UNIVERSITY OF OKLAHOMA GRADUATE COLLEGE

INTEGRATING DEEP LEARNING AND PERSISTENT HOMOLOGY FOR ENHANCED FINANCIAL RISK ASSESSMENT

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfilment of the requirements for the Degree of

DOCTOR OF PHILOSOPHY

By GHANESHVAR RAMINENI CHITTIBABU NAIDU

Norman, Oklahoma

2025

INTEGRATING DEEP LEARNING AND PERSISTENT HOMOLOGY FOR ENHANCED FINANCIAL RISK ASSESSMENT

A DISSERTATION APPROVED FOR THE

SCHOOL OF INDUSTRIAL AND SYSTEMS ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Kash Barker, Committee Co-Chair

Dr. Talayeh Razzaghi, Committee Co-Chair

Dr. Naveen Kumar, Graduate College Representative

Dr. Andrés González, Committee Member

Dr. Yifu Li, Committee Member

Acknowledgements

I would like to express my sincere gratitude to my advisors, Dr. Kash Barker and Dr. Talayeh Razzaghi, for their continuous support, valuable guidance, and encouragement throughout the course of my Ph.D. Their insights and mentorship were crucial in shaping this research. I am also thankful to my committee members for their constructive feedback and support at every stage. Thank you, Dr. Barker, for EVERYTHING! I would also like to thank my committee members for their valuable feedback and support.

I extend my appreciation to the faculty and staff of the School of Industrial and Systems Engineering, University of Oklahoma, for fostering a collaborative academic environment and providing the resources essential for this work.

My heartfelt thanks go to my lab mates, colleagues, and friends who contributed to this journey with their feedback and encouragement. In particular, I would like to thank Rachel for her help, my former lab mates Nafiseh, Leili, Reza, Alireza, and Safa, and my current lab mates Mahdi, Shaghayegh, and Alice for the camaraderie and encouragement. Special thanks to Dr. Cilali and (soon to be) Dr. Emre, we did it!

I am also grateful to my therapists, Dr. Mary George Ittoop, Dr. Abhishek, and Dr. Sadahana, whose guidance helped me stay balanced through challenging times.

A special mention goes to my bhais at 930 — Mayank, Rishab, and Harsh — for their friendship and constant encouragement. Thank you for taking care of me, I am grateful!

Thank you to all my cousins, aunts, uncles, and my grandmother for their unwavering support from afar. Thank you Crackheads and Tarak anna, for the inspiration and always being there.

Finally, I am deeply grateful to my wife, parents, sister, and brother-in-law for their unconditional love, patience, and belief in me, which gave me strength throughout this academic journey.

Table of Contents

List of Tables

List of Figures

Abstract

Accurately predicting bankruptcy is a critical challenge for financial institutions, businesses, and policymakers. Traditional bankruptcy prediction models rely on financial ratios and historical trends, but they often fail to capture complex patterns within financial data. With the increasing availability of large-scale financial datasets, there is a growing need for more advanced methodologies that can enhance predictive accuracy and provide deeper insights into financial risk assessment.

The necessity for improved bankruptcy prediction stems from the economic and societal consequences of corporate financial distress. Ineffective risk assessment can lead to cascading failures in financial markets, supply chains, and employment sectors. As financial systems become more intricate, the ability to predict and mitigate bankruptcy risk is crucial for maintaining stability, reducing losses, and informing strategic decision-making. Deep learning techniques offer a promising avenue for capturing the temporal dependencies in financial data, while topological data analysis provides structural insights that go beyond conventional feature engineering.

This dissertation presents a unique approach to bankruptcy prediction by integrating deep learning models with topological data analysis. While existing models rely primarily on statistical methods, this study incorporates Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) networks to analyze temporal financial trends. Additionally, persistent homology is leveraged to extract topological features, addressing gaps in traditional risk assessment models. By combining these methodologies, this research aims to improve prediction accuracy and enhance the interpretability of financial risk factors.

To validate the effectiveness of the proposed approach, multiple evaluation techniques are employed. Model performance is assessed using standard machine learning metrics such as accuracy, precision, recall, and F1-score. Comparative analysis is conducted against traditional bankruptcy prediction models, highlighting the advantages of integrating deep learning and topological insights. Robustness tests are performed to ensure the reliability of extracted features,

and explainability metrics are used to interpret the influence of topological data in bankruptcy forecasting.

The results of this study demonstrate significant improvements in bankruptcy prediction accuracy compared to conventional models. By integrating deep learning and PH, this research provides a novel framework that enhances financial risk assessment. The findings have valuable implications for financial institutions seeking to refine predictive analytics, reduce investment risks, and improve early warning systems. Exploring PH in temporal data is crucial for advancing financial forecasting and risk assessment, especially when combined with deep learning models, enabling the development of more sophisticated, data-driven decision-making strategies.

Chapter 1: Introduction and Motivation

When a company is in financial distress, owners, employees, and shareholders are adversely impacted, but so are clients that purchase from those companies. An ability to predict the financial distress of a company has wide ranging implications – from financial institutions making lending decisions, to investors and fund managers making investment decisions, to organizations looking to enter supply contracts. While this last implication is especially important to entities with vulnerable supply chain as the financial distress of a company may threaten their performance, and organizations across many industry sectors seek to contract with suppliers that are not at risk of bankruptcy. In 2019, an electric utility, PG&E Corporation in California, filed the biggest bankruptcy of the year losing over $71 billion in assets [Telford and Mufson 2019]. With billions of dollars and thousands of jobs at stake, it is crucial to be able to effectively categorize financially healthy and unhealthy institutions.

For the last 100 years, bankruptcy prediction models have played an important role in investment and supplier selection decision-making. At higher levels, these models receive significant attention by policy makers to design proactive policies with the aim of minimizing adverse impacts of bankruptcies on the economy. With today's increasingly complex economic systems and their deep interaction with a network of convoluted stakeholders and competitors, modelers have now access to substantially more relevant features that can help discover hidden financial distress patterns. As a result, more accurate bankruptcy prediction models can be developed to provide early signals of financial distress leading to bankruptcy.

This dissertation explores novel methodologies for bankruptcy prediction by integrating temporal deep learning models and topological data analysis (TDA). Specifically, it examines the potential of deep learning architectures such as Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) to learn temporal dependencies in financial data, and extends this framework by incorporating persistent homology (PH), a tool from TDA, to extract meaningful topological features that describe the underlying structure of financial

dynamics. Together, these methods aim to provide a more holistic understanding of corporate distress and significantly improve predictive performance beyond conventional models.

Chapter 3 presents the first study, which investigates the use of deep learning models on temporal financial data to enhance bankruptcy prediction accuracy. CNN and BiLSTM models are implemented to capture local patterns and long-term dependencies within sequential financial records. Feature engineering is used to select a set of 12 financial indicators that serve as input across time. The study evaluates the performance of these models using multiple variations of the dataset, showing how temporal modeling can reveal hidden signals that static models typically overlook. Through extensive experimentation, the findings demonstrate the strength of deep learning methods in recognizing early warning signs of financial instability.

Building on this foundation, Chapter 4 introduces a new layer of insight by incorporating PH, a mathematical approach from TDA. This method extracts shape-based, structural features from financial data, representing how changes in relationships between features evolve over time. These topological features are integrated with the original financial inputs to test whether they can boost the predictive power of CNN and BiLSTM models. The results highlight cases where the addition of topological descriptors not only improves prediction accuracy but also increases recall, suggesting a better sensitivity to identifying bankrupt firms. The study offers compelling evidence that topological insights capture information not accessible to traditional models or even deep learning alone.

Together, these two studies demonstrate a synergistic approach to bankruptcy prediction by combining data-driven learning and shape-aware modeling. This dissertation contributes to the growing literature on financial risk modeling by showcasing how hybrid techniques — temporal deep learning and TDA — can together provide more robust, interpretable, and accurate predictions. These insights can inform decision-making for financial institutions, regulators, and corporate managers, ultimately contributing to stronger early warning systems and more resilient financial ecosystems.

The remainder of this dissertation is organized as follows. Chapter 2 provides a review of the relevant literature on bankruptcy prediction and persistent homology, highlighting existing methodologies and gaps in the field. Chapter 3 presents the first study, which explores the use of temporal financial data combined with deep learning models such as BiLSTM and CNN for bankruptcy prediction. Chapter 4 builds upon this work by integrating topological features derived through PH to further enhance predictive performance. Chapter 5 offers concluding remarks, reflecting on the key findings and contributions of the research. Chapter 6 compiles the references used throughout the dissertation.

Chapter 2: Literature Review and Background

2.1 Bankruptcy Prediction

The initial models that were developed to study the prediction of failure of companies in the early 1920s were based on financial ratio analysis [Bliss 1923]. Ratio analysis helped gain insights into a company's liquidity, efficiency, profitability, and survivability. Factors like the size and industry of the company were considered when comparisons were made among different companies [Littleton 1926]. FitzPatrick [1932] compares 13 financial ratios of 19 pairs of successful and failed companies and observed that there existed a consistent change in the ratios at least three years prior to the failure point. Winakor and Smith [1935] investigate mean ratios of 183 failed companies for a period of ten years prior to failure and found that the rate of decline of these mean ratios was higher closer to the failure, especially the "current assets to total assets" ratio. Mervin [1942] focused on comparing mean ratios of continuing and discontinued companies between 1926 and 1936 and observe a pattern of increased differences in the ratios as the year of discontinuance approached. Recent works do not explore the deteriorating trends in the ratios and feature values closer to the bankruptcy point. We intend to capture this essence by exploring 3-year and 5-year data prior to bankruptcy (explained in section 4.2).

Until the mid-1960s, bankruptcy prediction studies continued to focus on ratio analysis and statistical models. Beaver [1966] introduced the first univariate model in classifying bankrupt companies, observing that the ratio of net income to total debt had the best predictive ability followed by the ratio of net income to sales. Beaver [1966] also suggests the use of multiple ratios simultaneously for predicting bankruptcy, leading to the multivariate models. Among the most popular techniques for predicting bankruptcy is the Z-score approach [Altman 1968], which provides a classification scheme based on five financial characteristics of a company. Such a score, which has seen several iterations since [Altman et al. 1977, 2014], provides more of a classification scheme than a means to calculate an implied probability of bankruptcy. A common heuristic suggests that a company with a Z-score < 1.81 is likely to be under financial distress or heading towards bankruptcy. The popular Altman's Z-score, provided in Eq. (3.1), served as a foundational means to understand the temporal nature of the dataset. The score is a function of five ratios, found in the first column of Table 3.1, of seven total variables. The Z-score was developed to describe

the bankruptcy risk of companies in the manufacturing sector. Barboza et al. [2017] explored six ratios, in addition to the ones used in calculating Altman's Z-score, that were inspired from Carton and Hofer [2006]. These six additional ratios are found in the second column of Table 3.1. For companies outside the manufacturing sector, a recent update to the Z-score [Altman et al. 2014], denoted Z'', is provided in Eq. (3.2). A Z''-score < 1.1 implies that a company is under financial distress. The first four variables from Table 3.1 are used in the calculation of the Z''-score.

With increasing efficiency of data collection, the availability of machine learning models has provided newfound sophistication and accuracy in predicting bankruptcy, at the cost of creating complex and less explainable models. While machine learning models help improve the accuracy of the predictions, feature selection techniques are especially useful in narrowing the vast number of variables that might be involved. Recent techniques that have been applied to bankruptcy prediction for several financial variables include neural networks [Cleofas-Sánchez et al. 2016; Heo and Yang 2014; López Iturriaga and Sanz 2015; Tsai et al. 2014; Wilson and Sharda 2018], regression techniques [Chandra et al. 2009], stochastic processes [Antunes et al. 2017] , over sampling with neural networks [H. Kim et al. 2022; Smiti and Soui 2020], feature selection engineering [Ben Jabeur et al. 2023] and regression trees [Booth et al. 2014; Liang et al. 2016; Wang et al. 2014], among others [Barboza et al. 2017; Hu 2020; S. Y. Kim and Upneja 2014; Min and Lee 2005; Uthayakumar et al. 2018, 2020; C. H. Wang and Chen 2021]. Some studies have inferred that a set of limited features is sufficient in predicting bankruptcy accurately [Barboza et al. 2017; Yu et al. 2014]. Barboza et al. [2017] developed several bankruptcy prediction models for data obtained from 10 types of industries and achieved an AUC of 92.97% for boosting, 92.48% for bagging, and 92.92% for random forest models, using just 11 attributes/features. However, very few of these works [Booth et al. 2014; López Iturriaga and Sanz 2015] address the influence of dynamic features using the time series data in their prediction models. Bellovary et al. [2007] and Wu et al. [2010] have reviewed the progress of a wide range of bankruptcy models in the past few decades.

Motivated by Barboza et al. [2017], integrating dynamic predictor variables into prediction models can improve prediction accuracy. In addition, bankruptcy may be related to many

explanatory variables (features) that previous studies have shown to be effectively modeled by non-linear techniques [Barboza et al. 2017; Mai et al. 2019]. Advanced machine learning techniques, such as neural networks and deep learning, have been successful in not only modeling non-linear decision boundaries but also modeling data with sequential predictor variables (such as time series). Two popular variants of deep learning methods that are especially effective for time series data are long short-term memory (LSTM) neural networks and convolutional neural networks (CNN). LSTM is a recurrent neural network model built upon feedback connections, and has been very successful for time series prediction and classification problems [Gamboa 2017; Graves and Schmidhuber 2005; H. Kim et al. 2022; Siami-Namini et al. 2019]. In this study, we particularly employed a bidirectional LSTM model (BiLSTM) that processes the input in two directions from past time periods to future periods and vice versa. CNN represents a class of deep neural networks with multiple layers between the input and output layers, employing the mathematical function of convolution in its operation. It has been used for many applications among facial recognition [Lawrence et al. 1997], image classification [Krizhevsky et al. 2017], fraud detection [Ghosh and Reilly 1994], speech recognition [Nal et al. 2014], and time series analysis [Yang et al. 2015]. Moreover, Nguyen et al. [2023] use SHAP technique to interpret their bankruptcy prediction results.

Studies have been conducted on financial analysis and bankruptcy prediction using neural networks and deep learning methods [Hosaka 2019; Kim and Kang 2010; Odom and Sharda 1990; Wilson and Sharda 2018; Yang et al. 1999]. Mai et al. [2019] focus on using deep learning models for bankruptcy prediction using textual disclosures. Qu et al. [2019] provide a review on the different machine learning and neural network models used in bankruptcy prediction. Limitations of these works include (i) not considering the impact of dynamic features to represent temporal company performance and (ii) the treatment of imbalance issues in the dataset. Since the number of companies that undergo bankruptcy is very small relative to those that do not, it is expected that a bankruptcy dataset would be imbalanced. As such, it is imperative that this imbalance is accounted for when building a predictive model. In addition, there have not been adequate computational experiments that evaluate the applicability of deep learning algorithms to handle imbalanced data in bankruptcy prediction problems.

2.2 Persistent Homology (PH)

Topological data analysis (TDA) has emerged as a powerful mathematical framework for uncovering intrinsic structural patterns in complex, high-dimensional datasets. Rooted in algebraic topology, TDA emphasizes the "shape" of data by identifying features such as connected components, loops, and voids that persist across multiple scales. Its strength lies in providing insights beyond those offered by traditional statistical or geometric methods, especially in scenarios where the data lacks an obvious parametric structure or contains noise. Applications of TDA have spanned numerous domains including biology, sensor networks, image processing, neuroscience, and economics, demonstrating its versatility in both qualitative and quantitative analyses [Carlsson, 2009]. One of the foundational tools within TDA is PH, which captures the birth and death of topological features as a filtration parameter varies, providing a multiscale summary of data topology.

PH stands out within TDA due to its ability to provide a concise yet informative summary of complex structures in the form of barcodes or persistence diagrams. Unlike conventional topological n a single scale or parameterization, PH systematically tracks topological features over a continuum of scales, enabling the quantification of both local and global structures [Zomorodian & Carlsson, 2004]. This makes PH particularly suitable for studying noisy or high-dimensional data, where such features may only emerge at specific scales. Its applications span a wide range, including shape recognition, biological data analysis, material science, and dynamic networks [Pun et al. 2022]. Furthermore, PH has been enriched by computational advancements and software tools such as Ripser, Dionysus, and GUDHI, which have facilitated its practical integration into modern data analysis pipelines [Otter et al. 2017].

More recently, PH has been applied not just as an exploratory tool but as a source of features for machine learning and predictive modeling. In this context, PH transforms complex structures into usable numerical representations—such as barcode statistics, persistence landscapes, or persistence images—that can be incorporated into classification, regression, or clustering models [Adams et al. 2017, Bubenik, 2015]. A significant advancement in this direction is the development of persistent-homology-based machine learning (PHML), which formalizes the

integration of topological summaries into learning frameworks like support vector machines, decision trees, and neural networks. For instance, [Pun et al. 2022] provide a comparative study on feature representations derived from PH and their application across multiple domains such as protein structure classification and image recognition. These methods have also demonstrated robustness in capturing nonlinear dependencies and invariances, making them well-suited for financial and temporal datasets.

In predictive analytics, PH has proven especially effective for modeling temporal financial data and understanding time-evolving systems. Unlike static descriptors, PH-based representations can account for structural transitions and trends over time, capturing subtle anomalies or shifts that might signal underlying systemic risks (Seversky et al. 2016). This is particularly valuable in fields such as finance, where early detection of distress or failure is paramount. The application of PH to time series and sequential data is facilitated by specialized filtrations such as sublevel set filtrations or temporal filtrations, allowing the capture of topological changes in evolving datasets [Pal et al. 2019]. When paired with machine learning models, these topological summaries offer a powerful hybrid approach for early warning systems, outperforming traditional metrics in several benchmarking tasks. As such, PH is increasingly recognized not just as a theoretical tool but as a practical component in the modern data science toolkit, particularly in domains requiring nuanced interpretation of complex temporal structures.

Chapter 3: Predicting Bankruptcy Risk with Bidirectional Long Short-Term Memory and Convolutional Neural Networks

3.1 Background

In an increasingly volatile and interconnected global economy, the ability to accurately predict corporate bankruptcy is of paramount importance. Bankruptcy not only disrupts business continuity but also has far-reaching consequences for creditors, investors, employees, and economic systems at large. While traditional bankruptcy prediction models rely heavily on financial ratios and static indicators, these methods often fail to capture the complex, temporal patterns and nonlinear relationships that characterize financial distress in modern enterprises. With the advent of high-dimensional data and more dynamic financial environments, there is a growing need for predictive approaches that are both adaptable and capable of uncovering hidden structures in data.

The contributions of this chapter are several. First, we provide a larger set of financial features using the COMPUSTAT database and identify the critical features relevant to bankruptcy prediction using the recursive feature elimination technique. Our methodological scheme enables us to explore 54 features from three datasets generated from a point in time relative to the occurrence of bankruptcy. Second, we construct two deep learning neural network models, BiLSTM and CNN, which are appropriate for both balanced and imbalanced time series datasets. Lastly, we demonstrate the effectiveness of the proposed models using COMPUSTAT data and explore the advantage of multiple temporal variations of the features that can discriminate the bankrupt and non-bankrupt companies at different periods prior to bankruptcy. Furthermore, we employ the cost-sensitive learning method embedded with BiLSTM and CNN for treating the imbalance problem associated with few bankrupt companies existing in the dataset. In addition, an approach to predict bankruptcy in advance was implemented by applying the above techniques to a dataset composed of three years of data collected two years in advance of the year when bankruptcy was experienced.

The rest of this chapter is arranged as follows. Section 2 elaborates on the literature survey on various bankruptcy models over the decades. Section 3 consists of the data analyzed in this

study and a brief description of deep learning models. Section 4 includes the implementation of feature selection, the classification results of the models, a discussion of comparative results, and model interpretation using SHAP values followed by concluding remarks and future work in Section 5.

3.2 Literature Review

Until the mid-1960s, bankruptcy prediction studies continued to focus on ratio analysis and statistical models. Beaver [1966] introduced the first univariate model in classifying bankrupt companies, observing that the ratio of net income to total debt had the best predictive ability followed by the ratio of net income to sales. Beaver [1966] also suggests the use of multiple ratios simultaneously for predicting bankruptcy, leading to the multivariate models. Among the most popular techniques for predicting bankruptcy is the Z-score approach [Altman 1968], which provides a classification scheme based on five financial characteristics of a company. Such a score, which has seen several iterations since [Altman et al. 1977, 2014], provides more of a classification scheme than a means to calculate an implied probability of bankruptcy. A common heuristic suggests that a company with a Z-score < 1.81 is likely to be under financial distress or heading towards bankruptcy. The popular Altman's Z-score, provided in Eq. (3.1), served as a foundational means to understand the temporal nature of the dataset. The score is a function of five ratios, found in the first column of Table 3.1, of seven total variables. The Z-score was developed to describe the bankruptcy risk of companies in the manufacturing sector. Barboza et al. [2017] explored six ratios, in addition to the ones used in calculating Altman's Z-score, that were inspired from Carton and Hofer [2006]. These six additional ratios are found in the second column of Table 3.1.

$$Z = 1.2A + 1.4B + 3.3C + 0.6D + 1.0E \tag{3.1}$$

For companies outside the manufacturing sector, a recent update to the Z-score [Altman et al. 2014], denoted Z'', is provided in Eq. (3.2). A Z''-score < 1.1 implies that a company is under financial distress. The first four variables from Table 3.1 are used in the calculation of the Z''-score.

$$Z'' = 6.56A + 3.26B + 6.72C + 1.05D \tag{3.2}$$

Table 3.1 Eleven variables analyzed by Barboza et al. [2017], including the five ratios used in the calculation of Altman's Z-score.

Variables from Altman [1968]	Variables from Carton and Hofer [2006]
$A = \frac{\text{net working capital}}{\text{total assets}}$	$F = \frac{\text{earnings before interest and taxes}}{\text{sales}}$
$B = \frac{\text{retained earnings}}{\text{total assets}}$	$G = \frac{\text{total assets}_t - t\text{otal assets}_{t-1}}{\text{total assets}_{t-1}}$
$C = \frac{\text{earnings before interest and taxes}}{\text{total assets}}$	$H = \frac{\text{sales}_t - \text{sales}_{t-1}}{\text{sales}_{t-1}}$
$D = \frac{\text{market value of shares} \times \text{number of shares}}{\text{total debt}}$	$I = \frac{\#\text{ employees}_t - \#\text{ employees}_{t-1}}{\#\text{ employees}_{t-1}}$
$E = \frac{\text{sales}}{\text{total assets}}$	$J = \left(\frac{\text{net income}}{\text{stockholder equity}}\right)_t - \left(\frac{\text{net income}}{\text{stockholder equity}}\right)_{t-1}$
	$K = \left(\frac{\text{market value/share}}{\text{book value/share}}\right)_t - \left(\frac{\text{market value/share}}{\text{book value/share}}\right)_{t-1}$

With increasing efficiency of data collection, the availability of machine learning models has provided newfound sophistication and accuracy in predicting bankruptcy, at the cost of creating complex and less explainable models. While machine learning models help improve the accuracy of the predictions, feature selection techniques are especially useful in narrowing the vast number of variables that might be involved. Recent techniques that have been applied to bankruptcy prediction for several financial variables include neural networks [Cleofas-Sánchez et al. 2016; Heo and Yang 2014; López Iturriaga and Sanz 2015; Tsai et al. 2014; Wilson and Sharda 2018], regression techniques [Chandra et al. 2009], stochastic processes [Antunes et al. 2017] , over sampling with neural networks [H. Kim et al. 2022; Smiti and Soui 2020], feature selection engineering [Ben Jabeur et al. 2023] and regression trees [Booth et al. 2014; Liang et al. 2016; Wang et al. 2014], among others [Barboza et al. 2017; Hu 2020; S. Y. Kim and Upneja 2014; Min

and Lee 2005; Uthayakumar et al. 2018, 2020; C. H. Wang and Chen 2021]. Some studies have inferred that a set of limited features is sufficient in predicting bankruptcy accurately [Barboza et al. 2017; Yu et al. 2014]. Barboza et al. [2017] developed several bankruptcy prediction models for data obtained from 10 types of industries and achieved an AUC of 92.97% for boosting, 92.48% for bagging, and 92.92% for random forest models, using just 11 attributes/features. However, very few of these works [Booth et al. 2014; López Iturriaga and Sanz 2015] address the influence of dynamic features using the time series data in their prediction models. Bellovary et al. [2007] and Wu et al. [2010] have reviewed the progress of a wide range of bankruptcy models in the past few decades.

3.3 Data and Methods

This section provides some background on the data used for bankruptcy prediction, as well as the two deep learning neural network methods along with the feature selection technique used in the analysis.

3.3.1 Data

The data used in this analysis are obtained from Standard & Poor's (S&P) COMPUSTAT database. The COMPUSTAT database contains fundamental financial, statistical, and market information for active and inactive publicly held companies from around the world. For most companies, the database has annual data since 1950 and quarterly data since 1962. The annual data includes all financial data from a company's Form 10-K, including company descriptions, balance sheet items, income statement data, and cash flow information, as well as standard industry classifiers, market identifiers, and S&P proprietary data. Some examples of these variables include a business description, stock ticker, market value of shares, cash on hand, and net sales. Similarly, the quarterly data includes all data from Form 10-Q's, like total assets, total liabilities, and retained earnings. In total, there are 974 non-screening variables for annual data and 674 for quarterly data. The COMPUSTAT database is updated daily. For this analysis, annual data from the North American companies, mostly from the United States of America and Canada, ranging from 1994 to 2020 were used. The data is queried based on a list of Standard Industrial Classification (SIC) codes, which the Security Exchange Commission uses to classify company industries. The sectors

considered for the analysis are shown in Table 3.3. The bankrupt label is assigned to companies that have filed for bankruptcy (e.g., bankruptcy Chapter 7 occurs when a company's assets are liquidated and the company ceases to exist or Chapter 11, or "reorganization bankruptcy," occurs when a company is reorganized to a different entity while continuing to operate).

Figure 3.1 represents the average Z-score plot for 7572 manufacturing companies (295 bankrupt and 7277 non-bankrupt) on the vertical axis and a time reference on the horizontal axis. The time reference depicts the most recent year of company data on the right-hand side. For non-bankrupt companies, the right-hand side represents the most recent available observation from the database, while for bankrupt companies, the right-hand side represents the last observation before the bankruptcy event. Time prior to the most recent year extends to the left. For example, for a bankrupt company, three time periods from the right represents three years prior to bankruptcy. Bankrupt companies are represented with red, non-bankrupt with green. The data for bankrupt companies after the bankrupt point were ignored. This was done to maintain uniformity among the bankrupt companies as few companies reorganize their financials and continue to operate, sometimes as a different entity, while others cease to exist altogether.

Table 3.2 provides an overview of the data, consisting of information about companies from 1994-2020, for the features in Table 3.5. Please refer Table 3.5 for the definitions of the features. The unit of measurement of some features such as 'lo', 'lt', 'ebit', and 'dltt' is $ million and that of 'growth in emp' and 'sale' is in units. The inf and -inf values occur when the denominator is a 0 while computing its value. We can observe that some features such as 'adjex_f', 'ebit', and 'cogs' are highly skewed. This is because the data includes companies ranging from small to large. Figure 3.1 represents the average Z-score plot for 7572 manufacturing companies (295 bankrupt and 7277 non-bankrupt) on the vertical axis and a time reference on the horizontal axis. The time reference depicts the most recent year of company data on the right-hand side. For non-bankrupt companies, the right-hand side represents the most recent available observation from the database, while for bankrupt companies, the right-hand side represents the last observation before the bankruptcy event. Time prior to the most recent year extends to the left. For example, for a bankrupt company, three time periods from the right represents three years prior to

bankruptcy. Bankrupt companies are represented with red, non-bankrupt with green. The data for bankrupt companies after the bankrupt point were ignored. This was done to maintain uniformity among the bankrupt companies as few companies reorganize their financials and continue to operate, sometimes as a different entity, while others cease to exist altogether.

Table 3.2 A statistical review of the data for a set of selected features.

Feature	Count	Mean	Std	Min	25%	50%	75%	Max
change in return on equity	161732	-	-	(-)inf	0.077	0.793	1.284	inf
lo	180589	336.419	2186.952	-27.06	0	1.641	40.9	104481
adjex_f	172759	1.277	4.28	0	1	1	1	822.5
lt	180434	2118.845	10378.5	0	7.355	60.614	606.192	460442
D	107394	inf	-	0	2.236	9.813	inf	inf
ib	180019	128.114	1119.308	-85162.2	-4.702	0.888	31.371	104821
ebit	180357	833.515	4001.36	-0.023	0	8.621	250.254	207174
dltt	132859	8.863	42.563	-1	-0.878	-0.125	3.766	2299
growth in emp	179061	257.174	1426.221	-25913	-2.141	4.249	72.381	71230
leverage	155837	inf	-	0	0.819	2.111	6.369	inf
sale	179799	2508.935	12858.86	-1965	9.8685	121.15	883.9	521426
cogs	179748	1745.547	9894.701	-366.645	7.643	75.325	574.048	435726.3

We can observe from the time series plot in Figure 3.1 that there is a distinct separation of average Z-score between the bankrupt and non-bankrupt companies in the manufacturing sector, especially close to the bankruptcy point. This separation motivated the use of time series methods to predict bankruptcy.

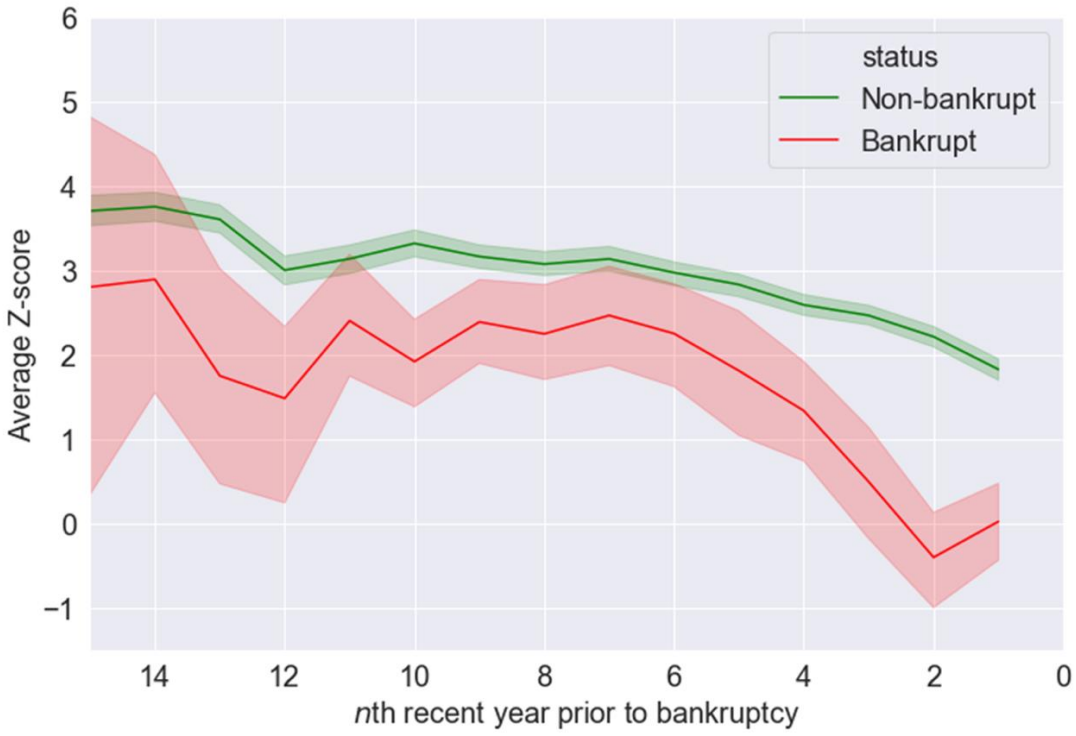


Figure 3.1 Average Z-score plot for manufacturing companies with Z-score in range [-15,15].

Table 3.3 shows the list of industry divisions with their related Standard Industrial Classification (SIC) codes. We note that companies from these industry sectors are considered for the analysis.

Table 3.3 List of the industry sectors and their corresponding SIC codes.

Range of SIC codes	Industry
A: 0100-0999	Agriculture, Forestry, and Fishing
B: 1000-1499	Mining
C: 1500-1799	Construction
D: 2000-3999	Manufacturing
E: 4000-4999	Transportation, Communications, Electric, Gas, and Sanitary Service
F: 5000-5199	Wholesale Trade
G: 5200-5999	Retail Trade

3.3.2 Recursive Feature Elimination

The recursive feature elimination (RFE) feature selection method utilizes a recursive approach to rank features based on their importance which was originally proposed by Guyon et al. [2002]. The algorithm calculates feature importance measures and eliminates the most irrelevant features at each iteration. At the end of the iteration process, a list of optimal features is obtained. Random forest classifier as a ranking method was used to implement this technique, where significant features are ranked as 1 and the rest are ranked as 0, thus eliminating them. The pseudo-code of the algorithm is shown in Algorithm 1.

Algorithm 3.1: Finding important features using a ranking method.

Inputs: T: Training set, $F = \{f_1, f_2, \dots, f_m\}$: a set of m features, $M(T, F)$: Ranking method

Output: R: List of features
for i in $\{1: m\}$
    Rank set F using $M(T, F)$
    $f^* \leftarrow$ Last ranked feature in F
    $R(m - i + 1) \leftarrow f^*$
    $F \leftarrow F - f^*$
end for
Return ***R***

3.3.3 Long Short-Term Memory Networks

Long short-term memory (LSTM) is an extended variation of recurrent neural network (RNN) architecture, proposed by [Hochreiter and Schmidhuber 1997]. In this section, the model of LSTM architecture for bankruptcy prediction is introduced. Unlike RNN, LSTM is powerful in handling the vanishing gradients and learn long-term dependencies in modelling the time series by using memory cells [How et al. 2016; Palangi et al. 2016]. A memory cell includes four units: an input gate, an output gate, a forget gate and a self-recurrent neuron. The interactions between neighboring memory cells and the memory cells themselves are controlled by the gates. The input gate controls the condition that the input signal changes the memory cell state. Instead, whether the state of memory cell can change the state of other memory cell is controlled by the output gate. Furthermore, the previous state of the memory cell will be remembered or forgotten by the forget

gate. The mathematical formulation of LSTM in Figure 3.2 is adapted from Bao et al. [2017]. Assume that a dataset is represented by a set $D = \{(X_l, y_l)\}$ where $l = 1, 2, \dots, n$, and $X_l = (x_1, x_2, \dots, x_T)_l$ represent an input sequence for sample l for an LSTM, $x_t \in \mathcal{R}^m$ is the m-dimensional input vector to the memory cell at time t, and $y_l \in \{0, 1\}$ is the associated class label; W_i, W_f, W_c, W_o, U_i, U_f, U_c, U_o, and V_o are weight matrices; b_i, b_f, b_c, and b_o are bias vectors; and h_t is the value of the memory cell at time t. We calculate the forward pass of the neural network through the following equations that describe LSTM cell vectors at time t. The values of the input gate (i_t) and the candidate state ($\tilde{C}_t$) of the memory cell at time t can be calculated with Eqs. (3.3) and (3.4).

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{3.3}$$

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{3.4}$$

The forget gate (f_t) and the state of the memory cell (C_t) at time t are formulated with Eqs. (3.5) and (3.6).

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{3.5}$$

$$C_t = i_t \odot \tilde{C}_t + f_t \odot C_{t-1} \tag{3.6}$$

Finally, the output gate (O_t) and memory cell (h_t) at time t are calculated with Eqs. (3.7) and (3.8).

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + V_o C_t + b_o) \tag{3.7}$$

$$h_t = o_t \odot \tanh(C_t) \tag{3.8}$$

In the above, $\odot$ represents element-wise multiplication, and σ and tanh are non-linear logistic sigmoid and hyperbolic tangent activation functions, respectively.

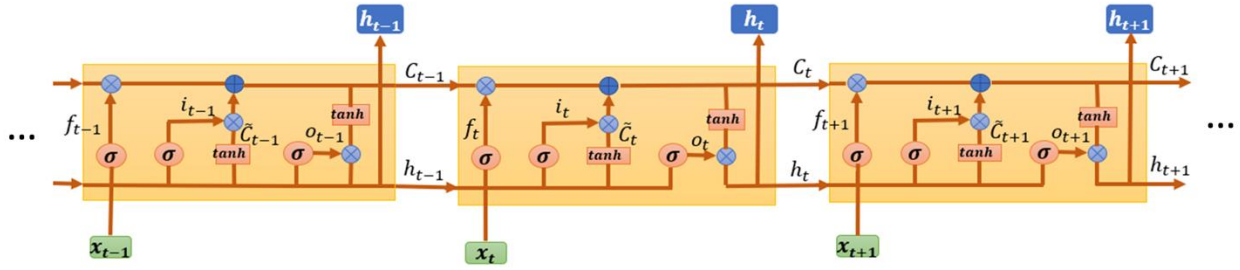


Figure 3.2 The sequential LSTM architecture overview. ⊗ and ⊕ denote element-wise multiplication and summation, respectively (adapted from Bao et al. [2017]).

3.3.4 Bidirectional LSTM

The classic LSTM is a unidirectional network that can only process data following one direction using past information. However, both past and future company status and attributes (temporal sequences) provide valuable information for accurate prediction of bankruptcy. Bidirectional LSTM (BiLSTM) is a well-known variant of LSTM first introduced by [Schuster and Paliwal 1997] that memorizes past and future information. BiLSTM consists of two hidden layers that are connected to the input and output. The sequence of operations in the first (forward) layer conforms to the same direction of the data sequence, while the other (backward) layer operates in the reverse direction. This structure enables the model to use additional information to process earlier time steps in reverse. Furthermore, BiLSTM has been successful in handling sequence data in many areas, including time series classification [Bao et al. 2017], natural language processing [Chen et al. 2017], speech recognition [Graves et al. 2013], and handwriting recognition [Graves et al. 2009]. Studies show BiLSTM performs better than LSTM for some applications that include time series data [Siami-Namini et al. 2019]. Therefore, BiLSTM is used in this study.

Figure 3.3 shows the methodology followed to construct the BiLSTM network. After comparing the performance of multiple architectures, the best performance was achieved by the BiLSTM architecture shown in Figure 3.4. The parameters of model are described in Table 3.9.

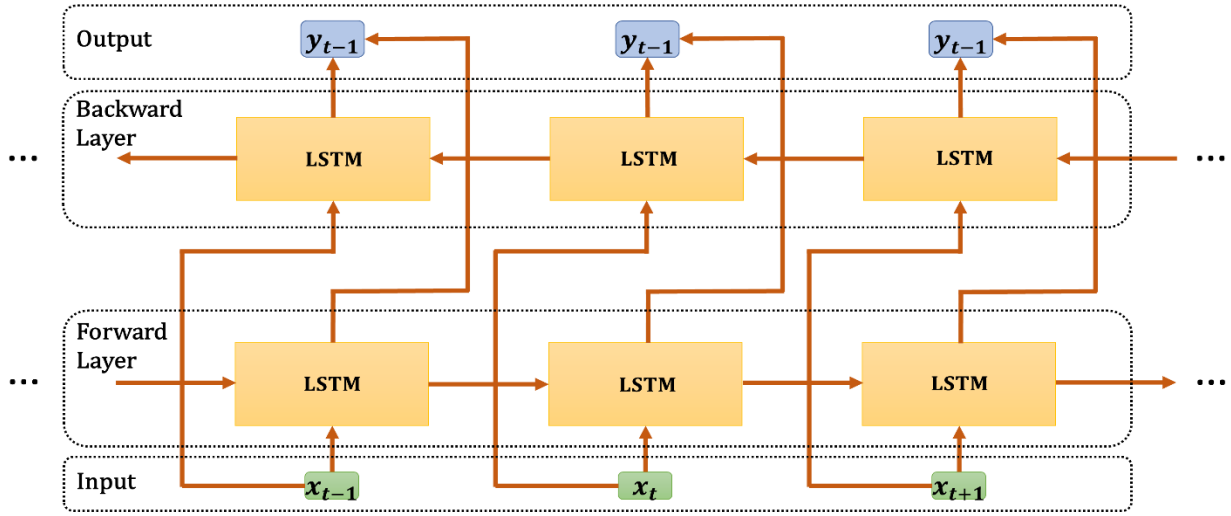


Figure 3.3 The sequential BiLSTM architecture overview adapted from [Fan et al. 2014].

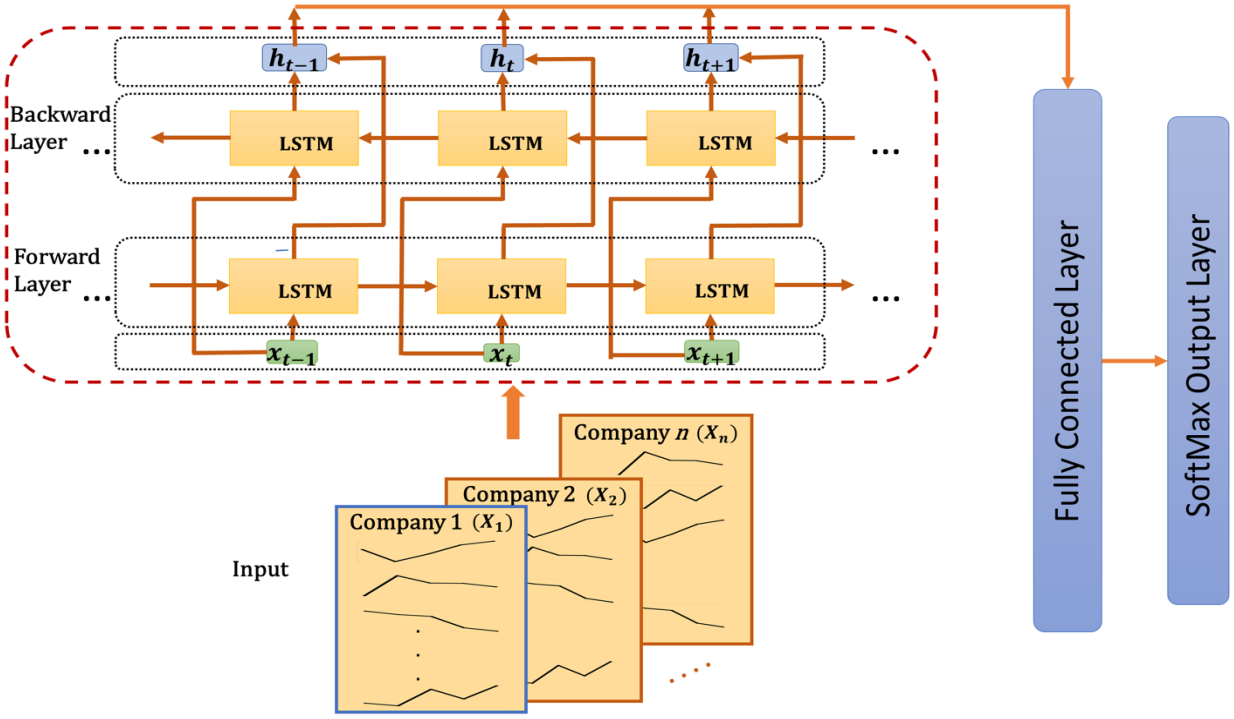


Figure 3.4 The architecture of the BiLSTM model for bankruptcy prediction.

3.3.5 Convolutional Neural Networks

A convolutional neural network (CNN) is a modification of the conventional artificial neural network (ANN), which has been successful in extracting high-level features from image [Cun et al. 1989], text [LeCun et al. 1998], and time series [Fuqua and Razzaghi 2020] data. In general, the steps involved in building a CNN model include (i) a convolution operation followed by activation function, (ii) a downsampling operation using pooling layers, (iii) flattening the pooled features, and (iv) computing a final score function using full-connected layers (i.e., conventional ANN layers). After feeding the input data into a convolutional layer, the local feature representation of data is created. Each convolutional layer is made up of multiple same collections of the identical neurons which are called filters (or kernels). Due to local connectivity of each neuron with only a subset of inputs, the CNN results into a model with sparse interactions and thus a reduced number of parameters sharing [LeCun et al. 1998]. Like ANN, an activation function is applied to each neuron in the convolutional layer to model the non-linearity in the output. For the activation function, we utilize the rectified linear unit (ReLU) function for all convolutional layers which is defined in Eq. (3.9).

$$g(x) = \max(0, x) \tag{3.9}$$

The most significant benefit of ReLU is that it is much faster in training compared to other activation functions [Krizhevsky et al. 2012]. As our objective is to predict the bankruptcy status of companies using multivariate time series data, we have developed a CNN model for this problem. Our CNN model is inspired by the notation introduced by Zheng et al. [2014]. We illustrate the architecture of our proposed CNN in Figure 3.5 which has been identified through extensive experiments for the appropriate number of filters, the size of the filters, and the number of pooling layers. The parameters of our proposed CNN model are summarized in Table 3.9.

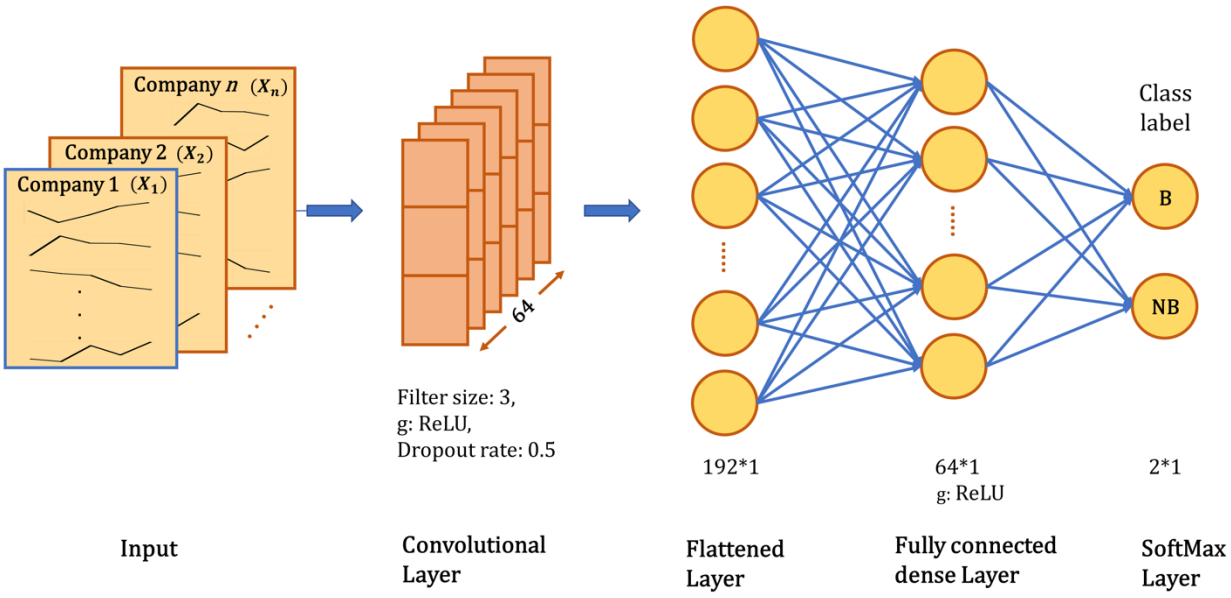


Figure 3.5 The architecture of our CNN model for bankruptcy prediction.

3.3.6 Performance Measures

Previous studies have used several measures to evaluate how well the model predicts the bankruptcy status of each company [Agarwal and Taffler 2007; Barboza et al. 2017; Mai et al. 2019]. These measures are computed from the confusion matrix in Table 3.4, where TP refers to the number of *true positive* company classifications, FP refers to *false positive* classifications, FN refers to *false negative* classifications, and TN refers to *true negative* classifications. TP and TN represent the number of accurately predicted bankrupt and non-bankrupt companies, respectively. FN represents the number of companies predicted as non-bankrupt while the actual status of the companies is bankrupt. While FP represents the number companies predicted as bankrupt when their actual status is non-bankrupt. It is important to focus on the FPs and FNs in such analysis as they can impact the companies' future. For example, if a company is labelled as FP, it can have a negative impact on the business of the healthy company. That is, if a company is predicted to be at risk for bankruptcy despite being a financially healthy one and such a prediction is made public, it could impact the decisions of the stakeholders or influence the investor faith in the company. Similarly, if a company is wrongly predicted as FN, it can lead to substantial loss of funds and

affect various stakeholders. The measures include the area under the receiver operating characteristics curve (AUC) together with sensitivity (or recall), specificity, G-mean, F-measure, precision, and accuracy, the calculation of which are found in Eqs. (3.10) -(3.15), respectively. Sensitivity represents the ratio of accurately predicted positive samples to all positive samples, while specificity is the ratio of accurately predicted negative samples to all negative samples. Precision is the ratio of accurately predicted positive samples to the total predicted positive samples. Sensitivity and precision focus on the bankrupt companies, while specificity focuses on the non-bankrupt companies. Accuracy represents the ratio of accurately predicted status (both positive and negative samples) to the total number of companies, giving us insight into the overall prediction performance.

Table 3.4 Confusion matrix for binary classification

		Actual	
		Bankrupt	Non-bankrupt
Predicted	Bankrupt	TP	FP
	Non-bankrupt	FN	TN

$$\text{Sensitivity (or Recall)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.10}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3.11}$$

$$\text{G} - \text{mean} = \sqrt{\text{sensitivity} \times \text{specificity}} \tag{3.12}$$

$$\text{F1} - \text{score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{3.13}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3.14}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{3.15}$$

3.4 Results and Discussion

3.4.1 Feature Selection

To develop accurate prediction models, the first step is to perform feature selection. In Section 3.1, we discussed two sets of features: the five ratios that made up the Altman [1968] Z-score and the expanded set of 11 ratios (inclusive of the five from Altman) used by Barboza et al. [2017], all of which are found in Table 3.1. We note that the features from Altman [1968] and Barboza et al. [2017] are developed from a financial perspective.

To expand upon these two sets of features, we further explored features available in the COMPUSTAT database using the RFE technique. We explored 54 total features from three datasets generated from a point in time relative to the occurrence of bankruptcy: (i) financial variables evaluated during the year when bankruptcy occurs (*year 0*), (ii) financial variables from one year prior to bankruptcy (*year 1*), and (iii) from two years prior to bankruptcy (*year 2*). We select these three static datasets due to the observed clear distinction between bankrupt and non-bankrupt companies (in terms of Z-score) in the latest years prior to bankruptcy (Figure 3.1). Moreover, early prediction of bankruptcy prior to its occurrence is more desirable. The RFE technique led to 29 significant features in the year 0 dataset, 46 features in the year 1 dataset, and 18 features in the year 2 dataset. Twelve features were common in those three datasets, as shown in Table 3.5, and 54 unique features made up the important features across the three datasets, as shown in Table 3.6.

Table 3.5 The 12 common features from the year 0, year 1, and year 2 datasets.

Feature	Definition
change in return on equity	$\left(\frac{\text{net income}}{\text{stockholder equity}}\right)_t - \left(\frac{\text{net income}}{\text{stockholder equity}}\right)_{t-1}$
lo	other liabilities total
adjex_f	cumulative adjustment factor by ex-date - fiscal
lt	total liabilities
D	$\frac{\text{market value of shares} \times \text{number of shares}}{\text{total liabilities}}$
ib	income before extraordinary items
ebit	earnings before interest and taxes
dltt	long term debt total
growth in emp	$\frac{\#\ \text{employees}_t - \#\ \text{employees}_{t-1}}{\#\ \text{employees}_{t-1}}$
leverage	$\frac{\text{market value of shares} \times \text{number of shares}}{\text{total debt}}$
sale	net sales or turnover
cogs	cost of goods sold

Table 3.6 List of 54 features obtained from the year 0, year 1, and year 2 datasets.

Feature	Definition
change in return on equity	$\left(\frac{\text{net income}}{\text{stockholder equity}}\right)_t - \left(\frac{\text{net income}}{\text{stockholder equity}}\right)_{t-1}$
lo	other liabilities total
adjex_f	cumulative adjustment factor by ex-date - fiscal
lt	total liabilities
D	$\frac{\text{market value of shares} \times \text{number of shares}}{\text{total liabilities}}$
ib	income before extraordinary items
ebit	earnings before interest and taxes
dltt	long term debt total
growth in emp	$\frac{\#\ \text{employees}_t - \#\ \text{employees}_{t-1}}{\#\ \text{employees}_{t-1}}$
leverage	$\frac{\text{market value of shares} \times \text{number of shares}}{\text{total debt}}$
sale	net sales or turnover
cogs	cost of goods sold
ib	income before extraordinary items

change in price to book	$\left(\frac{\text{market value/share}}{\text{book value/share}}\right)_t - \left(\frac{\text{market value/share}}{\text{book value/share}}\right)_{t-1}$
dp	depreciation and amortization
prcc_f	annual closing price per share of fiscal year
ivncf	investing activities net cash flow
dlc	debt in current liabilities
Current assets by current liabilities	$\frac{\text{current assets}}{\text{current liabilities}}$
nopi	nonoperating income
capx	capital expenditures
gp	gross profit
growth in sale	$\frac{\text{sales}_t - \text{sales}_{t-1}}{\text{sales}_{t-1}}$
xido	extraordinary items and discontinued operations
at	total assets
bkvlps	book value per share
productivity	$\frac{\text{earnings before interest and taxes}}{\text{total assets}}$
profitability	$\frac{\text{retained earnings}}{\text{total assets}}$
liquidity	$\frac{\text{net working capital}}{\text{total assets}}$
dpact	depreciation, depletion, and amortization
wcap	working capital (balance sheet)
txdc	deferred taxes
ceq	common/ordinary equity - total
che	cash and short-term investments
aox	other assets (sundry)
ivaco	other investing activities
fopo	other funds from operations
aoloch	net change of assets and liabilities (other)
txt	total income taxes
ao	other assets
cash by current liabilities	$\frac{\text{cash}}{\text{current liabilities}}$
sppiv	sales of property, plant and equipment and investments gain
dd1	long-term debt due in one year
csho	common shares outstanding
ch	cash
ap	accounts payable
spi	special items
B	$\frac{\text{retained earnings}}{\text{total assets}}$

A	$\frac{\text{net working capital}}{\text{total assets}}$
fincf	financing activities net cash flow
ceqt	tangible common equity
caps	capital surplus/share premium reserve
lco	total other current liabilities
re	retained earnings
E	$\frac{\text{sales}}{\text{total assets}}$

In this work, we focus our analysis and comparative evaluations based upon four different sets of features: (i) five features from Altman [1968], (ii) 11 features from Barboza et al. [2017], (iii) 12 features common to the three year-based datasets, and (iv) 54 unique features across the three datasets.

3.4.2 BiLSTM and CNN Models

Figure 3.6 shows the overview of our method. Two different variants of datasets are used to build the predictive models: (i) balanced, where the number of bankrupt and non-bankrupt companies are roughly equal, and (ii) imbalanced, where the imbalanced class ratio (the ratio of the number of non-bankrupt samples over the total number of bankrupt samples) is approximately 8.7:1. Further, we used three specific datasets for both balanced and imbalanced datasets. They consist of data up to (i) five consecutive years of data prior to bankruptcy which are used to derive two different temporal datasets including, (ii) the latest three years, and (iii) the two-year ahead dataset comprised of the oldest three years. We also note that the BiLSTM and CNN models are trained using four different sets of features as described in Section 3.4.1. As the number of years prior to bankruptcy increases, the data tend to have more empty values thus reducing the number of companies to explore.

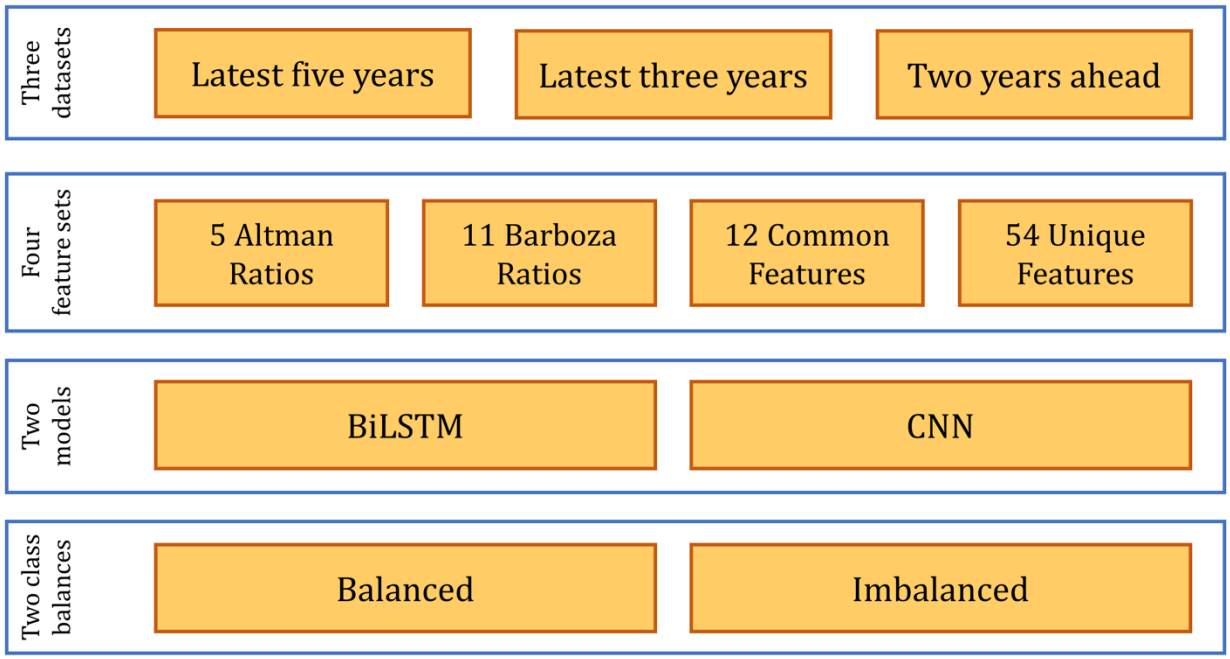


Figure 3.6 Summary of the numbers of datasets, feature sets, predictive models, and class balance sets analyzed.

Cleaning the data included eliminating features with more than 10% missing values, and highly correlated features (with an absolute value of the Pearson correlation coefficient, r, greater than 0.95). A cutoff of 0.95 was chosen to avoid highly redundant information. After this, the missing values of a company for a numeric feature were filled with the mean value of the feature for that company. This approach was implemented considering the wide range of companies, even within similar sectors. Considering only the companies that have at least five consecutive years of data prior to bankruptcy, we observed that the total number of bankrupt companies is 155. Also, the number of non-bankrupt companies with at least five recent consecutive years of data is 1348. The samples are selected from companies that filed/did not file for bankruptcy during the period 1994-2020. The numbers of bankrupt and non-bankrupt companies in the balanced and imbalanced datasets are found in Table 3.7. The balanced dataset is built by random undersampling, which randomly selects 158 non-bankrupt companies from the existing 1348 non-bankrupt companies.

Table 3.7 Number of companies in the balanced and imbalanced datasets.

	Balanced	Imbalanced
Non-bankrupt	158	1348
Bankrupt	155	155
Total companies	313	1503

Each dataset is divided into 80% (of companies) used for training and 20% used for testing, keeping the multiple years of data for each company together within the company's location in the training or testing set. Table 3.8 provides the number of companies (bankrupt/non-bankrupt) in the train and test sets of the balanced and imbalanced data.

Table 3.8 Number of companies in the train and test sets of the balanced and imbalanced data.

	Balanced			Imbalanced		
	Training	Testing	Total	Training	Testing	Total
Non-bankrupt	118	40	158	1064	284	1348
Bankrupt	134	21	155	139	16	155
Total companies	252	61	313	1203	300	1503

The train dataset is then used for training the two models, BiLSTM and CNN with the parameters shown in Table 3.9, followed by testing on the test dataset. For balanced datasets, the class weights for bankrupt and non-bankrupt companies during training is considered equal. For imbalanced datasets, the *'class_weight'* is set to *'balanced'* using Scikit-learn in Python when each model is trained (cost-sensitive learning). The class weight for positive class (bankrupt) and for negative class (or non-bankrupt) are calculated with $n/(2n^+)$ and $n/(2n^-)$, respectively. The sizes of the positive and negative classes are denoted by n^+ and n^-, respectively. The prediction on the test data provides the (complementary) probabilities of companies being bankrupt or non-bankrupt. These probabilities are then converted to a binary representation of 1 for bankrupt companies and 0 for non-bankrupt companies using a threshold probability of 0.5.

Table 3.9 Parameters used in the machine learning models.

Parameters	BiLSTM	CNN
Filters	-	64
Units	16	-
Filter size	-	3
Dropout rate	0.5	0.5
Activation function used in dense layer	ReLU	ReLU
Activation function at the output layer	SoftMax	SoftMax
Loss function	Categorical Cross Entropy	Categorical Cross Entropy
Optimizer	Adam	Adam
Metrics	AUC	AUC
Learning rate	0.001	0.001
Number of epochs	500	500
Batch size	20	20
Validation split	0.1	0.1

3.4.3 Model Evaluation and Discussion

In this section, we explain the experimental results using both BiLSTM and CNN in combination with different sets of features per different temporal dataset (as stated in Section 4.2). We implement our methods using Python version 3.7.4 with Keras [Chollet 2015] and TensorFlow libraries [Abadi et al. 2015] on an Intel core i7, 3.4 GHz with 16 Gb of RAM in a 64-bit platform.

For the sake of comparison and validation of our models, we utilized the raw original data without any transformation (e.g., normalization) for the analysis followed by previous studies conducted by Barboza et al. [2017], Cleofas-Sánchez et al. [2016], Heo and Yang [2014], and Tsai et al. [2014]. This was done to retain the originality of the data samples and to test the predictive capability of the models without depending on transformation techniques. As the study focuses on a wide range of companies, this approach also helped in preserving the divergence in magnitudes of the attributes' values. Tables 3.9-3.11 represent the average values of the performance measures over 100 iterations for both balanced and imbalanced datasets per different sets of features and temporal scales, three-year, five-year, and two-year ahead, respectively. For the sake of consistency and robustness of the results, all performance metrics including precision, recall,

specificity, AUC, F-measure, and accuracy for both BiLSTM and CNN models are an average over 100 iterations of the same data with different random seeds (bootstrap) as shown in these tables. It is clear from the boldface results in all tables that BiLSTM model with the five-year imbalanced data and 5 features yields the highest AUC (95.4%) followed by the three-year imbalanced data and 54 features with 95.2% AUC. For balanced datasets, BiLSTM produces the highest AUC values with 54 features in three-year and five-year datasets followed by 5 features in two-year ahead datasets. This is consistent with the boxplots in Figure 3.8 (a), (c), and (e). Among temporal datasets, the two-year ahead dataset shows the least accurate results for both balanced and imbalanced datasets as the observations come from further in the past before the bankruptcy onset, thus increasing the uncertainty. More specifically, for two-year ahead data, we can observe that BiLSTM model achieves an average AUC of 78.3% with 5 ratios for the balanced data and an average AUC of 81.5% with 54 features for the imbalanced data. Furthermore, by comparing different sets of features used, we conclude that the 54-feature set provides the best results followed by 5- and 12-feature sets, while the 11-feature set is the least effective in this experiment. This can also be perceived from all ROC plots in Figure 3.7. We note that CNN was only comparable with BiLSTM with 5 features when data is balanced. In most cases, the BiLSTM model outperforms the CNN model in terms of the average AUC, which is higher for the majority of the feature set and year combinations. This is particularly true for imbalanced datasets, which are more common to occur in bankruptcy prediction. Thus, considering the imbalanced nature of bankruptcy data, we can conclude that the BiLSTM model with 54 features provides the best prediction results in terms of AUC. The G-mean of this model (with 87.5%) is comparable with the G-mean of CNN with 5 features with balanced dataset (with 88.8%). Higher AUCs are obtained for the imbalanced datasets using cost-sensitive learning with CNN and BiLSTM compared to the balanced datasets in most cases. This is due to the loss of data from the majority class while creating the balanced datasets using undersampling. In Table 3.10 and Table 3.11 we can observe that the BiLSTM model for 54 feature set provides a highest average recall value for both balanced and imbalanced datasets, where recall is the ratio of TP (true positives) to sum of TP and FN (false negatives). The higher the recall value, the higher the true positives. Also, for the 12 and 54-features sets, the recall values for the BiLSTM model are higher compared to the CNN model. This reiterates that BiLSTM model with 54 features provides better results. In these tables, the average precision values

obtained by the 12 and 54-feature sets is higher compared to the 5-feature set in the imbalanced dataset for both the models. High precision value implies high true positives and low false negatives. Whereas, in the balanced dataset, the precision values among these sets are comparable. This shows that even though the average AUC values among these three feature sets are comparable, the 12 and 54-feature sets provide more accurate predictions.

Table 3.10 Comparison of BiLSTM and CNN using performance measures for three-year data using 5 features (Table 3.1), 11 features [Barboza et al. 2017], 12 common features (Table 3.5), and all 54 features (Table 3.6).

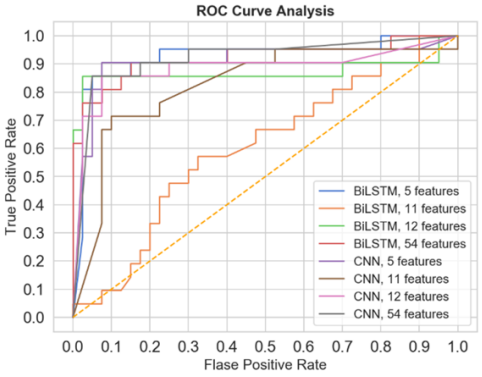
	5 Features		11 Features		12 Common Features		All 54 Features	
Balanced	BiLSTM	CNN	BiLSTM	CNN	BiLSTM	CNN	BiLSTM	CNN
Precision	0.791	0.789	0.442	0.526	0.771	0.768	0.745	0.798
Accuracy	0.875	0.881	0.562	0.653	0.863	0.861	0.861	0.880
Recall	0.891	0.917	0.862	0.790	0.885	0.882	0.941	0.898
Specificity	0.866	0.861	0.397	0.578	0.851	0.849	0.816	0.870
F-measure	0.835	0.846	0.583	0.618	0.822	0.818	0.829	0.842
G-mean	0.877	**0.888**	0.579	0.659	0.867	0.864	0.875	0.882
AUC	0.926	0.929	0.684	0.750	0.927	0.921	**0.942**	0.921
Imbalanced	BiLSTM	CNN	BiLSTM	CNN	BiLSTM	CNN	BiLSTM	CNN
Precision	0.559	0.573	0.167	0.216	0.755	0.787	0.736	0.705
Accuracy	0.955	0.956	0.940	0.939	0.972	0.973	0.973	0.969
Recall	0.777	0.768	0.032	0.063	0.707	0.697	0.782	0.752
Specificity	0.964	0.967	0.991	0.988	0.986	0.989	0.983	0.981
F-measure	0.645	0.653	0.052	0.092	0.723	0.729	0.753	0.716
G-mean	0.863	0.860	0.113	0.188	0.832	0.825	**0.875**	0.855
AUC	0.951	0.941	0.761	0.718	0.940	0.934	**0.952**	0.945

Table 3.11 Comparison of BiLSTM and CNN using performance measures for five-year data using 5 features (Table 3.1), 11 features [Barboza et al. 2017], 12 common features (Table 3.5), and all 54 features (Table 3.6).

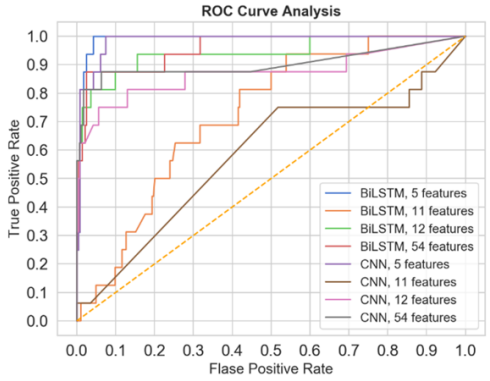
	5 Features		11 Features		12 Common Features		All 54 Features	
Balanced	BiLSTM	CNN	BiLSTM	CNN	BiLSTM	CNN	BiLSTM	CNN
Precision	0.791	0.782	0.455	0.468	0.761	0.746	0.762	0.765
Accuracy	0.876	0.880	0.583	0.571	0.857	0.843	0.869	0.863
Recall	0.894	0.929	0.833	0.815	0.879	0.863	0.932	0.905
Specificity	0.866	0.853	0.445	0.437	0.845	0.832	0.833	0.840
F-measure	0.836	0.847	0.587	0.573	0.814	0.796	0.836	0.825
G-mean	0.879	**0.890**	0.604	0.553	0.861	0.845	0.880	0.870
AUC	0.925	0.922	0.697	0.673	0.918	0.906	**0.947**	0.900
Imbalanced	BiLSTM	CNN	BiLSTM	CNN	BiLSTM	CNN	BiLSTM	CNN
Precision	0.517	0.565	0.124	0.085	0.737	0.794	0.727	0.747
Accuracy	0.949	0.955	0.941	0.939	0.970	0.968	0.971	0.972
Recall	0.733	0.746	0.022	0.029	0.697	0.615	0.763	0.750
Specificity	0.961	0.967	0.992	0.990	0.985	0.988	0.983	0.984
F-measure	0.602	0.638	0.036	0.039	0.707	0.675	0.739	0.738
G-mean	0.836	0.846	0.079	0.084	0.825	0.767	**0.864**	0.856
AUC	**0.954**	0.930	0.749	0.635	0.923	0.899	0.943	0.930

Table 3.12 Comparison of BiLSTM and CNN using performance measures for two-year ahead data using 5 features (Table 3.1), 11 features [Barboza et al. 2017], 12 common features (Table 3.5), and all 54 features (Table 3.6).

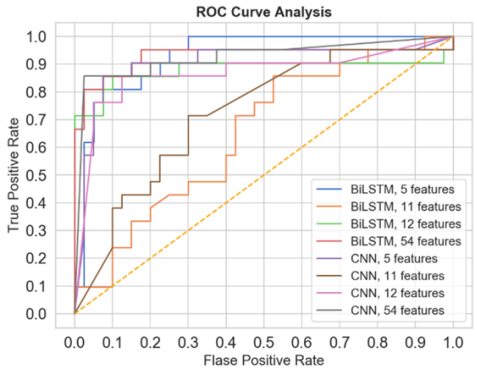
	5 Features		11 Features		12 Common Features		All 54 Features	
Balanced	BiLSTM	CNN	BiLSTM	CNN	BiLSTM	CNN	BiLSTM	CNN
Precision	0.590	0.560	0.411	0.411	0.506	0.474	0.496	0.493
Accuracy	0.727	0.699	0.519	0.515	0.647	0.608	0.638	0.632
Recall	0.788	0.759	0.812	0.830	0.756	0.790	0.807	0.812
Specificity	0.694	0.666	0.357	0.343	0.588	0.508	0.545	0.533
F-measure	0.672	0.641	0.545	0.548	0.604	0.589	0.613	0.611
G-mean	**0.737**	0.707	0.533	0.526	0.663	0.628	0.660	0.654
AUC	**0.783**	0.751	0.600	0.588	0.696	0.699	0.717	0.703
Imbalanced	BiLSTM	CNN	BiLSTM	CNN	BiLSTM	CNN	BiLSTM	CNN
Precision	0.244	0.323	0.033	0.092	0.244	0.158	0.328	0.129
Accuracy	0.890	0.9223	0.942	0.94	0.934	0.940	0.935	0.946
Recall	0.491	0.385	0.005	0.018	0.126	0.034	0.228	0.014
Specificity	0.913	0.953	0.994	0.992	0.979	0.990	0.975	0.998
F-measure	0.323	0.344	0.009	0.028	0.157	0.053	0.262	0.024
G-mean	**0.665**	0.596	0.019	0.064	0.306	0.115	0.450	0.050
AUC	0.803	0.795	0.672	0.616	0.771	0.696	**0.815**	0.771



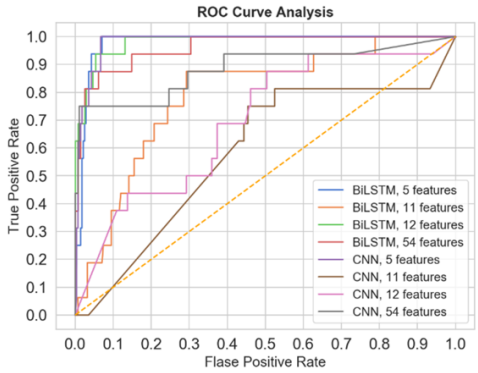
(a) 3-year balanced data



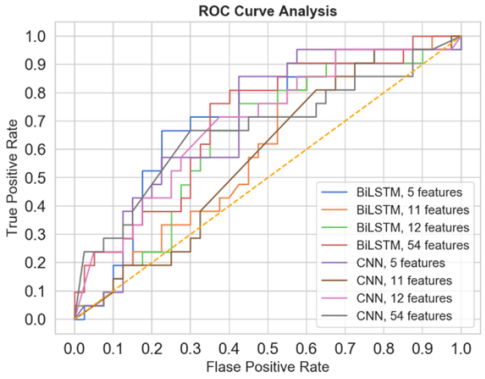
(b) 3-year imbalanced data



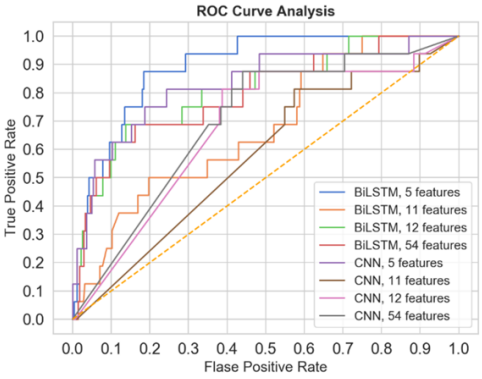
(c) 5-year balanced data



(d) 5-year imbalanced data



(e) 2-year ahead balanced data

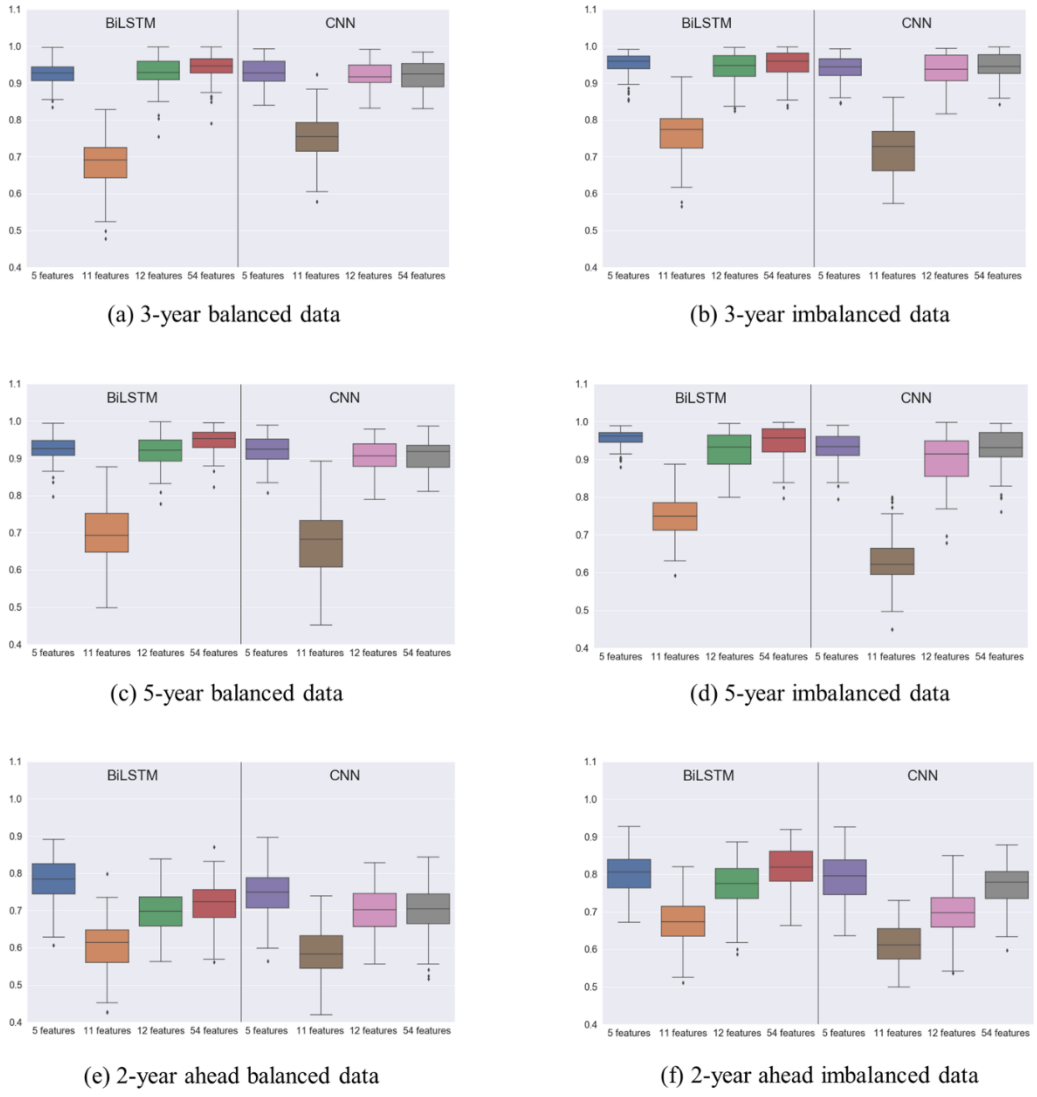


(f) 2-year ahead imbalanced data

Figure 3.7 ROC curves of the different models and datasets.

For measuring and comparing the characteristic of both CNN and BiLSTM, we used box plots for AUC measure as shown in Figure 3.8, which have been obtained over 100 iterations on the same data for each algorithm. As the figure shows, BiLSTM with 5 features is more robust than the other

algorithms for both balanced and imbalanced datasets. A small standard deviation is observed in BiLSTM with 5 features followed by 54 features for 3-year and 5-year datasets for imbalanced case. BiLSTM with 54 features better results each time for three- and five-year balanced dataset, while the performance of both BiLSTM and CNN with 11 features is inferior in most cases. Therefore, integrating temporal data with more features into risk models could improve the accuracy of models.



(a) 3-year balanced data

(b) 3-year imbalanced data

(c) 5-year balanced data

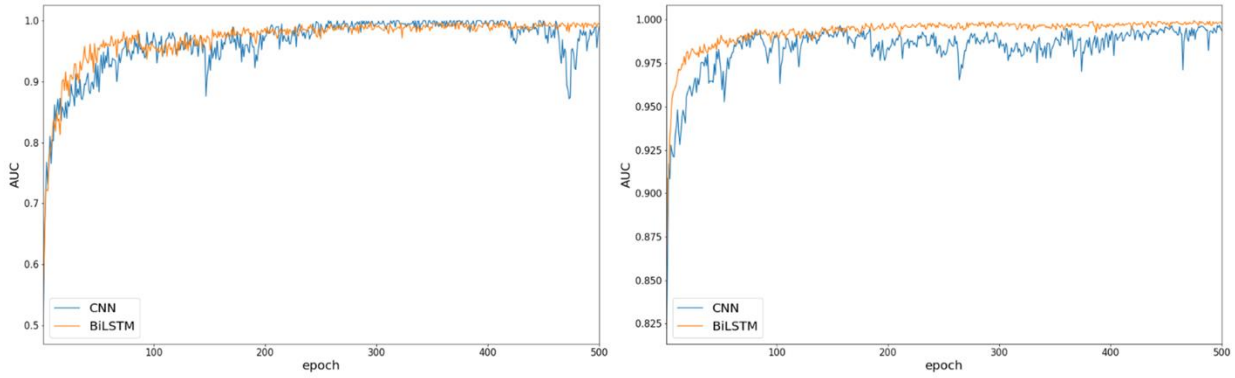
(d) 5-year imbalanced data

(e) 2-year ahead balanced data

(f) 2-year ahead imbalanced data

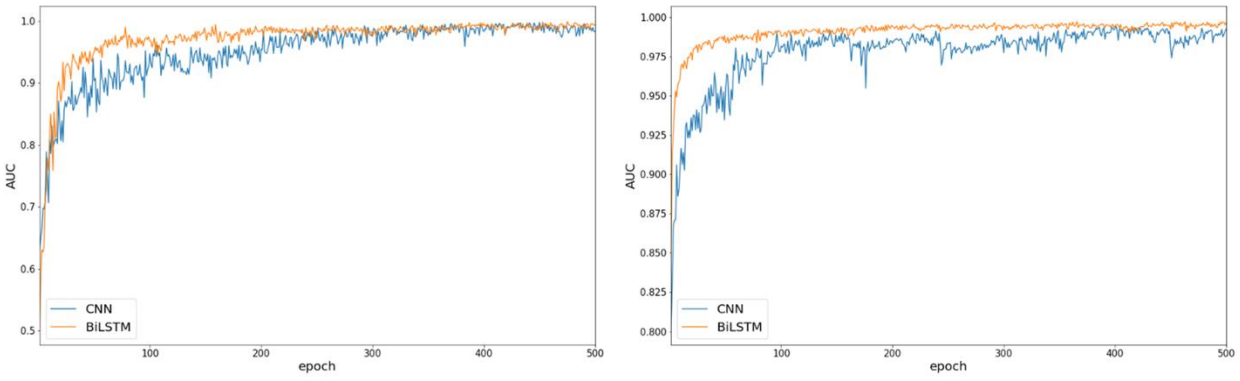
Figure 3.8 Comparing the robustness of BiLSTM versus CNN across different sets of features for balanced and imbalanced datasets. Each boxplot denotes variability of the AUC (vertical axis) for different feature sets.

Keeping in mind that the data analyzed here and the studies subsequently mentioned for comparison are different, better results were obtained using the 5 features from Altman [1968] in contrast with Barboza et al. [2017]. Although we were not able to replicate similar results using the 11 features from Barboza et al. [2017], our BiLSTM model shows, on average, approximately 4% higher AUC compared to their best model. Specifically, much better results were obtained with respect to AUC scores using the 12- and 54- feature sets obtained from the feature selection technique as described in Section 4.1, compared to other feature sets. The plots of the test AUC versus the number of the epochs for the 3-year data with 12 features and the 5-year data with 54 features, are shown in Figure 3.9 and Figure 3.10, respectively. We excluded the detailed results for brevity. One can gather from Figure 3.9 and Figure 3.10 that BiLSTM performs better outcomes than CNN, and it also converges faster compared to CNN. Furthermore, the BiLSTM equipped with class weights converges slightly faster using imbalanced data in contrast to balanced data.



(a) Balanced data (b) Imbalanced data

Figure 3.9 AUC vs epoch plots for 5-year data with 54 features.

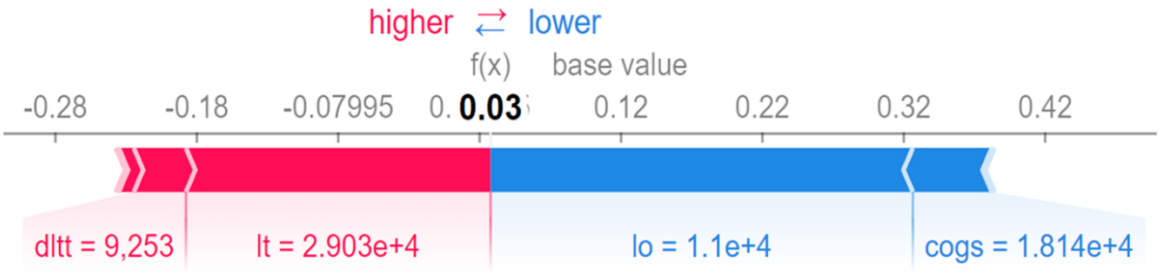


(a) Balanced data (b) Imbalanced data

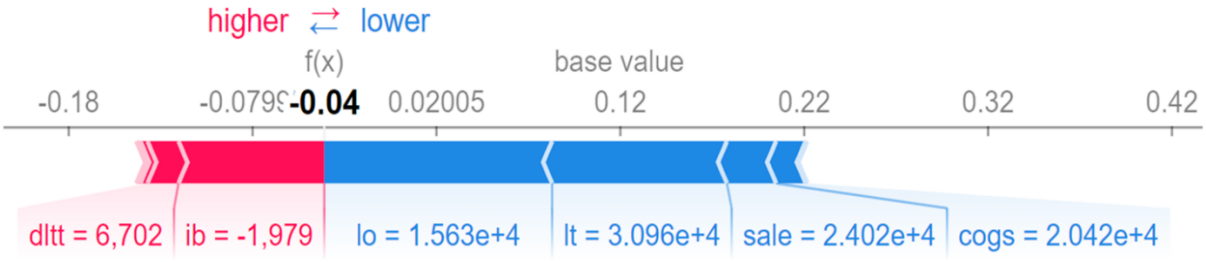
Figure 3.10 AUC vs epoch plots for 3-year data with 12 common features.

3.4.4 Model Interpretations

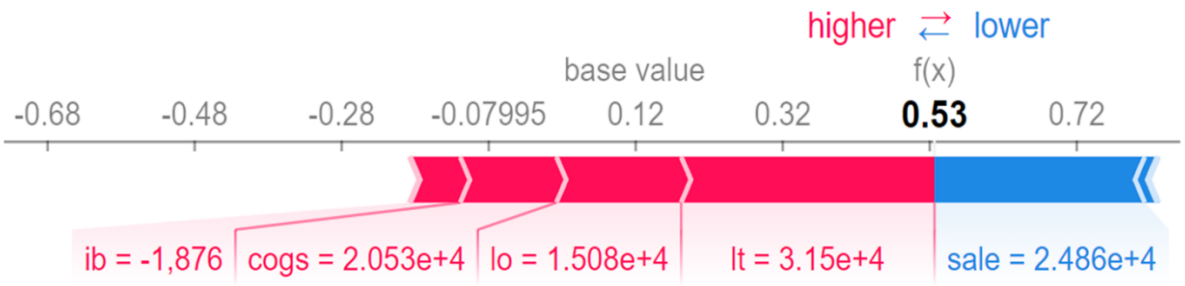
In this section, we interpret the models using SHapely Additive exPlanations (SHAP) technique [Lundberg and Lee 2017]. SHAP values represent the importance value of each feature which are calculated using a game theoretic approach. For the sake of brevity, we consider the combination of a CNN model with three-year imbalanced dataset and 12 common features for the interpretation. We use the SHAP library in Python.



(a) 2 years before bankruptcy



(b) 1 year before bankruptcy



(c) Bankruptcy year

Figure 3.11 SHAP values of a test case (a company) for the three different years.

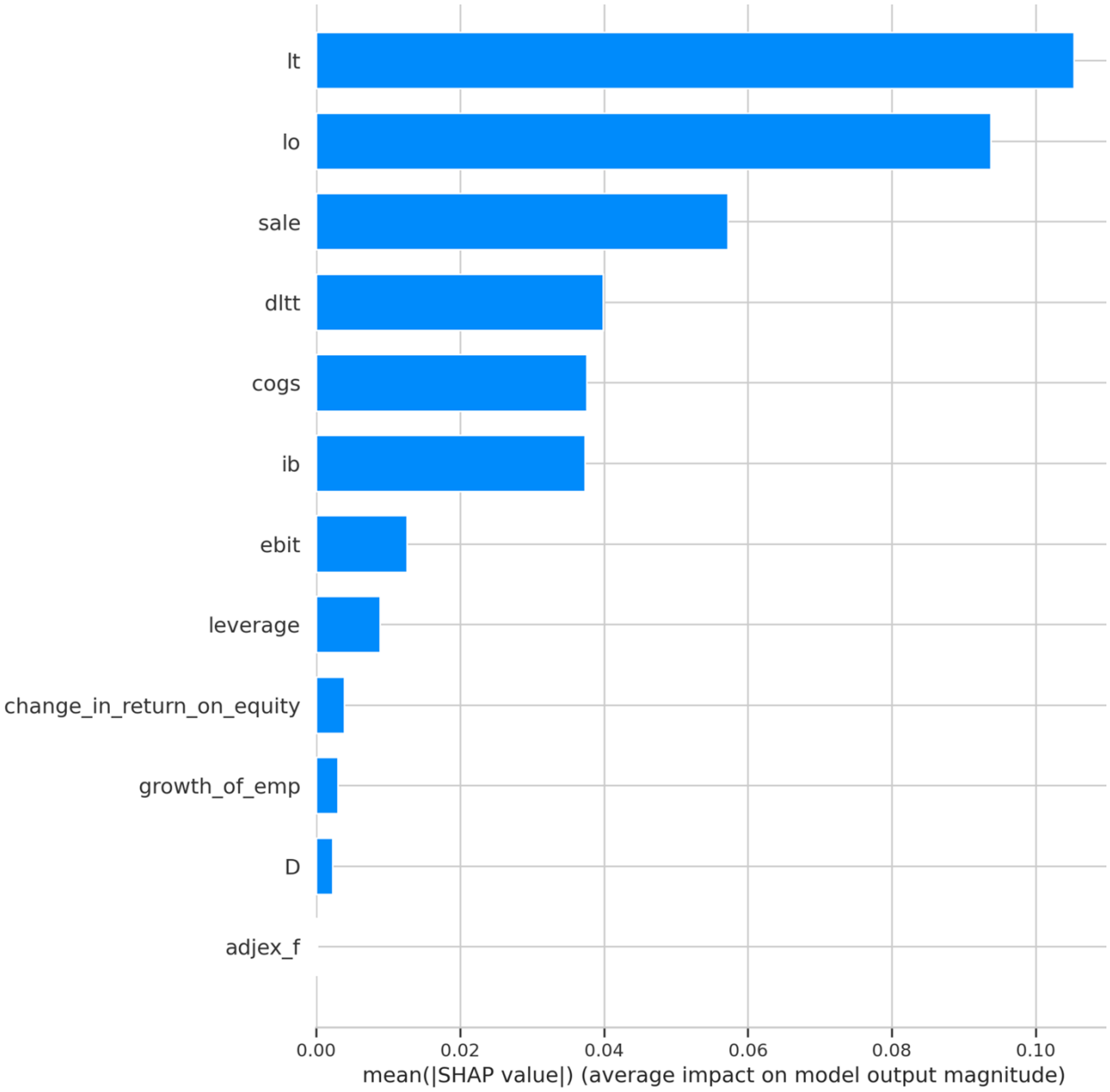


Figure 3.12 Feature importance bar plot for the test data set.

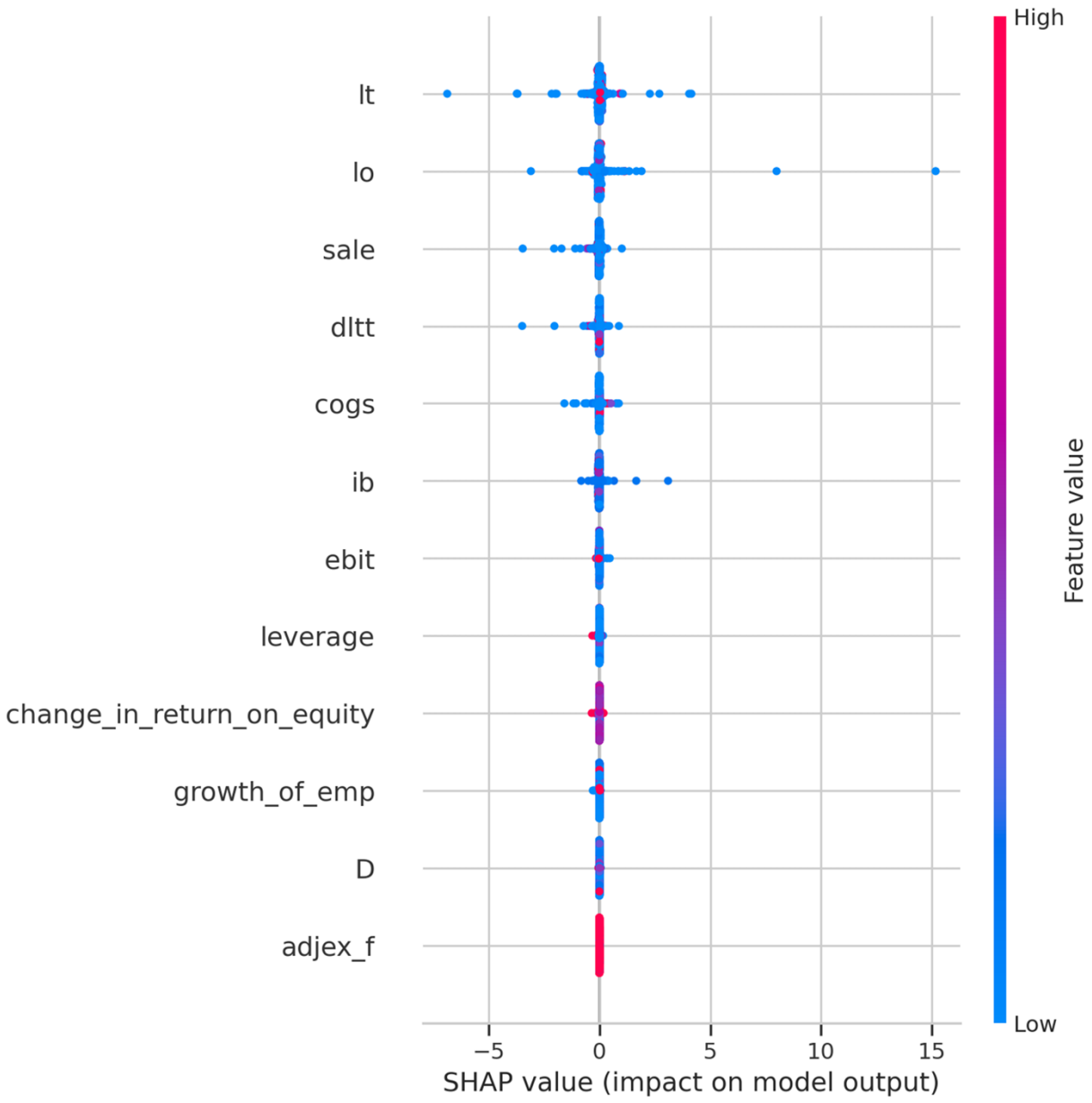


Figure 3.13 SHAP value summary plot for the test data set.

In Figure 3.11, the base value of 0.12 represents the expected value of the bankrupt class. The *f(x)* value in bold represents the model's estimated probability value for the test sample. Features in red represent those that contributed to increasing the probability value, while those in blue contributed to decreasing it. Figure 3.11 shows the significant features along with their values of

features for three cases: (a) two years before bankruptcy, (b) one year before bankruptcy, and (c) bankruptcy year. The length of the display of the features is proportional to their contribution. Figure 3.12 represents the bar plot displaying the average impact of the features on the model, calculated using the SHAP values, sorted in the order of highest to lowest. We can observe that 'lt' and 'lo' are the biggest contributors followed by 'sale', 'dltt', 'cogs', and 'ib', while the bottom six features are not impactful. Figure 3.13 represents the summary plot of the SHAP values for the 12 features. This plot can be used to explore the data, interpret patterns, and draw conclusions. The horizontal axis represents the SHAP value/impact value while the right-hand side vertical axis represents the legend for illustrating the range of values of the features, and each dot represents the SHAP value for each feature in test dataset. That is, the blue dots are lower in magnitude, and the red ones are larger. We can observe from Figure 3.13 that the red dots are centered closer to 0 which indicates their lower impact compared to the many dispersed blue dots that are lower in magnitude but have higher impact on the model. We can also observe that the features on the left-hand side are sorted based on their average impact, similar to that in Figure 3.12.

3.5 Conclusion

Financial risk analysis is a significant aspect to many financial and business institutions and its beneficiaries. It is necessary that any institution should avoid a situation of poor financial status, such as bankruptcy or liquidation of assets to protect its beneficiaries. Thus, the study of financial risk analysis has been in focus since a long time and continuous to be so. In our experiment, we attempt to predict the status of companies using various financial features. Our computational results demonstrate that temporal data helps in efficiently capturing the behavior of companies using advanced machine learning algorithms. In particular, we examined the effectiveness of two well-known deep learning algorithms for the bankruptcy prediction problem. We showed that our BiLSTM would improve predictive accuracy using past and future time steps. We used recursive feature elimination technique to select important features from a large set of features and compared with traditional financial indicators.

We observed that the prediction results using 54 common features is the best among the four different sets of features analyzed followed by 5 and 12 features. The results obtained using

the 11 features demonstrate less accuracy. To show the advantages of our BiLSTM and CNN methods, we have performed an extensive experimental study using both balanced and imbalanced datasets with different temporal settings (three-year, five-year, and two-year ahead) and various sets of features. Highlighting some of the best prediction results (in terms of AUC scores) obtained from the comprehensive study, we can observe that BiLSTM yielded an average AUC of 0.942 for balanced and 0.952 for imbalanced data with 54 features for the three-year data. For the five-year balanced data, an average AUC of 0.947 for 54 features and an average AUC of 0.954 for the imbalanced data and 5 features was achieved using BiLSTM model in both cases. This suggests that BiLSTM is a better prediction model in comparison with the CNN model for the data and features used in this experiment.

To discover the effectiveness of the models in predicting the bankruptcy status in advance (by two years), we used a two-year ahead dataset. For the balanced data, and average AUC of 0.783 was obtained using BiLSTM for the 5 features set. And for the imbalanced data, an average AUC of 0.815 was achieved using the BiLSTM model for the 54 features set. However, it should be noted that the results of the two-year ahead data are inferior compared to other temporal datasets.

There are many aspects in which this work can be further explored. One of the extensions would be to explore the quarterly data of companies on these models. Textual data can also be used for deep learning models to predict the financial stability of companies [Mai et al. 2019]. More advanced feature selection techniques can be implemented with the help of deep learning models [Jiménez et al. 2020, Niu et al. 2020] rather than the static recursive feature elimination technique. An interesting advantage of having the probability scores for the prediction is that the cutoff level can be customized to analyze risk of a company. For example, if a company's probability of being bankrupt is 0.3, this can be used to flag the company as risky or under financial distress. Further future efforts can incorporate bankruptcy prediction into a supplier selection and portfolio optimization framework, as well as into a supply chain network reliability analysis.

Chapter 4: Improving Bankruptcy Prediction Using Persistent Homology on Temporal Financial Data

4.1 Background

Corporate bankruptcy prediction is vital for financial risk management, investment decisions, and regulatory oversight. Conventional financial ratio-based methods, while simple and interpretable, often fail to capture complex, evolving patterns present in financial data over time. As discussed in Chapter 3, it is important in incorporating temporal and structural data features to improve predictive accuracy. This project aims to combine financial indicators with advanced topological data analysis (TDA) features, leveraging the power of deep learning models—specifically, Bidirectional Long Short-Term Memory (BiLSTM) networks and Convolutional Neural Networks (CNN)—to enhance bankruptcy prediction capabilities.

TDA, specifically persistent homology (PH), offers a powerful framework for uncovering hidden structures in high-dimensional data. Unlike traditional methods that rely on fixed metrics or statistical summaries, PH captures the evolving shape of data by tracking how topological features—such as connected components, loops, and voids—appear and disappear across a range of scales. These features are represented as persistence diagrams, which reflect the birth and death of structural patterns in the data. The longer a feature persists, the more topologically significant it is. This multi-scale view enables PH to identify nonlinear relationships, structural stability, and subtle anomalies that might be missed by pointwise financial indicators. In essence, PH does not just ask *what* values exist in the data, but *how* features relate and evolve geometrically across different thresholds.

When applied to temporal financial data, PH becomes especially valuable. As firms progress toward bankruptcy, their financial feature space often undergoes gradual deformations or abrupt disruptions—reflected not just in numbers, but in how features connect or separate over time. By analyzing changes in persistence diagrams across years—either through visual comparison, distance metrics (Wasserstein, Bottleneck), or persistence landscapes—we can quantify the evolution of financial structure. For instance, growing instability may be reflected in

the fragmentation of connected components (in dimension 0) or the formation of unusual loops and voids (in dimensions 1 and 2). These topological changes act as early indicators of distress, often before traditional metrics show red flags. Thus, PH provides a structurally-aware, mathematically grounded approach to track temporal changes, making it a powerful complement to deep learning models in financial risk prediction.

The rest of this article is structured as follows: Section 4.2 delves into the fundamentals of TDA and PH, providing a detailed overview of its principles and applications in network analysis. Section 4.3 outlines the methodology employed in this study, including the specific techniques and tools used to apply PH to networks representing financial data. Section 4.4 presents the different data combinations experimented and results, showcasing practical examples and the insights gained from analyzing different combinations of data and features explored, Section 4.5 concludes with a discussion of the findings, understanding the scope of TDA, and directions for further research.

4.2 Understanding PH

Algebraic topology is a branch of mathematics that applies tools from algebra to study the properties of space, shape, or structure. TDA is a tool developed to apply the concepts of topology to data exploration. TDA has emerged as a robust and powerful method for analyzing data, of various forms – point cloud, networks, and time series, from a geometrical perspective that could help extract imperceptible information. By applying principles from topology, TDA allows us to study the geometric and structural features of data sets. PH and Mapper algorithms are the most common TDA techniques. Mapper algorithm helps visualize high-dimensional data by building simplicial complexes. While PH helps with the generating and capturing the persistence of homological components using a filtration process.

PH is a method within TDA that captures and analyzes topological components across multiple scales. It identifies components that persist over a range of spatial resolutions, providing insights into the data's structure. PH works by constructing simplicial complexes, which are higher-dimensional analogs of graphs, from the data. As the scale parameter changes, PH tracks the birth

and death of topological features such as connected components, loops, and voids. The persistence of these features is recorded in a persistence diagram (PD) or barcode, which summarizes the lifespans of the features. This allows for the distinction of significant patterns from noise, offering a deeper understanding of the data's topological characteristics. Filtration techniques like Rips filtration, Alpha complex, weight rank clique filtration (WRCF) are some of the commonly explored approaches.

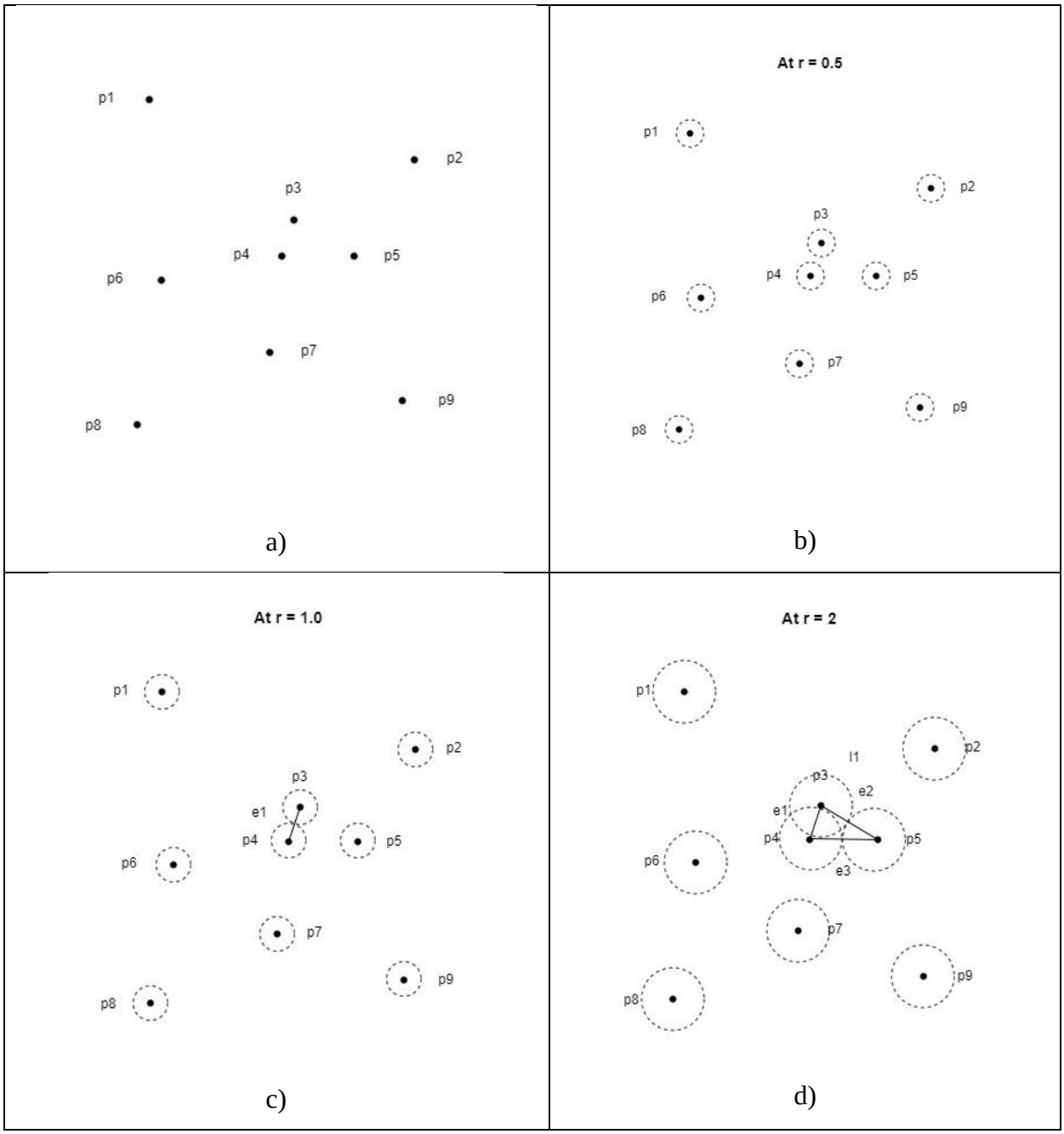
4.2.1 Filtration Process in PH:

PH is a central tool of TDA, useful for analyzing the shape and structure of complex data. PH captures "persistent" topological features that remain stable through different scales of observation, revealing subtle structural patterns often missed by traditional analysis techniques.

Consider the illustrative example of a simple point cloud composed of nine points labeled p1 through p9. Initially, at radius r=0, the vector space simply consists of these nine isolated points, known as 0-simplices (individual components). To analyze their topological structure, we perform a process known as filtration. Filtration involves gradually increasing the radius (r) of imaginary circles drawn around each point as shown in steps in Figure 4.1. As we progressively increase r, intersections between circles lead to the birth and death of new topological features such as edges (1-simplices) and loops (2-simplices):

- At r = 0.5, no intersections occur, and all points remain isolated. Thus, we still have nine separate connected components.
- Increasing to r=1.0, circles around points p3 and p4 intersect, forming an edge labeled e1, known as a 1-simplex. When this edge e1 appears, the individual points p3 and p4 merge into a single connected component. In topological terms, the two separate points (0-simplices) "die," and a new edge (1-simplex) is "born." Thus, the vector space now has eight simplices: seven remaining 0-simplices (isolated points) and one 1-simplex (edge).
- At r=2.0, additional intersections occur, forming edges e2 and e3. Crucially, these three edges (e1, e2, e3) collectively create a closed loop labeled *l1*, a 2-simplex. At this stage, the three individual edges are considered to "die," and the loop *l1* is "born," representing a higher-dimensional topological feature—a cyclic structure.

- As the radius increases further to r=3.0, the loop *l1* disappears (or "dies") due to additional connections forming between points and edges that disrupt its distinct structure. This process of increasing r continues until no additional topological features appear or vanish, fully capturing the underlying structure of the data at all scales.



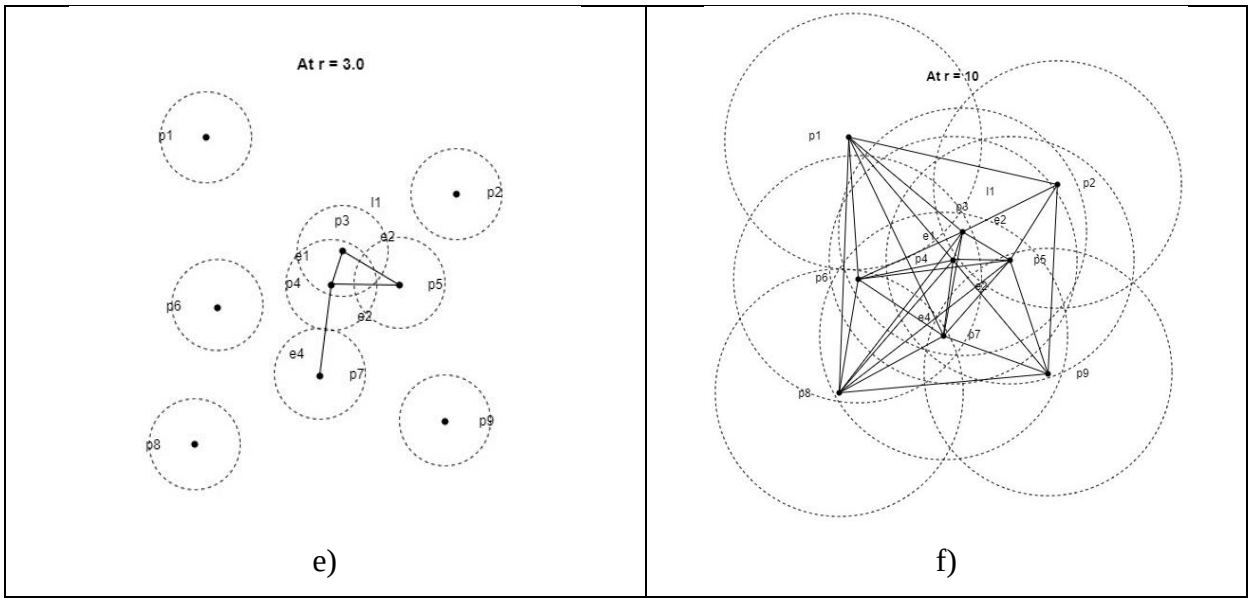


Figure 4.1 Filtration steps involved in persistent homology.

4.2.2 Persistence Diagram (PD) and Persistence Barcode (PB):

PH utilizes two primary tools to capture and analyze the topological structure of data: Persistence Diagrams (PD) and Persistence Landscapes (PL). PD is a graphical representation that plots the birth and death times of topological features as points on a two-dimensional plane, where each point represents a feature. The x-coordinate of a point indicates the scale or radius at which a feature appears ('birth'), and the y-coordinate indicates the scale at which the feature disappears ('death'). This plotting results in a scatter plot that provides a concise summary of the data's topological structure, highlighting the most persistent features that are robust across different scales.

For example, consider a set of features with the following birth-death pairs: Edge1: (1,3), Edge2: (2,4), Edge3: (2,6), Edge4: (3,5). Figure 4.2 shows the PD and PB for the example birth-death pairs.

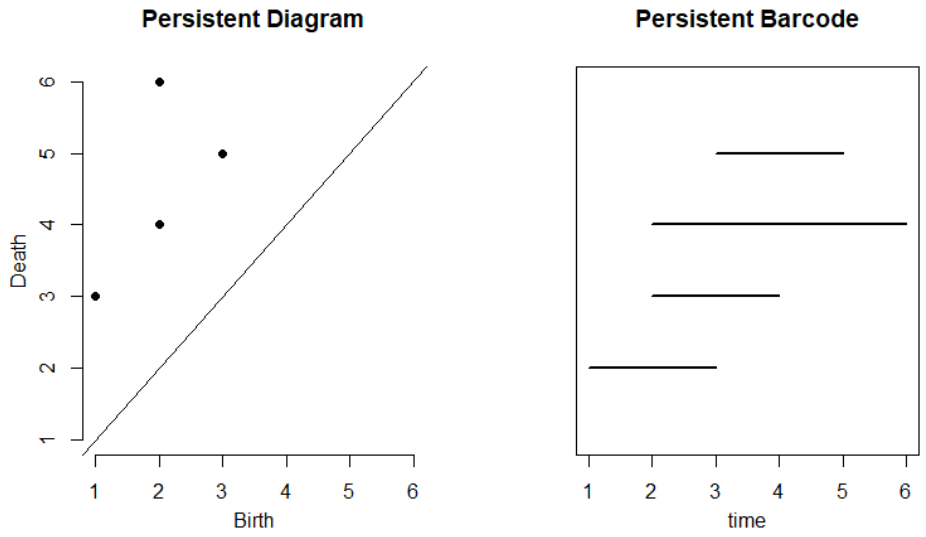


Figure 4.2 Persistence diagram and barcode of the example birth-death pairs.

In the PD, these pairs are marked as points, with each point's placement indicating the lifespan and significance of the corresponding feature. Features that persist over a larger range (from birth to death) are considered structurally significant, as they represent stable attributes of the underlying data topology. A related visual tool is the Persistence Barcode (PB), a line plot where each line or 'barcode' represents a feature horizontally, spanning from its birth to its death. Longer bars indicate features with greater persistence, underscoring their importance in the topology of the data.

4.2.3 Persistence Landscape (PL):

Persistence Landscapes (PL) transform the information from PD into a more analyzable form by converting topological feature data into a series of piecewise-linear functions. Each point (b, d) in the PD translates into a triangular peak within the landscape as shown in Figure 4.3. This peak forms through a function f(b, d), where it rises at the feature's birth (b), reaches its highest point at the midpoint of its lifespan, and falls back to zero at its death (d). This transformation allows each feature—whether an edge, loop, or more complex structure—to be represented as a continuous, piecewise-linear graph, reflecting its lifecycle within the data space as shown in Figure 4.4.

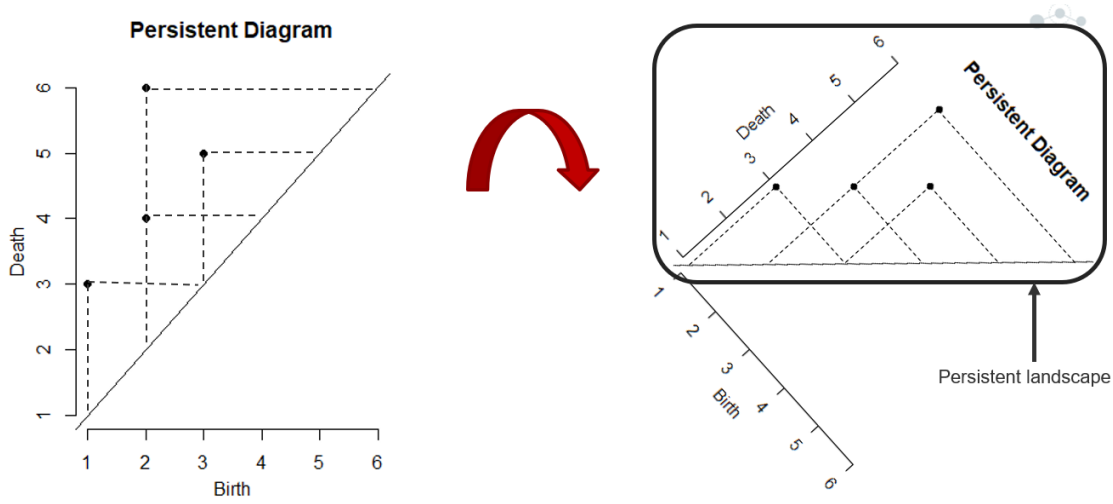


Figure 4.3 Representation of a persistence landscape using a persistence diagram.

The mathematical definition of a persistence landscape function for a point (b, d) can be expressed as shown in Eq **(4.1)**. Here, $\lambda_k(t)$ refers to the value of the landscape function at a given scale *t*, with *k* denoting the layer number. This function is zero both before b and after d, peaking at t = (b+d)/2, thus capturing the existence and persistence of a topological feature.

$$\lambda k(t) = \begin{cases} t - b & if\ b \leq t \leq \dfrac{b+d}{2} \\ d - t & if\ \dfrac{b+d}{2} < t \leq d \\ 0 & otherwise \end{cases} \qquad \textbf{(4.1)}$$

As multiple features overlap in time, they generate multiple layers of these landscape functions, where the first layer (λ1) represents the highest/top layer, and subsequent layers (λ2, λ3, and so on) represent progressively lower peaks, each capturing less prominent but still significant topological structures, as shown in Figure 4.5. This structuring of data into layers allows for a comprehensive quantitative analysis of the topological features present in the dataset.

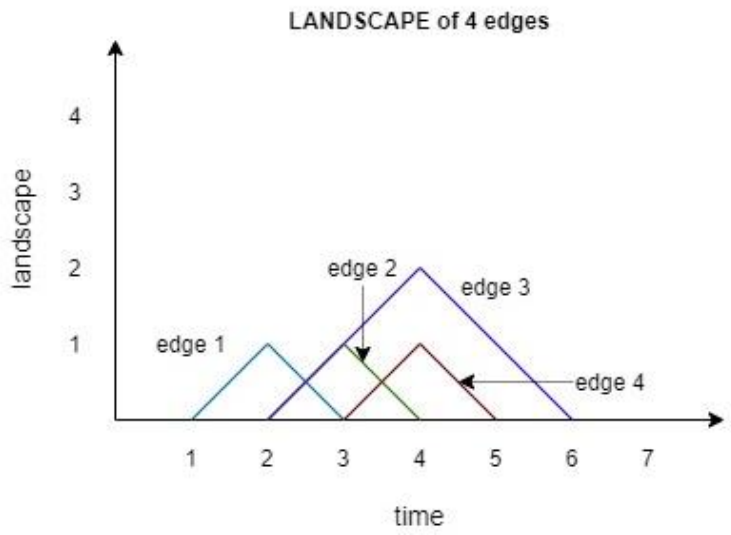


Figure 4.4 Persistence landscape of the 4 edges.

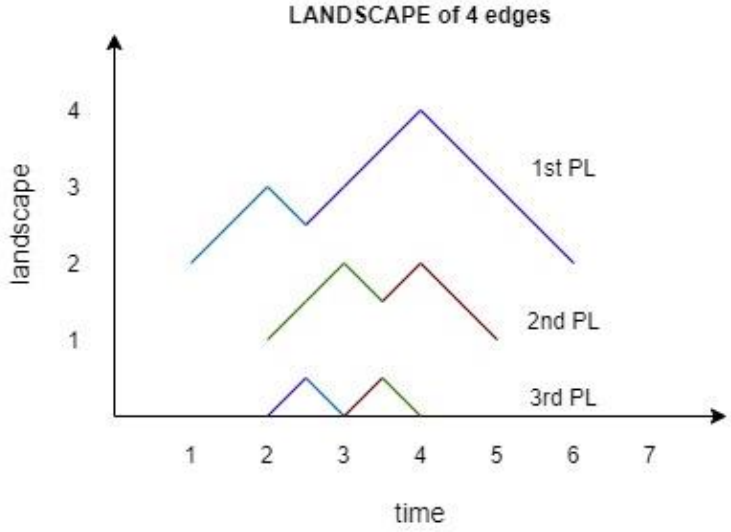


Figure 4.5 Different layers within a persistence landscape.

Persistence Landscapes offer several advantages over traditional PDs and Persistence Barcodes (PB), particularly in the realm of statistical analysis and machine learning integration. PLs provide a continuous, piecewise-linear representation of the topological features, making them suitable for a variety of computational techniques [Bubenik, 2015]. This continuous nature facilitates the application of standard statistical tools and supports the use of machine learning

algorithms for predictive modeling, enhancing the robustness of analyses. Furthermore, PLs enable the definition of norms and distances between different datasets, allowing for a rigorous quantitative comparison of their topological properties. This capability is particularly valuable in fields requiring detailed comparisons of complex data structures, such as bioinformatics, shape analysis, and network vulnerability assessment.

By converting topological data into a format amenable to advanced analytical techniques, Persistence Landscapes contribute significantly to more precise and effective data analysis, opening new avenues for leveraging the full potential of TDA in practical applications.

4.2.4 Extension to Multiple Dimensions (0,1,2):

This filtration process and subsequent landscape analysis are performed independently across different homological dimensions—dimension 0 (connected components), dimension 1 (loops), and dimension 2 (voids). Each dimension produces its own set of birth-death pairs, resulting in separate persistence diagrams and landscapes. By analyzing multiple dimensions simultaneously, PH reveals a richer, multi-layered view of the underlying data, capturing complex and subtle patterns often invisible to conventional statistical methods.

4.2.5 Wasserstein and Bottleneck distances

To quantify the difference between topological features captured by Persistence Diagrams (PDs), two commonly used metrics are the Wasserstein distance and the Bottleneck distance. These distances allow us to compare the evolution of topological structures in different datasets or across time (Cohen-Steiner et al. 2007, Edelsbrunner & Harer, 2010).

The p-Wasserstein distance measures the overall cost of optimally transporting one persistence diagram to another. It is defined for $p \geq 1$ and is sensitive to global variations in the diagrams. Formally, given two persistence diagrams D1 and D2, the p-Wasserstein distance is defined in Eq **(4.2)**, where γ ranges over all bijections between the points in D_1 and D_2, and $\|\cdot\|$

denotes a suitable norm, often the L∞ or Euclidean norm. This distance accounts for both small and large changes across all features [Chazal et al. 2016].

$$w_p(D_1, D_2) = \left(\inf_{\gamma} \sum_{x \in D_1} \| x - \gamma(x) \|^p \right)^{1/p} \quad \textbf{(4.2)}$$

The Bottleneck distance is a special case of Wasserstein distance where $p \to \infty$, it captures the largest single deviation between matched points in two persistence diagrams. It is particularly effective in detecting prominent topological differences. Given persistence diagrams D_1 and D_2, the Bottleneck distance is defined in Eq **(4.3)**, where γ again ranges over all bijections between the diagrams. This metric highlights the most significant topological shift, focusing on the worst-case scenario of all feature pairings (Cohen-Steiner et al., 2007).

$$\lim_{p \to \infty} W_p(D_1, D_2) = d_B(D_1, D_2) \quad \textbf{(4.3)}$$

4.2.6 Applications of PH

PH has wide-ranging applications in various fields of research due to its ability to analyze the topological features of complex data. In neuroscience, PH is used to study the structure and connectivity of brain networks, helping to identify patterns related to different neurological conditions or developments [Petri et al., 2014; Robinson & Turner, 2013). In biology, it aids in analyzing the shape and structure of biomolecules, such as proteins and DNA, providing insights into their functions and interactions (Pun et al., 2018). Additionally, PH is applied in materials science to investigate the properties of porous materials and nanostructures, enabling the design of materials with desired characteristics (Feng & Porter, n.d.). In the field of sensor networks, PH helps optimize coverage and detect anomalies, ensuring robust and efficient network performance (Munch, 2013). The versatility of PH in revealing hidden patterns and structures makes it an

invaluable tool across diverse domains, driving advancements and innovations in scientific research.

4.3 Methodology

This section outlines the methodology used to evaluate the contribution of data extracted upon applying PH. PH was implemented on networks representing financial data and then using deep learning techniques to predict the bankruptcy status for different combinations of datasets. The objective is to provide a detailed approach to network representation, feature extraction, applying machine learning techniques, and hyperparameter tuning, ensuring the readers can reproduce the study and its findings.

To capture the underlying shape and structure of multivariate financial data over time, we applied PH, a core technique within TDA. Financial data from each year was transformed into distance matrices (using Euclidean distances between features). These matrices were used to compute persistence diagrams—topological summaries describing the data's shape at various scales. Specifically, we computed diagrams for homological dimensions 0, 1, and 2.

Rather than directly applying temporal persistence methods, this study introduced a structured way to quantify how the topology of financial data evolves over time by computing pairwise distances between persistence diagrams (PDs) across years. As shown in the illustration, each firm has a sequence of persistence diagrams corresponding to each financial year leading up to the bankruptcy event (Year 1). To capture the temporal evolution, we computed inter-year distances using Wasserstein and Bottleneck metrics, which measure how much the shape of topological features (like connected components, loops, and voids) change from one year to another. These distances serve as topological signatures of instability or risk progression. The figure illustrates three core types of distance-based features:

- Distances d1: Measures the distance between year 5 PD and each of the other years (years 4, 3, 2, and 1). This captures how far the firm has moved from its earliest financial state, potentially highlighting long-term deviation.
- Distances d2: Computes the distance between consecutive years (year 5 to 4, 4 to 3, and so on). This approach captures the rate of change or volatility in financial structure year by year, helping to detect sudden disruptions or instability.
- Distances d3: Compares the bankrupt year (year 1 PD) to all previous years. This highlights how topologically distinct the bankruptcy year is from the firm's historical financial patterns—capturing late-stage divergence as bankruptcy approaches.

By using Wasserstein distance, we accounted for cumulative or average changes across all matched features in the diagrams, whereas Bottleneck distance emphasized the single largest deviation, flagging abrupt topological shifts. Together, these measures formed a comprehensive set of temporal distance features, encoding the structural evolution of financial health. In addition to these distances, we also extracted persistence landscape features, such as peak height, area under the peak, and quantiles of persistence values. These landscape-derived statistics offered a compact representation of persistence diagrams and served as a complementary topological perspective to the distance features. Altogether, this approach enabled the model to detect both gradual drifts and sudden topological shifts that precede bankruptcy, making it well-suited for early financial risk prediction.

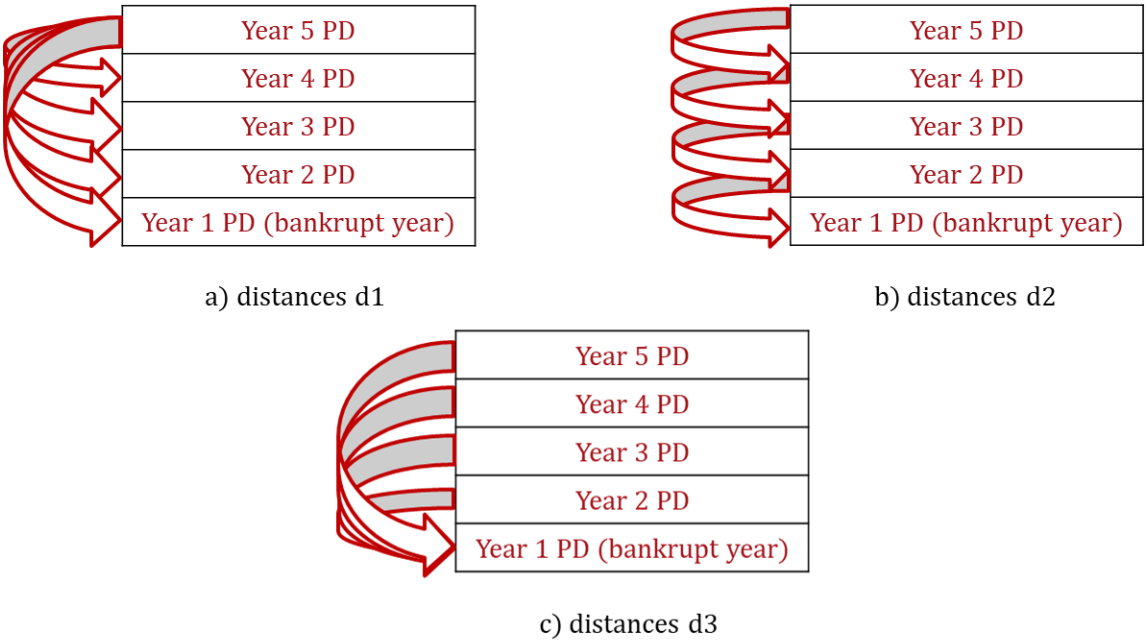


Figure 4.6 Illustration explaining the different distance combinations between the years.

4.3.1 Network Representation of Financial Data

Each year's financials of a company were represented as a network where the features represent the nodes, and the edge represents the connection between the features. The data was scaled per company per feature using the StandardScalar() technique. The edge weight was calculated using Euclidean distance between the features' values after scaling the data. By this approach we get a fully connected network for each year of a company. Network representation allows for exploring the topological aspects of financial data by transforming multivariate financial data into a structured format that can be analyzed using PH. This approach captures the underlying shape and structure of the data, highlighting changes in structural patterns that are crucial for detecting subtle signs of financial distress.

4.3.2 PH features

PH was applied to each of the above networks to extract the topological features where the distance matrix acts as the input to PH. Vietoris-Rips filtration technique was employed as the input represents distance matrix. The findings of applying PH are then converted to PL which

makes the feature extraction easy. The different features extracted are area, number of peaks, height of the maximum peak - under the PL layers, different percentiles of the PL, Wasserstein and Bottleneck distances between the base network and the interdicted network. The feature extraction was extended up to 5 layers deep and for components of three dimensions (0 - connected components, 1- holes/loops, 2 - voids). These features represent the independent variables, and the bankruptcy status represents the target variable.

4.3.3 Machine learning approaches

Two deep learning techniques were explored to predict bankruptcy status. CNN and BiLSTM models with extensive hyper tuning. Each model was tuned using a primary seed to identify optimal hyperparameters, which were then fixed. Subsequently, models were evaluated across 50 new randomized train-test splits to ensure results were robust and generalizable. The performance was assessed with multiple metrics including Accuracy, Precision, Recall, Specificity, F1 Score, Geometric Mean (G-Mean), and Area Under ROC Curve (AUC). The following section details the experiment's execution across multiple scenarios.

4.4 Data Combinations and Results

4.4.1 Data combinations explored

Multiple data combinations were generated using the financial data and the extracted topological data to explore the efficiency of prediction models. Different performance metrics were recorded, and the best results are highlighted. The study compared the PH approach with basic data on node and edge availability, demonstrating PH's advantages in understanding vulnerabilities. The case study effectively showed PH's ability to aid the predictive capacity of the models.

This study systematically explored seven distinct data combinations to assess the effectiveness of integrating financial and topological features derived through PH in predicting bankruptcy. The configurations tested included: df_none, which utilized only traditional financial data for a baseline comparison; df_d1, df_d2, and df_d3, each enriching financial data with topological distance features extracted from between different years of the temporal data (distances

d1, d2, and d3, respectively); df_d1d2d3, which combined features from all three distances to maximize data depth; df_landscape, incorporating persistence landscapes that offer a continuous and statistically tractable representation of topological features; and df_all, which integrated all available financial and topological data to explore the full potential of combined analytics. These combinations were designed to evaluate how different layers of data integration could enhance the models' ability to capture and predict complex financial patterns that precede bankruptcy.

For each data combination, the study employed Bayesian optimization [J. Wu et al. 2019] for hyperparameter tuning, aiming to find the optimal settings that would maximize the performance of predictive models. This method is particularly effective in managing the trade-offs between exploration of the parameter space and exploitation of the well-known configurations. It significantly reduces the number of trials needed to achieve near-optimal configurations, making it highly efficient compared to traditional grid or random search methods. By leveraging Bayesian optimization, the study not only refined the model's parameters to better fit the complex data structures but also ensured that each model variant was tuned under the same rigorous, systematic approach to maintain consistency across evaluations. The best parameters for the BiLSTM and CNN models are shown in the Table 4.1and Table 4.2.

Table 4.1 Hyper parameters for BiLSTM model for 100 epochs of training

df_version	filters	recurrent dropout	dropout	dense_units	activation	optimizer	learning_rate
df_none	128	0	0.5	16	relu	rmsprop	0.01
df_d1	112	0.2	0.4	56	relu	rmsprop	0.01
df_d2	128	0.1	0.5	16	relu	rmsprop	0.01
df_d3	80	0	0.3	64	relu	rmsprop	0.01
df_d1d2d3	48	0.2	0.1	16	relu	rmsprop	0.01
df_landscape	64	0	0.3	40	relu	rmsprop	0.01
df_all	80	0	0.1	24	relu	rmsprop	0.01

Table 4.2 Hyper parameters for CNN model for 100 epochs of training

df_version	filters	kernel_size	dropout	dense_units	activation	optimizer	learning_rate
df_none	80	2	0.3	48	relu	adam	0.001
df_d1	80	5	0.2	64	relu	rmsprop	0.01
df_d2	32	5	0.2	40	relu	rmsprop	0.01
df_d3	32	5	0.4	48	relu	rmsprop	0.01
df_d1d2d3	80	2	0.3	32	relu	rmsprop	0.001

df_landscape	64	2	0.2	64	relu	adam	0.001
df_all	64	3	0.2	32	relu	adam	0.001

4.4.2 *Results*

The BiLSTM model showed superior F1-score, G-mean, and recall on dataset *df_d3*, compared to *df_none*, highlighting that the topological distances from bankrupt year are particularly informative. The CNN model achieved its best performance on dataset *df_landscape*, which comprises financial and non-topological engineered features. This suggests CNN's local pattern recognition capabilities are better utilized on structured financial indicators and engineered features, rather than purely topological descriptors.

The integration of topological features significantly improved bankruptcy prediction performance in the five-year ahead data setting, particularly when using the BiLSTM model which is highlighted in Table 4.3. The best-performing dataset was *df_d3*, which included financial features combined with topological distance features derived from distance d3 (voids)—specifically comparing the bankrupt year (Year 1) with previous years. This configuration achieved the highest recall (0.643), F1-score (0.740), and G-mean (0.795) among all tested combinations. These improvements demonstrate that incorporating the final year topological distance features enables BiLSTM to better capture temporal and structural irregularities leading to bankruptcy. Notably, *df_landscape*, which contained persistence landscape features, ranked second with a recall of 0.639, F1-score of 0.732, and G-mean of 0.792. In contrast, *df_none* (financial data only) recorded lower values with recall (0.614), F1-score (0.714), and G-mean (0.775), suggesting that financial indicators alone are less informative for long-term bankruptcy signals compared to topology-enhanced features. On the other hand, CNN models, which are adept at extracting local patterns, showed their strongest performance with the *df_landscape* dataset as seen in Table 4.4. CNN achieved a recall of 0.721, F1-score of 0.751, G-mean of 0.838, and AUC of 0.919, significantly outperforming its baseline *df_none* counterpart (recall = 0.626, F1-score = 0.713, G-mean = 0.781, AUC = 0.924). This highlights that CNN's convolutional architecture effectively leverages structured, non-topological engineered features such as landscape peaks and

patterns. Notably, CNN's performance on df_d3 was relatively poor compared to BiLSTM, with much lower recall (0.188), F1-score (0.226), and G-mean (0.269), reinforcing that the temporal modeling strength of BiLSTM is better suited for learning from topological distance patterns, while CNN benefits more from smoothed, shape-oriented features like persistence landscapes.

When evaluating two-year ahead bankruptcy prediction, performance generally declined across all models and data combinations, which is expected as the prediction horizon increases uncertainty. Still, the inclusion of topological features yielded noticeable gains in key metrics for both BiLSTM and CNN models which can be observed in the Table 4.5 and Table 4.6. For BiLSTM, *df_d1* achieved the highest recall (0.267) and F1-score (0.328), outperforming the financial-only baseline df_none, which had a recall of 0.199 and F1-score of 0.270. This demonstrates that even at an extended forecast horizon, early structural signals captured by topological distances improve the model's sensitivity to bankruptcy signals. Similarly, df_d2 and df_d3 showed better recall values (0.263 and 0.246 respectively) than df_none, confirming that topological distances help capture early-stage deterioration. However, precision remained relatively low across all configurations, reflecting the inherent challenge in identifying bankrupt firms well in advance. With CNN, performance on two-year ahead data was weaker overall, but the dataset df_all, which integrated all financial and topological features, produced the best results with a recall of 0.340, F1-score of 0.379, and G-mean of 0.559. This was a significant improvement over df_none (recall = 0.132, F1-score = 0.201, G-mean = 0.338), underscoring that even CNN, which was less responsive to distance-based features in the five-year setup, benefits from the full spectrum of financial and PH-derived attributes when predicting further into the future. Importantly, df_landscape also performed competitively for CNN with a recall of 0.198 and F1-score of 0.274, demonstrating once again that persistence landscape features retain predictive value across both temporal windows. These results validate that incorporating topological data—especially year-wise distances and landscape functions—enhances the predictive ability of deep learning models for both near- and long-term bankruptcy risk.

This project successfully demonstrates the combined effectiveness of financial, topological, and deep learning approaches in bankruptcy prediction. Overall, this integrated

approach offers significant promise for practical financial forecasting and risk management applications.

Table 4.3 Average performance metrics of 50 iterations for 5-year data using BiLSTM model for hyper tuned parameters.

df_version	Precision	Accuracy	Recall	Specificity	F1 Score	G-Mean	AUC
df_none	0.887	0.949	0.614	0.987	0.714	0.775	**0.930**
df_d1	0.877	0.949	0.606	0.989	0.705	0.770	0.909
df_d2	0.860	0.939	0.539	0.985	0.648	0.723	0.891
df_d3	**0.891**	0.954	**0.643**	0.990	**0.740**	**0.795**	0.928
df_d1d2d3	0.873	0.947	0.577	0.989	0.686	0.752	0.876
df_landscape	0.875	0.953	**0.639**	0.989	**0.732**	**0.792**	0.922
df_all	0.865	0.946	0.572	0.989	0.680	0.748	0.871

Table 4.4 Average performance metrics of 50 iterations for 5-year data using CNN model for hyper tuned parameters.

df_version	Precision	Accuracy	Recall	Specificity	F1 Score	G-Mean	AUC
df_none	0.855	0.950	0.626	0.987	0.713	0.781	**0.924**
df_d1	0.194	0.903	0.112	0.994	0.125	0.149	0.768
df_d2	0.857	0.948	0.621	0.986	0.706	0.778	0.907
df_d3	0.382	0.912	0.188	0.995	0.226	0.269	0.837
df_d1d2d3	0.854	0.950	0.635	0.987	0.723	0.789	0.902
df_landscape	0.794	0.950	**0.721**	0.977	**0.751**	**0.838**	0.919
df_all	0.766	0.937	0.556	0.980	0.631	0.725	0.877

Table 4.5 Average performance metrics of 50 iterations for 2-year ahead data using BiLSTM model for hyper tuned parameters.

df_version	Precision	Accuracy	Recall	Specificity	F1 Score	G-Mean	AUC
df_none	0.468	0.894	0.199	0.973	0.270	0.429	0.810
df_d1	0.454	0.890	**0.267**	0.961	**0.328**	**0.500**	0.783
df_d2	0.428	0.887	0.263	0.959	0.319	0.495	0.785
df_d3	0.432	0.889	0.246	0.962	0.304	0.474	0.783
df_d1d2d3	0.432	0.889	0.224	0.965	0.287	0.456	0.775
df_landscape	0.478	0.895	0.210	0.974	0.284	0.441	0.781
df_all	0.438	0.890	0.252	0.963	0.315	0.485	0.764

Table 4.6 Average performance metrics of 50 iterations for 2-year ahead data using CNN model for hyper tuned parameters.

df_version	Precision	Accuracy	Recall	Specificity	F1 Score	G-Mean	AUC
df_none	0.511	0.897	0.132	0.985	0.201	0.338	0.736

df_d1	0.462	0.895	0.175	0.977	0.246	0.397	0.754
df_d2	0.475	0.895	0.170	0.978	0.240	0.388	0.733
df_d3	0.292	0.888	0.069	0.982	0.109	0.232	0.710
df_d1d2d3	0.313	0.888	0.080	0.980	0.121	0.247	0.728
df_landscape	0.487	0.895	0.198	0.975	0.274	0.429	0.730
df_all	0.467	0.890	**0.340**	0.953	**0.379**	**0.559**	0.689

The analysis was implemented using both R and Python programming environments. R was primarily used for representing financial data as networks and for extracting topological features using the TDA package [Fasy et al, 2025] for PH related computations. Additional TDA libraries such as Ripser, and Dionysus were also employed to perform advanced PH analyses. Python was used for all machine learning tasks, including model training and evaluation, leveraging libraries such as scikit-learn, Keras [Chollet 2015] and TensorFlow libraries [Abadi et al. 2015]. The study utilized financial data from COMPUSTAT database. The computing for this project was performed at the OU Supercomputing Center for Education & Research (OSCER) at the University of Oklahoma (OU). Access to high-performance computing resources via the OSCER Supercomputer enabled large-scale experimentation and hyperparameter tuning. OSCER Associate Director for Remote & Heterogeneous Computing Horst Severini and OSCER Research Computing Facilitator Thang Ha provided valuable technical expertise.

4.5 Conclusion

Topological data analysis (TDA), and in particular persistent homology (PH), offers a powerful lens for understanding the geometric and structural evolution of complex systems. In the context of bankruptcy prediction, financial deterioration is rarely abrupt; it is a gradual process characterized by subtle, often nonlinear changes over time. Traditional predictive models typically analyze static snapshots of financial health, overlooking the rich temporal structure embedded in multiyear data. Motivated by this gap, this study adopts a topological perspective to explore how the shape of financial data evolves across time—capturing underlying relationships that may act as early indicators of corporate distress.

The methodological approach involved representing each year of financial data as a network based on feature similarities. PH was then applied to each annual network to compute

persistence diagrams, offering a multiscale summary of topological features across dimensions 0, 1, and 2. Rather than applying temporal persistence directly, the study quantified inter-year changes using Wasserstein and Bottleneck distances between persistence diagrams—resulting in three types of distance-based features: d1 (between Year 5 and all others), d2 (consecutive years), and d3 (between Year 1 and previous years). In parallel, persistence landscape features were extracted from each diagram, capturing shape descriptors such as peak height, area under the curve, and persistence quantiles. These topological features were combined with financial ratios to construct enriched datasets, which were then fed into deep learning models for classification.

The results showed that incorporating topological features led to measurable improvements in bankruptcy prediction. Specifically, the CNN model achieved its best performance with the *df_landscape* dataset, which combined financial data with persistence landscape features. On the other hand, the BiLSTM model achieved the highest overall metrics with *df_d3*, which included distance-based features comparing the bankrupt year to earlier years. This configuration consistently yielded the best recall, F-score, and G-mean, particularly in the 5-year prediction window. Hyperparameter tuning further optimized model performance, confirming that topological descriptors not only enhance predictive accuracy but also add robustness across different time horizons and data imbalance settings.

Future work can expand on this framework in several directions. First, instead of relying solely on pairwise distance comparisons, the application of true temporal persistence techniques—which track the evolution of topological features across time more continuously—could offer a richer understanding of dynamic changes in financial health. Another promising direction involves integrating graph neural networks (GNNs) with PH-derived node- or edge-level features to create more expressive models for analyzing dynamic financial networks. Additionally, incorporating unsupervised learning techniques such as clustering or anomaly detection using persistent homology may uncover latent patterns in firm behavior beyond binary outcomes. Exploring quarterly financial data can also enhance temporal resolution, potentially capturing finer-grained shifts that annual data may miss. Finally, improving the performance of two-year ahead prediction models remains a critical objective, as these models have strong potential for practical deployment

in early warning systems—allowing stakeholders to intervene well before financial distress develops. Collectively, these directions reinforce the value of combining TDA with deep learning to advance predictive modeling in complex, evolving financial environments.

Chapter 5: Closing Remarks

Bankruptcy prediction remains a critical challenge in finance, with wide-reaching impacts on stakeholders, supply chains, and the broader economy. Early and accurate identification of firms at risk can enable timely intervention and informed decision-making. This research addresses this vital problem by exploring how temporal financial data and advanced machine learning models can improve bankruptcy detection, especially in real-world settings where class imbalance and data complexity are common.

The first part of the study focused on identifying key financial indicators that could signal bankruptcy risk. Motivated by observable trends in financial trajectories closer to bankruptcy, the analysis emphasized the five most recent years of financial data, as they provided the strongest signals for distinguishing between bankrupt and non-bankrupt firms. Using recursive feature elimination, two key sets—12 and 54 features—were selected based on their predictive relevance. These refined feature sets were then used to train deep learning models, namely CNN and BiLSTM, to capture temporal patterns in firm-level financial behavior. Among these, the BiLSTM model consistently outperformed CNN, particularly for imbalanced datasets, due to its ability to model long-term dependencies in both forward and backward directions. Across various temporal prediction settings, BiLSTM achieved the highest AUC and recall scores, especially when using the 54-feature set with 3- and 5-year data. To improve interpretability, SHAP analysis was applied to both models, highlighting the most influential features contributing to predictions and enhancing model transparency.

The second part of this research incorporated persistent homology (PH) to extract topological features that characterize the evolving structure of financial data. PH was chosen for its unique ability to capture shape-driven patterns and multi-scale interactions in multivariate time series, which are often overlooked by traditional financial features. Using the persistence barcodes generated from dimensions 0, 1, and 2, the study derived persistence landscapes—summarizing key topological features such as the height, area, and count of peaks across multiple layers. In addition, Wasserstein and Bottleneck distances were computed to quantify changes between these persistence diagrams over time. These formed three types of distance-based features: d1 captured

differences between the 5th year and each of the remaining years; d2 compared consecutive years; and d3 measured the change between the 1st year (bankrupt year) and previous years. Among all feature combinations, the BiLSTM model with *df_d3* (financial + distance 3 features) achieved the best overall performance, with the highest recall, F-score, and G-mean. The *df_landscape* dataset (financial + landscape features) provided the second-best performance for BiLSTM, while for CNN, *df_landscape* consistently outperformed other configurations—achieving the highest recall and G-mean in the 5-year prediction window. These findings demonstrate that integrating PH-based insights, especially through persistence landscapes and distance metrics, adds significant predictive value to temporal financial risk modeling.

In conclusion, this research presents a robust framework that combines temporal deep learning, feature selection, and topological data analysis to enhance bankruptcy prediction. The integration of PH not only improves performance but also opens new directions for interpreting complex financial systems. This work lays the foundation for applying TDA in temporal analytics, particularly in network-based scenarios, such as supply-demand disruptions, clustering analysis, and feature dynamics. Overall, the findings contribute meaningful advancements to the field of financial risk modeling and network analytics.

This work opens several avenues for future exploration. One potential direction is evaluating the resilience of supply-demand networks before and after disruptions such as bankruptcies or economic shocks. Additionally, the impact of firm bankruptcy on broader supply chain connectivity can be studied to understand cascading failures. Clustering analysis could reveal communities or patterns in financial distress, while component importance metrics may help identify key firms or features driving network behavior. Lastly, since nodes in our temporal networks represent features, this framework can be extended to explore feature importance over time, contributing to more interpretable and dynamic bankruptcy prediction models.

References

Abadi, M., Agarwal, A., & Barham, P. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*.

Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Hanson, E., Motta, F., & Ziegelmeier, L. (2017). Persistence Images: A Stable Vector Representation of Persistent Homology. In *Journal of Machine Learning Research* (Vol. 18). http://jmlr.org/papers/v18/16-337.html.

Agarwal, V., & Taffler, R. (2007). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking and Finance*, *32*(8), 1541–1551.

Altman, E. (1968). Financial Ratios, Discriminant Aanalysis and The Prediction of Corporate Bankruptcy. *The Journal of FINANCE, XXIII*, 21.

Altman, E. I., Haldeman, R. G., & Narayanan, P. K. (1977). A new model identify bankruptcy risk of corporations. In *Journal of Banking and Finance* (Vol. 1, pp. 29–54).

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2014). Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. *Journal of International Financial Management and Accounting*, *28*(2), 131–171.

Antunes, F., Ribeiro, B., & Pereira, F. (2017). Probabilistic modeling and visualization for bankruptcy prediction. *Applied Soft Computing Journal*, *60*, 831–843.

Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long- short term memory. *International Committee of the Red Cross*, 1–24.

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, *83*, 405–417.

Beaver, W. H. (1966). Financial Ratios As Predictors of Failure Authors (s): William H . Beaver Source : Journal of Accounting Research , Vol . 4 , Empirical Research in Accounting : Selected Published by : Wiley on behalf of Accounting Research Center , Booth School of Busi. *Journal of Accounting Research*, *4*(1966), 71–111. http://www.jstor.org/stable/2490171

Bellovary, J. L., Giacomino, D. E., & Akers, M. D. (2007). A Review of Bankruptcy Prediction Studies: 1930 to Present. *Journal of Financial Education*, *Winter 2007*, 1–42.

Ben Jabeur, S., Stef, N., & Carmona, P. (2023). Bankruptcy Prediction using the XGBoost Algorithm and Variable Importance Feature Engineering. *Computational Economics*, *61*(2), 715–741.

Bliss, J. H. (1923). *Financial and Operating Ratios in Management*. New York: The Ronald Press. https://www.abebooks.com/first-edition/Financial-Operating-Ratios-Management-James-Bliss/22644116731/bd

Booth, A., Gerding, E., & McGroarty, F. (2014). Automated trading with performance weighted random forests and seasonality. *Expert Systems with Applications*, *41*(8), 3651–3661.

Bubenik, P. (2015). Statistical Topological Data Analysis using Persistence Landscapes. In *Journal of Machine Learning Research* (Vol. 16).

Carlsson, G. (2009). Topology and Data. In *Bulletin of the American Mathematical SocietyY* (Vol. 46, Issue 2). https://www.ams.org/journal-terms-of-use

Carton, R. C., & Hofer, C. W. (2006). *Measuring Organizational Performance: Metrics for Entrepreneurship and ... - Robert B. Carton, Charles W. Hofer - Google Books*.

https://books.google.com/books?hl=en&lr=&id=hg35tP2rfxoC&oi=fnd&pg=PR1&ots=G592ei7ZMv&sig=Va18bgR4ATgOoZcxMtjNY3VQtEc#v=onepage&q&f=false

Chandra, D. K., Ravi, V., & Bose, I. (2009). Failure prediction of dotcom companies using hybrid intelligent techniques. *Expert Systems with Applications*, *36*(3 PART 1), 4830–4837.

Chazal, F., de Silva, V., Glisse, M., & Oudot, S. (2016). *The Structure and Stability of Persistence Modules*.

Chen, T., Xu, R., He, Y., & Wang, X. (2017). Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. *Expert Systems with Applications*, *72*, 221–230.

Chollet: keras - Google Scholar. (n.d.). Retrieved June 16, 2021, from https://scholar.google.com/scholar?cluster=17868569268188187229&hl=en&oi=scholarr

Cleofas-Sánchez, L., García, V., Marqués, A. I., & Sánchez, J. S. (2016). Financial distress prediction using the hybrid associative memory with translation. *Applied Soft Computing Journal*, *44*, 144–152.

Cohen-Steiner, D., Edelsbrunner, H., & Harer, J. (2007). Stability of persistence diagrams. *Discrete and Computational Geometry*, *37*(1), 103–120.

Cun, Y. Le, Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Handwritten Digit Recognition with a Back-Propagation Network. *Proceedings of the Advances in Neural Information Processing Systems*, *39*, 9.

Edelsbrunner, H., & Harer, J. (n.d.). *COMPUTATIONAL TOPOLOGY AN INTRODUCTION*.

Fan, Y., Qian, Y., Xie, F.-L., & Soong, F. K. (2014). TTS synthesis with bidirectional LSTM based recurrent neural networks. *Fifteenth Annual Conference of the International Speech Communication Association*.

Fasy, B. T., Kim, J., Lecci, F., Maria, C., Millman, D. L., & Rouvreau., V. (2025). *Package "TDA" Title Statistical Tools for Topological Data Analysis*.

Feng, M., & Porter, M. A. (n.d.). *PERSISTENT HOMOLOGY OF GEOSPATIAL DATA: A CASE STUDY WITH VOTING*. www.math.ucla.edu/

Fitzpatrick, P. J. (1932). *A comparison of the ratios of successful industrial enterprises with those of failed companies.*

Fuqua, D., & Razzaghi, T. (2020). A cost-sensitive convolution neural network learning for control chart pattern recognition. *Expert Systems with Applications*, *150*.

Gamboa, J. (2017). Deep Learning for Time-Series Analysis. *ArXiv*.

Ghosh, S., & Reilly, D. L. (1994). Credit card fraud detection with a neural-network. *Proceedings of the Hawaii International Conference on System Sciences*, *3*, 621–630.

Graves, A., Jaitly, N., & Mohamed, A. (2013). *Hybrid Speech Recognition With Deep Bidirectional LSTM*. 273–278.

Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., & Schmidhuber, J. (2009). A novel connectionist system for unconstrained handwriting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(5), 855–868.

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, *18*(5–6), 602–610.

Guyon, I., Weston, J., & Barnhill, S. (2002). Gene selection for cancer classification using Support Vector Machine. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *46*, 389–422.

Heo, J., & Yang, J. Y. (2014). AdaBoost based bankruptcy forecasting of Korean construction companies. *Applied Soft Computing Journal*, *24*, 494–499.

Hochreiter, S., & Schmidhuber, J. (1997). *Long Short-Term Memory*. *1780*, 1735–1780.

Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications*, *117*, 287–299.

How, D. N. T., Loo, C. K., & Sahari, K. S. M. (2016). Behavior recognition for humanoid robots using long short-term memory. *International Journal of Advanced Robotic Systems*, *13*(6), 1–14.

Hu, Y. C. (2020). A multivariate grey prediction model with grey relational analysis for bankruptcy prediction problems. *Soft Computing*, *24*(6), 4259–4268.

Jiménez, F., Palma, J., Sánchez, G., Marín, D., Francisco Palacios, M. D., & Lucía López, M. D. (2020). Feature selection based multivariate time series forecasting: An application to antibiotic resistance outbreaks prediction. *Artificial Intelligence in Medicine*, *104*(August 2019).

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. *ArXiv Preprint ArXiv:1404.2188*.

Kim, H., Cho, H., & Ryu, D. (2022). Corporate Bankruptcy Prediction Using Machine Learning Methodologies with a Focus on Sequential Data. *Computational Economics*, *59*(3), 1231–1249.

Kim, M. J., & Kang, D. K. (2010). Ensemble with neural networks for bankruptcy prediction. *Expert Systems with Applications*, *37*(4), 3373–3379.

Kim, S. Y., & Upneja, A. (2014). Predicting restaurant financial distress using decision tree and

AdaBoosted decision tree models. *Economic Modelling*, *36*, 354–362.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems 25*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, *60*(6), 84–90.

Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997). Face recognition: A convolutional neural-network approach. *IEEE Transactions on Neural Networks*, *8*(1), 98–113.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2323.

Liang, D., Lu, C. C., Tsai, C. F., & Shih, G. A. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research*, *252*(2), 561–572.

Littleton, A. C. (1926). The 2 to 1 Ratio Analyzed'. *Certified Public Accountant*, 244–246.

López Iturriaga, F. J., & Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks. *Expert Systems with Applications*, *42*(6), 2857–2869.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, *274*(2), 743–758.

Mervin, C. (1942). Financing small corporations: in five manufacturing industries, 1926-36.

National Bureau of Economic Research.

Min, J. H., & Lee, Y. C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*, *28*(4), 603–614.

Munch, E. (2013). Applications of Persistent Homology to Time Varying Systems by. *Duke University*.

Nguyen, H. H., Viviani, J. L., & Ben Jabeur, S. (2023). Bankruptcy prediction using machine learning and Shapley additive explanations. In *Review of Quantitative Finance and Accounting* (Issue 0123456789). Springer US.

Niu, T., Wang, J., Lu, H., Yang, W., & Du, P. (2020). Developing a deep learning framework with two-stage feature selection for multivariate financial time series forecasting. *Expert Systems with Applications*, *148*, 113237.

Odom, M. D., & Sharda, R. (1990). A neural network model for bankruptcy prediction. *IJCNN. International Joint Conference on Neural Networks*, 163–168.

Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., & Harrington, H. A. (2017). A roadmap for the computation of persistent homology. *EPJ Data Science*, *6*(1).

Palangi, H., Ward, R., & Deng, L. (2016). Distributed Compressive Sensing: A Deep Learning Approach. *IEEE Transactions on Signal Processing*, *64*(17), 4504–4518.

Petri, G., Expert, P., Turkheimer, F., Carhart-Harris, R., Nutt, D., Hellyer, P. J., & Vaccarino, F. (2014). Homological scaffolds of brain functional networks. *Journal of the Royal Society Interface*, *11*(101).

Pun, C. S., Lee, S. X., & Xia, K. (2022). Persistent-homology-based machine learning: a survey and a comparative study. *Artificial Intelligence Review*, *55*(7), 5169–5213.

Pun, C. S., Xia, K., & Lee, S. X. (2018). Persistent-homology-based machine learning and its applications - A survey. *ArXiv*.

Qu, Y., Quan, P., Lei, M., & Shi, Y. (2019). Review of bankruptcy prediction using machine learning and deep learning techniques. *Procedia Computer Science*, *162*(Itqm 2019), 895–899.

Robinson, A., & Turner, K. (2013). Hypothesis Testing for Topological Data Analysis. *Journal of Applied and Computational Topology*, *1*(2), 241–261. http://arxiv.org/abs/1310.7467

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, *45*(11), 2673–2681.

Seversky, L. M., Davis, S., & Berger, M. (n.d.). *On Time-series Topological Data Analysis: New Data and Opportunities*.

Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2019). A Comparison of ARIMA and LSTM in Forecasting Time Series. *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, 1394–1401.

Smiti, S., & Soui, M. (2020). Bankruptcy Prediction Using Deep Learning Approach Based on Borderline SMOTE. *Information Systems Frontiers*, *22*(5), 1067–1083.

Telford, T., & Mufson, S. (2019). *PG&E, the nation's biggest utility company, files for bankruptcy after California wildfires*. Washingtonpost.Com. https://www.washingtonpost.com/business/2019/01/29/pge-nations-biggest-utility-company-files-bankruptcy-after-california-wildfires/

Tsai, C. F., Hsu, Y. F., & Yen, D. C. (2014). A comparative study of classifier ensembles for bankruptcy prediction. *Applied Soft Computing Journal*, *24*, 977–984.

Uthayakumar, J., Metawa, N., Shankar, K., & Lakshmanaprabu, S. K. (2018). Intelligent hybrid model for financial crisis prediction using machine learning techniques. *Information Systems and E-Business Management*, *0123456789*.

Uthayakumar, J., Metawa, N., Shankar, K., & Lakshmanaprabu, S. K. (2020). Financial crisis prediction model using ant colony optimization. *International Journal of Information Management*, *50*(October 2018), 538–556.

Wang, C. H., & Chen, J. Z. (2021). Combining hidden Markov models with probabilistic Bayes networks to conduct business forecasting and risk simulation. *Soft Computing*, *25*(13), 8773–8784.

Wang, G., Ma, J., & Yang, S. (2014). An improved boosting based on feature selection for corporate bankruptcy prediction. *Expert Systems with Applications*, *41*(5), 2353–2361.

Wilson, R. L., & Sharda, R. (2018). Bankruptcy prediction using neural networks. *Proceedings of the 2nd International Conference on Inventive Systems and Control, ICISC 2018*, *11*, 248–251.

Winakor, A. H., & Smith, R. F. (1935). Changes in the financial structure of unsuccessful industrial corporations. *University of Illinois Press*, *51*.

Wu, J., Chen, X. Y., Zhang, H., Xiong, L. D., Lei, H., & Deng, S. H. (2019). Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization. *Journal of Electronic Science and Technology*, *17*(1), 26–40.

Wu, Y., Gaunt, C., & Gray, S. (2010). A comparison of alternative bankruptcy prediction models. *Journal of Contemporary Accounting and Economics*, *6*(1), 34–45.

Yang, J. B., Nguyen, M. N., San, P. P., Li, X. L., & Krishnaswamy, S. (2015). Deep convolutional neural networks on multichannel time series for human activity recognition. *IJCAI*

International Joint Conference on Artificial Intelligence, *2015-Janua*, 3995–4001.

Yang, Z. R., Platt, M. B., & Platt, H. D. (1999). Probabilistic neural networks in bankruptcy prediction. *Journal of Business Research*, *44*(2), 67–74.

Yu, Q., Miche, Y., Séverin, E., & Lendasse, A. (2014). Bankruptcy prediction using Extreme Learning Machine and financial expertise. *Neurocomputing*, *128*, 296–302.

Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2014). Time series classification using multi-channels deep convolutional neural networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8485 LNCS*, 298–310.

Zomorodian, A., & Carlsson, G. (2004). Computing persistent homology. *Proceedings of the Annual Symposium on Computational Geometry*, *274*, 347–356.