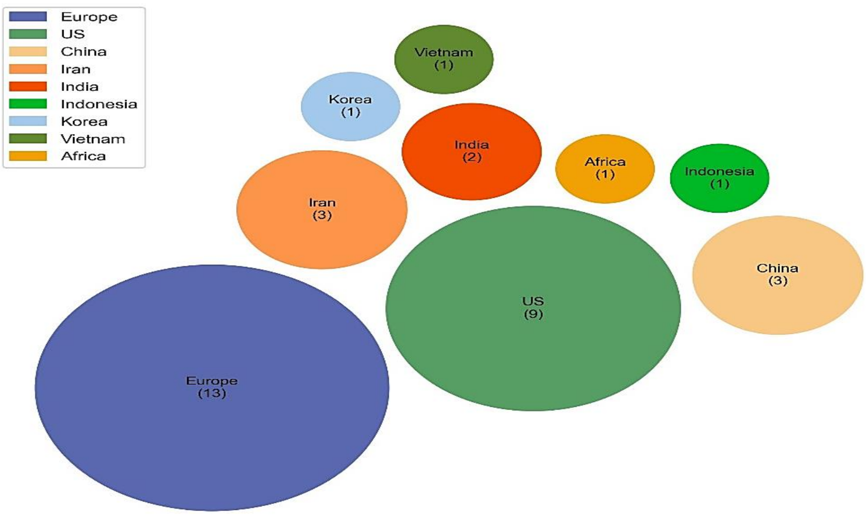

Figure 4: Regional Representation in FD Prediction Research

Overview of Variables Utilized in FD Studies

Variables are essential in predicting FD, providing insights into a company's financial health and performance. Each article has taken a different set of variables depending on the study. Most studies used financial variables, whereas very few used non-financial variables (34, 35) and economic variables (36). Table 2 highlights financial, non-financial, and economic variables considered in the selected 34 studies and further classified under the different ratios. Financial variables are predominantly used in predicting FD and measuring the company's financial health and performance. These financial variables are classified under liquidity, leverage, capital structure, solvency, earnings coverage, activity or efficiency, profitability, market ratios, earnings per share analysis, or shareholder ratios. Non-financial variables are based on non-financial data such as company age, management experience, industry competition, and marketing strategy. These non-financial variables can provide insights into a company's operations and management practices, which can be used to identify potential warning signs of FD. However, very few articles have highlighted this. Economic variables, including interest rates, inflation, the unemployment rate, and GDP growth, can notably impact a company's financial health and performance, and these were used to identify potential risks and opportunities. Therefore, the variables that significantly impact FD prediction provide insights into a company's financial health, operations, management practices, and the broader economic conditions in which the company operates. By incorporating a variety of financial, non-financial, and economic variables into ML models, analysts can identify patterns and relationships that can be employed for FD prediction with greater accuracy.

Table 2: Overview of Variables Utilized in FD Studies

Ratios	Description	Author Citation
Financial Variables		
Liquidity ratios		
Current Ratio	Current Assets / Current Liabilities	(7,10,34,36–43)
Quick Ratio OR Acid Test Ratio	Cash + Marketable Securities + Net Accounts Receivable / Current Liabilities OR (Current Assets - Inventories) / Current Liabilities	(10,38,40–42,44)
Net Working Capital Ratio	Net Working Capital (Current Assets – Current Liabilities)/ Total Asset	(36,45,46)
Cash Ratio	Cash and Cash Equivalents + Marketable Securities/Current Liabilities	(38,40–42)
Cash Flow Ratio OR Operating Cash Flow Ratio	Operating Cash Flow /Period-End Current Liabilities OR Net Operating Cash Flow / Current Liabilities	(41,44)
Leverage, Capital Structure, Solvency and Earnings Coverage Ratios		
Leverage		
Financial Leverage Ratio, OR Equity Multiplier OR Assets to Equity Ratio	Total Assets/ Total Equity OR Assets / Equity	(38,40,47)
Capital Structure and Solvency Ratios		
Debt to equity ratio	Total Liabilities / Total Equity OR Debt / Equity	(41–44)
Long-term debt-to-equity ratio OR Long-Term Debt to Capital Employed	Total Debt – Current Liabilities /Total Equity OR Long term debt / Long term debt+ Preferred Stock + Common Stock	(10,42)
Debt to EBITDA	Debt / EBITDA	(42)

Debt to total assets ratio	Total Liabilities /Total Assets OR Debt / Assets	(34–36,38,39,41–43,46,48)
Earnings Coverage Ratios		
Interest Coverage (Times Interest Earned) Ratio	Earnings Before Interest and Taxes (EBIT)/ Interest Expense	(34,41,44)
Activity Ratios / Efficiency Ratio		
Working Capital Turnover	Net Sales / (Current Assets - Current liabilities)	(38)
Days Purchases in Payables	Average Accounts Payable / Average Daily Credit Purchases (Annual Credit Purchases ÷ 365) OR Sales/Balance of Accounts Payable	(39)
Current Asset Turnover	Net Sales / Current Assets	(38,40,41)
Fixed Asset Turnover Ratio	Sales / Average Net Property, Plant and Equipment	(10,41)
Accounts Payable Turnover Ratio	Annual Credit Purchases / Average Accounts Payable OR Account Payable to Equity	(42)
Accounts Receivable Turnover Ratio OR Trade Receivable Turnover	Net Annual Credit Sales/Average Gross Accounts Receivable OR Net Sales / Trade Receivables	(38,40,44)
Inventory Turnover Ratio	Annual Cost of Goods Sold/Average Inventory	(39,40,44)
Total Asset Turnover Ratio	Sales / Average Total Assets OR Net Sales / Assets	(10,34,36–41,45,46)
Profitability Analysis		
Gross Profit Margin Percentage	Gross Profit/ Net Sales	(10,38,41)
Net Profit Margin Percentage	Net Income / Net Sales OR Net Profit / Net Sales Revenue	(10,39,41)
Operating Profit Margin Percentage	Operating Income / Net Sales	(10,41)
Return on Equity (ROE)	Net Income / Average Total Equity OR Net Profit / Total Equity	(2,36,40,41,43,47)
Return on Assets (ROA)	Net Income / Average Total Asset OR Net Profit / Assets	(2,34,36,38–41,45–48)
Market Ratios and Earnings Per Share Analysis, or Shareholder Ratio		
Total Liabilities/Total Shareholders' Equity		(39)
Market/Book ratio	Market Price per Share / Book Value per Share	(10,42)
Book Value per Share	Total Stockholders' Equity – Preferred Equity / Number of Common Shares Outstanding	(43)
Price/Earnings (P/E) ratio	Market Price per Common Share / Basic Earnings per Share (annual)	(42,43)

Macroeconomic Variables	
Bank non-performing loans to total gross loans ratio (percent)	(2)
Unemployment (percent of total labor force)	(2)
Population annual growth (percent)	(2)
Consumer per index (CPI)	(2,36)
Gross domestic product (GDP)	(2,36)
Federal funds rate (FFR)	(36)
Non -Financial Variables	
Legal Forms	(34)
Age	(10,34)
Size	(10,35)

Industry-Specific Sampling Methods and Approaches

Table 3 presents detailed information regarding the study period, sectors under scrutiny, total sample size (encompassing both failed and non-failed companies), and the employed sample design across the selected studies. These studies primarily focused on the manufacturing and banking sectors. The samples in these studies were bifurcated into two categories: failed companies and non-failed companies, with varied nomenclatures employed to denote these classifications, such as bankruptcy firm and non-bankruptcy firm (49), business failure and non-business failure, bankruptcy and healthy firm/active firm (50, 51), FD and non-FD, and failed and non-failed firm (52). The study duration varied across papers, although most studies incorporated data until 2018. These findings highlight the necessity for research utilizing recent data. Some researchers used a matched paired sampling method, where an equal number of failed and non-failed companies were paired. This approach resulted in a balanced dataset. On the other hand, most of the studies utilized an unmatched sampling design, particularly in real-time scenarios, where there is no equal number of failed companies and their non-failed counterparts. These studies employed the oversampling approach using the SMOTE algorithm and random sampling method with unbalanced data sets.

Table 3: Industry-Specific Sampling Methods and Approaches

Author Citation	Study Period	Company /Sector/Industry	Total Sample Size	Sample Design
(2)	2010-2017	-	119 developed and developing countries	-
(3)	2013-2017	Manufacturing	6833	Unmatched Paired Design - Random Sampling
(7)	1980-2017	US Restaurant Industry	2747	Unmatched Paired Design - Random Sampling
(8)	2010-2016	Banks	-	Matched Paired Design
(10)	2006-2016	Companies Were Filed Under IBC	112	Matched Paired Design
(34)	2014-2019	Wholesale and Retail Trade: Repair of Motor Vehicles and Motorcycles	2,13,931	Unmatched Paired Design - Oversampling approach -SMOTE algorithm
(35)	1961-2020	Publicly Traded US Companies	20,235	-

(36)	1980-2016	Construction Industry - Building Construction Heavy Construction Special Trade Construction	1378	Unmatched Paired Design - Oversampling approach -SMOTE algorithm
(37)	2006-2014	Iranian Firm	174	Matched Paired Design
(38)	1992-2017	Companies Listed on The Tehran Stock Exchange	218	Unmatched Paired Design - Random Sampling
(39)	2009-2018	Construction Industry	-	Unmatched Paired Design - Oversampling approach -SMOTE algorithm
(40)	2015,2016,2 017	Manufacturing, Service and Trade Sectors	240	-
(41)	2011-2021	Health Sector - Hospitals, Biotechnology and Pharmaceuticals, And Service Manufacturers	206	Matched Paired Design
(42)	2010-2021	Listed Companies in Vietnam	3277	Unmatched Paired Design - Oversampling approach -SMOTE algorithm
(43)	2014-2021	Telecommunication, Oil and Gas, Materials, Industrials, Health Care Financials, Consumer Services, Consumer Goods, and Basic Materials	61	Unmatched Paired Design
(44)	2000-2019	Electronics Industry	2946	Unmatched Paired Design - Random Sampling
(45)	2011-2019	Companies Listed on The Tehran Stock Exchange	1521	Unmatched Paired Design - Random Sampling
(46)	1985-2020	North American corporate credit market	18,858	Unmatched Paired Design – Random sampling strategy during model training
(47)	2017 and 2018	Electrical And Engineering Industry	Electrical industry- 754 non-financial corporations Engineering industry- 233 non-financial corporations	-
(48)	2010-2018	US Firms	1824 publicly traded firms in the United States.	Unmatched Paired Design - Oversampling approach -SMOTE algorithm

(49)	1999-2009	Industrial, Electronic, Shipping, Tourism, And Retail Companies	6819	Unmatched Paired Design - Random Sampling
(50)	2016-2018	Spanish Companies - Agriculture Financial Activities Manufacturing Industry Accommodation + Wholesale Trade Technology Construction	6888	Unmatched Paired Design
(51)	2001-2018	Italian SMEs	-	Unmatched Paired Design
(52)	2011-2019	Banks	32,287	Unmatched Paired Design
(53)	2002-2013	Polish Companies	43,405	Unmatched Paired Design - Oversampling approach -SMOTE algorithm
(54)	2000-2017	Banks	59	Unmatched Paired Design - Random Sampling
(55)	2014-2018	Manufacturing	5500	Unmatched Paired Design - Random Sampling
(56)	2000-2012 2007-2013	Polish Firms	1st year -7027 2ndyear-10173 3rdyear-10503 4th year - 9792 5thyear – 5910	Unmatched Paired Design - Oversampling approach -SMOTE algorithm
(57)	1999-2009	Taiwanese Companies	6819	Unmatched Paired Design - Undersampling Method
(58)	2000-2012	Polish Firms	42,627 Rows	-
(59)	2002 and 2012	Belgian Small and Medium Enterprises (SME's)	3728	Matched Paired Design
(60)	1990-2016	-	-	Unmatched Paired Design -Oversampling approach -SMOTE algorithm
(61)	2015-2019	Education, energy, furniture, transportation, mining, automotive, textile, and tourism.	392	Unmatched Paired Design - Random Sampling
(62)	2006- 2023	Listed US companies	-	-

Current Status on Adoption of ML Models

Recently, there has been a notable increase in the utilization of ML models for predicting corporate FD. This trend gained momentum after 1990 when Ramesh Sharda and Marcus D Odom introduced a neural network model to predict bankruptcy and compared it with statistical models. Their study demonstrated that the neural network model exhibited greater resilience when compared to the discriminant analysis method, mainly when dealing with reduced sample sizes (20), which marked the initial step in adopting ML models. This paved the way for researchers to develop various models to predict FD. The accuracy of these models was studied to assist decision-makers in making informed decisions. In contrast, some researchers conducted comparative studies between machine-learning models and traditional models, demonstrating that the machine-learning models outperformed the traditional ones (10, 43, 51). Thus, the current study seeks to discover the trends in ML models used between 2020 and 2023 and explore the study's further scope.

Figure 5 presents the most prominent ML models employed in the chosen studies, while Figure 6 provides an overview of the models used in different countries. Among SVM, RFC, ANN, DT, KNN, GBC, NB, XGBC, DL, and MLP models, SVM emerged as the most prevalent employed in 16 studies, underscoring its widespread applicability and effectiveness in predicting FD. RFC and ANN followed closely, utilized in 13 and 12 studies, respectively, highlighting their prominence in research applications. The use of ML models across the research studies reveals significant trends in their utilization. The researchers were trying to improve the accuracy of the model.

Figure 6 reflects the dynamic landscape of ML models used for FD prediction from 2020 to 2023. In 2020, Europe and the US employed models such as ANN, DL with Long Short-Term Memory (DL-LSTM), and SVM. By 2021, Europe expanded its repertoire to include models like GBC, DT, and KNN. In 2022, there was a further diversification of models, with an emphasis on DL with Neural Networks (DL-NN) and XGBC. The year 2023 saw the continued use of various models, including RF and Deep Neural Networks (DNN). The study also reveals that studies from developed countries have embraced AI at a higher rate while developing countries are in the early stages of adopting ML models.

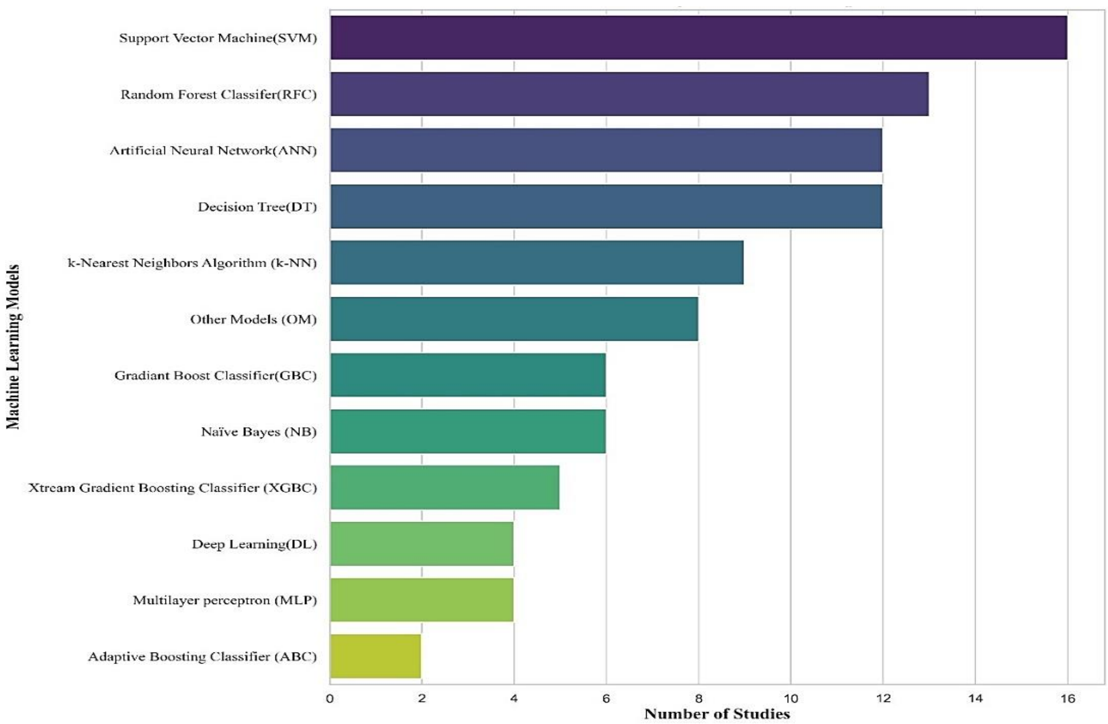

Figure 5: Models Used in the Study

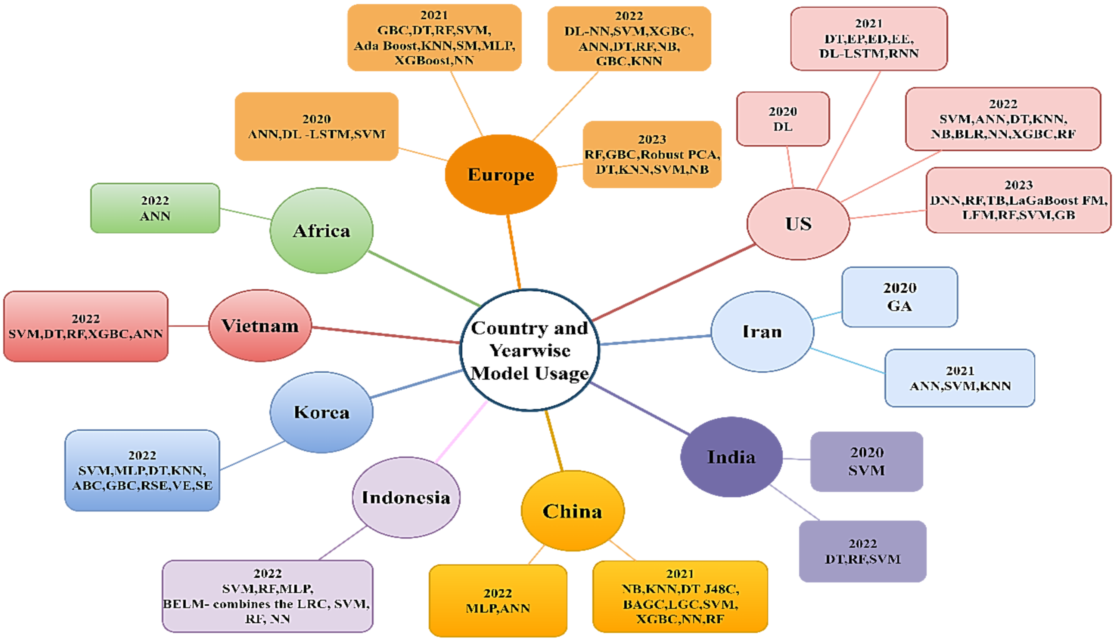

Figure 6: Country-wise Status on the Usage of ML Models

Discussion

This literature review examines the utilization of ML models in predicting FD, adhering to PRISMA Extension for Scoping Reviews (PRISMA-scr) guidelines. Analyzing 34 articles published between 2020 and 2023 from Science Direct, EBSCO, and ProQuest databases, this study focuses on recent publications to provide an updated overview of research trends, insights, and developments in this field. It is the first scoping review to address FD prediction with ML, emphasizing the evolution from traditional to machine-learning models. The review identified a notable increase in ML algorithm use for FD prediction starting in 2022, with several studies gaining recognition in prestigious journal rankings like the ABDC Journal and SJR. Citation analysis revealed widespread scholarly engagement, with many articles receiving more than ten citations on Google Scholar by November 1, 2023. Geographically, a substantial volume of research originated from developed countries, notably Europe (13 studies) and the United States (9 studies), focusing on financial variables such as profitability, leverage, liquidity, and efficiency ratios within sectors like manufacturing and banking.

Methodologically, researchers employed various sampling methods, predominantly favoring unmatched techniques due to data availability constraints. SVM emerged as the prominent model for FD prediction, followed by RFC and ANN, with continuous advancements noted, particularly in Europe and the United States. However, developing countries are in the early stages of ML model adoption for FD prediction, facing challenges in technological infrastructure and methodological innovation. The financial landscape's constant evolution, influenced by economic, regulatory, and technological changes, underscores the dynamic nature of ML applications in financial analysis. Continuous monitoring and analysis of emerging trends and methodologies are essential to adapting ML models effectively in this evolving field.

Conclusion

In conclusion, this scoping review highlights the dynamic evolution and expanding applications of ML models in predicting FD. SVM, RFC, and ANN have demonstrated efficacy in analyzing complex financial data sets, with notable advancements observed in developed economies. The study emphasizes the impact of economic, regulatory, and technological factors on ML model evolution within financial research. The concentration of research efforts in developed economies reveals opportunities for enhanced collaboration and knowledge transfer to facilitate the broader adoption of ML techniques globally. Future research should explore emerging ML methodologies in diverse financial contexts to enhance predictive accuracy and decision-making in financial management and risk assessment. Ongoing monitoring of emerging trends and methodologies is crucial for navigating the complexities of global financial systems and developing robust, data-driven strategies to mitigate financial risks and enhance economic resilience.

Implications

This scoping review substantially benefits researchers, policymakers, and financial institutions alike. Researchers gain a synthesized view of current ML research for predicting FD, identifying trends, and uncovering areas for future exploration. Policymakers can leverage these insights to refine financial regulations and enhance risk management strategies, informed by a nuanced understanding of ML model capabilities and limitations in predicting bankruptcy. Financial institutions stand to enhance their risk assessment practices and strategic decision-making by adopting advanced ML techniques, bridging academic research with practical applications to effectively navigate financial challenges. This review informs theoretical advancements and sheds light on unexplored ML methodologies, crucial for refining management and decision-making frameworks across sectors and fostering innovation in financial risk management strategies.

Limitations

There are a few limitations to this review. The study is a scoping literature review, not a full systematic review of the empirical findings. The papers have been selected from three significant databases and possibly missed some critical information from other databases. Also, the study was limited from 2020 to 2023. The study excluded the research reported in the grey literature. The study is also limited to studies published in English, which may have led to a language bias.

Future Scope

The outcome of this scoping review provides a foundation for future research on ML models in predicting FD. While existing ML models have displayed potential in this area, there is a continuous demand for novel and enhanced models capable of adapting to evolving economic and technological landscapes. To tackle these challenges, further research shall focus on developing new ML models integrating more comprehensive data sources and advanced algorithms, such as Deep learning models. Further studies should investigate various ML models' effectiveness and applicability across different geographic locations, including developing countries. Moreover, future research can delve into identifying factors influencing the accuracy of ML models in FD prediction, encompassing aspects such as data quality and quantity, model complexity, and feature selection. Extensive research is required concerning the impact of external factors, including economic and regulatory changes, on the accuracy of ML models for predicting FD. Additionally, there is a growing need for research addressing the ethical implications related to the use of ML models in FD prediction, including concerns about privacy, bias, and transparency. To expand on this scoping review, researchers can conduct a systematic literature review on ML models, evaluating their efficacy in predicting FD.

Abbreviations

ML: Machine Learning
FD: Financial Distress
ANN: Artificial Neural Network
SVM: Support Vector Machine
DT: Decision Tree
XGBC: Extreme Gradient Boosting Classifier
GBC: Gradient Boosting Classifier
RF: Random Forest
NN: Neural Network
GA: Genetic Algorithm
KNN: K-Nearest Neighbor Algorithm
LSTM: Long Short-Term Memory
MLP: Multilayer Perceptron
DL: Deep Learning

Acknowledgment

The authors acknowledge CHRIST (Deemed to be University) for its support and resources that contributed to the success of this research.

Author Contributions

All authors made an equal contribution.

Conflict of Interest

The authors state no conflicts of interest about this manuscript submission.

Ethics Approval

Not applicable.

Funding

The current research has not received any specific grant from funding agencies that belong to public, not-for-profit, or commercial sectors.

References

1. Wulandari Y, Musdholifah, Kusairi S. The impact of macroeconomic and internal factors on banking distress. International Journal of Economics and Financial Issues. 2017;7(3):429–36.
2. Bitetto A, Cerchiello P, Mertzanis C. Measuring financial soundness around the world: A machine learning approach. International Review of Financial Analysis. 2023;85(March 2022 102451):1–14.
3. Horak J, Vrbka J, Suler P. Support Vector Machine Methods and Artificial Neural Networks Used for the Development of Bankruptcy Prediction Models and their Comparison. Journal of Risk and Financial Management. 2020;13(3):60.
4. ElBannan MA. On the prediction of financial distress in emerging markets: What matters more? Empirical evidence from Arab spring countries. Emerging Markets Review. 2021;47(February 2020):100806.
5. Farooq U, Jibran Qamar MA, Haque A. A three-stage dynamic model of financial distress. Managerial Finance. 2018;44(9):1101–16.
6. Platt HD, Platt MB. Comparing financial distress and bankruptcy. Journal of Risk and Financial Management. 2006;1(1):1–27.
7. Kim SY, Upneja A. Majority voting ensemble with a decision trees for business failure prediction during economic downturns. Journal of Innovation and Knowledge. 2021;6(2):112–23.
8. Siswoyo B, Abas ZA, Pee ANC, Komalasari R, Suyatna N. Ensemble machine learning algorithm optimization of bankruptcy prediction of bank. IAES International Journal of Artificial Intelligence. 2022;11(2):679–86.
9. Farooq M, Hunjra Imran A, Ullah S, Al-Faryan Saleh Abdulaziz M. The determinants of financial distress cost : A case of emerging market The determinants of financial distress cost : A case of emerging market. Cogent Economics & Finance. 2023;11(1):1–22.
10. Gupta V. Bankruptcy Prediction Using Machine Learning Techniques: Evidence on Indian Companies Under Insolvency And Bankruptcy Code. The Journal of Prediction Markets. 2022;16(2):77–100.
11. Beaver WH. Financial Ratios As Predictors of Failure. Journal of Accounting Research. 1966;4(1966):71–111.
12. Altman EI. Financial Ratios, Discriminant Analysis And The Prediction Of Corporate Bankruptcy. The Journal of Finance. 1968;XXIII(4):589–609.
13. Ohlson JA. Financial Ratios and the Probabilistic Prediction of Bankruptcy. Journal of Accounting Research. 1980;18(1):109.
14. Altman EI, Loris B. A Financial Early Warning System for Over-the-Counter Broker-Dealers. The Journal of Finance. 1976;31(4):1201–17.
15. Lin WY, Hu YH, Tsai CF. Machine learning in financial crisis prediction: A survey. IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews. 2012;42(4):421–36. Available from: https://ieeexplore.ieee.org/document/6069610.
16. Black WK, Calavita K, Pontell HN. The Savings and Loan Debacle of the 1980s: White-Collar Crime or Risky Business? Law & Policy. 1995 Jan 1;17(1):23–55. Available from: https://onlinelibrary.wiley.com /doi/10.1111/j.1467-9930.1995.tb00138.x.
17. Elhoseny M, Metawa N, Sztano G, El-hasnony IM. Deep Learning-Based Model for Financial Distress Prediction. Annals of Operations Research. 2022 May 25:1-13. https://doi.org/10.1007/s10479-022-04766-5
18. Rapanyane MB, Sethole FR. The rise of artificial intelligence and robots in the 4th Industrial Revolution: implications for future South African job creation. Contemporary Social Science. 2020;15(4):489–501.
19. Frydman H, Altman EI, Kao D -LI. Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress. The Journal of Finance. 1985;40(1):269–91.
20. Odom MD, Sharda R. A neural network model for bankruptcy prediction. In: IJCNN International Joint Conference on Neural Networks. 1990; 163–8.
21. Ravi Kumar P, Ravi V. Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review. European Journal of Operational Research. 2007;180(1):1–28.
22. Shi Y, Li X. An overview of bankruptcy prediction models for corporate firms: A systematic literature review. Intangible Capital. 2019;15(2):114–27.
23. Sun J, Li H, Huang QH, He KY. Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. Knowledge Based Systems. 2014;57:41–56.
24. Leary DEO. Using Neural Networks to Predict Corporate Failure. International Journal of Intelligent Systems in Accounting, Finance & Management. 1998;7:187–97.
25. Oyewo B, Ajibola O, Ajape M. Characteristics of consulting firms associated with the diffusion of big data analytics. Journal of Asian Business and Economic Studies. 2020;28(4):281–302.
26. Weber P, Carl KV, Hinz O. Applications of Explainable Artificial Intelligence in Finance—a systematic review of Finance, Information Systems, and Computer Science literature. Management Review Quarterly. 2024;74:867-907.
27. Arksey H, O'Malley L. Scoping studies: Towards a methodological framework. International Journal of Social Research Methodology: Theory and Practice. 2005;8(1):19–32.
28. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Medical Research Methodology. 2019;18(143):147–60.
29. Levac D, Colquhoun H, K O'Brien K. Scoping studies: advancing the methodology. Implementation Science. 2010;5:69:1–18.
30. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. Annals of Internal Medicine. 2018;169(7):467–73.

31. Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Research Synthesis Methods. 2020;11(2):181–217.
32. Valizadeh A, Moassefi M, Nakhostin-Ansari A, Hosseini Asl SH, Saghab Torbati M, Aghajani R, et al. Abstract screening using the automated tool Rayyan: results of effectiveness in three diagnostic test accuracy systematic reviews. BMC Medical Research Methodology. 2022 Dec 1;22(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9161508/.
33. Polanin JR, Pigott TD, Espelage DL, Grotpeter JK. Best practice guidelines for abstract screening large-evidence systematic reviews and meta-analyses. Research Synthesis Methods. 2019;10(3):330–42.
34. Tumpach M, Surovičová A, Juhászová Z, Marci A, Kubaščíková Z. Prediction of the bankruptcy of slovak companies using neural networks with SMOTE. Ekonomicky casopis. 2020;68(10):1021–39.
35. Sigrist F, Leuenberger N. Machine learning for corporate default risk: Multi-period prediction, frailty correlation, loan portfolios, and tail probabilities. European Journal of Operational Research. 2023;305(3):1390–406.
36. Jang Y, Jeong I, Cho YK. Identifying impact of variables in deep learning models on bankruptcy prediction of construction contractors. Engineering, Construction and Architectural Management. 2021;28(10):3282–98.
37. Bateni L, Asghari F. Bankruptcy Prediction Using Logit and Genetic Algorithm Models: A Comparative Analysis. Computational Economics. 2020;55(1):335–48.
38. Jandaghi G, Saranj A, Rajaei R, Ghasemi A, Tehrani R. Identification of the Most Critical Factors in Bankruptcy Prediction and Credit Classification of Companies. Iranian journal of Management Studies. 2021;14(4):817–34.
39. Jeong J, Kim C. Comparison of Machine Learning Approaches for Medium-to-Long-Term Financial Distress Predictions in the Construction Industry. Buildings. 2022;12(10):1–15.
40. Aydin N, Sahin N, Deveci M, Pamucar D. Prediction of financial distress of companies with artificial neural networks and decision trees models. Machine Learning with Applications. 2022;10:100432.
41. Özparlak G, Özdemir dilidüzgün M. Corporate Bankruptcy Prediction Using Machine Learning Methods: The Case Of The USA. International Journal of Management Economics and Business. 2022;18(4):1007–32.
42. Tran Long K, Le Anh H, Nguyen Hien T, Nguyen Trung D. Explainable Machine Learning for Financial Distress Prediction: Evidence from Vietnam. Data. 2022;7(160):1–12.
43. Muparuri L, Gumbo V. On logit and artificial neural networks in corporate distress modelling for Zimbabwe listed corporates. Sustainability Analytics and Modeling. 2022;2(May):100006.
44. Chen YS, Lin CK, Lo CM, Chen SF, Liao QJ. Comparable studies of financial bankruptcy prediction using advanced hybrid intelligent classification models to provide early warning in the electronics industry. Mathematics. 2021;9(20):1–26.
45. Shafiee M, Fakhari H. Evaluation of Back Propagation-Artificial Neural Network (BP-ANN) Fit Rate and Types of Vector Machine Algorithms in Estimating the Bankruptcy Prediction of Companies Listed on Tehran Stock Exchange. Turkish Journal of Computer and Mathematics Education. 2021;12(14):1854–68.
46. Radovanovic J, Haas C. The evaluation of bankruptcy prediction models based on socio-economic costs. Expert Systems with Applications. 2023;227:1–18. Available from: https://doi.org/10.1016/j.eswa.2023.120275.
47. Jencova S, Petruska I, Lukacova M, Abu-Zaid J. Prediction of bankruptcy in non-financial corporations using neural network. Montenegrin Journal of Economics. 2021;17(4):123–34.
48. Garcia J. Bankruptcy prediction using synthetic sampling. Machine Learning with Applications. 2022;9(May 100343):1–11.
49. Brenes RF, Johannssen A, Chukhrova N. An intelligent bankruptcy prediction model using a multilayer perceptron. Intelligent Systems with Applications. 2022;16:1–18.
50. Pérez-Pons ME, Parra-Dominguez J, Hernández G, Herrera-Viedma E, Corchado JM. Evaluation metrics and dimensional reduction for binary classification algorithms: A case study on bankruptcy prediction. Knowledge Engineering Review. 2022;37(4):1–18.
51. Perboli G, Arabnezhad E. A Machine Learning-based DSS for mid and long-term company crisis prediction. Expert Systems with Applications. 2021;174(July 2020):114758.
52. Kristóf T, Virág M. EU-27 bank failure prediction with C5.0 decision trees and deep learning neural networks. Research in International Business and Finance. 2022;61:1–17.
53. Smiti S, Soui M. Bankruptcy Prediction Using Deep Learning Approach Based on Borderline SMOTE. Information Systems Frontiers. 2020;22(5):1067–83.
54. Shrivastav SK, Janaki Ramudu P. Bankruptcy prediction and stress quantification using support vector machine: Evidence from Indian banks. Risks. 2020;8(2):1–22.
55. Vochozka M, Vrbka J, Suler P. Bankruptcy or success? The effective prediction of a company's financial development using LSTM. Sustainability. 2020;12(18):1–17.
56. Tabbakha A, Routa JK, Sahoob KS, Jha N. Bankruptcy Prediction using Robust Machine Learning Model. Turkish Journal of Computer and Mathematics Education. 2021;12(1):3060–73.
57. Wang H, Liu X. Undersampling bankruptcy prediction: Taiwan bankruptcy data. PLoS One. 2021;16(7 July):1–18.
58. Muslim MA, Dasril Y. Company bankruptcy prediction framework based on the most influential features using XGBoost and stacking ensemble learning. International Journal of Electrical and Computer Engineering. 2021;11(6):5549–57.
59. Shetty S, Musa M, Brédart X. Bankruptcy Prediction Using Machine Learning Techniques. Journal of Risk and Financial Management. 2022;15(1):1–10.
60. Pietrzak M. Can financial sector distress be detected early? Borsa Istanbul Review. 2022;22(6):1132–44.

61. Aker Y, Karavardar A. Using Machine Learning Methods in Financial Distress Prediction: Sample of Small and Medium Sized Enterprises Operating in Turkey. Ege Academic Review. 2023;23(2):145–61. Available from: https://www.proquest.com/docview/2813696278/CB4AFB49FFE64267PQ/1?accountid=38885.
62. Lohmann PC, Möllenhoff S. How Do Bankruptcy Risk Estimations Change in Time? Empirical Evidence From Listed US Companies. Finance Research Letters. 2023;104389:1–15. Available from: https://doi.org/10.1016/j.frl.2023.104389.