Two-Step Classification for Solving Data Imbalance and Anomalies in an Altman Z-Score-based Bankruptcy Prediction Model

Abdul Syukur, Arry Maulana Syarif, Ika Novita Dewi, Aris Marjuni
Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia

***Abstract*—Differences in bankruptcy regulations with varying value parameters cause data anomalies when implemented in the Altman Z-Score model. Another common problem in bankruptcy predictions is imbalanced data; the number of companies that fall into the bankruptcy category is much smaller than those that do not. Therefore, a novel method was proposed to address data imbalance and anomalies in an Altman Z-Score-based bankruptcy prediction model. The proposed method employs a two-step classification controlled with data binning. Assumption values were used to set the proportion of distress and non-distress classes. Quartile calculation-based data binning is then used to ordinally rank the non-distress category into three classes. Furthermore, a two-step classification was performed using the Long-Short Term Memory (LSTM) method, followed by a rule-based classification method. The LSTM method predicts output in the form of one class representing the distress zone and three classes representing non-distress zone subcategories. The results are then processed using a rule-based classification to summarize the output into a two-class classification, where all data not in the distress zone class is part of the non-distress zone. The performance evaluation shows promising results, with outcomes closely matching the source bankruptcy data. These findings strengthen the evidence that the Altman Z-Score is a powerful tool for bankruptcy prediction and demonstrate that the proposed method can improve the Altman Z-Score model in handling differences in data value parameters.**

Keywords*—Bankruptcy prediction; Altman Z-Score; data imbalance and anomaly; data binning; two-steps classification; LSTM; rule-based classification

I. Introduction

Bankruptcy prediction is a major topic in the field of finance and an interesting and challenging research area in artificial intelligence, including machine learning and deep learning. The task of bankruptcy prediction is to measure the financial condition of a company, with the prediction output identifying companies that will go bankrupt within a certain period and those that will not. Despite various regulations and bankruptcy prediction tools, the Altman Z-Score remains a reliable method for measuring and anticipating bankruptcy risk in various industrial sectors, such as predicting bankruptcy in the automotive sector [1], tourism and hotel sectors [2], the supply chain sector [3], or the banking sector [4]. However, differences in bankruptcy parameters present challenges in developing bankruptcy prediction models [5]. These challenges also apply to the implementation of the Altman Z-Score model. Differences in bankruptcy regulations and varying value parameters cause data anomalies when implemented in the Altman Z-Score model. For example, in the public dataset of Taiwanese Bankruptcy Prediction – UCI machine learning repository collected from the Taiwan Economic Journal for the years 1999 to 2009, there are 6,819 observations, of which 220 are bankrupt companies and 6,599 are non-bankrupt companies. This results in a ratio of 3:95% for bankrupt companies (the distress zone) and non-bankrupt companies (the non-distress zone), respectively. However, when the data were calculated using the Altman Z-Score formula, there was a significant difference in the amount of data in the two classes. The calculation results show that the ratio changes to 51:49. Differences in scoring values in bankruptcy parameters cause data anomalies when the data are implemented in different bankruptcy prediction tools, leading to incorrect predictions. Another common problem in bankruptcy predictions is imbalanced data. The number of companies that fall into the bankruptcy category is much smaller than those that do not. Works by [6-7] and data found in the public dataset of Taiwanese Bankruptcy Prediction show that only around 3% of the total data falls into the class of bankrupt companies. Under these conditions, simply by categorizing all inputs into the non-bankrupt company class, the bankruptcy prediction algorithm will have an accuracy above 90%. Imbalanced data in bankruptcy prediction is a crucial problem [8-9].

Based on the description above, we propose a novel method to improve the Altman Z-Score model for bankruptcy prediction. The proposed model addresses data anomalies caused by differences in bankruptcy parameter scoring from other prediction tools, as well as the data imbalance problems commonly found in bankruptcy predictions. Classification techniques have become popular for solving bankruptcy prediction problems, with non-linear classification models demonstrating better accuracy than linear models [10]. Additionally, artificial intelligence and machine learning algorithms, particularly deep learning algorithms, have rapidly advanced in solving prediction and classification problems, including bankruptcy prediction as developed by [11-13]. The superiority of deep learning algorithms in these areas has motivated the development of a non-linear classification-based bankruptcy prediction method using Long Short-Term Memory (LSTM).

The proposed model uses value assumptions ranked ordinally through data binning techniques. The ranking results are used to divide the dataset into four classes based on bankruptcy potential. This data is then used as input for an LSTM model to learn and classify the potential bankruptcy of companies. The output target of the proposed model is a two-class classification that categorizes companies into either a non-distress zone or a distress zone. The four-class classification results from the LSTM are subsequently reclassified into two classes using rule-based classification techniques.

This study employs a two-step classification process using two classification models, which can be the same or different methods, working sequentially to determine the output. The first classification model reduces the problem's dimensions by producing an output that serves as the input for the second classification model. A similar two-step classification approach was used by [14], where the first classification detected problems, followed by the second classification to assess the problem's severity. The two-step classification model has demonstrated significant improvements in training and performance [15-16]. Details of the proposed model are organized as follows: Section II reviews relevant bankruptcy prediction research. Section III describes the model and methodology used in this study. Section IV discusses the experimental results. Finally, Section V presents the conclusions drawn from the proposed bankruptcy prediction model.

II. Related Works

Linear analysis for predicting bankruptcy was first introduced by Edward I. Altman in 1968 through the Z-Score formula, designed to forecast a company's likelihood of going bankrupt within two years. Since then, non-linear methods like Neural Networks and decision trees have gained traction in bankruptcy prediction [17-18]. Originally, the Z-Score relied on linear analysis, employing ordinal ranking for small datasets, with a focus on explainability and clarity [19]. Another approach involves a linear regression method that utilizes the Least Absolute Shrinkage and Selection Operator (Lasso) regression technique for feature selection, and ridge regression for dataset training [7]. In a different scenario, which involved ambient temperature and seasonal changes, [20] combined linear and non-linear regressions using quasi-Poisson regression analysis. This analysis, based on the assumption that variance is a linear function of the mean, aims to establish the relationship between dependent and independent variables, followed by modeling the association using the distributed lag non-linear model (DLNM). Comparative studies between generalized linear models (GLMs) and generalized additive models (GAMs) demonstrate that non-linear relationships in statistics and economics significantly enhance the discriminatory power in bankruptcy prediction [21].

Machine learning approaches, known for their reliability in handling large datasets and supporting nonparametric learning models, are well-suited for tackling complex non-linear problems [5]. Examples of non-linear classification models used in bankruptcy prediction include Random Forests [6], Decision Trees [7], Gradient Boosting [8], Support Vector Machine [9], and Artificial Neural Network and Ada Boost [12]. Deep learning, which leverages Artificial Neural Networks within the machine learning framework, has demonstrated superior performance compared to shallow machine learning models like simple Artificial Neural Networks and Decision Trees [22]. Despite their lack of explainability, machine learning and deep learning models, which operate in a black-box mode and require substantial datasets, yield effective results in bankruptcy prediction [7, 21, 23].

Time series analysis delves into unraveling the essence of a phenomenon by scrutinizing a sequence of data points across a specified timeframe. Bankruptcy prediction presents a multifaceted challenge within the domain of gray systems [24], often addressed through time series analysis to construct regression models [25]. Regression analysis, in turn, quantifies the correlation between a dependent variable and independent variables. Crucially, pinpointing relevant independent variables is pivotal for nonlinear analysis in bankruptcy prediction [21]. The impact of dataset size on prediction accuracy is gauged by gradually reducing the volume of training data. This approach not only determines the smallest dataset size that yields optimal predictions but also offers a potential remedy for imbalanced data by selectively discarding majority-class instances. Adequate data volume fosters robust data, thereby facilitating pattern recognition by LSTM networks. However, existing literature fails to provide conclusive evidence regarding the minimum data required for developing a bankruptcy prediction system using deep learning algorithms. For instance, [6] scrutinized bankruptcy data from 6819 Taiwanese enterprises spanning 1999 to 2009, while [21] analyzed data from 2635 companies over the period 2000 to 2014. Furthermore, [26] leveraged data from 3728 Belgian Small and Medium Enterprises (SMEs) between 2002 and 2012.

The selection of attributes to represent a company's bankruptcy significantly influences the performance of the bankruptcy prediction model. Building a simple model with minimal features is a crucial aspect of developing such a model, focusing on selecting highly relevant attributes as features [7]. Additionally, variations in regulations concerning bankruptcy determination and differences in economic characteristics across company locations, which serve as research subjects, influence the choice of attributes for predicting bankruptcy. For instance, in previous studies [7, 11, 12], varying numbers of features were employed: 110, 64, and 28, respectively. Despite the potential of machine learning algorithms to handle a large number of features [27], prioritizing the control of selected attributes as features remains essential to ensure the cost efficiency of the model.

Imbalanced datasets significantly impact overfitting as the model lacks sufficient input samples from minority classes [8]. Deep learning-based bankruptcy prediction has employed various techniques to address this challenge, including the Synthetic Minority Oversampling Technique (SMOTE), Stacked Autoencoder algorithm, and softmax classifier, resulting in high prediction accuracy even with imbalanced datasets [12]. Additionally, SMOTE was utilized by [28], to mitigate imbalanced data issues, coupled with LSTM for bankruptcy prediction. In a construction industry case, [29] enhanced LSTM bankruptcy prediction by incorporating construction market and macroeconomic variables alongside accounting variables. Furthermore, the study in [11] utilized three non-financial variables for bankruptcy prediction in the restaurant industry, noting their significant contribution to prediction accuracy. Both [11] and [28] underscored the importance of non-financial variables in enhancing prediction accuracy. Moreover, feature selection techniques play a crucial role in refining bankruptcy prediction models. [30] employed Genetic Algorithm (GA) to select features and control the number of units in the LSTM layers. Addressing outliers is another critical aspect for accurate model development. Outliers in the data can notably impair the performance of bankruptcy prediction models. Two prevalent approaches to handle outliers are omission and winsorization [13]. The research in [21] successfully employed winsorization to tackle outliers, while [5] opted to eliminate variables with over 1000 outliers.

Based on a search for research literature related to the experiments in this research, it can be concluded that outliers and data imbalances are common problems that are the focus of most bankruptcy prediction research using classification techniques, including bankruptcy classification based on the Altman Z-Score model. However, the phenomenon of differences in bankruptcy regulations with varying value parameters causing data anomalies when implemented in the Altman Z-Score model has not received attention and is still an open and challenging gap, that needs to be studied further as explained in this research.

III. METHODOLOGY

The quartile calculation-based data binning technique has been proposed to address imbalanced data and data anomaly issues in bankruptcy prediction using Altman Z-Score and LSTM. This method comprises two primary stages. The initial stage involves data preprocessing, which includes financial dataset preparation, outlier handling, Altman Z-Score calculation, quartile calculation-based data binning, and feature selection. The subsequent stage involves the classification process, which is divided into a four-class classification using LSTM, followed by two-class classification rules to ascertain distress zone and non-distress zone outputs. Fig. 1 illustrates the flow diagram of the proposed prediction model.

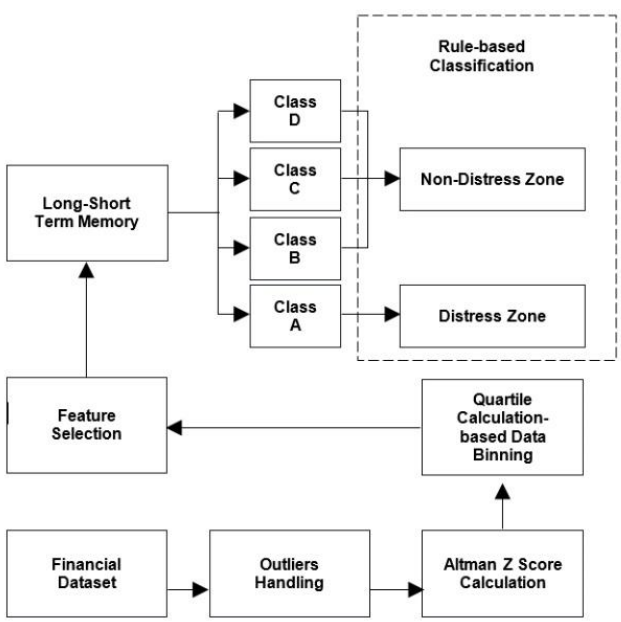

Fig. 1. The proposed bankruptcy prediction model diagram.

A. Data Preprocessing

The data preprocessing phase comprises five stages: financial dataset preparation, outliers handling, Altman Z-Score calculation, quartile calculation-based data binning, and feature selection. The dataset was initially sourced from the Taiwanese Bankruptcy Prediction – UCI machine learning repository, compiled from the Taiwan Economic Journal spanning the years 1999 to 2009 (Taiwanese Bankruptcy Prediction, 2020). Company bankruptcy was determined according to the business regulations of the Taiwan Stock Exchange. The sample consists of 96 financial ratios and 6,819 observations, with 220 representing bankrupt companies and 6,599 representing non-bankrupt companies. While the dataset does not exhibit outlier problems, there was imbalanced data within the distress zone and the non-distress zone classes, with a ratio of 3:95%. However, upon calculating the data using the Altman Z-Score formula, a significant difference in the data volume between the two classes emerged. The calculations reveal that 3,467 companies are in the distress zone, while 3,352 companies are in the non-distress zone, resulting in a revised ratio of 51:49% respectively.

Assumptions derived from the prevalence of imbalanced data, often encountered in bankruptcy prediction problems, serve to reconcile disparities between Taiwan Stock Exchange regulations and the Altman Z-Score framework in classifying companies into distress or non-distress zones. A baseline proportion of 5:95% is adopted to establish assumptions for the distress and non-distress zones, respectively. Following computation via the Altman Z Score formula, the data is arranged in ascending order. Subsequently, the bottom 5% of the data, representing companies with the smallest values, is considered to belong to the distress zone class.

Let $D_N = \{(X_i, Y_i)\}_{i=1}^{N}$ is a dataset of N rows ordered by class Y, where $X = (X_1, X_2, \dots, X_p)$ represents the p attributes of D_N and Y is the class of D_N with $Y = F(X)$. Suppose the dataset D_N contains m rows of distress classes and n rows of non-distress classes, then the dataset D_N can be split into two independent sub-datasets, which are:

$$D_m = \{(X_i, Y_i)\}_{i=1}^{m}$$

$$D_n = \{(X_i, Y_i)\}_{i=1}^{n}$$

with splitting ratio: $\gamma = m/N$

where,

$$m + n = N$$

$$D_m \cup D_n = D_N \tag{1}$$

If it is assumed that 5% of the classes are distress, then the sub-datasets are as follows:

$$D_{5\%} = \{(X_i, Y_i)\}_{i=1}^{0.05N}$$

$$D_{0.95} = \{(X_i, Y_i)\}_{i=1}^{0.95N}$$

with the splitting ratio of $\gamma = 5\%$ (2)

The splitting results indicate an imbalance in these sub-datasets, posing a challenge for classification purposes. Out of the 6819 data points sorted based on the Altman Z Score calculation, assuming a value of 5%, the distress zone class comprises 349 companies with the smallest Z values. Conversely, the non-distress zone class encompasses the remaining 6,470 companies. Rather than resorting to under-sampling or over-sampling methods to address the data imbalance, an alternative approach could involve segmenting 95% of the non-distress zone data into multiple categories using data binning. Data binning transforms continuous data into categorical data, with each category mirroring the size of the distress zone class. Consequently, the length of each bin would be determined by dividing the amount of data in the non-distress zone category by the size of the distress zone class.

Classes based on the length of bins, which are sub-categories of the non-distress zone, are supposed to have a balanced proportion relative to the size of the distress zone class. However, with an imbalanced data proportion of 5% in distress zone class and 95% in non-distress zone class, the number of subcategories within the non-distress zone becomes less controlled. Binning the non-distress zone class size of 6,470 with a non-distress zone class size of 349 results in 18 subcategories within the non-distress zone. Having 19 classes, comprising 1 distress zone class and 18 classes from the 18 subcategories of the non-distress zone, will complicate the classification process in prediction. The data binning approach will be used to address the imbalance in sub-datasets. This means that the initial supposition of the non-distress sub-dataset will be split into three equal parts restrained by values as follows:

$$Q_j = \left| \frac{j(n+1)}{3} \right|_{j=1}^{2} \tag{3}$$

Let D_1 represent the $D_{95\%}$ of the initial supposition of the non-distress subcategory, comprising $n = 6{,}478$ rows. Subsequently, the data binning approach will produce the subcategories of $(D_{1i})_{i=1}^{3}$, which are D_{11}, D_{12}, and D_{13}, separated by $Q_1 = 2{,}508$, and $Q_2 = 4{,}665$ with a size of 2157 rows for each subcategory. Considering the distress zone category as Class A and the three subcategories of the non-distress zone as Class B, C, and D, Class A comprises 341 rows, indexed from 1 to 349. Meanwhile, Classes B, C, and D encompass 2,157 rows each, indexed from 350 to 2507, 2508 to 4664, and 4665 to 6819, respectively.

At this point, the results still exhibit imbalance issues. Under sampling is conducted to ensure that the sizes of classes B, C, and D are proportional to class A. Some rows within classes B, C, and D are deleted at specific index intervals determined using quartile calculation. The interval value for the calculation is derived by dividing the highest size by the lowest size among the four classes. Thus, the resulting interval value is 2157 divided by 349, which equals six. Consequently, each class B, C and D comprises 360 rows of data, with index numbers (350, 356, 362, ..., 2504), (2505, 2511, 2517, ..., 4659), and (4660, 4666, 4672, ..., 6814), respectively. Finally, the data preprocessing stage yields class A, B, C, and D with 349, 360, 360, and 360 rows, respectively. Class A represents the distress zone category, while the non-distress zone category is divided into three subcategories: class B, C, and D, representing the non-distress zone to the strong non-distress zone. This technique facilitates the proportional distribution of data across sorted values, ensuring representation from the smallest to the largest values. Fig. 2 illustrates the flow diagram of the data preprocessing stage.

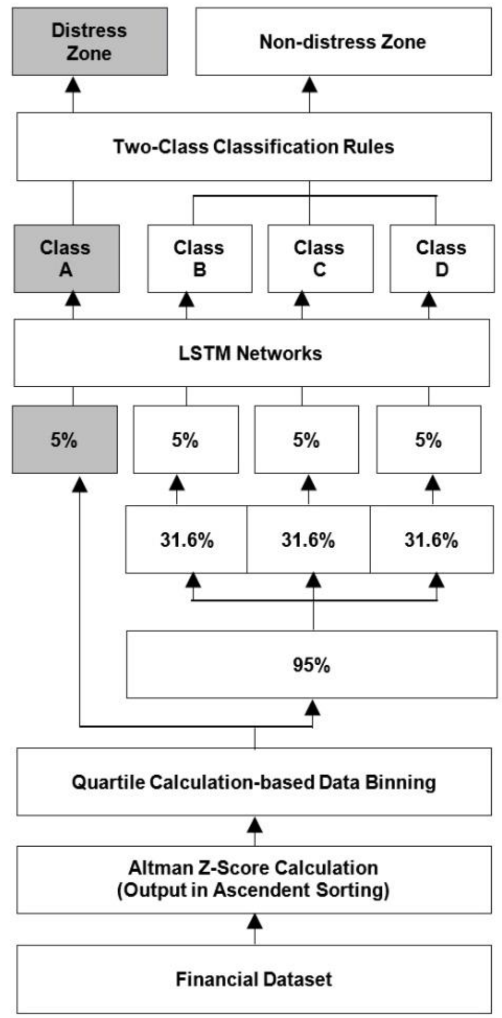

Fig. 2. The data preprocessing diagram.

The new dataset consists of 29 attributes and 1 predicted class of financial ratios. These attributes encompass a range of metrics including Operating Profit Rate, Pre-tax Net Interest Rate, After-tax Net Interest Rate, Net Value Per Share (B), Net Value Per Share (A), Net Value Per Share (C), Revenue Per Share (in Yuan), Operating Profit Per Share (in Yuan), Total Asset Growth Rate, Total Debt/Total Net Worth, Debt Ratio %, Net Worth/Assets, Operating Profit/Paid-in Capital, Net Profit Before Tax/Paid-in Capital, Revenue Per Person, Working Capital to Total Assets, Current Assets/Total Assets, Current Liability to Assets, Inventory/Working Capital, Long-term Liability to Current Assets, Retained Earnings to Total Assets, Total Assets to GNP Price, Gross Profit to Sales, Working Capital, EBIT, Market Value of Equity, Book Value of Total Debt, Sales, and Total Assets.

Fig. 3. The feature selection diagram.

Feature selection is a pivotal step, significantly enhancing model performance, curbing overfitting, and expediting training times. Choosing the most pertinent features is crucial for constructing efficient and interpretable models. In the realm of deep learning, a valuable approach is selecting features based on their weights. The Select Feature by Weight technique was employed for feature selection, prioritizing the weights assigned to each feature during training to pinpoint and preserve the most impactful ones, thereby augmenting model performance. Features with higher absolute weights hold greater influence in the model. Refer to Fig. 3 for the flow diagram illustrating the feature selection process.

The process began with normalizing the values of features within the dataset, followed by assigning weights to the features and ranking them. Normalization methods are preprocessing techniques used to standardize feature values within a dataset, ensuring they are all on the same scale. In this research, Z-transformation methods were applied. Z-score normalization entails transforming variable values to have a mean of 0 and a standard deviation of 1.

Suppose we possess a dataset comprising n subjects. Let X= 〖{x〗 _1,x_2,…,x_j,…,x_m} represent the set of normalized feature values within the chosen data. The formula for Z-transformation is as follows:

$$Z = \frac{X-\mu}{\sigma}, \tag{4}$$

where:

Z is the Z-score,

X is the original data point,

μ is the mean of the dataset,

σ is the standard deviation of the dataset

The feature selection method evaluates the relevance of each feature to both the target feature and the prediction model, deciding whether to include or exclude it from the prediction process. Feature weighting determines the magnitude of influence or importance that each feature has in relation to the target feature and the prediction model. The resulting importance score directs the utilization of the feature's magnitude in prediction, influencing its impact on the overall model.

Consider the following linear regression model: it utilizes normalized feature values available in the selected dataset. The formula for Z-transformation is as follows:

$$\hat{y} = w_0 + w_1 \cdot x_1 + w_2 \cdot x_2 + \dots + w_n \cdot x_n \tag{5}$$

where:

$\hat{y}$ is the predicted output,

w_0 is the intercept,

$w_1, w_2, \dots w_n$ are the weights corresponding to features $x_1, x_2, \dots, x_n$, respectively.

The selected features are determined using the formula: $selected\ features = \{x_i | w_i \geq T\}$, where T is a predetermined threshold value. The correlation matrix of key financial indicators used in the bankruptcy prediction model is displayed in the following heatmap. Fig. 4 shows the heatmap, which represents the correlations between features. All the colored cells indicate the correlation between two features, with the color of the cell denoting the strength of the correlation. A correlation value less than zero indicates a negative correlation, while a zero value indicates no correlation.

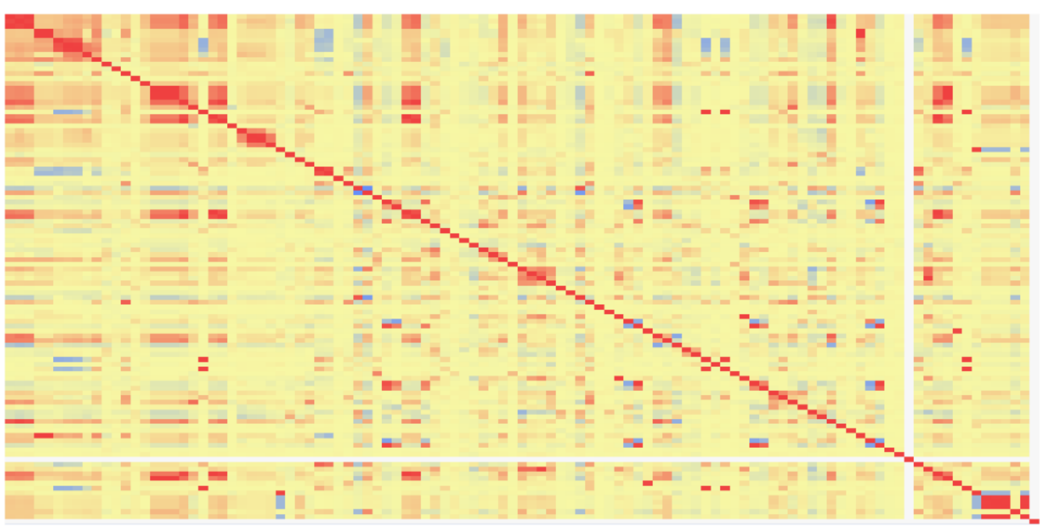

Fig. 4. The heatmap of the correlation matrix of key financial indicators used in the bankruptcy prediction model.

B. Two-Step Classification

Deep learning is a powerful and versatile approach for making predictions across various domains, including bankruptcy prediction. These models are built upon neural networks, which consist of layers of interconnected nodes (neurons). A typical deep neural network includes an input layer, multiple hidden layers, and an output layer. The input layer of the neural network receives raw features or data points related to the prediction task. Each node in the input layer corresponds to a specific feature, and these values are fed into the network. Connections between nodes in different layers are associated with weights. During training, the model adjusts these weights to learn optimal patterns from the input data. Each node in the hidden layers performs a weighted sum of its inputs and passes the result through an activation function. ReLU is commonly used for hidden layers, while Softmax is applied to the output layer.

The flow diagram for bankruptcy prediction begins with learning representations. As data passes through the layers, the network learns hierarchical representations. Lower layers capture simple patterns, while higher layers combine these patterns to form more complex and abstract features. This hierarchical learning enables the model to automatically extract relevant features from the input data without explicit feature engineering. Next is the loss function measurement, where the output of the network is compared to the actual target values using a loss function. The loss quantifies the difference between the predicted and actual values. The goal during training is to minimize this loss by adjusting the weights of the connections through a process called backpropagation. The optimization algorithm stage follows, using methods such as stochastic gradient descent (SGD) to iteratively update the weights based on the calculated gradients of the loss function. This process allows the model to converge towards a set of weights that minimizes the prediction error on the training data. The next stage is data training, where the network is trained on a labeled dataset containing both input features and corresponding target labels. The model iteratively adjusts its parameters to improve its predictive performance. The final stage is testing and evaluation. Once trained, the model is evaluated on a separate test set to assess its generalization performance. Various metrics, depending on the prediction task, are used to measure performance, such as accuracy, precision, recall, and F1-score. Fig. 5 shows the architecture of the LSTM used for the proposed bankruptcy prediction.

The deep learning architecture begins with initializing the weights (W^1, W^2) and biases (b^1, b^2). Next, an input vector X with n features is provided to the input layer, where:

$$X = [x_1, x_2, \dots, x_n]:$$

$$Z^{[1]} = X \cdot W^{[1]} + b^{[1]}$$

$$A^1 = ReLU(Z^1) \quad (6)$$

where:

$Z^{[1]}$ is the weighted sum of inputs, and A^1 is the output after applying the ReLU activation function.

Next is the definition of the relationship between the hidden layer and the output layer. Given A^1, the hidden layer activations:

$$Z^{[2]} = A^1 \cdot W^{[2]} + b^{[2]}$$

$$A^{[2]} = \sigma\left(Z^{[2]}\right), \quad (7)$$

where:

$Z^{[2]}$ represents the weighted sum of the hidden layer activations, and A^2 is the output after applying the softmax activation function (σ).

The softmax activation function (σ) is applied element-wise to the output of the second layer:

$$A_{i,j}^{[2]} = \frac{e^{z_{i,j}^2}}{\sum_{k=1}^{C} e^{z_{i,j}^2}}\ , \quad (8)$$

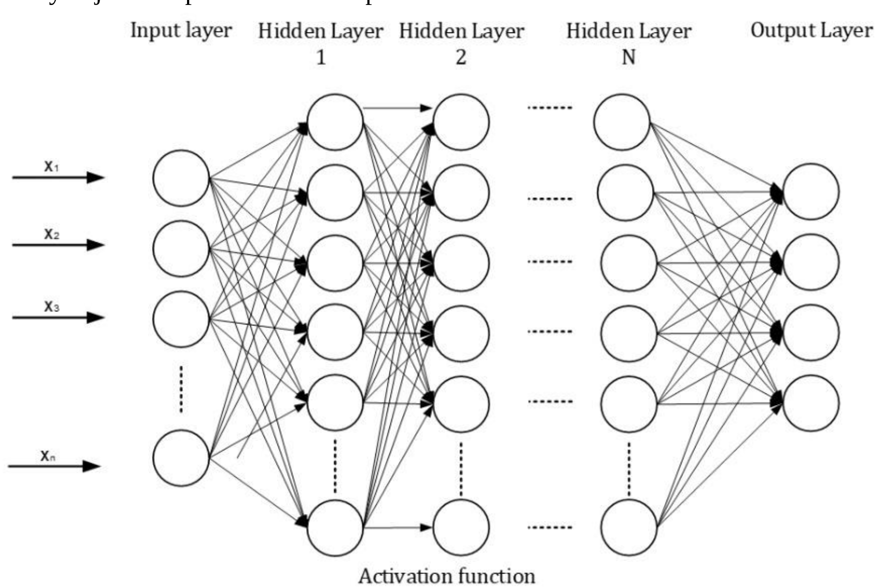

Fig. 5. The architecture of LSTM used for the proposed bankruptcy prediction.

where:

C is the number of classes. Additionally, the loss function is applied using categorical cross-entropy as follows:

$$J = -\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{C} Y_{i,j} \cdot log(A_{i,j}^{[2]})$$

Backpropagation

$$dZ^{[2]}=A^{[2]} - Y$$

$$dW^{[2]} = \frac{1}{m}A^{[1]T} \cdot dZ^{[2]}$$

$$db^{[2]} = \frac{1}{m}\sum_{1=1}^{m} dZ^{[2]}$$

$$dZ^{[1]} = (W^{[2]})^T \cdot dZ^{[2]} \odot ReLU'(Z^{[1]})$$

$$dW^{[1]} = \frac{1}{m}X^T \cdot dZ^{]1]}$$

$$db^{[1]} = \frac{1}{m}\sum_{i=1}^{m} dZ^{]1]}$$

$\odot$ denotes element-wise multiplication. (9)

Finally, the parameter update using gradient descent is applied to optimize the weights and biases as follows:

$$W^{[2]} = W^{[2]} - \alpha \cdot dW^{[2]}$$

$$b^{[2]} = b^{[2]} - \alpha \cdot db^{[2]}$$

$$W^{[1]} = W^{[1]} - \alpha \cdot dW^{[1]}$$

$$b^{[1]} = b^{[1]} - \alpha \cdot db^{[1]} \quad , \tag{10}$$

The final classification output comprises two classes: distress and non-distress categories. The non-distress category undergoes division using data binning techniques, resulting in three classes: B, C, and D. Class A represents the distress category. Consequently, any incorrect predictions among classes B, C, and D are rectified since they fall within the non-distress category. Following the prediction of the four classes, a rule is applied for the final classification. If the prediction outcome is class A, the data is classified into the distress zone category; otherwise, it is classified into the non-distress zone category. The formula for the two-class classification rule, with BP denoting bankruptcy prediction, is as follows:

$$BP = \begin{cases} Distress\ ,\ Class\ A \\ Non-distress, others \end{cases} \tag{11}$$

IV. Results and Discussion

A classification-based bankruptcy prediction model is proposed by utilizing the Altman Z Score on different bankruptcy regulations, and the data binning method is used to overcome the problem of imbalanced data. The extreme data imbalance commonly found in bankruptcy prediction datasets is assumed to define companies falling into the distress zone category. The baseline for companies in the distress zone comprises 5% of the data with the smallest Z Score values in the dataset, with the remainder falling into the non-distress zone category. Of the data in the non-distress zone category, 95% is further divided into three sub-categories using the data binning method to represent varying levels of non-distress, from non-distress to strong non-distress. To address the issue of disproportionate data among the three non-distress zone sub-categories and the distress-zone class, under-sampling and quartile calculation techniques are employed. Data is deleted at specific intervals until the amount of data in each of the three non-distress zone sub-categories is proportional to the amount of data in the distress-zone class. This method ensures the representativeness of Z Score values across classes, maintaining the order from smallest to largest within each class. Instead of balancing the data by selecting 330 of 6599 non-distress zone data randomly as done by [5], binning data into subcategories and then cutting to select a portion of the data within each subcategory is more precise in representing the data by selecting a small portion of a large amount of data.

Data preprocessing yields four classes: one distress zone class (Class A) and three sub-categories of the non-distress zone (Classes B, C, and D). Feature selection, employing feature weighting techniques, identifies 29 features from 96 financial ratios defined by the Taiwan Stock Exchange, along with Classes B, C, and D, each comprising 360 data points. The dataset is split 70:30 for training and testing in the LSTM network, resulting in 1,000 training data points and 429 testing data points. Evaluation of the LSTM network's performance is conducted using precision and recall metrics. Tables I and II present the evaluation results for both training and testing data.

The accuracy achieved in evaluating training data reached 86.3%. Incorrect predictions between class A and class B are the most frequently found. 43 data points that should be included in class A are predicted to be class B. This condition also occurred in 21 instances where data intended for class B was predicted as class A, 30 instances where data meant for class C are predicted to be class B, and 24 instances where data designated for class D are predicted to be class C. This is because the data binning technique, which converts categorical data into ordinal data, makes the distance between successive classes closer. For example, the distance between class A and class B is closer than between classes C and D. Compared with classes A and D, classes B and C are closer to the previous and following classes. However, the incorrect prediction in classes B and C turns out to be stronger in their previous class. Incorrect predictions from class B to class A occur in 21 instances, and there are no incorrect predictions to class C. Incorrect predictions from class C to class B occur in 30 instances, and 10 instances to class D. In terms of recall, the incorrect prediction between classes A and B still dominates among the other classes. The recall evaluation value shows that 21 data points, which should be included in Class A, are predicted by the system as part of class B. Vice versa, 43 data points in class B are predicted as class A. This shows that the system still has difficulty distinguishing the characteristics of classes A and B. There are similar patterns in predicting bankruptcy between training data evaluation and test data evaluation, with incorrect predictions dominated in classes A and B. However, compared with training data evaluation, there is an increase in accuracy in test data evaluation. The accuracy achieved in evaluating test data reached 89.98%, representing an increase of 3.68%.

TABLE I. PRECISION AND RECALL RESULTS FOR THE DATA TRAINING EVALUATION

	True Class A	True Class B	True Class C	True Class D	Class Precision
Class A	219	43	2	3	82.02%
Class B	21	179	0	0	89.5%
Class C	4	30	226	10	83.7%
Class D	0	0	24	239	90.87%
Class Recall	89.75%	71.03%	89.68%	94.84%	

TABLE II. PRECISION AND RECALL RESULTS FOR THE DATA TEST EVALUATION

	True Class A	True Class B	True Class C	True Class D	Class Precision
Class A	97	8	0	0	92.38%
Class B	6	100	5	0	90.09%
Class C	2	0	81	0	97.59%
Class D	0	0	22	108	83.08%
Class Recall	92.38%	92.59%	75%	100%	

At this point, the classification process was still underway, transitioning from a four-class classification to a binary classification, consisting of distress and non-distress categories. This process serves as the cornerstone of the proposed prediction model. The final classification output consists of two categories: distress and non-distress. Class A denotes the distress category, while classes B, C, and D represent the non-distress category. Following the initial four-class classification, additional rules were integrated to facilitate the transition to the two-class classification, which predicts distress and non-distress categories. Misclassifications within classes B, C, and D do not constitute errors as these classes are encompassed within the non-distress category. The proposed bankruptcy prediction model, employing a two-stage classification approach, starting with a four-class classification and proceeding to a two-class classification, demonstrates a notable enhancement in performance. In the four-class classification, the accuracy attained for both training and test data evaluation stood at 86.3% and 89.98%, respectively. Furthermore, the subsequent two-class classification stage yields improved accuracy in both training and test data evaluation, reaching levels of 92.70% and 96.27%, respectively. Tables III and IV present the evaluation outcomes for training and test data during the two-class classification stage. However, in the four-class classification, the system encounters challenges in distinguishing between class B, class A, and class C, resulting in a recall performance of 71.03%. Given that class B and class C fall within the non-distress category, post-two-class classification processing strengthens the overall recall performance within the non-distress category to 93.65%. This trend is mirrored in precision performance metrics within the non-distress category, exhibiting an increase in values up to 96.59%.

There are consistent patterns observed in predicting bankruptcy across both training data evaluation and test data evaluation in both four-class and two-class classifications. Notably, there is an uptick in accuracy during test data evaluation, reaching 96.27%. This marks a 3.57% increase from the accuracy observed during training data evaluation, which stood at 92.7%. By first binning the data to determine 1 distress zone category and three non-distress zone sub-categories and applying four-class classification using an LSTM network followed by two-class classification followed by a two-class classification to summarize predictions into distress and non-distress zone classes, prediction accuracy can be increased to more than 90% and outperform the accuracy achievements of the method proposed by study [5] which is below 90%. Table V illustrates the accuracy comparison between four-class and two-class classifications, while Fig. 6 presents a visual comparison of accuracy across these classifications.

TABLE III. PRECISION AND RECALL RESULTS FOR THE DATA TRAINING EVALUATION IN THE TWO-CLASS CLASSIFICATION

	True Distress Zone	True Non-Distress Zone	Class Precision
Distress Zone (Class A)	219	48	82.02%
Non-Distress Zone (Class B, C, D)	25	708	96.59%
Class Recall	89.75%	93.65%	

TABLE IV. PRECISION AND RECALL RESULTS FOR THE DATA TEST EVALUATION IN THE TWO-CLASS CLASSIFICATION

	True Distress Zone	True Non-Distress Zone	Class Precision
Distress Zone (Class A)	97	8	92.38%
Non-Distress Zone (Class B, C, D)	8	316	97.53%
Class Recall	92.38%	97.53%	

TABLE V. ACCURACY FROM FOUR-CLASS CLASSIFICATION TO TWO-CLASS CLASSIFICATION

	Four-class Classification	**Two-class Classification**
Training data	86.3%	92.7%
Test data	89.98%	96.27%

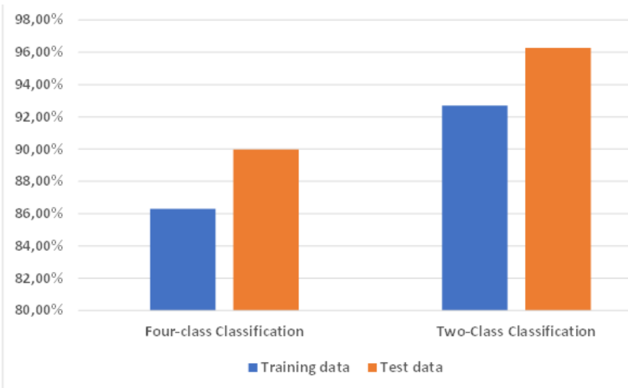

Fig. 6. Accuracy from four-class classification to two-class classification.

Experiments demonstrate that the proposed method yields promising results in crafting a bankruptcy prediction model capable of addressing regulatory disparities in assessing a company's bankruptcy risk, as well as tackling imbalanced data issues. In binary classification scenarios, employing data binning techniques can mitigate imbalanced data concerns by partitioning the predominant data category into multiple classes or subcategories via ordinal ranking. Once the data category is segmented into multiple classes, incorporating rules to revert the workflow to binary classification significantly enhances accuracy, precision, and recall performance. Conversely, the application of data binning techniques necessitates a ranking framework, and the Altman Z Score has been validated as a suitable benchmark for ranking data in bankruptcy prediction tasks.

V. CONCLUSION

The bankruptcy prediction model has demonstrated proficiency in addressing challenges related to variations in data value parameters and imbalanced data. Experimental findings bolster the assertion that the Altman Z-Score serves as a robust tool for predicting bankruptcy. Moreover, the proposed method enhances the Altman Z-Score model's ability to handle variations in data value parameters. Segmenting the non-distress zone category into multiple classes through data binning has effectively elucidated the distinguishing characteristics of companies with high Z scores in the non-distress zone, juxtaposed with those in the distress zone.

The proposed method is proven to be reliable as a bridge over different regulations and parameter values with the Altman Z-Score model in determining bankruptcy. The differences produce bankruptcy data anomalies, namely the characteristics of extreme imbalance data in company bankruptcy data which disappear when recalculated using the Altman Z-Score model. This problem can be bridged using the proposed method so that the imbalance data pattern which is an inherent characteristic of company bankruptcy data can be maintained with good accuracy of bankruptcy prediction results. However, the system encounters difficulty in discerning companies within the non-distress zone with rankings proximate to the distress zone. Conversely, while employing an LSTM-based four-class classification for bankruptcy prediction yields promising outcomes, identifying the precise cause of the model's accuracy weakness remains challenging. Current indications suggest that enhancing feature selection performance is imperative to enable the system to differentiate between distress classes and non-distress subcategories with minimal value disparities. Additionally, further investigation is warranted to determine the optimal addition or reduction of financial ratio variables as features.

REFERENCES

[1] A. Sareen, and S. Sharma, “Assessing financial distress and predicting stock prices of automotive sector: robustness of altman z-score,” Vision, vol. 26, no. 1, pp. 11-24, 2022. Doi: 10.1177/09722629219909.

[2] E. Goh, S. M. Roni, and D. Bannigidadmath, “Thomas cook (ed): using altman's z-score analysis to examine predictors of financial bankruptcy in tourism and hospitality businesses,” Asia Pacific Journal of Marketing and Logistics, vol. 34, no.3, pp. 475-487, 2022. Doi: 10.1108/APJML-02-2021-0126.

[3] R. Alcalde, C. Alonso de Armiño, S. García, “Analysis of the economic sustainability of the supply chain sector by applying the altman z-score predictor,” Sustainability, vol. 14, 2022. Doi: 10.3390/su14020851.

[4] J. Elia, E. Toros, C. Sawaya, and M. Balouza, “Using altman z-score to predict financial distress: evidence from lebanese alpha banks,” Management Studies and Economic Systems, vol. 6, no. 1/2, pp. 47-57, 2021.

[5] R. F., Brenes, and A. Johannssen, “An intelligent bankruptcy prediction model using a multilayer perceptron,” Intelligent Systems with Applications, vol. 16, 2022. Doi: 10.1016/j.iswa.2022.200136.

[6] H. Wang, and X. Liu, “Undersampling bankruptcy prediction: taiwan bankruptcy data,” PLoS ONE, vol. 16, no. 7, 2021. Doi: 10.1371/journal.pone.0254030.

[7] M. S. Park, H. Son, C. Hyun, and H. J. Hwang, “Explainability of machine learning models for bankruptcy prediction,” IEEE Access, vol. 9, pp. 124887-124899, 2021. Doi: 10.1109/ACCESS.2021.3110270.

[8] A. Tabbakh, J. K. Rout, K. S. Sahoo, and N. Z. Jhanjhic, “Bankruptcy prediction using robust machine learning model,” Turkish Journal of Computer and Mathematics Education, vol. 12, no. 10, pp. 3060-3073, 2021.

[9] G. Lombardo, M. Pellegrino, G. Adosoglou, S. Cagnoni, P. M. Pardalos, and A. Poggi, “Machine learning for bankruptcy prediction in the american stock market: dataset and benchmarks,” Future Internet, vol. 14, no. 8. 2022. Doi: 10.3390/fi14080244.

[10] A. Ptak-Chmielewska, “Bankruptcy prediction of small- and medium-sized enterprises in poland based on the lda and svm methods,” Statistics in Transition New Series, vol. 22, no. 1, pp. 179–195. 2021. Doi: 10.21307/stattrans-2021-010.

[11] R. Becerra-Vicario, D. Alaminos, E. Aranda, and M. A. Fernández-Gámez, “Deep recurrent convolutional neural network for bankruptcy prediction: a case of the restaurant industry,” Sustainability, vol. 12, no. 12, 5180, 2020. Doi: 10.3390/su12125180.

[12] S. Smiti, and M. Soui, “Bankruptcy prediction using deep learning approach based on borderline smote,” Information Systems Frontiers, vol. 22, pp. 1067–1083, 2020. Doi: 10.1007/s10796-020-10031-6.

[13] D. Ogachi, R. Ndege, P. Gaturu, Z. Zoltan, “Corporate bankruptcy prediction model, a special focus on listed companies in kenya,” Journal

of Risk and Financial Management, vol. 13, no. 3, 2020. Doi: 10.3390/jrfm13030047.

[14] T. S. Tran, V. P. Tran, H. J. Lee, J. M. Flores, and V. P. Le. "A two-step sequential automated crack detection and severity classification process for asphalt pavements," International Journal of Pavement Engineering , vol. 23, no. 6, 2022. Doi: 10.1080/10298436.2020.1836561.

[15] J. H. Park, and P. Fung, "One-step and two-step classification for abusive language detection on twitter," in proceedings of the first workshop on abusive language online, Aug. 2017, pp. 41–45. Doi: 10.18653/v1/W17-3006.

[16] T. Losel, and Y. Kim, "A comparison of 2 step classification with 3-class classification for webpage classification," in 13th International Conference on Information and Communication Technology Convergence (ICTC), Oct. 2022, pp. 654-657. Doi: 10.1109/ICTC55196.2022.9952595.

[17] V. García, I. Ana, J. Marqués, S. Sánchez, and H. J. Ochoa-Domínguez, "Dissimilarity-based linear models for corporate bankruptcy prediction," Computational Economics, vol. 53, pp. 1019–1031, 2019. Doi: 10.1007/s10614-017-9783-4.

[18] G. Jandaghi, A. Saranj, R. Rajaei, A. Ghasemi, and R. Tehrani, "Identification of the most critical factors in bankruptcy prediction and credit classification of companies," Iranian Journal of Management Studies (IJMS), vol. 14, no. 4, pp. 817-834, 2021. Doi: 10.22059/IJMS.2021.285398.673712.

[19] E. I. Altman, E. Hotchkiss, W. Wang, "Corporate financial distress, restructuring, and bankruptcy: analyze leveraged finance, distressed debt, and bankruptcy," John Wiley & Sons, Hoboken, NJ, USA, 2019.

[20] A. Tobías, M. Casals, M. Saez, M. Kamada, and Y. Kim, "Impacts of ambient temperature and seasonal changes on sports injuries in Madrid, Spain: a time-series regression analysis," BMJ Open Sport & Exercise Medicine, vol. 7, 2021. Doi: 10.1136/bmjsem-2021-001205.

[21] C. Lohmann, S. Möllenhof, T. Ohliger, "Nonlinear relationships in bankruptcy prediction and their efect on the proftability of bankruptcy prediction models," Journal of Business Economics, vol. 93, pp. 661–1690, 2023. Doi: 10.1007/s11573-022-01130-8.

[22] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," Electron Markets, vol. 31, pp. 685–695, 2021. Doi: 10.1007/s12525-021-00475-2.

[23] M. Mai, S. Tian, C. Lee, and L. Ma, "Deep learning models for bankruptcy prediction using textual disclosures," European Journal of Operational Research, vol. 274, pp. 743–58, 2018. Doi: 10.1016/j.ejor.2018.10.024.

[24] Y.-C. Hu, P. Jiang, H. Jiang, and J.-F. Tsai, "Bankruptcy prediction using multivariate grey prediction models," Grey Systems: Theory and Application, vol. 11, no. 1, pp. 46-62, 2021. Doi: 10.1108/GS-12-2019-0067.

[25] B. Wei, and N. Xie, "Parameter estimation for grey system models: A nonlinear least squares perspective," Communications in Nonlinear Science and Numerical Simulation, vol. 95, 2021. Doi: 10.1016/j.cnsns.2020.105653.

[26] S. Shetty, M. Musa, and X. Brédart, "Bankruptcy prediction using machine learning techniques," Journal of Risk and Financial Management, vol. 15, no. 1, 2022. Doi: 10.3390/jrfm15010035.

[27] H. Son, C. Hyun, D. Phan, and H. J. Hwang, ''Data analytic approach for bankruptcy prediction,'' Expert Syst. Appl., vol. 138, no. 112816, 2019. Doi: 10.1016/j.eswa.2019.07.033. Doi: 10.1016/j.eswa.2019.07.033.

[28] H. Kim, H. Cho, and D. Ryu, "Corporate bankruptcy prediction using machine learning methodologies with a focus on sequential data," Computational Economics, vol. 59, pp. 1231–1249, 2022. Doi: 10.1007/s10614-021-10126-5.

[29] Y. Jang, I. Jeong, and Y. K. Cho, "Business Failure Prediction of Construction Contractors Using a LSTM RNN with accounting, construction market, and macroeconomic variables," Journal of Management in Engineering, vol. 36, no. 2, 2020. Doi: 10.1061/(ASCE)ME.1943-5479.0000733.

[30] A. Al Ali, A. M. Khedr, M. El Bannany, and S. Kanakkayil, " GALSTM-FDP: A time-series modeling approach using hybrid ga and lstm for financial distress prediction," International Journal of Financial Studies, vol. 11, no. 1, 2023. Doi: 10.3390/ijfs11010038.