E-ISSN: 2676-6167 | P-ISSN: 2538-5100

Journal of Applied Research on Industrial Engineering

www.journal-aprie.com

J. Appl. Res. Ind. Eng. Vol.9, No. 3 (2022) 274–290.

Paper Type: Research Paper

Evaluation of the Hybrid Method of Genetic Algorithm and Adaptive Neural-Fuzzy Network (ANFIS) Model in Predicting the Bankruptcy of Companies Listed on the Tehran Stock Exchange

Mohammad Amin Rahbar *

Pyame Noor Tehran-North University, Hamedan, Iran; rahbaramin@ymail.com.

Citation:

Rahbar, M. A. (2022). Evaluation of the hybrid method of genetic algorithm and adaptive neural-fuzzy network (ANFIS) model in predicting the bankruptcy of companies listed on the Tehran stock exchange. *Journal of applied research on industrial engineering, 9*(3), 274-290.

Received: 24/08/2021 Reviewed: 29/09/2021 Revised: 03/10/2021 Accepted: 23/11/2021

Abstract

One of the most important issues in financial, economic, and accounting matters is the phenomenon of bankruptcy and its prediction. There is presented a hybrid method of Genetic Algorithm (GA) and Adaptive Neural-Fuzzy Network (ANFIS) model to evaluate predicting the bankruptcy of companies listed on the Tehran Stock Exchange. The statistical population of this research is the successful and bankrupt manufacturing companies in Tehran Stock Exchange and in this research, there is a different way as opposed to previous and purposeful research and all companies can prevent their possible bankruptcy with accurate forecasting. In this way, the statistical population includes 136 companies consisting of bankrupt and non-bankrupt companies. In order to construct prediction models, four variables were first selected: 1) independent sample t-test, 2) Correlation Matrix (CM), 3) Step-by-step Diagnostic Analysis (SDA), and 4) Principal Component Analysis (PCA). The final financial ratios were selected from 19 financial ratios that using selected financial ratios and a hybrid model of ANFIS and GA and the results of the proposed model and its comparison with the hybrid model of GA and Group Method of Data Handling (GMDH) shows the high capability of the proposed GA-ANFIS model in bankruptcy prediction modeling and its superiority over Group Method of Data Handling with GA-GMDH method. The results also show that the CM-GA-ANFIS model is known as the best model for predicting bankruptcy of companies listed on the Tehran Stock Exchange. The main reason for choosing the model (GA-ANFIS) is that in addition to the fact that for the first time a combination of two methods ANFIS and GA is used to predict the bankruptcy of companies, and also in none of the studies conducted in both areas which further highlights the need for the present study.

Keywords: Bankruptcy prediction, Tehran stock exchange, Genetic algorithm and adaptive neural-fuzzy networks.

1 | Introduction

In recent years, bankruptcy has grown exponentially and has expanded significantly. Financial managers and those involved in the financial and economic field can, by recognizing the causes of bankruptcy and understanding it correctly before it occurs, inform the owners of the business about it and examine the ways to prevent it. Bankruptcy predictions were first proposed in the scientific field in 1932. Assessing the financial health of companies and estimating their default risk is important for a wide range of people, company management, employees, shareholders, customers, investors, creditors, competitors, lenders and even governments, because the wrong decisions in relationship with the detection of companies' financial health will cause countless problems and losses for these

Corresponding Author: rahbaramin@ymail.com

http://dx.doi.org/10.22105/jarie.2021.254142.1204

people. This justifies the need for more detailed research to increase the ability to detect or predict corporate bankruptcy [1]. There are many benefits to evaluate the health or financial risk of companies, including:

- *By providing the necessary warnings, companies can be alerted to the occurrence of bankruptcy so that according to these warnings, the managers of the companies can take appropriate measures [2].*
- *Providing a tool to identify favorable investment opportunities.*
- *Providing a model to improve the process of credit and credit risk control (non-repayment of credit) of financial institutions.*

Although most of the models used in bankruptcy research have performed well in predicting corporate bankruptcy, in general, re-applying these models to different datasets from the original data (different data in terms of time or different economic conditions or systems) have not been able to replicate previous successes. The reason is that, over time and changing conditions, the variables used in the models lose their efficiency and other variables may become more predictive. In general, there are two research processes for conducting bankruptcy research in the literature: the first process includes research that examines the bankruptcy to try to find the influencing factors and predictor variables and signs of bankruptcy. But the second group includes studies in which a new model for prediction is proposed or the accuracy of prediction of different models is compared with each other [3].

The present study belongs to the above two categories. In addition to comparing the proposed new method with a number of artificial intelligence methods and traditional statistical methods, it applies some of the most widely used variable selection methods to find the most appropriate predictors of bankruptcy. In recent years, special attention has been paid to artificial neural networks as the modern method of modelling. These models are used to predict and classify in cases where classical statistical methods cannot be used due to their limitations. Adaptive Neural-Fuzzy Network (ANFIS) model has shown great ability to solve complex problems. This mathematical model can provide a suitable model for managers, analysts and investors to identify or even fully describe the process used. On the other hand, combining Genetic Algorithm (GA) with this model can improve its performance. Considering the mentioned factors, it seems that the proposed ANFIS model has a good ability in modeling the prediction process. It should be noted that so far, the ANFIS model and its combination with GA has not been used to model bankruptcy prediction. Therefore, studying this method and showing its ability to predict bankruptcy can start a new field of research in future studies. Since in order to decide on the performance of the proposed model, we need to compare it with other models, the GMDH group method algorithm model and its combination with the GA are used to compare with the proposed model. The reason for choosing the model (GA-ANFIS) is that in addition to being used for the first time, a combination of two methods, ANFIS and GAs, is used to predict the bankruptcy of companies. In none of the previous research on bankruptcy, there has been no effort to provide an appropriate prediction mode. The process of selecting variables affecting the bankruptcy simultaneously, which further highlights the need for the present study. The main purpose of this article is to present a suitable model for predicting the bankruptcy of Tehran Stock Exchange companies. The statistical population of the study is the companies listed on the Tehran Stock Exchange. According to the latest report available on the Tehran Stock Exchange website (www.irbourse.com) and the stock exchange information (www.irportfolio.com), 495 companies have been classified in 37 industries, of which 435 companies are manufacturing companies and 60 companies are non-manufacturing. The definition of bankruptcy in this study is based on article 141 of the Commercial Code, according to which a bankrupt company has an accumulated loss equal to 50% of the company's capital. However, article 141 of the Trade shall not lead to the expedited liquidation or liquidation of the Company and shall be terminated only at the request of the interested groups [4]. According to the information of the companies, during this period, 78 companies were subject to article 141 of Commerce and as a result were declared bankrupt in that year. Due to the lack of all the required information, 68 companies from bankrupt companies were considered as a research sample. In order to obtain a paired sample, 68 non-bankrupt companies were selected using random sampling during the study period. Non-bankrupt companies were matched with bankrupt companies in terms of data collection year.

Rahbar| J. Appl. Res. Ind. Eng. 9(3) (2022) 274-290

It should be noted that due to the limited number of companies in any particular industry, it is not possible to select the entire sample from one industry. But as much as possible, active and bankrupt companies were tried to be industry-compliant in order to carry out the prediction process. Thus, the sample under study includes 136 companies (68 active companies and 68 bankrupt companies). Prediction of financial distress by providing the necessary warnings can alert companies to the occurrence of financial distress and by providing the necessary warnings can alert companies to the occurrence of financial distress; bankruptcy should be vigilant so that they can take appropriate action in response to these warnings. Ability to predict corporate bankruptcy as well from the point of view of investors as well as from the social point of view, as it is a clear sign of the lack of proper allocation of resources, it is of great importance enjoys. By becoming aware of the possibility of bankruptcy of companies, investors and business units can take action Take precautions. Financial therapy and corporate bankruptcy it led to wasting resources and not taking advantage of investment opportunities some companies are successful and some are unsuccessful. Unsuccessful companies cause capital owners to worry and investors are looking for the right decision tools. Due to reasons mentioned above perediction of bankruptcy is of paramount importance. The main objective of this paper is to provide an appropriate model for predicting bankruptcy of Tehran Stock Exchange companies. The results of this research can provide a suitable model for assessing financial health and the possibility of continuity of companies. So, the main research question is this js it possible to predict the bankruptcy of companies listed on the Tehran Stock Exchange using the combined method of GA and ANFIS model? Which companies with these models are more likely to go bankrupt? Top model in terms of power what is predictability?

It is important to note that for the first time through two methods of emphysema and GA is used to predict the entry of companies, which has not been investigated in any of the researches in both fields:

- *Provide an appropriate forecasting model.*
- *The process of selecting variables that affect input bankruptcy it has not been done as usual, which is what necessity has done, which has now become more apparent.*

This research consists of 6 steps to achieve the desired result. First, the sample population and the time period of the research are determined, and then the highest financial ratios are determined. In the third step, the variable selection process is performed using four methods of independent t-test, Correlation Matrix (CM), step-by-step detection analysis and principal factor analysis. Following this, bankruptcy prediction is modeled by combining in proposed new method of GA and ANFIS. Afterwards, the results are analyzed and a detailed comparison between the installed models (GA-GMDH) is performed using the concepts of predictive accuracy, first and second type prediction errors and the area under the received performance characteristic curve (ROC) curve. In the last step the model for detecting the activity or financial crisis of companies listed on the Tehran Stock Exchange, the most powerful variable selection model and also the best set of financial ratios predicting bankruptcy are determined.

2 | Literature Review

There are several factors that can affect the occurrence of bankruptcy. Rising interest rates and high debts are among the factors that increase the likelihood of bankruptcy. Beaver was the first researcher to work in bankruptcy prediction [5]. He chose six out of 30 financial ratios, which he considered to be the best measure of any company's health, and many researchers after him were able to achieve significant success in this field by presenting their statistical models. Today, several prediction models have been introduced by researchers. These models are classified into different groups based on the method of model making, financial ratios selected for model making, type of companies under review (manufacturing companies, banks, etc.) to identify a bankrupt company, etc. [6].

Using these models, creditors, investors and managers will be able to predict the bankruptcy of companies a few years before it occurs and, according to the results, take the necessary measures to prevent losses. Unlike some other areas of academic research, the results of bankruptcy prediction research and its resulting models were welcomed in the capital markets. Altman's [7] research and the model he proposed are good evidence for this claim.

Confirmatory Factor Analysis (CFA) technique has been used to evaluate the significance of regression weight (factor loading) of different constructs of the questionnaire in predicting the relevant items [6]. Based on model can determine projection points of inputs, outputs, and flexible measures in the presence of integer data in two-stage network DEA [8]. Kaviani and Fakhrehosseini [9] provide three sections: introduction to fuzzy logic, portfolio management, and the function of fuzzy model in portfolio management with research approach. Other types of methods were to measure the impact of stock liquidity on future investment in companies listed on the Tehran Stock Exchange and showed that stock liquidity has a negative and significant effect on the company's future investment [10]. The proposes in the paper that integer-valued FNDEA approach is based on envelopment form of CCR that evaluates the relative efficiency of DMUs and determines the status of flexible measures in the presence of integer data in basic and general two-stage network structures [8]. A great amount of cash goes to January and due to receiving bonuses, salary and pension, households have greater cash compared with other months [11]. According to the research conducted by Sarmadi [12], there is a significant relationship between intellectual capital and two performance indices (return on shareholders' equity and return on sale) in petrochemical companies listed on Tehran Stock Exchange.

Table 1. Primary models of Altman [7].

Model				**Model Name**	
$Z_1 - \text{Score} = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 0.999X_5$ (1)				1968	
$Z_2 - \text{Score} = 0.717X_1 + 0.847X_2 + 3.1X_3 + 0.42X_4 + 0.998X_5$ (2)				1983	
$Z_3 - \text{Score} = 6.5X_1 + 3.26X_2 + 6.72X_3 + 1.05X_4$ (3)				1993	
X_1 =Capital in circulation to total assets					
X_2 = Profit (loss) accumulated on all assets					Altman
X_3 = Profit before interest and taxes (operating profit and loss) on all assets				Variables	
X_4 = Equity market value of total debts					
X_5 = Net sales (income) to total assets					
$SP - Score = 1.3X_1 + 3.07X_2 + 0.66X_3 + 0.04X_4$ (4)				1978	
X_1 = Capital in circulation to total assets					
X_2 = Profit before interest and taxes (operating profit and loss) on all assets				Variables	Springate
X_3 = Profit before tax on current debts					
X_4 = Net sales (income) to total assets					
$\text{ZM} - \text{Score} = -4.531(\text{ROA}) + 5.679(\text{FINL}) + 0.004(\text{LIQ}) - 4.236$ (5)				1984	
ROA= Net profit to total assets (return on assets)					
FINL= Total debts to total assets (financial leverage)				Variables	Ziemsky
LIQ= Current assets to current liabilities (liquidity)					
Ziemsky	Springate	Altman			Model
Range ZM	Range SP	Range Z3	Range Z2	Range Z1	Range
ZM ≤0/5	SP ≤ 0/862	Z3 ≤ 1/1	Z2 ≤ 1/23	Z1 ≤ 1/8	Bankruptcy
-	-	1/1 < Z3≤ 2/6	1/23 < Z2 ≤ 2/99	1/8 < Z1 ≤ 2/99	Distress
ZM >0/5	SP>0/862	Z3> 2/6	Z2 > 2/99	Z1 > 2/99	Financial safety

3 | Research Method

3.1 | Theoretical Models of Bankruptcy

Considering that artificial intelligence methods have been used in this research, *Table 1* refers to the most important bankruptcy researches based on these methods.

3.2 | Bankruptcy Prediction Models Using GA-ANFIS Model

ANFIS is the integration of adaptive neural networks and fuzzy system representing the advantages of both methods including setting fuzzy rules to predict bankruptcy, providing a tangible knowledge base for users, reducing optimal search space (due to fuzzy system), executing the appropriate training process and adjusting the required parameters of the structure automatically (due to the operation of neural networks). In this hybrid structure, the GA presents the optimal number of neurons in each hidden layer and how they are connected, resulting in the optimal set of appropriate coefficients of quadratic (polynomial) functions of the predictor variables of the dependent variable. The final output of this model is obtained as a polynomial of financial ratios that has the best ability to differentiate between bankrupt and active companies. The final output of the hybrid model is a polynomial that results in the least training and testing errors simultaneously.

In order to achieve a suitable structure of the ANFIS model to perform the prediction process, various structures of the ANFIS model were implemented and the results were compared with each other. As a result of analyzing the obtained results, the ANFIS structure with the parameters shown in *Table 2* was considered to perform the prediction process using the financial ratios selected by the variable selection methods.

Table 2. Parameters for the proposed ANFIS.

Clustering Method	Cluster Radius	Admission Rate	Rejecting Rate	Membership Functions of Prector Variables	Output Membership Function	Training Algorithm	No of Repitions	Error Tolerance
Subtractive clustering	0.01	0.5	0.15	Gausean	linear	hybrid	5	0

In order to improve the performance of ANFIS, to build a fuzzy inference system, two methods of fuzzy c-mean clustering (grid partition) and subtractive clustering are generally used fuzzy c-mean clustering, first introduced by Bezdek, can be considered as a generalized state of hard partitioning (Hard Partitioning) [13]. The fuzzy c-mean clustering method is capable of working on ambiguous states that occur in the data, that is, a state in which clusters are created through the existence of subjects who belong to an intermediate state in terms of different clusters.

Subtractive clustering, introduced by Chiu [14], is a rapid algorithm for estimating the number of clusters, their radii and their center. Examining the results of different ANFIS structures, it was concluded that ANFIS networks using subtractive clustering are more accurate than fuzzy c-mean clustering and have better predictive performance. Error tolerance is a criterion for evaluating the error of training data and stopping the training algorithm. Usually, the value of this criterion is considered zero.

Evolutionary methods such as GAs are widely used in various stages of neural network design due to their unique capabilities in finding optimal values and the ability to search in unpredictable spaces. In this hybrid structure, the GA presents the optimal number of neurons in each latent layer and how they are connected, resulting in the optimal set of appropriate coefficients of quadratic (polynomial) predictors of dependent variable functions. In the hybrid structure, there are generally three main objective functions: training error, prediction error and number of neurons. Therefore, the optimal structure is obtained based on trade-offs between these objective functions. The Mean Square Error (MSE) is considered as the measure of predictive error [15]. In the GA-ANFIS structure, the representation of genes and chromosomes is considered as a symbolic (non-binary) string in which each of the variables is represented alphabetically (a, b ...). To generalize ANFISs, the constraint of using the adjacent layer to build the next layer must be removed, and all previous layers (including the input layer) are used to build the new layer. In fact, all layers (layer two and three) and not necessarily layer three are used to make the output neuron chromosome. After defining chromosomes and genes, crossover and mutation operators are used to produce two children from two parents. The roulette wheel method is used to select two chromosome parents.

3.3 | The Summarized Steps of the Research

In the first step, the sample population and time period of the research are determined. Then, in the second step, the most widely used initial financial ratios are determined by studying previous research. Initial financial ratios are adjusted according to the economic conditions of Iran, the results of national research and the availability of the required information, and the appropriate initial financial ratios are determined for modeling the bankruptcy of Tehran Stock Exchange companies. In the third step, using independent t-test methods, CM, step-by-step detection analysis and principal factor analysis, the variable selection process is performed. Then, in the fourth step, the bankruptcy prediction modeling is performed by the proposed new method of GA and ANFIS.

As a result of this process, 8 prediction models will be obtained. In step five, the results are analyzed and the exact comparison between the fitted models is performed using the concepts of predictive accuracy, first and second type errors of prediction and the area under the Received Operating Characteristic (ROC) curve. In the last step, the most appropriate model for detecting activity or the financial crisis of companies listed on the Tehran Stock Exchange, the most capable variable selection model and the best set of financial ratios predicting bankruptcy are determined. To perform the above process, various software such as SPSS 16, Eviews 10, toolboxes and coding in MATLAB environment have been used.

According to article 141 of trade, a total of 78 bankrupt companies were identified during the years under review, the number of which is shown in *Table 2* by year of bankruptcy. A number of these companies went public after going bankrupt, and some continued to operate on the stock exchange, and their symbol is still on the stock exchange board. In this case, in the first year that a company is subject to article 141 of the trade, regardless of its continuation or non-continuation in the stock exchange, that company is considered as a bankrupt company.

Table 3. The number of bankrupt companies in terms of year in the research period.

95	94	93	92	91	90	89	88	87	86	85	84	Year
0	3	3	5	13	6	8	6	12	12	4	6	Number

As the bankrupt companies identified, only the information and financial statements of 68 companies were available in the required year and were adjusted on a one-year basis (*Table 3*). Because a paired sample was used in the present study, 68 non-bankrupt companies were randomly selected to match the bankrupt companies. Non-bankrupt companies were matched with bankrupt companies in terms of data collection year. Thus, the sample under study includes 136 companies (68 active companies and 68 bankrupt companies).

Table 4 shows the number of bankrupt companies related to each industry along with the share of that industry in the total number of bankrupt companies. It can be seen that the textile, machinery and equipment industries have the highest number of bankrupt companies, respectively.

Table 4. Bankrupt companies in terms of industries.

Industry	No of Bankrupt Companies	Percentage of Bankrupt Companies(%)
Textiles	16	23.5
Machinery and equipment	13	19.2
Plaster and cement	5	7.4
Electrical machinery	6	8.8
Automobile and manufacturing parts	3	4.4
Food products and driks except sugar	9	13.2
Griculture and animal breeding	9	13.2
Metal mineral extraction	7	10.3

Since in any sample-based inference technique, calculating the percentage of classification accuracy in the same sample on which the model is estimated, the power of the model is greater than it actually is, such an error must be removed by the holdout sample or test sample for this purpose, the sample consisting of 136 companies is divided into two parts: training sample and test sample (holdout sample). There are no defined instructions for dividing the original sample into two sub-samples of training and testing. However, some researchers consider the ratio of experimental sample to test sample to be 60 to 40 or 70 to 30. In the present study, the sample is randomly divided into 96 companies (48 active companies and 48 bankrupt companies) for training and 40 companies (20 active companies and 20 bankrupt companies) for testing the models.

After selecting the sample used in the research to deduce the financial ratios, we need the financial statements of the companies, such as profit and loss statements and balance sheet. This information is possible in the library of the Exchange Organization through Rahavard and Tadbir Pardaz software.

4 | Selecting a Variable in Bankruptcy Prediction

Two different approaches to the problem of variable selection have been developed: a) accurate or exact techniques (enumerative techniques) that can guarantee the optimal solution but are only applicable to small sets of variables, and b) heuristic techniques that can be used to ensure optimal solutions in a reasonable interval. Of course, it should be noted that these techniques do not guarantee an optimal solution. Variable selection methods can be divided into the following two categories in terms of the type of combination with prediction and classification models.

Separation models (filtering). These models select a smaller set from a series of variables, based on statistical properties such as mean, correlation, etc. These models, which are often based on statistical concepts, select variables before performing the prediction process and based on their theoretical theory. Some of the most widely used methods in predicting bankruptcy are step-by-step discriminant analysis, independent t-test, factor analysis, principal factor analysis, CM, and logistic regression.

Wrapper models. This group of models is called wrapper, because the variable selection process and learning algorithm (prediction) are performed simultaneously. These models often include meta-heuristic methods for selecting an effective set of variables. In this method, the prediction model is taught with different subsets of primary variables. Finally, a subset of variables that maximize prediction accuracy or minimize error are selected as the final variables.

4.1 | Calculation of Principal Components Using Covariance Matrix

Based on the proposed definition of Principal Component Analysis (PCA), the purpose of this analysis is to transfer the X data set in M-dimension to the L-dimensioned Y data set. It is assumed that the matrix X consists of N columnar vectors (X1… XN). Therefore, the dimensions of the X matrix are M × N (ibid). The first step is to normalize the data and calculate their experimental mean.

The experimental mean vector on the matrix rows is obtained as follows:

$$u[m] = \frac{1}{N}\sum_{i-1}^{N} X[m, i]. \tag{1}$$

Then the distance matrix is obtained from the mean as follows:

$$Uh - X = B. \tag{2}$$

So that h is a vector of size N × 1 with a value of 1 in each of the elements.

The covariance matrix C with dimensions M × M is obtained as follows:

$$C = E[B\ddot{A}B] = E[B \bullet B^{*}]B = \frac{1}{N} B \bullet B^{*}. \tag{3}$$

Where E is the arithmetic mean, xxx external multiplication, [B] ^ * conjugate transposition of matrix B. The eigenvalues and eigenvectors of the covariance matrix (C) are obtained as follows:

$$D = CV - V. \tag{4}$$

V is a matrix of eigenvectors and C is a diagonal matrix whose diameter elements are eigenvalues.

Each eigenvalue corresponds to an eigenvector. This means that the matrix V is a matrix M × M whose columns are eigenvectors and the eigenvector VP is located in the column PTH and the eigenvalue PTH, Cp = λp, p corresponds to it. The rearrangement of eigenvectors is based on the size of the corresponding eigenvalues. That is, based on the order in which the eigenvalues decrease, the eigenvectors are rearranged, meaning:

$$q£pÞ\lambda q£\lambda p. \tag{5}$$

As mentioned, the goal of this method is to reduce the variables to the minimum possible number, so that as much information as possible is present in the generated variables. So, the question is how many components we can replace with the original variables of the problem without losing a lot of information?

5 | Step-by-step Discriminant Analysis

Multiple Discrimination Analysis (MDA) is a multivariate method that classifies entities into distinct groups based on their characteristics. Multiple discriminant analysis works by first determining the first differentiation function in such a way that it separates the groups as much as possible. GA is one of the repetitive algorithms that has been applied in various studies [16] by the researchers. The second differentiation function is determined in such a way that it does not correlate with the first differentiation function and creates more separation from the groups. The algorithm continues in the same way until the number of differentiation functions reaches the maximum possible number, i.e., m-1 (m is the number of groups). In many cases, the dependent variable consists of two groups or classifications such as female and male, high and low, low risk and high risk [17].

The assumptions of the MDA are as follows:

- *There is not much correlation between quantitative variables.*
- *The mean and variance of each quantitative variable are not related.*
- *The correlation between both variables in each group is equal.*
- *The value of each quantitative variable follows the normal distribution (ibid).*

The significance test of the differentiation function is determined by measuring the distance between the gravity centers of the groups, and this is done by comparing the function of distributing the discrimination values for two or more groups. If the overlap of the distribution values is low, the differentiation function distinguishes the groups more precisely, and if the distribution overlap is higher, it indicates that the differentiation function has performed poorly in separating the groups. *Figs. 1* and *2* show the two cases mentioned. The common area between the two distribution functions indicates the incorrect classification of statistical units, which is referred to as the gray area, and means that the future status of companies located in this area is unclear.

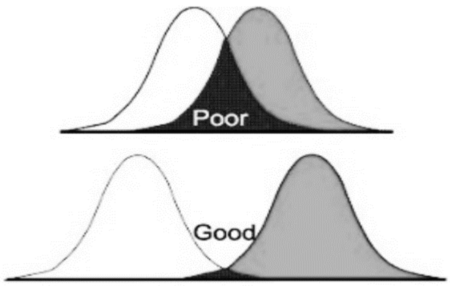

Fig. 1. Differentiation function of good and poor.

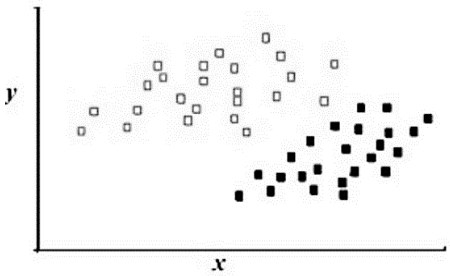

Fig. 2. Bi-nodal distribution.

One of the most widely used discrimination analysis capabilities is the selection of a subset of predictor variables from a wide range of variables so that these variables can efficiently and effectively improve the prediction or classification process.

5.1 | Perform Step-by-step Discrimination Analysis

Variable selection by independent sample t-test. In the case under study, t-test is an independent sample to answer the question of whether the average of a financial ratio in the group of bankrupt companies is statistically different from that of the same financial ratio in the group of non-bankrupt companies. The statistical hypothesis of the t-test in order to select the financial ratios affecting bankruptcy is defined as follows:

- *H0: The average financial ratio X is equal between the active and bankrupt groups.*
- *H1: The average financial ratio X is not equal between the active and bankrupt groups.*

A 95% confidence level is considered for the analysis of test results. The larger the statistic obtained from the test, the greater the probability to reject the null hypothesis. After performing t-test of independent samples on each of the 19 initial financial ratios, 11 variables were selected as predictors of bankruptcy with the null hypothesis assumed to be rejected. The final set of predictor variables with the largest value of T-statistics are: profit before interest and tax on total assets (X_1), profit and loss accumulated on total assets (X_{15}), total debt on total assets (X_{17}), fast assets to current liabilities (X_6), cash to total debt (X_7), sale to current assets (X_{12}), and total debt to equity (X_{16}). Variable selection by CM method. In this method, if the correlation of a financial ratio with the dependent variable

Rahbar| J. Appl. Res. Ind. Eng. 9(3) (2022) 274-290

(bankruptcy) is significantly non- zero, that financial ratio is selected as one of the variables predicting bankruptcy. For this purpose, the following hypothesis test is used:

- *H0: The correlation coefficient between financial ratio X and the variable indicating the state of activity or bankruptcy of companies (variable y) is equal to zero.*
- *H1: The correlation coefficient between financial ratio X and the variable indicating the soundness or bankruptcy of companies (variable y) is significantly non- zero.*

The test of the above hypothesis is carried out for each of the initial financial ratios at a 95% confidence level. If the null hypothesis for a given financial ratio is rejected, it means that the financial ratio has a significant relationship (positive or negative) with the occurrence of bankruptcy, as a result, the financial ratio is selected as the financial ratio affecting bankruptcy. If the null hypothesis for a given financial ratio is accepted, that financial ratio has no significant effect on the bankruptcy. CM analysis was performed by SPSS software on 19 initial variables of the study and the results showed that 11 financial ratios have a significant correlation with the bankruptcy variable of companies. Like the t-test method, of the 11 variables selected, the 7 variables with the highest correlation values are selected as the final set, which are: profit before interest and tax on total assets (X_1), accumulated profit and loss on Total Asset (X_{15}), Total Debt to Total Asset (X_{17}), fast Assets to Current Debt (X_6), Total Equity Debt (X_{16}), Cash to Total Debt (X_7), and interest Cost to gross profit (X_8).

6 | Variable Selection by PCA Method

As mentioned, the purpose of PCA is to reduce data and remove redundant and correlated variables from the modeling process and replace them with new uncorrelated variables. Usually a certain number (but not all components) include the maximum variability in the sample, which can replace the initial variables and enter the model. In this study, by performing PCA by SPSS 16 software on 19 primary variables, component research is selected to perform bankruptcy prediction. To answer the question whether it is possible to perform PCA method on research data and whether it will have the desired results, Kolmogorov-Smirnov test and Bartlett`s test of sphericity is used. The results obtained from these two tests (KMO = 0.726, Bartlett's test = 2.726, DF = 190, Sig. = 0.00) indicate the desirable capability of the PCA method on the research data. It should be noted that among the above four methods, the ratios selected by the two methods of t-test of independent samples and CM have the most overlap with each other. Among the 7 variables selected by these two methods, 6 have common financial ratios. *Table 5* lists the financial ratios selected by the four methods of variable selection.

Table 5. Set of initial financial ratios.

Financial Ratio	Type of Financial Ratios	Code	Financial Ratio	Type of Financial Ratios	Code
Sales to capital in circulation		X_{11}	Profit before interest and tax to total assets	Profitability	X_1
Sales to current asset		X_{12}	Current asset to total asset		X_2
Sales to equity	Performance or activity	X_{13}	Fast asset to total asset		X_3
Sales to fast asset		X_{14}	Capital in circulation to total asset		X_4
Accumulated profit and loss to total assets		X_{15}	Current asset to current dept	Liquidity	X_5
Total debt to equity		X_{16}	Fast assets to current liabilities		X_6
Total debt to total assets	Debt solvency ratios (leverage or capital structure)	X_{17}	Cash to total debt		X_7
Current debt to equity		X_{18}	Cost of interest on gross profit	Cover	X_8
Long-term debt to equity		X_{19}	Sales to total asset Sales to liquidity	Performance or activity	X_9 X_{10}

7 | Predictive Models

7.1 | Group Method of Data Handling (GMDH)

One type of artificial neural networks having been proven to be effective in modeling and prediction is self-organizing networks. One type of self-organizing modelling is polynomial networks. Polynomial networks are the result of a combination of linear regression techniques and artificial neural networks. Polynomial network training algorithm can be divided into several categories. One of these training algorithms is GMDH.

The GMDH expresses the main idea, which is the procedure of random division of model input data into two sets of training and adaptation. In this algorithm, models constructed from data with similar statistical properties must behave similarly. Otherwise, the effective parameter in this behavioral difference must be corrected. The GMDH algorithm was first used by McKee and Lensberg [18] in 1996 to model complex systems involving a series of data with multiple inputs and one output. In fact, the main purpose of this method is to build a function in a network based on the quadratic transfer function. The main advantage of GMDH over conventional neural networks is the acquisition of a mathematical model in terms of polynomials for the process under study. This mathematical model can be used to identify or even describe a complete process.

In general, to model complex systems involving sets of data with multiple inputs and single outputs, VKG polynomials can be used:

$$y = a_o + \sum_{i-1}^{n} a_i x_i + \sum_{i-1}^{n} \sum_{j-1}^{n} a_{ij} x_i x_j + \sum_{i-1}^{n} \sum_{j-1}^{n} \sum_{k-1}^{n} a_{ij} x_i x_j x_k + \ldots, \tag{6}$$

where x= (x1, x2… xn) are the input vectors, y is the output of model and ai are polynomial coefficients. VKG polynomials are approximated using quadratic polynomials. These quadratic polynomials are based on the binary combinations of network inputs. The GMDH algorithm has been introduced using this idea as a learning method for modeling complex systems.

The GMDH neural network has the structure of a multilayered, forward-facing network consisting of a set of neurons that arise from the junction of different input pairs through a quadratic polynomial. Each layer in this network consists of one or more processor units, each of which has two inputs and one output. These units practically play the role of the components of the model and are assumed as a quadratic polynomial (*Eq. (7)*).

$$\widehat{y}_n = a_o + a_1 x_1 + a_2 x_2 + a_3 x_1 x_2 + a_4 x_1^2 + a_5 x_2^2. \tag{7}$$

The unknown parameters of the GMDH algorithm are the polynomial coefficients of *Eq. (7)*. To calculate the output value y for each input vector (xi1, xi2… xin) based on *Eq. (7)*, the mean squares error (*Eq. (8)*) must be minimized.

$$e = \sum_{i+1}^{n} (\widehat{y}_i - y_i)^2. \tag{8}$$

To find the minimum error value, the partial derivative of *Eq. (8)* is used. By placing *Eq. (7)* in this partial derivative, a matrix equation (AA=y) is obtained. In this equation, the matrix A corresponds to *Eq. (9)*.

$$A = \begin{pmatrix} 1 & x_{1p} & x_{1p} & x_{1p}^2 & x_{1p}^2 & x_{1p} & x_{1p} \\ 1 & x_{2p} & x_{2q} & x_{2p}^2 & x_{2q}^2 & x_{2p} & x_{2q} \\ 1 & x_{np} & x_{nq} & x_{np}^2 & x_{nq}^2 & x_{np} & x_{nq} \end{pmatrix}. \tag{9}$$

One solution for this matrix equation (AA=y) is to use the Singular Value Decomposition (SVD) method. If the unknown SVD method is used, it is calculated from *Eq. (10)*.

Rahbar | J. Appl. Res. Ind. Eng. 9(3) (2022) 274-290

$$a = (A^T A)^{-1} A^T y. \tag{10}$$

In *Eq. (5)*, AT is the transposition of matrix A. Using this method, the unknown a can be calculated in any case. If the matrix (AA) is not inverted, the Thikhonov method will be used to solve the equation.

When building a GMDH network, all the generated input combinations are sent to the first layer of the network, then the outputs from this layer are sorted and selected to enter the next layer, which is sent to layer 2 with all the selected output combinations. This trend continues until layer (n + 1) produces a better result than layer n. When the n+1 layer is found to be worse than the layer n, the process stops [19].

7.2 | Comparison of Performance of Models Made from the Point of View of Predicting Models

The results of 12 prediction models made on the training sample (68 bankrupt companies, 68 non-bankrupt companies) and the test sample (20 bankrupt companies, 20 non-bankrupt companies) are shown in *Tables 6* and *7*, respectively.

Table 6. Performance of model prediction on the training samples (in percent).

Model Ranking	Second Type Error	First Type Error	Classification Accuracy	Prediction Model	Variable Selection Model
2	0	14.58	92.71	GA-ANFIS	t-test
1	0	0	100	GA-GMDH	
2	2.08	12.5	92.71	GA-ANFIS	CM
1	0	0	100	GA-GMDH	
2	2.08	18.75	91.67	GA-ANFIS	SDA
1	0	0	100	GA-GMDH	
2	10.41	27.08	81.25	GA-ANFIS	PCA
1	0	0	100	GA-GMDH	

To perform the ranking, we first compare the prediction models in terms of the accuracy of the correct classification. If the classification accuracy of the models is equal, we consider the model as the superior model that has less error of the first type. According to the above process, prediction models are ranked for each of the variable selection methods [19]. As can be seen in *Table 6*, most of the built models have good performance in classifying bankrupt companies in Tehran Stock Exchange companies in the training sample.

The results reported in *Table 6* indicate the superiority of the GA-GMDH model in 4 cases of using variable selection methods in the training sample. However, considering the results of the GA-ANFIS hybrid model in the experimental sample of *Table 7*, it can be said that this model outperforms the GA-GMDH model. To justify this, it can be said that the GA-GMDH hybrid model suffers from the phenomenon of over-fitting. This phenomenon means that a predictive model works very well in the educational set but has poor performance in classifying the educational sample. This model has correctly classified all bankrupt and non-bankrupt companies in the training sample, therefore the accuracy of classification of this model in all cases is equal to 100%. Therefore, it can be concluded that the GA-ANFIS hybrid model is in the first place in predicting the bankruptcy of the research sample.

Table 7. Performance of model prediction on the testing samples (in percent).

Model Ranking	AUC	Second Type Error	First Type Error	Classification Accuracy	Model of Variable Selection	Prediction Model
1	0.884	5	20	87.5	GA-ANFIS	t-test
2	0.836	5	30	82.5	GA-GMDH	
1	0.908	5	10	90	GA-ANFIS	CM
2	0.911	5	15	90	GA-GMDH	
1	0.863	5	20	87.5	GA-ANFIS	SDA
2	0.848	15	20	82.5	GA-GMDH	
1	0.842	15	20	82.5	GA-ANFIS	PCA
2	0.756	25	25	75	GA-GMDH	

7.3 | Comparison of the Performance of the Constructed Models from the Point of View of Variable Selection Methods

Each method of variable selection has its own logic and concepts. In *Figs. 3* and *4*, the classification accuracy and error of the first type of the models constructed for each variable selection method for the test sample are plotted.

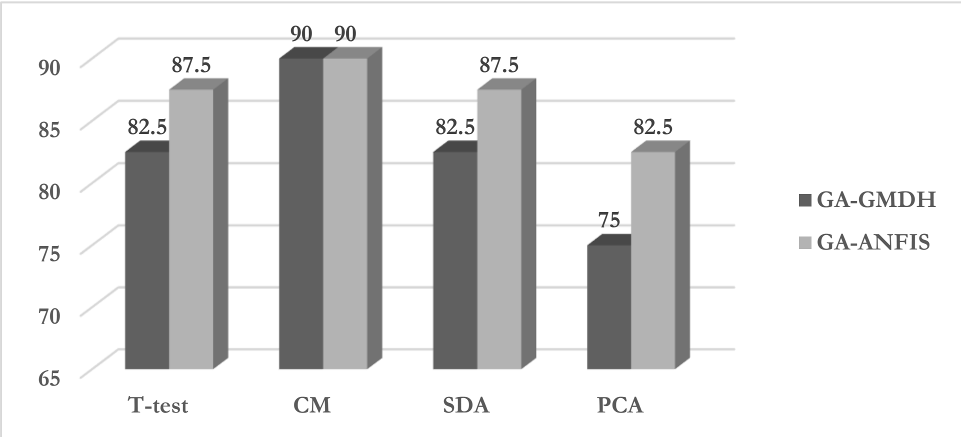

Fig. 3. Model classification accuracy by methods of selecting variable.

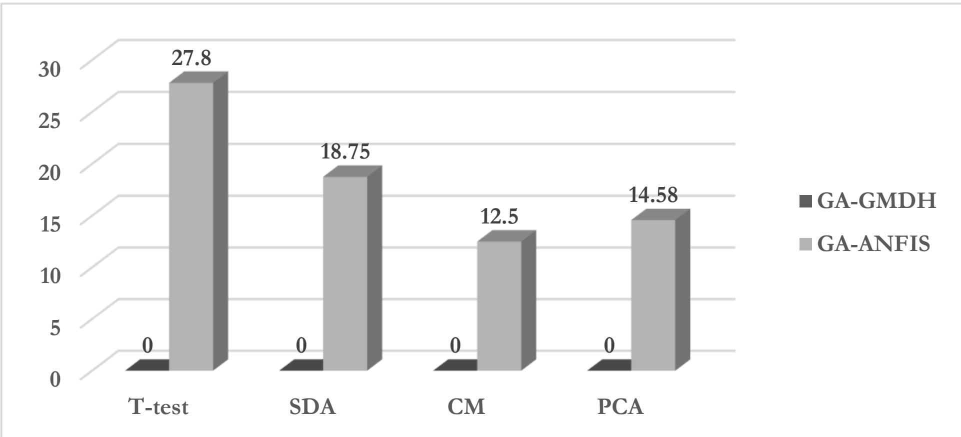

Fig. 4. First type error of models in terms of methods of variable selection.

Rahbar | J. Appl. Res. Ind. Eng. 9(3) (2022) 274-290

According to the above diagrams, it can be said that the CM method leads to the highest classification accuracy in the two models GA-GMDH and GA-ANFIS, compared to other selected methods. The methods of selecting the t-test and SDA variables for the GA-ANFIS prediction model have the same results. However, since the classification accuracy of SDA-GA-ANFIS model is equal to t-test model and the error of the first type is less, it can be said that the method of selecting the SDA variable will be in the second place. Comparison of t-test and PCA diagrams clearly shows the superiority of the method of selecting the t-test variable to PCA. In justifying the results, it can be said that since the logic of the PCA method is based only on the correlation between predictor variables (financial ratios) regardless of the dependent variable (bankruptcy), this method can not be a suitable method for selecting a variable in the prediction problem of bankruptcy. Therefore, t-test and PCA are in the third and fourth ranks, respectively. In *Fig. 8*, the AUC values (the area under the graph) are plotted for prediction methods. The results of this diagram are a confirmation of the above-mentioned contents and conclusions.

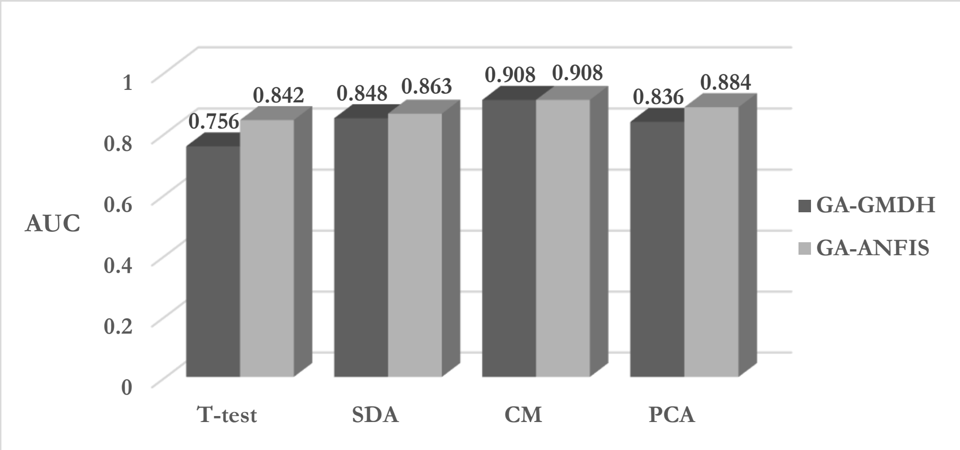

Fig. 5. The area of model Roc curve in terms of methods of variable selection.

8 | Choosing the Best Bankruptcy Prediction Model

Bankruptcy prediction is a matter of classification so that the input of this problem can be modeled as a vector of financial and non-financial variables. With this vector, companies can be classified into two categories: bankrupt and non-bankrupt. In this section, the best bankruptcy prediction model made by the research is selected. This model has the highest power to differentiate companies from each other, in other words, it has the highest classification accuracy. On the other hand, it has the lowest error of the first type and the highest level under the ROC curve compared to other models. *Fig. 6* shows the prediction performance of the constructed models, including classification accuracy and type I error.

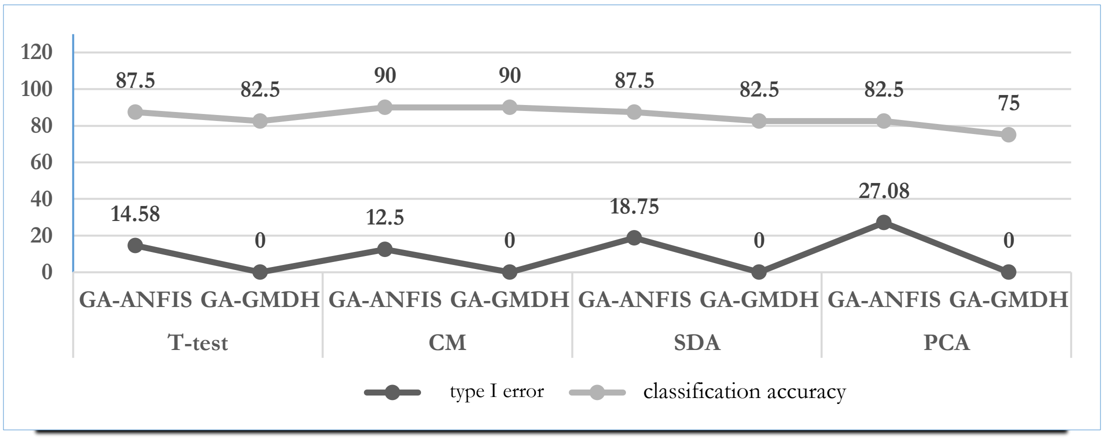

Fig. 6. Classification accuracy and first type error of the model.

Using SPSS 20 software, the ROC curve was plotted for the constructed models. This curve is shown in *Fig. 7*.

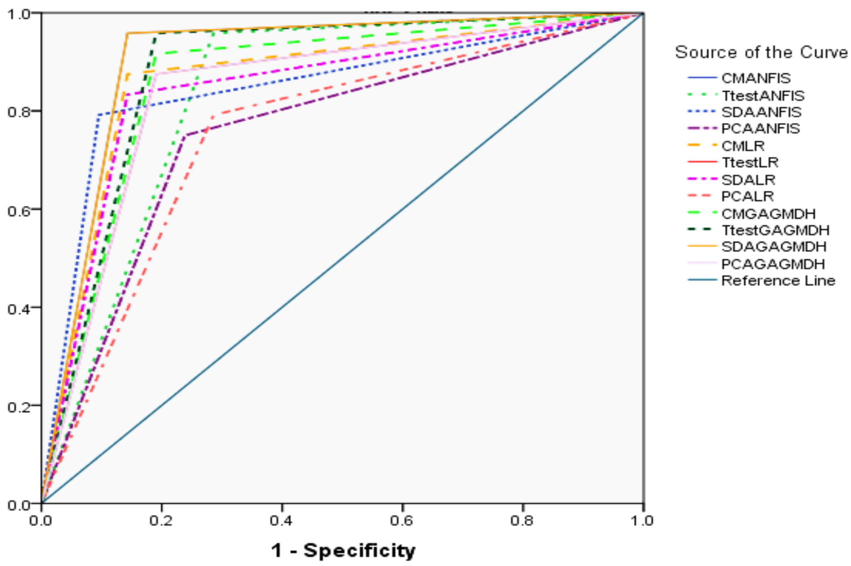

Fig. 7. Roc curve of prediction model.

According to *Figs. 6* and *7*, the two models CM-GA-ANFIS and CM-GA-GMDH with similar performance are selected as the best models for bankruptcy prediction. The CM-GA-ANFIS model correctly classifies 92.7% and 90% of companies in the training and testing sample, and also has the first type error of 10% in the test sample. The CM-GA-GMDH prediction model also has 100 and 90% prediction accuracy in training and testing samples, respectively. From 20 active companies in the test sample, this model classified 3 wrongly in the group of bankrupt companies and therefore has a first type error of 15%. In the ROC curve, the closer the graph of a model is to the left and top of the axes, the higher the resolution of that model.

On the other hand, the results of GA-GMDH model in the experimental and training sample indicate the occurrence of overfitting, which is not desirable in the modeling and predicting process. According to the above explanations, it can be said that despite the similar performance of the two models, the GA-ANFIS model has a relative advantage over the GA-GMDH model. Therefore, the CM-GA-ANFIS model is considered as the best model for predicting research bankruptcy.

9 | Conclusion

In most studies, the identification of bankruptcy predictor variables has not been considered as it is addressed in the present study. The above reasons motivated us to identify the influential variables in the field of bankruptcy prediction and model a new method (GA-ANFIS) that has not been used in the issue of bankruptcy prediction, for predicting bankruptcy in Tehran stock exchange market.

In this study, first using three methods of variable selection, independent sample t-test, CM, stepwise diagnostic analysis and a principal component data extractin method, the variable selection process was performed before prediction. To perform the prediction process, the ANFIS modeling method was used in combination with GA and another research goal is to show the capability of this model in solving bankruptcy prediction problems. In order to evaluate the performance of the proposed research model, the GA-GMDH and its comparison with the proposed research method for modeling bankruptcy prediction was used.

The proposed model combines the GA-ANFIS has a high capability in modeling bankruptcy prediction. Comparison of the results of this model with the group model of data manipulation with GA (GA-

GMDH) proves this claim. The CM method has more ability in selecting variables that affect the bankruptcy prediction of companies compared to other methods of variable selection. This method is a set of financial ratios x1. Profit before interest and tax to total assets, x_2. Accumulated profit and loss to total assets, x3. Total debt to total assets x4. Fast assets to current liabilities, x5. Total debt to equity, x6. Cash to total debt, x7. Introduces the cost of interest on gross profit as the variables that describe the bankruptcy of companies.

The best bankruptcy prediction model developed in the present study is the combination of GA-ANFIS model with CM variable selection method (CM-GA-ANFIS). This model re-modifies the set of financial ratios selected by the CM method, and finally the financial ratios of "profit before interest and taxes on total assets", "total debt to total assets", "and fast assets to current liabilities", "and introduces the cost of interest on gross profit as the final predictors of bankruptcy of Tehran Stock Exchange companies. Comparing the results of this model with other models made in other countries indicates that this model is among the best designed models in terms of efficiency and accuracy of prediction.

Comparison of the proposed GA-GMDH model with other models of bankruptcy prediction, such as other types of neural networks and intelligent methods, and even meta-heuristic methods, etc. can be done in the continuation of this research, to determine the strength or weakness of this technique compared to other techniques. Combining the GMDH polynomial network model with other algorithms such as PSO, Ant colony, genetic programming, etc. and examining its efficiency and effectiveness in predicting corporate bankruptcy.

The GMDH model is one of the neural-regression networks and its output is a polynomial of predictor variables that predicts the dependent variable. This model can be considered in future research as a variable selection method in combination with another prediction method using a definition other than article 141 of trade law to identify bankrupt companies.

References

[1] Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review. *European journal of operational research*, *180*(1), 1-28

[2] Tsai, C. F. (2009). Feature selection in bankruptcy prediction. *Knowledge-based systems*, *22*(2), 120-127.

[3] Newton, G. W. (2009). *Bankruptcy and insolvency accounting, practice and procedure* (Vol. 1). John Wiley & Sons.

[4] Fakhrehosseini, S. F., & Aghaei Meybodi, O. (2019). Prediction and identification of companies with high bankruptcy probability in Tehran stock exchange (different analysis of models). *Journal of decisions and operations research*, *4*(2), 100-111. (**In Persian**). https://dx.doi.org/10.22105/dmor.2019.179504.1111

[5] Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of accounting research*, 71-111. https://doi.org/10.2307/2490171

[6] Jolliffe, I. T. (2002). *Principal component analysis for special types of data* (pp. 338-372). Springer New York.

[7] Altman, Edward I (2000), Predicting Financial Distress of Companies: Revisiting the Z-Score and ZETA Models. In Bell, A. R., Brooks, Ch., Prokopczuk, M (Eds.), *Handbook of research methods and applications in empirical finance* (pp. 428-456). https://doi.org/10.4337/9780857936080.00027

[8] Hosseini Monfared, S. N., Hosseinzadeh Lotfi, F., Mozaffari, M. R., & Rostamy Malkhalifeh, M. (2021). Classifying flexible and integer data in two-stage network data envelopment analysis. *Journal of applied research on industrial engineering*, *8*(3), 270-289.

[9] Kaviani, M., & Fakhrehosseini, S. F. (2020). Application of fuzzy logic in portfolio management: evidence from Iranian researches. *Journal of fuzzy extension and applications*, *1*(2), 108-111.

[10] Nadafi, Z., & Pourali, M. R. (2020). The effect of stock liquidity on companies future investment: a study of the Iranian capital market. *Innovation management and operational strategies*, *1*(3), 269-283. (In Persian). http://www.journal-imos.ir/article_125869.html?lang=en

[11] Darvishinia, R., Ebrahimzadeh Shermeh, H., & Barzkar, S. (2019). Development of a forecasting model for investment in Tehran stock exchange based on seasonal coefficient. *Journal of applied research on industrial engineering*, *6*(4), 333-366.

[12] Sarmadi S. (2013). *Investigating of relationship between intellectual capital and financial performance of petrochemical companies listed in Tehran stock exchange*. Retrieved from https://dx.doi.org/10.2139/ssrn.2251620

[13] Bezdek, J. C. (1976). A physical Interpretation of Fuzzy ISODATA. *IEEE trans syst. man, cybern*, 6(5), 387-389. https://doi.org/10.1109/TSMC.1976.4309506

[14] Chiu, S. L. (1994, June). A cluster estimation method with extension to fuzzy model identification. *Proceedings of 1994 IEEE 3rd international fuzzy systems conference* (pp. 1240-1245). IEEE.

[15] Varetto, F. (1998). Genetic algorithms applications in the analysis of insolvency risk. *Journal of banking & finance*, *22*(10-11), 1421-1439.

[16] Molaei, S., & Cyrus, K. M. (2014). Robust design of maintenance scheduling considering engineering insurance using genetic algorithm. *International journal of research in industrial engineering*, *3*(1), 39-48.

[17] Jones, S., & Hensher, D. A. (2007). Modelling corporate failure: a multinomial nested logit analysis for unordered outcomes. *The British accounting review*, *39*(1), 89-107.

[18] McKee, T. E., & Lensberg, T. (2002). Genetic programming and rough sets: a hybrid approach to bankruptcy classification. *European journal of operational research*, *138*(2), 436-451.

[19] Altman, E. I. (2006). *Corporate financial distress and bankruptc*. John Wiley & Sons, Inc.

Rahbar | J. Appl. Res. Ind. Eng. 9(3) (2022) 274-290