Winter 1-31-2023

An Artificial Intelligence Framework to Contractor Financial Prequalification

Salah Elgamal
salahelgamal@aucegypt.edu

THE AMERICAN UNIVERSITY IN CAIRO
الجامعة الأمريكية بالقاهرة

Graduate Studies

An Artificial Intelligence Framework to Contractor Financial Prequalification

A thesis submitted by:

Salah Elgamal

To the Construction Engineering Department

Under the supervision of:

Dr. Ossama Hosny

Graduate Program Director & Professor

Declaration of Authorship

I, Salah Elgamal, declare that this thesis titled, "An artificial intelligence framework to contractor prequalification" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Salah ElGamal

Date:

25th of August 2022

Abstract

Financial distress in the construction industry always causes major disruptions that usually result in a rippling effect on the economy. Avoiding such defaults is a top priority for employers to meet their demands. Artificial Intelligence (AI) models have provided increased accuracy in predicting financial distress compared to statistical, fuzzy and logistic regression models, and other classification models. The main objective of this work is to support project employers in pre-qualifying contractors by predicting the status of construction contractors during a bid analysis to disqualify contractors with a high probability of experiencing financial distress during the project duration. Eight financial indicators & six macroeconomic variables were used in the analysis. The selected variables were proven to be highly correlated with the output values as provided in the literature while maintaining variables with diverse effects on the output. This work employs multiple models including artificial neural networks (ANN), support vector machines (SVM), and logistic regression using different tools (Python & NeuralTools) based on collected financial statements and macroeconomic indicators. The results show that the ANN model developed using python achieved higher performance measures than SVM (radial basis function & linear kernel functions), logistic regression & ANN developed using NeuralTools. The results also show that adding macroeconomic variables to financial ratios as input variables significantly enhance the accuracy and F-1 score of the model. Accordingly, the developed model is effective in predicting financial distress for construction companies.

Keywords: Financial Distress, Financial Ratios, Economic Indices.

Acknowledgments

My sincere appreciation and gratitude to my advising professor, Dr. Ossama Hosny for his guidance throughout my master's journey and during this research.

I would like to thank Eng. Ahmed El Shanawy for supporting me in the data collection phase and reviewing the selected financial ratios and macroeconomic indicators.

I would like to acknowledge and appreciate my colleague’s guidance through my research, Ahmed Shiha, by providing insights and suggestions on the model development phase.

This work is dedicated to my family and friends, their unconditional encouragement is the key driver for my commitment to complete this work.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Name	**Abbreviation**
Financial Ratios	FR
Macroeconomic Variables	MV
North American Industry Classification System	NAICS
Artificial Neural Networks	ANN
Multiple Discriminant Analysis	MDA
Multiple Criteria Decision Analysis	MCDA
Support Vector Machines	SVM
Logit Analysis	LA
Logistic Regression	LR
Asset Turnover	AT
Current Ratio	CR
Working Capital to Total Revenues	WCTR
Retained Earnings to Total Assets	RETA
Total Debt to Total Assets	DA
Total Debt to Total Equity	DE
EBIT to Total Assets	EBITTA
EBIT to Total Revenues	EBITTR
Consumer Price Index	CPI
Gross Domestic Product	GDP
Total Business Failures	TBF
Total Employees	TE
Long short-term memory - recurrent neural network	LSTM-RNN

Chapter 1: Introduction

1.1 Background

A construction project, as defined by PMBOK (2017) is a temporary process to create a unique product where the project could be stand-alone or be part of a program or portfolio (PMBOK, 2017). Construction contractors are susceptible to many risks and these risks may affect the contractor either on the project level, the company level, or both; posing a risk for contractor failures to perform projects and therefore putting the contractor in a position for dissolution or filing bankruptcy which are forms of financial distress.

Due to the unique nature of projects, cash flow could form a liquidity challenge for contractors since the cash lifecycle is comprised of the following: a working period where the contractor covers the cost of work until he submits the invoice; then another period initiates where the invoice is reviewed and revised by the engineer and the owner then compensates the contractor; the total number of days could reach up to 60 or more for full cash to cash conversion. However, it is important to note that the payments received by the Contractor for his finished goods are often subject to retainage which also increases the need for financing either by debt, equity, or both. (Ihab, 2014).

It is also important to note that there are different types of financial distress that a firm might face as shown in table 1; other than bankruptcy; a company might be acquired with poor financial status, delisted, and some types of mergers (Chen, 2012).

Table 1 Definition of financial distress code (Chung, 2016)

Events	Description
Non-sufficient funds (NSFs)	It is a term used in the banking industry to indicate that a demand for payment (a check) cannot be honored because insufficient funds are available in the account on which the instrument was drawn
Bankruptcy	It is a legal status of a person or other entity that cannot repay the debts it owes to creditors
Going concern	It is a business that functions without the threat of liquidation for the foreseeable future, usually regarded as at least within 12 months
Restructuring	It is the corporate management term for the act of reorganizing the legal, ownership, operational, or other structures of a firm for the purpose of making it more profitable, or better organized for its present needs
Bailout	It is a colloquial term for giving financial support to a firm or country which faces serious financial difficulty or bankruptcy
Net assets is negative	Net assets of the firm are negative, and there are no plans to increase its capital

Source Taiwan Economic Journal

Table 2 explains the number of firms and establishments available in the US annually between 2000 to 2019, the number of employees working for these firms and establishments; and finally, the last two columns indicate the number of firms and establishments that cease all operations as defined by the US census bureau (2022). It is noticeable from the high number of firm deaths that the industry encounters high risks that force contracting companies into financial distresses.

Table 2 Statistics for firms and establishments (US Census, 2022)

Year	Firms	Establishments	Employees	Firmdeath_firms	Firmdeath_establishments
2000	4,905,920	6,295,525	113,339,093	405,496	421,880
2001	4,916,644	6,336,676	113,827,275	408,907	426,387
2002	4,931,112	6,391,857	111,629,580	434,669	449,930
2003	4,992,095	6,457,382	112,319,850	407,126	418,847
2004	5,085,833	6,565,627	113,994,503	405,744	417,810
2005	5,168,533	6,677,154	115,303,461	420,154	431,425
2006	5,257,435	6,838,865	119,106,296	451,311	461,740
2007	5,294,245	6,918,368	119,684,436	458,737	468,469
2008	5,233,330	6,875,295	119,437,627	487,263	497,858
2009	5,092,532	6,738,298	113,559,895	503,996	516,924
2010	5,019,625	6,671,187	111,191,855	434,298	446,876
2011	4,996,662	6,667,358	112,781,943	414,629	427,592
2012	5,043,234	6,710,043	115,178,250	374,363	386,695
2013	5,068,464	6,746,894	117,366,195	380,912	388,629
2014	5,106,214	6,825,393	119,997,966	380,185	388,630
2015	5,151,871	6,909,168	123,127,141	380,745	390,563
2016	5,219,436	7,037,030	125,653,010	387,455	397,679
2017	5,252,059	7,071,934	127,304,992	403,778	415,332
2018	5,288,540	7,098,769	129,588,481	417,399	431,957
2019	5,324,658	7,151,577	131,788,812	458,835	470,653

The definitions provided below are vital to differentiate between the terms provided in table 2 (US Census Bureau, 2022).

Firm: a business organization comprised of one or more local establishments in the same geographic area

Establishment: an economic unit that generates services or goods, often at a single physical location and engaged in one activity primarily. (U.S. Bureau of Labor Statistics)

Contractor Prequalification's objective is to disqualify unfit contractors that are predicted to experience financial distress over the project's lifetime. This technique improves the quality of the bid and promotes effective decision-making by the employer. Financially disqualified contractors might pose a risk to the employer for default or bankruptcy; both would affect the employer negatively even if the contractor is insured.

Previous research shows the parameters that are studied which contribute to the financial status of the contractor. Figure 1 illustrates the variables that affect the prediction model for contractor default developed by Al-Sobei (2005). The author divided the selection criteria into three categories based on: the contractor selected for the project, the contract characteristics, and the nature of the project. The input variables enlisted under those categories are processed into the prediction model. The output is the likelihood of contractor default. Following the likelihood of contractor default, recommended actions are proposed to mitigate and decrease the likelihood of contractor default (Al Sobiei, 2005).

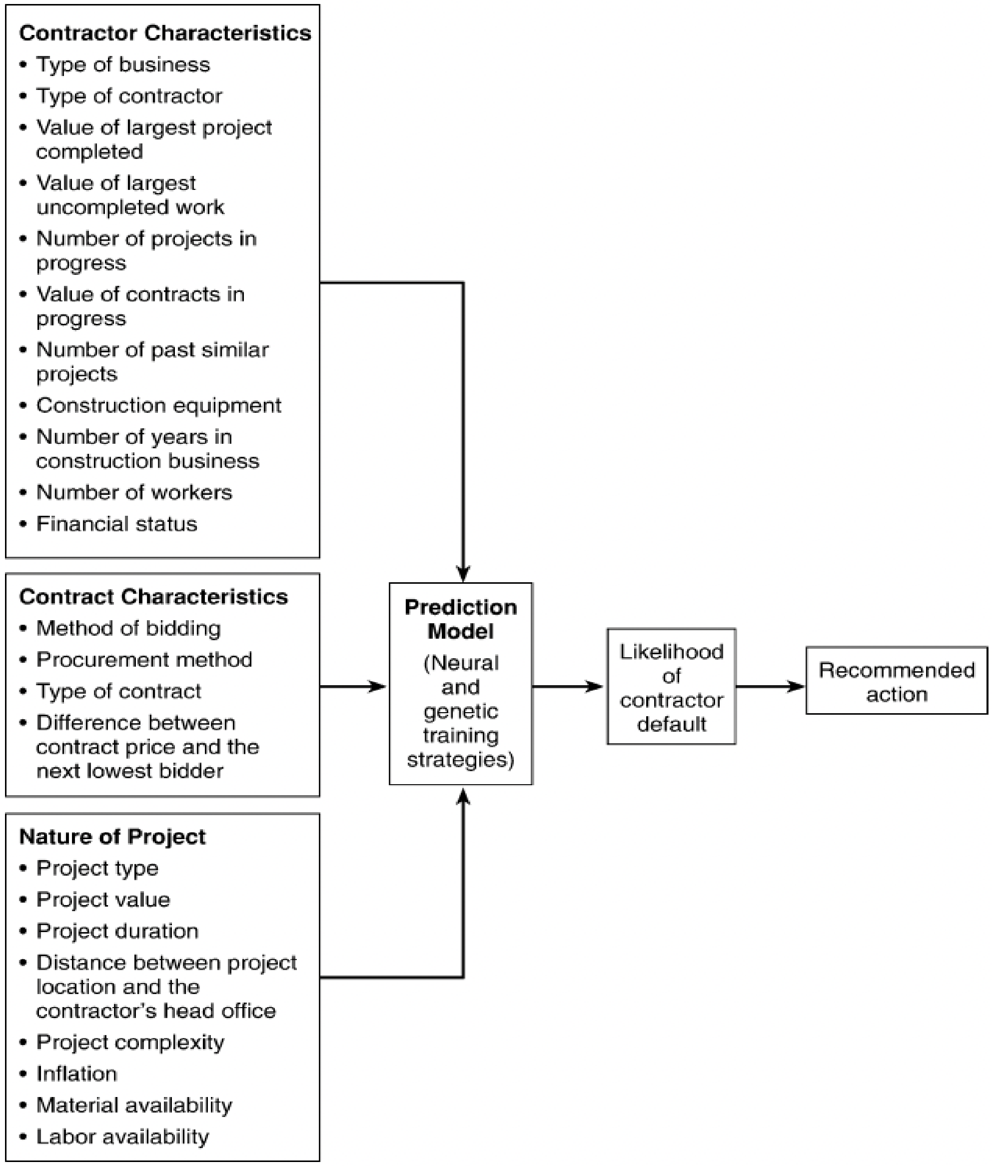


Figure 1 Contractor Characteristics for Prequalification (Al-Sobiei, 2005)

Generally, disqualification is decided based on numerous grounds which might be divided into two categories; qualitative which includes reputation, past experience, quality of work performed, contractor's organizational structure, and many other factors, and quantitative factors that reduce subjectivity through the use of company financial data and macroeconomic data.

Figure 2 shows the criteria recommended by Holt (1994) to be considered for qualifying contractors: a criteria layer consisting of five criteria groups: contractor's organization, financial consideration; management resource; past experience and past performance

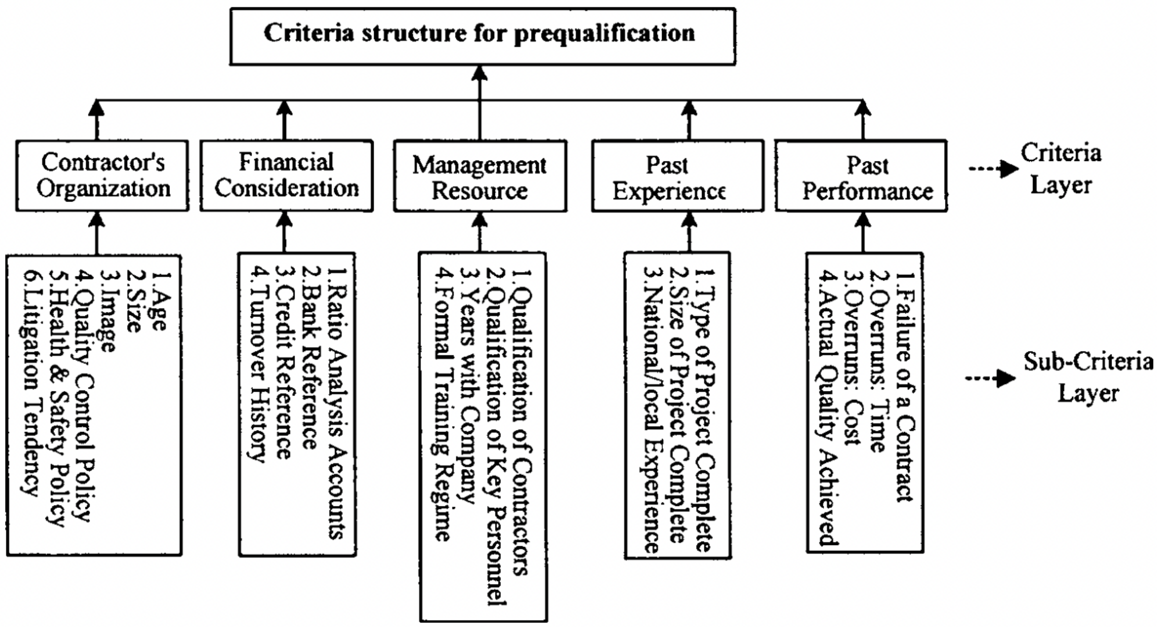


Figure 2 Prequalification criteria structure (Holt, 1994)

Beaver (1966) utilized financial ratios to predict bankruptcy in multiple industries for a range of over a period of five years based on multiple financial ratios illustrated in figure 3 believed to mostly affect the company's solvency state using a dichotomous model which identifies specific cut-off scores for each of the input variables and an overall score is calculated which indicates the status of the studied construction company.

GROUP I (CASH-FLOW RATIOS)
1. Cash flow to sales
2. Cash flow to total assets
3. Cash flow to net worth
4. Cash flow to total debt

GROUP II (NET-INCOME RATIOS)
1. Net income to sales
2. Net income to total assets
3. Net income to net worth
4. Net income to total debt

GROUP III (DEBT TO TOTAL-ASSET RATIOS)
1. Current liabilities to total assets
2. Long-term liabilities to total assets
3. Current plus long-term liabilities to total assets
4. Current plus long-term plus preferred stock to total assets

GROUP IV (LIQUID-ASSET TO TOTAL-ASSET RATIOS)
1. Cash to total assets
2. Quick assets to total assets
3. Current assets to total assets
4. Working capital to total assets

GROUP V (LIQUID-ASSET TO CURRENT DEBT RATIOS)
1. Cash to current liabilities
2. Quick assets to current liabilities
3. Current ratio (current assets to current liabilities)

GROUP VI (TURNOVER RATIOS)
1. Cash to sales
2. Accounts receivable to sales
3. Inventory to sales
4. Quick assets to sales
5. Current assets to sales
6. Working capital to sales
7. Net worth to sales
8. Total assets to sales
9. Cash interval (cash to fund expenditures for operations)
10. Defensive interval (defensive assets to fund expenditures for operations)
11. No-credit interval (defensive assets minus current liabilities to fund expenditures for operations)

Figure 3 Tested financial ratios (Beaver, 1996)

Following Beaver, Altman (1968) utilized five ratios over a period of 1-2 years to predict bankruptcy in manufacturing companies. He developed the Z model which is an equation that includes these five ratios in a multiple discriminant model (MDA) where the output of the Z is compared to 3 different ranges, a bankrupt range, an operational range, and a range with the uncertainty of whether the company is bankrupt or stay operational. Eq. 1 illustrates the 5 financial ratios employed in his research and the weights added to each ratio. (Altman, 1968)

X1 = Working Capital / Total Assets

X2 = Retained Earnings / Total Assets

X3 = Earnings Before Interest and Tax / Total Assets

X4 =Market Value Equity / Book Value of Total Debt

X5 =Sales / Total Assets

$$\mathbf{Z= 0.12\ X1 + 0.14\ X2 + 0.33\ X3 + 0.006\ X4 + 0.999\ X5} \quad (1)$$

However, it is to be noted that the construction industry is distinct from the manufacturing industry in many aspects, such as that construction projects are unique and different which increases the risks of failures and defaults, increases uncertainties, and variations in quantities. Therefore, researchers analyzing construction companies used different ratios than the ones used by Altman (1968).

Russel (1996) provided a stochastic statistical model using stepwise regression that tackles an important limitation in earlier models, where he included external macroeconomic factors in addition to internal financial factors in his model. This addition changes the model from being a partially instant model; since the weights are not updated based on external factors, to a more instantaneous model where the changes in the economy are reflected in the model; using a total sample of 430 financial statements that represent 120 contractors (49 failed and 71 non-failed). His research findings concluded that failed contractors possess a negative trend and larger volatility in the percentage of net worth, gross profit, and net working capital.

1.2 Problem Statement

Defaults, bankruptcies, and other forms of distress of construction contractors are a significant risk for project stakeholders causing contract defaults for owners and bankruptcy for the contractors. The effect of such defaults disrupts the economy and causes major losses to project owners. It is substantial to predict whether the contractor is susceptible to financial distress before bidding on a project to allow employers to eliminate unfit contractors with a high probability of facing financial distress.

1.3 Research Objective

The objective of this research is to create a tool that classifies financial distress among construction contractors in the following year; This will be achieved through:

1. Review the available models and their accuracy to predict the financial status of a construction contracting company.
2. Determine the variables affecting the contractor's financial distress and study the significance and correlation of variables to the output.
3. Construct a model that inputs financial ratios only or financial ratios and macroeconomic variables, and classifies the output either as distressed or non-distressed cases.

4. Evaluate the combined effect of financial ratios and economic indices on the accuracy of distress prediction before financial distress.
5. Apply several techniques in developing the distress prediction models (ANN, SVM, and logistic regression), compare and evaluate their accuracy in predicting construction companies future financial status.

1.4 Research Methodology

This research passed through the following stages

1. Reviewing the available literature on distress prediction using various models such as ANN, Fuzzy, SVM, multiple discriminant analysis, and logistic regression and comparing the different accuracies provided by these models.
2. Researching the significant financial and macroeconomic input variables and collecting such data for financially distressed and non-distressed contractors
3. Calculating the correlation between the input variables financial ratios (FR) and macroeconomic indicators (MV).
4. Studying the correlation and significance between the input variables (FR & MV) and the output value.
5. Develop several distress prediction models using different techniques (ANN, SVM, and logistic regression models using python and ANN model using NeuralTools first using financial ratio inputs (FR) and then using financial ratios and macroeconomic variable (FR & MV) inputs.

6. Measure the effect on the performance measures (accuracy and F1-score) when using macroeconomic variables in addition to financial ratios versus financial ratios only.

Figure 4 illustrates the research methodology followed throughout data collection, model development, and evaluating the performance measures with different iterations encountered through multiple trials of data collection and generating performance measures which constitute accuracy and F1-score.

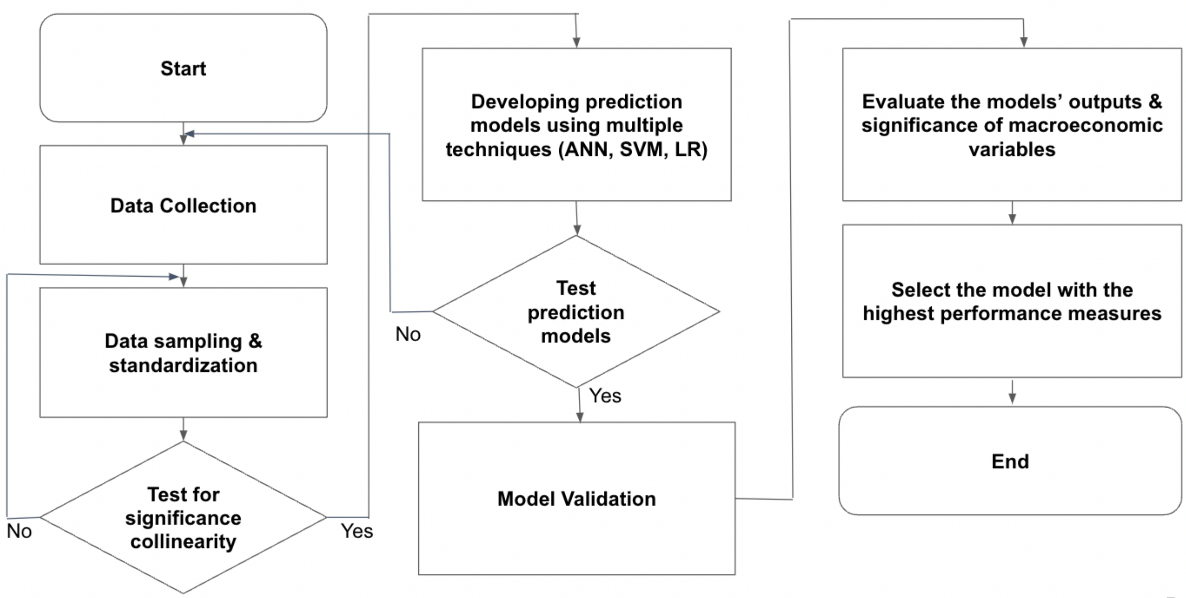


Figure 4 ANN research methodology flowchart

1.5 Thesis Organization

This research consists of six chapters. The following is a brief description of each chapter:

Chapter 1 – Introduction: This chapter provides general background on distress prediction and the importance of prequalification in the construction field followed by the problem statement, research objectives, methodology, and thesis organization.

Chapter 2 – Literature Review – This chapter summarizes literature findings and the different types of models used to predict distress for construction contractors and the different variables used for prediction (financial ratios, macroeconomic indicators, cash flow).

Chapter 3 - Data Collection and Sampling - this chapter introduces the collected data consisting of input variables (Financial Ratios & Macroeconomic Variables) and illustrates the removal of outliers and missing data points.

Chapter 4 - Model Development - this chapter illustrates the developed models (ANN, SVM, and logistic regression) and includes a brief description of the model parameters used.

Chapter 5 - Discussion and Analysis - this chapter discusses the results provided in chapter 4, and provides a comparison between the different tools and techniques used in this research.

Chapter 6 - Conclusion and Recommendations - this chapter summarizes the outcomes of this research and compares the output parameters of the different models; the contribution of this research to the available literature; in addition to providing recommendations for further research.

Chapter 2: Literature Review

This chapter provides an overview of the different models used to predict and classify the different forms of financial distress using different tools and techniques.

2.1 Dichotomous Classification Test

The first model predicted to classify bankruptcy in companies was introduced by Beaver (1966); where the model indicates a certain value for each ratio as a cut-off score for the classification; a limitation within this model is the linear assumption for all ratios. (Ismail, 2014)

2.2 Multiple Discriminant Analysis (MDA)

MDA is the first statistical modeling technique to be utilized in bankruptcy prediction by Altman (1968). It is a statistical model used when the input variables are discrete, while the output variable is categorical. The model produces a statistical significance report that shows whether the input variables are significant, then using a scaled vector, the relative contribution of each variable is determined which results in the output function. The data inputs of this model assume that the variables are normally distributed which may decrease the accuracy of the model based on the input data and the time of simulation (Altman, 1968). Table 3 shows the scaled vector which determines the ranking of the input variables.

Table 3 Contribution of factors and their ranking (Altman, 1968)

RELATIVE CONTRIBUTION OF THE VARIABLES

Variable	Scaled Vector	Ranking
X_1	3.29	5
X_2	6.04	4
X_3	9.89	1
X_4	7.42	3
X_5	8.41	2

X1 = Working Capital / Total Assets

X2 = Retained Earnings / Total Assets

X3 = Earnings Before Interest and Tax / Total Assets

X4 =Market Value Equity / Book Value of Total Debt

X5 =Sales / Total Assets

However, this model is not encouraged to be used in this study, due to the assumptions presumed by the statistical nature of MDA such as linearity, normality, and independence of predictor variables which is not necessarily the case for the economic indices tackled in the following sections. (Zhang, 1999).

2.3 Logit Analysis (LA)

LA started to be used more than MDA in the 1980s which is a linear maximum likelihood method used to determine company failure (Ismail, 2014). The model utilizes parameter values by combining multiple company parameters into a multivariate probability score; the output is represented as the company's probability of failure (Ohlson, 1980). LA has several drawbacks, although costs of type 1 and type 2 errors are minimized, they assume equal misclassification cost, the choice of cut-off score is considered robust and multicollinearity affects the model's performance negatively. (Doumpos & Zopoudinis, 1999).

2.4 Neural Networks (NN)

Neural networks is a simulation technique that mimics the thought process of the brain; wherein the neurons are all connected through a network transferring the inputs to layers that process the information and output a categorical value based on previous experiences "training data set"(Aggarwal, 2018). Although a major drawback of neural networks is the possibility of getting stuck in a local minimum, Zhang (1999) approached this problem by increasing the processing power as the author increased the number of training sets to 50 and trained them 50 times; randomly selected by weights to decrease the probability of being trapped in a local minimum. Also, cross-validation was performed in his model which determines the robustness of the model through statistical analysis. (Zhang, 1999). However, his work did not consider the effect of external factors such as gross domestic product (GDP), interest rates, and other factors that affect financial performance.

Figure 5 shows the architecture developed by Zhang (1999) which consists of 3 layers, an input layer that includes financial ratios obtained through data collection, a hidden layer which is comprised of neuron connections that are multiplied by certain weights that are updated each trial to improve the prediction power of the model, and the output layer which is a classification either bankrupt or non-bankrupt.

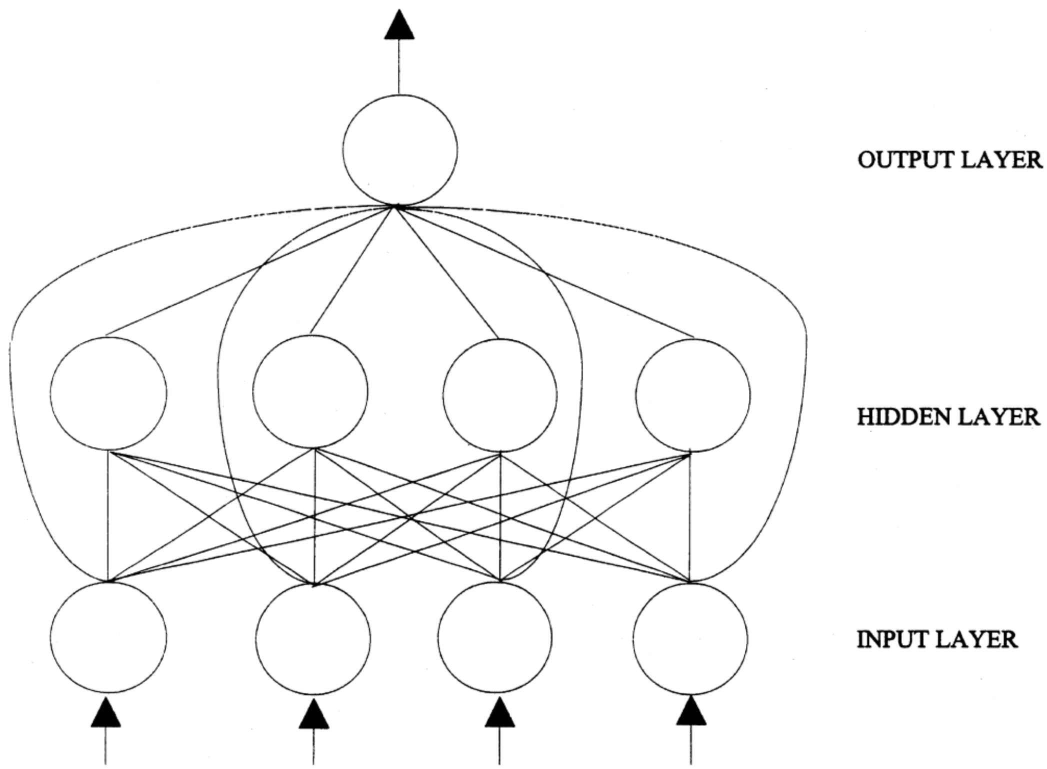


Figure 5 ANN typical model architecture (Zhang, 1999)

Although neural networks were applied to bankruptcy prediction in the construction field in 1999, there have been limited applications since that time in the literature for bankruptcy prediction. Although there are lots of different architectural formations for ANNs, most researchers focused on changing the type of the neural network while maintaining an architecture with 1 hidden layer as provided in figure 5.

Neural networks were verified to give better results with respect to multiple discriminant analysis. Also, it is illustrated by Zhang (1999), that complex problems that predict unseen objects in the test sample would require more complex architectures which would result in an increase in the number of hidden layers. (Zhang, 1999)

Alternatively, Jang (2019) introduced a long short-term memory - recurrent neural network (LSTM-RNN) model that accounted for the time series aspect of distress prediction, the model employed iterative functions that utilize previous information from earlier time intervals integrated with current time interval data to produce the output. The study used 12 financial ratios in its analysis, 3 construction market variables, and 3 macroeconomic variables. A feedforward neural network and a support vector machine were used as a benchmark to evaluate the performance of the proposed LSTM-RNN. The results show that this model generates higher accuracies compared to its benchmark values.

Figure 6 represents the LSTM iterations, where figure 6 (b) consists of multiple parallel neurons which resemble the ordinary neural network; however, the time series feature is added to this network which resembles the different values of input and output at different times; where h represents the output, x represents the input and f represents the activation function used in the network. While the left section of figure 6 (a) shows the different times where the network learns from the stored memory which is an added feature to the LSTM compared to artificial neural networks.

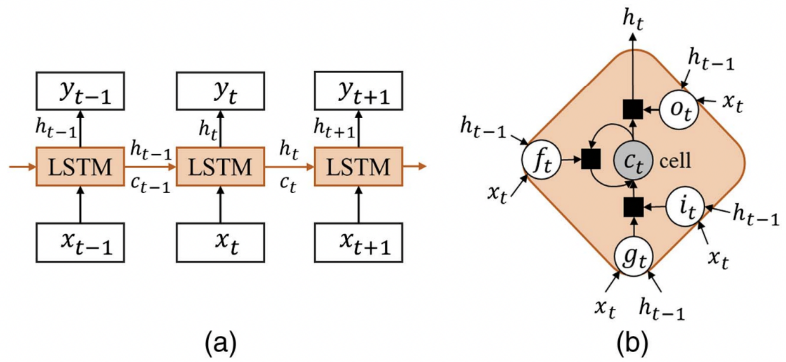


Figure 6 LSTM iteration diagram (Jang, 2019)

Figure 7 demonstrates the implementation of the methodology provided by Jang (2019), which explains the different layers and their constituents; where the input is in the form of a time series with results at different instants for the same input variable, then the LSTM layer proceeds with adjusting the weights according to the descent function and to predict the value after passing through the softmax activation function and to finally predict the output (y_t).

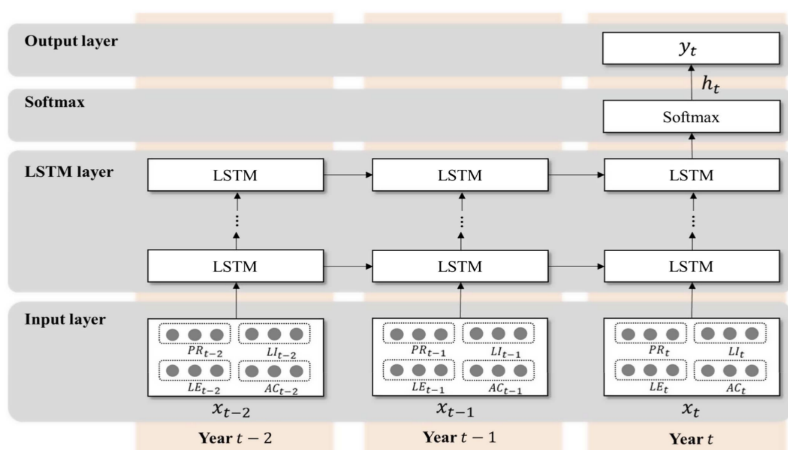


Figure 7 Architecture of Business Failure Prediction using LSTM RNN (Jang 2019)

Another method utilized by Chen (2012), discusses a different configuration of neural networks applied in corporate distress; using self-organizing neural networks incorporating Hyper-rectangular composite neural networks (HRCNN), fuzzy and self-organizing mapping optimization (SOMO). This technique provides a self-mapping feature where the HRCNN performs supervised decision directed learning and the model integrates the fuzzy concept in the HRCNN layer by measuring the similarity or distance between inputs and the hyper-rectangular area. This study was employed on a total of 1615 financial reports and a ratio of 0.5 to 0.5 for the distressed versus non-distressed group. (Chen, 2012)

Figure 8 shows the fuzzy-based rectangular composite neural network architecture where the inputs shown in "X" are transferred from the input to the hidden nodes where fuzzy membership functions illustrated in "m(.)" are transferred through rules to the output.

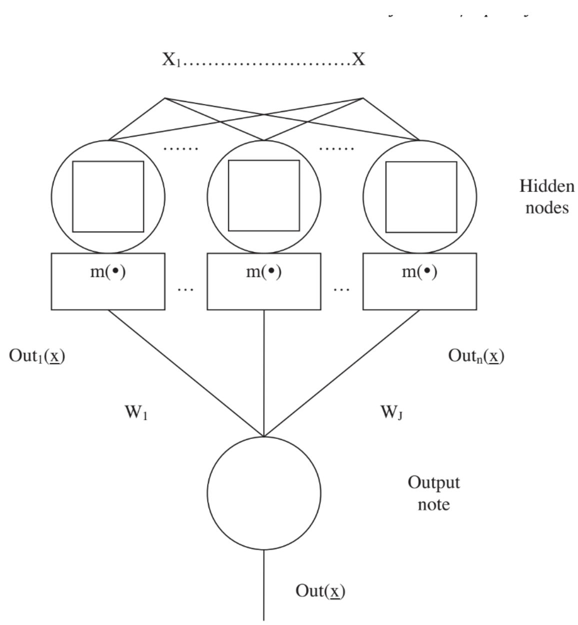


Figure 8 Fuzzy-based hyper rectangular composite neural network (Chen, 2012)

Similarly, a model developed to predict contractor default in Saudi Arabia (Al-Sobiei, 2005) aimed to use artificial neural networks and genetic algorithms with input variables from three categories: nature of project; contract characteristics, and contractor characteristics; each set of those categories contain multiple variables that are inserted in a model. This study employs a sample of 21 contractor defaults and 33 operational contractors for the training set; while 5 projects were retained for the testing set where 2 projects were defaulted by the contractor and 3 projects were completed successfully. The model included a genetic algorithm as an alternative to ANN training. Genetic algorithm is used to predict the risk of contractor default in construction projects undertaken for the Saudi armed forces.

Table 4 summarizes the training statistics provided by Al-Sobiei (2005) for both neural and genetic algorithms developed. The table shows a high correlation coefficient and determination coefficient for both techniques over the training sample. However, the mean squared error for the genetic algorithm is much higher than the neural network; which could be explained by overfitting as the genetic algorithm is susceptible to being stuck at local minimum values even after the use of mutation techniques which proves the neural networks to be better at minimizing cost functions using stochastic techniques such as the gradient descent method.

Table 4 Comparison between training ANN and Genetic Algorithm (Al-Sobiei, 2005)

Training strategy	R-squared	Average error	Correlation coefficient r	Mean squared error
Neural	0.98	0.05	0.99	0.01
Genetic	0.96	2.09	0.98	62.33

Chung (2016) developed two models to predict financial distress (Genetic Programming and Cerebellar neural network). He used Taiwanese data from the stock market; this study utilized a large data set comprising 240 non-distressed companies and 120 distressed companies. The model provides different loss functions to converge and update the weights to increase the predictive power.

The architecture shown in figure 9 represents the common features of an artificial neural network. In addition, this model includes a memory cell associated with the hidden layer; which is a feature of the recurrent neural networks to update the weights linked to the previous iteration in stochastic learning.

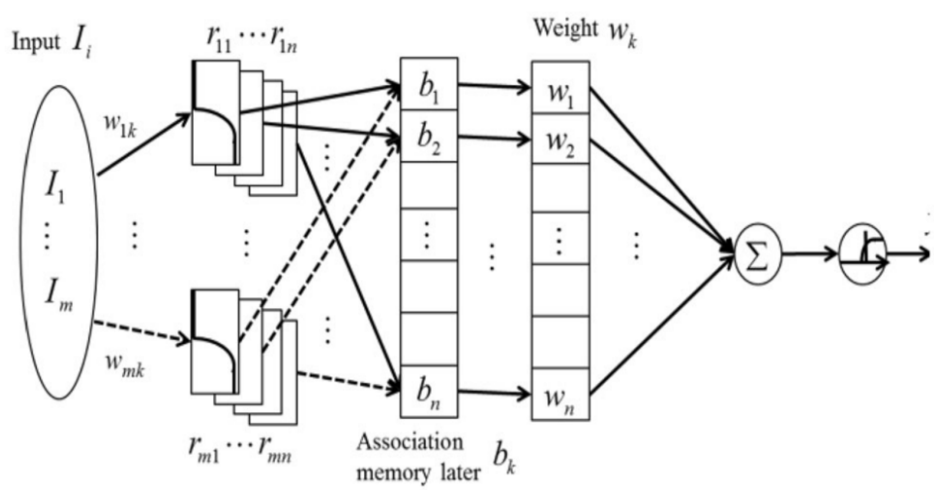


Figure 9 Cerebellar Model Architecture (Chung, 2016)

Table 5 displays the discriminant analysis on learning samples for Cheng (2016) and table 6 represents the discrepant analysis on testing samples. The accuracy of both models (Genetic Programming and Cerebellar neural network) in the training set achieved 100% which is due to the high number of samples provided; while for the testing set, the accuracy of the cerebellar model achieved higher accuracy (97.5%) than the genetic programming model (90.84%) with a higher type 1 & 2 errors for the neural networks model with respect to genetic programming which are the classification errors for both groups where type 1 error is the predicted value produced equals 1 while the actual value was 0 and type 2 error represents the case where a 0 value is predicted while the actual value is 1.

Table 5 Discriminant analysis on learning samples (Chung 2016)

	Actual	Predicted	Error	
Genetic programming (samples: 240)				
NBF	160	160	0	$T = 100\ \%$
BF	80	80	0	
Cerebellar model neural network (samples: 240)				
NBF	160	160	0	$T = 100\ \%$
BF	80	80	0	

BF bankrupt firm, *NBF* non-bankrupt firm, *T* correctly predicted

Table 6 Discriminant analysis on testing samples (Chung, 2016)

	Actual	Predicted	Error	
Genetic programming (samples: 120)				
NBF	80	75	5	$\alpha = 4.16$ %
BF	40	34	6	$\beta = 5.00$ %
				$T = 90.84$ %
Cerebellar model neural network (samples: 120)				
NBF	80	78	2	$\alpha = 1.67$ %
BF	40	39	1	$\beta = 0.83$ %
				$T = 97.50$ %

BF bankrupt firm, *NBF* non-bankrupt firm, α type I error, β type II error, T correctly predicted

2.5 Support Vector Machines (SVM)

Similar to neural networks, SVM shares the same method for learning and testing methods for pattern recognition, while neural networks utilize multi-layer connections and multiple activation functions to resolve non-linear issues. SVM employs non-linear mapping to make the data linear and separable which is developed by the kernel function (Ren, 2012).

Lam (2009), developed an SVM model to predict contractor prequalification applied to 3-different datasets; the first dataset is formed of hypothetical input data designated for a pilot model, the second dataset is obtained through normalized practical application of pre-qualification cases and the third dataset is a recollection of past documented datasets to check the model further by generalizing potentials of proposed frameworks. Table 7 displays a summary of the three datasets.

Table 7 SVM constructs (Lam, 2009)

Summary of research constructs		Summary of data for modeling	
Construct ID	Main purpose	Data ID	Key details
Construct 1	Pilot modeling	Data-I	105 hypothetically simulated datasets, 10 input variables (6-point numerical scale, normalized) and 1 output variable (binary scale)
Construct 2	Modeling practical datasets and verification of optimal design	Data-II	74 practical datasets from recent cases, 8 input variables (5-point numerical scale, normalized) and 1 output variable (binary scale)
Construct 3	Further validation and comparison with neural network outcomes	Data-III	84 practical datasets from secondary source, 11 input variables (5-point scale, normalized) and 1 output variable (binary scale)

Figure 10 demonstrates the methodology followed by Lam (2009) to determine the prequalification status of construction companies. After data collection, cross-validation is applied to the x and y values and the data sets are separated into training and testing, the training set is then initialized after choosing one of the kernel functions (either sigmoid, polynomial, or radial base) all models are evaluated and if they pass; the different kernel functions are compared and the optimal function is considered for re-codification of predictors. (Lam, 2009)

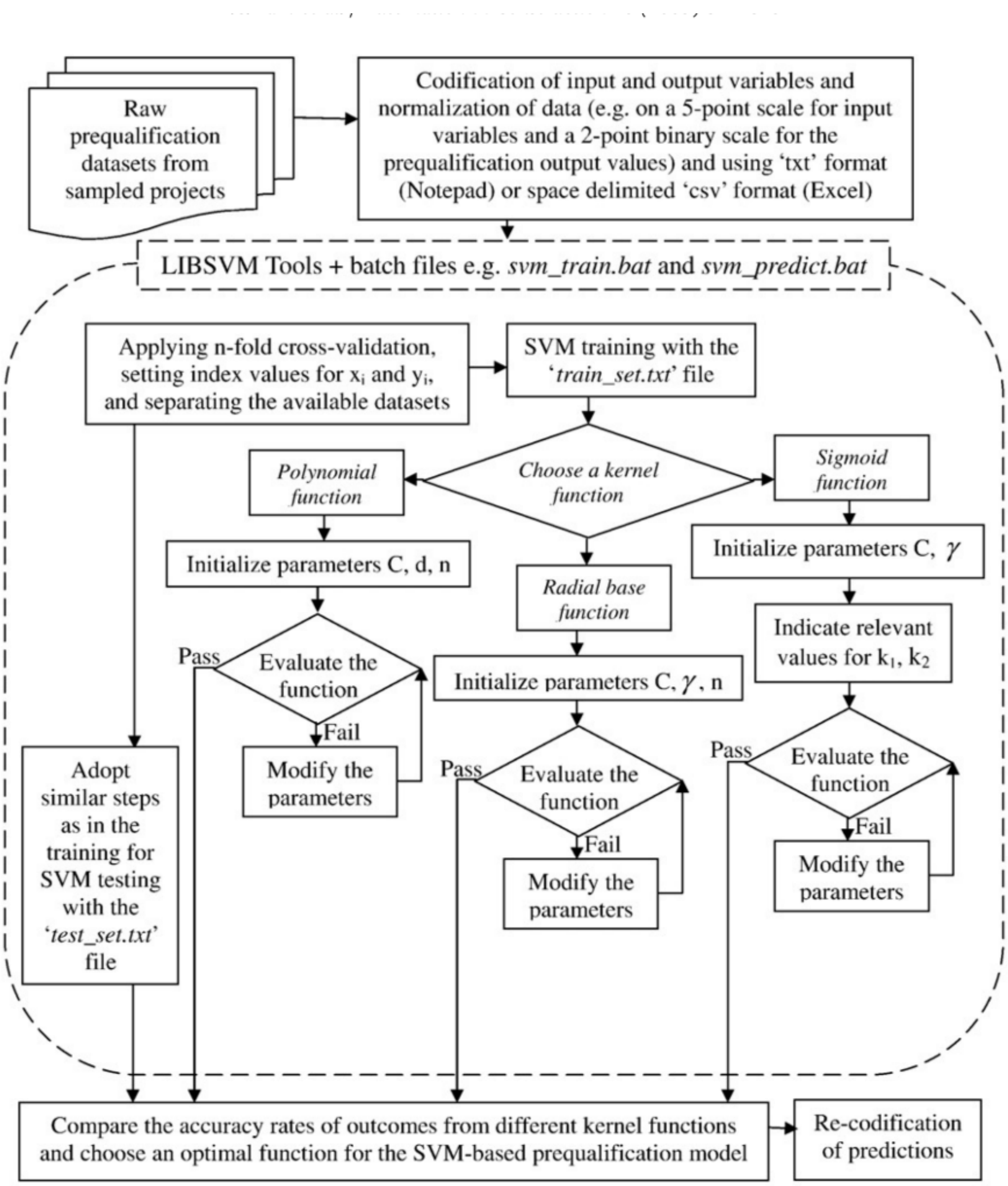


Figure 10 SVM qualification model (Lam, 2009)

Cheng (2014) employed a model to predict contractor default using evolutionary least squares SVM (ELSIM) by hybridizing differential evolution, SMOTE (synthetic minority over-sampling technique), and least square support vector machines, then compared the output of this model with an ordinary SVM model and an ANN model.

Figure 11 represents the optimization algorithm used to minimize the error between the predicted and actual output where mutation and crossover are utilized to generate new values with different error values and as the algorithm reaches the stoppage criteria, the optimal solution is generated. (Cheng et al, 2014).

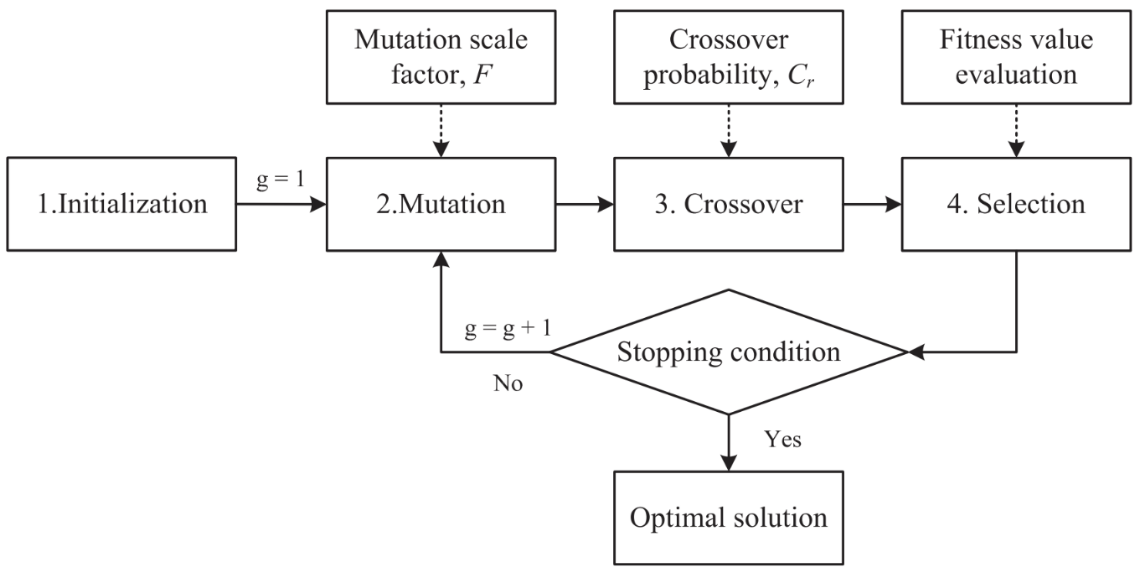


Figure 11 Differential evolution optimization algorithm (Cheng et al, 2014)

The accuracy of the output produced from the three models enlisted in table 8. It can be realized that the ELSIM model achieved the highest accuracy through the differential evolutionary algorithm which improves minimization and therefore increases accuracy. Also, it is noticeable that using 7 financial variables obtained higher accuracy compared to the 20 financial variables which might have contributed to high multicollinearity between the larger number of inputs (Cheng et al, 2014).

Table 8 Accuracy results of ELSIM, SVM, and ANN (Cheng et al, 2014)

Average prediction results obtained from the 5-fold cross validation process.

Case	Models	ELSIM	SVM	ANN
1. Using 20 financial variables	Training result (AUC)*	0.984	0.917	0.930
	Testing result (AUC)*	0.985	0.917	0.750
2. Using 7 financial variables	Training result (AUC)*	0.983	0.925	0.914
	Testing result (AUC)*	0.989	0.922	0.752

* Area under curve

2.6 Fuzzy Logic

Many studies approached the qualification issue from different perspectives, using multiple models such as fuzzy logic, case-based reasoning, and other models. Fuzzy sets were first introduced by Nguyen (1985), who evaluated submitted tenders based on cost, past experience of the tenderer, and present bid information. Usually, most of the parameters used in the input variables are categorical which requires a subjective input. Following Nguyen's model, Plebankiewicz (2009) presented multiple criteria that show the qualities required for a contractor to meet most of the employers' objectives. Table 9 shows the criteria which are studied in Plebankiewicz's work with each sub-criteria.

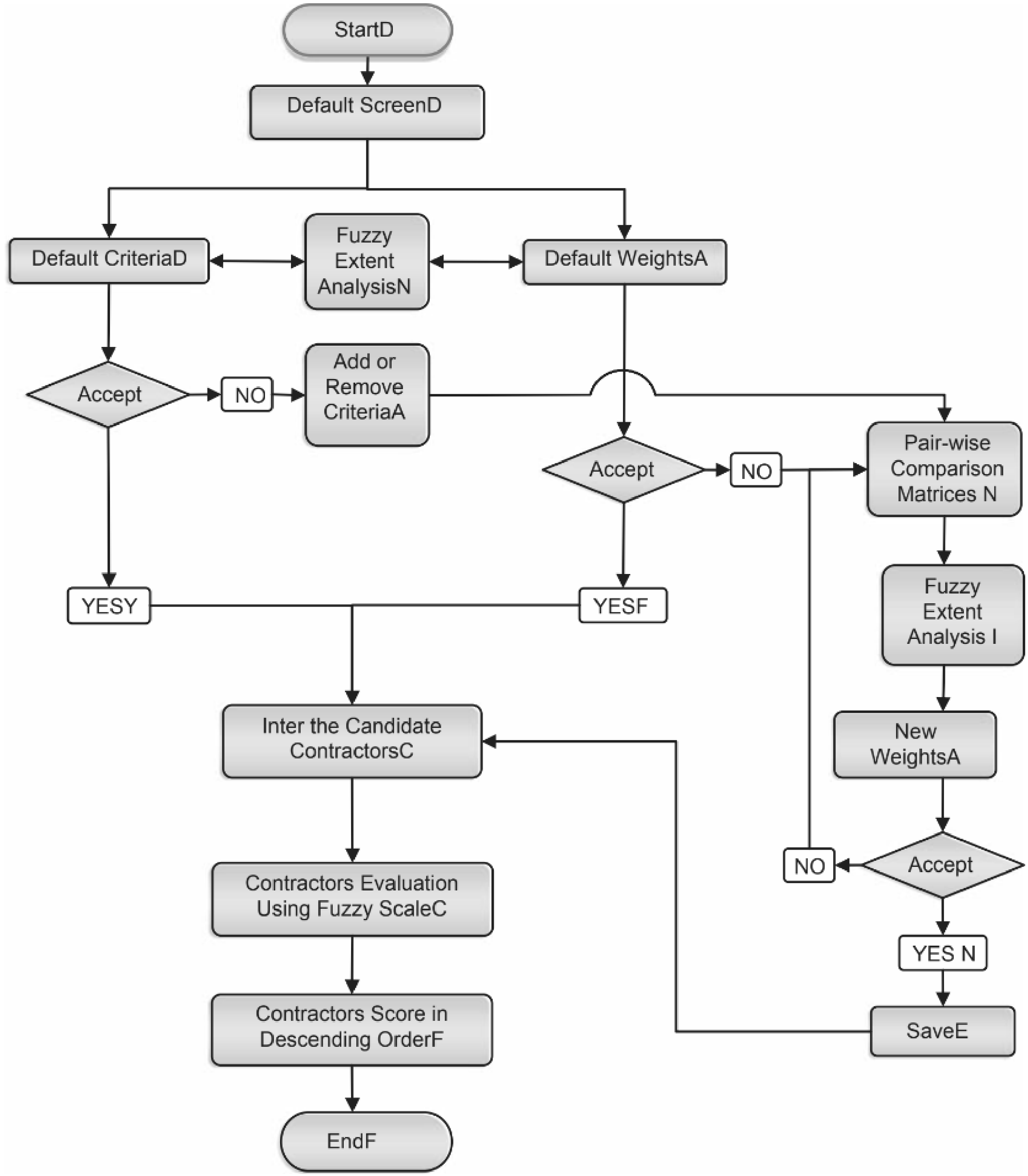
Table 9 Prequalification Criteria (Plebankiewicz, 2009)

Criteria	Example subcriteria
Financial standing	1. Financial stability 2. Turnover, profit, obligations, amounts due 3. Owned financial funds
Technical ability	1. Experience 2. Plant and equipment 3. Personnel
Management capability	1. Past performance and quality 2. Quality control policy 3. Quality management system 4. Project management system 5. Experience of technical personnel 6. Management knowledge
Health and safety	1. Accidents 2. Health and safety management system 3. Insurance policy
Reputation	1. Past failures in completed projects 2. Number of years in construction 3. Past client relationships 4. Cooperation with contactors

As discussed in the previous paragraph, since these inputs are usually subjective; the applicability of the model would depend on the sample of results given to the user which might restrict the model applicability and comparability to multiple iterations that might cause inaccuracy in prediction.

Hosny et al (2013) employed a fuzzy-AHP model by developing 6 criteria to determine the financial status of the contractor such as Contractor's Organization, Financial Considerations, Technical Capability, Past Experience, Past Performance, and Reputation. Following the criteria, variables under each criteria are selected and the methodology included an expert survey where each of the variables selected had a value depending on the contractor selected. Following that, the Fuzzy model outputs the estimated value score for each contractor based on the expert inputs and the cut-off scores. (Hosny et al, 2013).

Figure 12 illustrates the methodology followed for the integrated model of Hosny et al. (2013). As the model starts by default screening, the criteria and weights are initiated with the fuzzy extent analysis then if accepted, the candidate contractor is entered, while if rejected, criteria are added or removed, and/or a process starting with the pairwise comparison matrices is initiated and then fuzzy extent analysis is conducted again, after that the weights are updated and if accepted, the process continues with contractor comparison using the fuzzy scaler and the output is displayed as a contractor score as a range from 0 to 1 for poor and excellent performance respectively.



-

Figure 12 Flowchart of the integrated model (Hosny et al, 2013)

2.7 Macro-economic indicators

Macroeconomic data are significant to the construction industry as construction contractors are critically affected by fluctuations in the economy (Sang et al. 2014). Construction contracts are closely related and highly vulnerable to macroeconomic shifts (Arditi et al. 2000). Russel (1996) concluded that an excess of 60% of construction contractor failures is due to economic factors that include (Russel et al, 1996):

1. Monthly and annual average prime interest rates
2. Consumer price index
3. Gross national product (GNP) measured in current dollars
4. Constant dollars
5. Deflator
6. New business incorporation
7. Total business failures
8. Failure rate
9. Number of construction contractor failures
10. Number of construction workers
11. Number of construction administrative employees
12. Total employees in construction
13. Value of new construction put in place measured in current dollars and deflator (monthly and annual average)
14. Value of construction contracts (monthly and annual average)
15. Holding of construction loans, number of corporate construction income tax forms returned, and items on corporate construction tax returns such as assets, liabilities, receipts, deductions, and net income.

Russel's (1996) work introduces a modeling technique that incorporates the internal variables affecting a construction contracting company to insolvency or bankruptcy with external variables that might increase or reduce the probability of financial distress.

Following Ruessl's (1996) work, an LSTM model performed by Jang (2019) incorporated the following variables:

1. Consumer price index
2. Gross Domestic Product
3. Federal funds rate
4. Employment in construction
5. Construction spending, housing starts

2.8 Summary of distress prediction models

Table 10 shows different techniques used in distress prediction and the variables used in previous research

Table 10 Summary of distress prediction models (Jang, 2019)

Authors	Techniques	Variables
Kangari et al. (1992)	MRA	ROA,revenues to net working capital,Current ratio, total liabilities to net worth,return on net worth, total assets to revenues
Abidali and Harris (1995)	MDA	EBIT to current liabilities,EBIT to net assets, EBIT to net capital employed,EBIT to share capital and reserve
Russell and Zhai (1996)	MRA	Trend - prime interest rate, future position - new construction value in-place, trend - new construction in place, future position - net worth to total asset, trend - net working capital to total asset-gross profit to total asset, volatility

Al-Sobiei et al. (2005)	ANN, GA	Type of contractor,Financial status,project type, project value, project duration, labor availability, material availability,value of largest project completed, value of largest uncompleted work, number of projects in progress, type of contract, value of contracts in progress, number of past similar projects, construction equipment, number of year in construction business, number of workers, method of biding, procurement method, difference between contract price and the next lowest bidder, distance between project location and the contractor's head office, project complexity, inflation
Chen (2012)	SFNN	ROA, net worth turnover ratio, after-tax rate of return,profit margin, operating margin, operating profit to paid-in capital ratio, pre-tax net profit to paid-in capital ratio, earnings per share, operating profit, growth rate, after-tax net profit growth rate, inventory turnover ratio, revenue growth rate, growth rate of total assets, growth in the total returns on assets, equity ratio, debt to assets ratio, long-term funds to fixed assets ratio, dependence on borrowing,receivable turnover ratio, total assets turnover ratio, fixed assets turnover ratio, current ratio, acid-test ratio, and times interest earned ratio
Horta and Camanho (2013)	SVM	ROS, ROA, ROE, current ratio, the working capital over total assets, current asset turnover, company main activity,company size, headquarter geographic location
Adeleye et al. (2013)	LR	Ability to generate sales or net worth, age of a company cost, leverage, non-routine business transactions, market factor, type of trade

Hosny, Ossama et al (2013)	Fuzzy-AHP	Capacity of Contractor, Health and Safety Program, Length of Time in Business, Company Image, Ratio analysis accounts, Credit rating, Bank Arrangements and bonding, Project Control, Experience with Company, Qualification of Key, Persons, Plant, and Equipment, Size of Projects Type of Projects, Experience in Local Area, Actual Quality Achieved, Cost Overrun, Time Overrun, Failure to have Completed Contracts, Past Owner/Contractor Relationship Litigation Tendency
Tserng et al. (2014)	LR	ROA, ROE, ROS,Current ratio, quick ratio, sales to net worth, net working capital to total asset, current asset to net assets, total liabilities to net worth, retained earnings to sales, debt ratio, accounts receivable turnover, accounts payable turnover, quality of inventory, fixed assets to net worth, turnover of total assets revenue to fixed assets, profits to net working capital, book to market ratio, revenue to net working capital, times interest earned ratio
Cheng et al. (2014)	SMOTE, LS-SVM, DE	ROA, ROE, ROS,Current ratio, quick ratio, net working capital to total assets, current asset to net assets, fixed assets to net worth, accounts payable turnover,total liabilities to net worth, retained earnings to sales, debt ratio, times interest earned, revenues to net working capital, accounts receivable turnover, sales to net worth, quality of inventory, turnover of total assets, revenues to fixed assets, profits to net working capital

Cheng and Hong (2015)	FKNC, SMOTE, FA	ROA, ROE, ROS,quality of inventory,Current ratio, quick ratio, net working capital to total assets, current asset to net assets, total liabilities to net worth, times interest earned, revenues to net working capital, accounts receivable turnover, accounts payable turnover, fixed assets to net worth, turnover of total assets, revenues to fixed assets, profits to net working capital,sales to net worth,retained earnings to sales, debt ratio
Tserng et al. (2015)	LS-SVM, Grey system theory	Current ratio, quick ratio, net working capital to total assets, current assets to net assets, total liabilities to net worth, retained earnings to sales, times interest earned, revenues to net working capital, accounts receivable turnover, accounts payable turnover, sales to net worth, quality of inventory, turnover of total assets, revenues to fixed assets,ROA, ROE, ROS, profits to net working
Jang et al. (2019)	LSTM RNN, SMOTE+Tomek	ROA, ROE, ROS,debt ratio, current ratio, current assets to net assets, working capital turnover,working capital to total asset, total liabilities to net worth, equity turnover, total asset turnover,retained earnings to sales

2.9 Accuracy comparison

Table 11 summarizes the average accuracy generated by previous works and the method utilized in their research. It is noticeable that the average accuracy provided by the models provided in the table increases where artificial intelligence models provide a higher accuracy compared to statistical models. Also, artificial intelligence models that incorporate a memory cell such as the cerebellar model neural networks (CMNN) developed by Chung (2016) and the LSTM neural networks developed by Jang (2019) show relatively higher accuracy with respect to artificial intelligence models, where ANN models developed do not surpass 85.5%.

Table 11 Testing Set average accuracy from literature

Year	Author	Method	Avg. Accuracy
1994	Severson et al.	Discrete Choice Modeling	87.5%
1996	Russel & Zhai	Stepwise Regression	78%
1999	Zhang	ANN	77.27%-84.09%
		Logistic Regression	75%-81.82%
2008	Chen et al.	CBR	88.47
2009	Lam et al.	SVM	83.3-92.3%
2012	Chen	SFNN	85.1%
		HRCNN	80.1%
2013	Adeleye et al.	Z-Score	72%
2014	Cheng et al.	ELSIM	98.9%
		ANN	75.2%
		SVM	92.2%
2016	Chung	CMNN	97.5%
2019	Jang et al.	LSTM	98.2%
		Feedforward NN	85.5%
		SVM	95.6%

2.10 Research gap

Prequalification is vital to ensure that project stakeholders are not affected negatively by contractors' financial distress. The majority of the available literature focuses on the use of financial ratios while limited research applying macroeconomic variables has been conducted. Therefore this research investigates the use of macroeconomic variables by comparing the efficiency of different techniques on two datasets where the first dataset contains financial ratios and macroeconomic variables (FR & MV) the other dataset is composed of financial ratios only (FR). In addition this work introduces a relatively specific variable to the construction market which is the GDP in the construction sector that aids the model in predicting the financial status of construction contractors.

Chapter 3: Data Collection and Sampling

3.1 Data Collection

Financial data are not often readily available, especially on distressed companies. Due to the unavailability of data in Egypt and several other countries; this research was applied to publicly listed construction companies in the US stock exchange. The main concept can be applied to other countries in case data were made available. The data was collected from multiple sources; to begin with, the financial data on non-distressed companies were collected using Thomson Reuters' Eikon for the period 2000-2022. (Thomson Reuters, 2022) While for distressed companies, the United States Securities and Exchange Committee (SEC) in addition to the US census (US Census, 2022) were used to obtain the lists of distressed companies using the 10-K and 10-Q forms which are annually and quarterly reports submitted periodically were obtained for those companies (SEC, 2022).

3.2 Data Sampling

Following data collection, data sampling was required to remove companies with diverse services that are not related to contracting construction projects. An example of the companies that were excluded are design firms and companies that generate revenues from industries that are not related to the contracting scope of this research.

3.3 Representative Sample Size

Based on the statistics provided by the North American Industry Classification System (NAICS, 2022); the number of companies listed in the construction sector in the US were around 1,500,000 as indicated in table 12. The statistical Z-model was used to determine the number that would indicate a representative sample of the companies working in the US on a selected standard error of 10%.

Table 12 Number of construction US Businesses in 2022 (NAICS, 2022)

Codes	Titles	Total Marketable US Businesses
23	Construction	1,514,282
2361	Residential Building Construction	508,034
236115	New Single-Family Housing Construction (except For-Sale Builders)	362,729
236116	New Multifamily Housing Construction (except For-Sale Builders)	30,226
236117	New Housing For-Sale Builders	6,596
236118	Residential Remodelers	108,483
2362	Nonresidential Building Construction	64,022
236210	Industrial Building Construction	7,631
236220	Commercial and Institutional Building Construction	56,391
2371	Utility System Construction	24,667
237110	Water and Sewer Line and Related Structures Construction	17,653
237120	Oil and Gas Pipeline and Related Structures Construction	4,241
237130	Power and Communication Line and Related Structures Construction	2,773
2372	Land Subdivision	41,759
237210	Land Subdivision	41,759
2373	Highway, Street, and Bridge Construction	28,377
237310	Highway, Street, and Bridge Construction	28,377
2379	Other Heavy and Civil Engineering Construction	7,814
237990	Other Heavy and Civil Engineering Construction	7,814

equation 1 represents the statistical formula used to calculate the representative sample size by Slovin (1960) based on the standard error, and population.

The collected sample size accumulated to 108 companies which exceed the minimum number of samples required to represent the population (100 samples).

$$n = \frac{N}{1 + Ne^2} \quad (1)$$

n = sample size

N (population) = 1,514,282

e = standard error (10%)

$$n = \frac{1{,}514{,}282}{1+(1{,}514{,}282*0.1)} = 100\ samples$$

3.4 Data Pre-processing

The collected data could not be processed before examining the overall suitability of the data set in the model. For example, cases that were encountered with missing values in their financial statements were removed from this study to prevent overfitting the results.

3.5 List of Construction Contractors Selected

Table 13 displays the non-distressed companies used in this research and their status as disclosed by the securities and exchange commission (SEC, 2022).

Table 13 Non-distressed contractors

#	Company Name	Status
1	AECOM	Non-distressed
2	Aegion Corp	Non-distressed
3	Alset Ehome International Inc	Non-distressed
4	Ameresco, Inc.	Non-distressed
5	AMERILINK CORP	Non-distressed
6	AMREP Corp	Non-distressed
7	AMREP INC	Non-distressed
8	APi Group Corp	Non-distressed

#	Company Name	Status
9	Arcosa Inc	Non-distressed
10	Argan Inc	Non-distressed
11	AXIOM CORP.	Non-distressed
12	Bishop Capital Corp	Non-distressed
13	China Cgame Inc	Non-distressed
14	Comfort Systems USA Inc	Non-distressed
15	Concrete Pumping Holdings Inc	Non-distressed
16	Construction Partners Inc	Non-distressed
17	DBM Global Inc	Non-distressed
18	Dycom Industries, Inc.	Non-distressed
19	Edd Helms Group Inc	Non-distressed
20	EMCOR Group Inc	Non-distressed
21	ENCOMPASS SERVICES CORP	Non-distressed
22	Energy Services of America Corp.	Non-distressed
23	ENGlobal Corp.	Non-distressed
24	ENTRX CORP	Non-distressed
25	Forestar Group Inc	Non-distressed
26	FTE Networks Inc	Non-distressed
27	FURMANITE CORP	Non-distressed
28	Granite Construction Inc	Non-distressed
29	Great Lakes Dredge & Dock Corp	Non-distressed
30	HOVNANIAN ENTERPRISES INC	Non-distressed
31	Howard Hughes Corp	Non-distressed
32	IES Holdings Inc	Non-distressed
33	INEI CORP	Non-distressed
34	Infrastructure and Energy Alternatives Inc	Non-distressed
35	Innovate Corp	Non-distressed
36	Installed Building Products Inc	Non-distressed
37	IREX CORP	Non-distressed
38	Jacobs Engineering Group Inc	Non-distressed
39	KBR Inc	Non-distressed
40	LianDi Clean Technology Inc	Non-distressed
41	Limbach Holdings Inc	Non-distressed
42	MasTec Inc. (FL)	Non-distressed
43	Matrix Service Co.	Non-distressed
44	MYR Group Inc	Non-distressed
45	Orion Group Holdings Inc	Non-distressed
46	Pernix Group Inc	Non-distressed

#	Company Name	Status
47	Primoris Services Corp	Non-distressed
48	Quanta Services Inc	Non-distressed
49	Reliant Holdings Inc	Non-distressed
50	SERVIDYNE, INC.	Non-distressed
51	Sterling Construction Company Inc	Non-distressed
52	TopBuild Corp	Non-distressed
53	Tutor Perini Corp	Non-distressed
54	UNITED HOMES INC	Non-distressed
55	UTILX CORP	Non-distressed
56	Valmont Industries Inc	Non-distressed
57	Westower Corp	Non-distressed
58	Williams Industrial Services Group Inc	Non-distressed

Table 14 displays the distressed companies used in this research and their status as disclosed by the securities and exchange commission (SEC, 2022).

Table 14 Distressed Contractors

#	Company Name	Status
1	AFFORDABLE GREEN HOMES INTERNATIONAL	Permanently Revoked
2	ALL AMERICAN GROUP INC	Acquired
3	Alternate Energy Holdings, Inc.	Inactive
4	AMERICAN INTERNATIONAL CONSOLIDATED, INC.	Forfeited
5	AQUENTIUM, INC. AND SUBSIDIARIES	Revoked
6	Arguss Communications Inc	Acquired
7	Atkinson (Guy F.) Co. of California	Failed
8	Axiom Corp. and Subsidiary	Inactive
9	CAPITAL GROUP ONE INC	Dissolved
10	Canadian Rockport Homes Int'l, Inc.	Revoked
11	China Housing & Land Development, Inc	Dissolved
12	CIAO GROUP INC.	Inactive
13	CONTEMPRI HOMES INC	Dissolved
14	Corrpro Cos., Inc.	Acquired
15	DAHUA INC	Acquired
16	DAW Technologies Inc.	Failed
17	Dominion Bridge Corp.	Failed

#	Company Name	Status
18	ELINE ENTERTAINMENT GROUP INC	Acquired
19	EMCON	Acquired
20	EXTENSIONS, INC	Inactive
21	Firemans Contractors, Inc.	Operational
22	Flour City International Inc.	Failed
23	FORTRESS GROUP INC	Acquired
24	GLOBAL DIVERSIFIED INDUSTRIES, INC.	Active
25	Golden Autumn Holdings Inc.	revoked
26	GOLDEN OPPORTUNITY DEVELOPMENT CORP	Permanently Revoked
27	HATHAWAY CORP	Revoked
28	HEARTLAND, INC	Dissolved
29	InfraSource Services Inc	Acquired
30	Lifestyle Innovations, Inc. and Subsidiaries	Dissolved
31	MORGAN COOPER, INC.	Merger
32	MORRISON KNUDSEN CORP	Acquired
33	Northtech Industries Inc	Dissolved
34	OmniAmerica Inc.	Acquired
35	Orleans Home Builders	Failed
36	PORTER MCLEOD NATIONAL RETAIL INC	Failed
37	PREMIER PACIFIC CONSTRUCTION, INC	Dissolved
38	PROSPECT GLOBAL RESOURCES INC	Acquired
39	REALCO INC	Forfeited
40	Schuff International, Inc.	Failed
41	SERVIDYNE, INC	Acquired
42	Sprout Tiny Homes, Inc	Failed & Acquired
43	Stone & Webster Inc.	Acquired
44	THE MAJESTIC COMPANIES, LTD.	Dissolved
45	Turner Corp.	Acquired
46	UpSnap, Inc.	Failed
47	USA Bridge Construction of New York Inc.	Failed
48	USABG Corp.	Failed
49	VRDT Corp	Acquired
50	WHITEHALL LTD INC	Failed

Although the total number of operational companies is much lower than the distressed companies; the number of data points generated from some of the operational companies spans over 17 years (2000-2016) whereas the distressed companies generate only 1 data point prior to the distress year.

3.6 Economic Indices Selected

Table 15 shows the selected economic indicators to be studied in this research; the selection of these variables is based on the previously studied variables by Russel (1996) and Jang (2019). In addition to the variables mentioned in the literature, the gross domestic product by industry (construction) is utilized in this research which was selected to highlight the changes in the construction field that might not be correlated with the other indicators.

Table 15 Macroeconomic Input Variables

	Macroeconomic Variable	**Abbreviation**
1	Average Prime Interest Rate (Russel 1996 & Jang 2019)	IR
2	Consumer Price Index (Russel 1996 & Jang 2019)	CPI
3	Gross Domestic Product by Industry (Construction)	GDPC
4	Gross National Product (Russel 1996)	GNP
5	Total Business Failures (Russel 1996 & Jang 2019)	TBF
6	Total Employees (Russel 1996 & Jang 2019)	TE

3.7 Descriptive Statistics

This section represents the statistics for the input variables of both financially distressed and non-distressed groups; the mean, standard deviation, minimum and maximum values. Table 16 indicates that some of the ratios have high variance and this is mainly attributed to the distressed companies as for the working capital to total revenues

(WCTR) ratio, the revenues could sometimes be much lower than the working capital and as for the retained earnings to total assets (RETA) ratio, the value of total assets for some of the distressed companies is very low compared to operational companies.

The negative values are also due to negative retained earnings for the RETA ratio, current liabilities that are greater than current assets, as well as the earnings before interest and taxes (EBIT), which is susceptible to showing negative earnings from the income statement.

Table 16 Descriptive Statistics for financial ratios

Ratio Name	Minimum	Maximum	Mean	St Dev.
Working Capital to Total Revenues	-133.44	19.80	-0.42	9.409794957
Retained Earnings - Total to Total Assets	-927.12	0.82	-8.18	73.97005768
EBIT to Total Assets	-81.56	0.48	-0.89	6.582507066
Asset Turnover	-0.43	13.77	1.63	1.315339852
Total Debt Percentage of Total Assets	0.00	599.51	1.63	1.315339852
Current Ratio	0.00	48.01	1.96	4.146696533
EBIT to Total Revenues	-81.56	0.48	-0.89	6.582507066
Debt to Equity	-59.95	292.62	2.79	23.91912207

Table 17 displays statistics for the company's macroeconomic variables instead of financial ratios; it is noticeable that the values of the total business failures and the total number of employees are greater than the other variables since they represent several instances while the other variables are indicators of the economy's performance.

Table 17 Descriptive statistics for macroeconomic variables

	Economic Factor	Minimum	Maximum	Mean	St Dev.
1	IR	0.03	0.08	0.05	0.02
2	CPI	152.4	271	208.75	35.17
3	GDPC	3.3	5.05	4.14	0.51
4	GNI	7732	21640.51	14043.22	4101.22
6	TBF	374,363	503,996	419035.2	32858.13
7	TE	99,215,837	131,788,812	11,5434,245	8,302,157

Figure 13 shows the distribution of the different financial ratios with respect to the minimum, maximum, lower quartile, upper quartile, and the median bisecting the former, while the scale is very large to account for the outlier values provided, where some of the outliers were unaccounted for with respect to the values provided from the table. The upper and lower quartiles are not observed which ensures that the majority of the data points lie within the marked "X" values in the diagram.

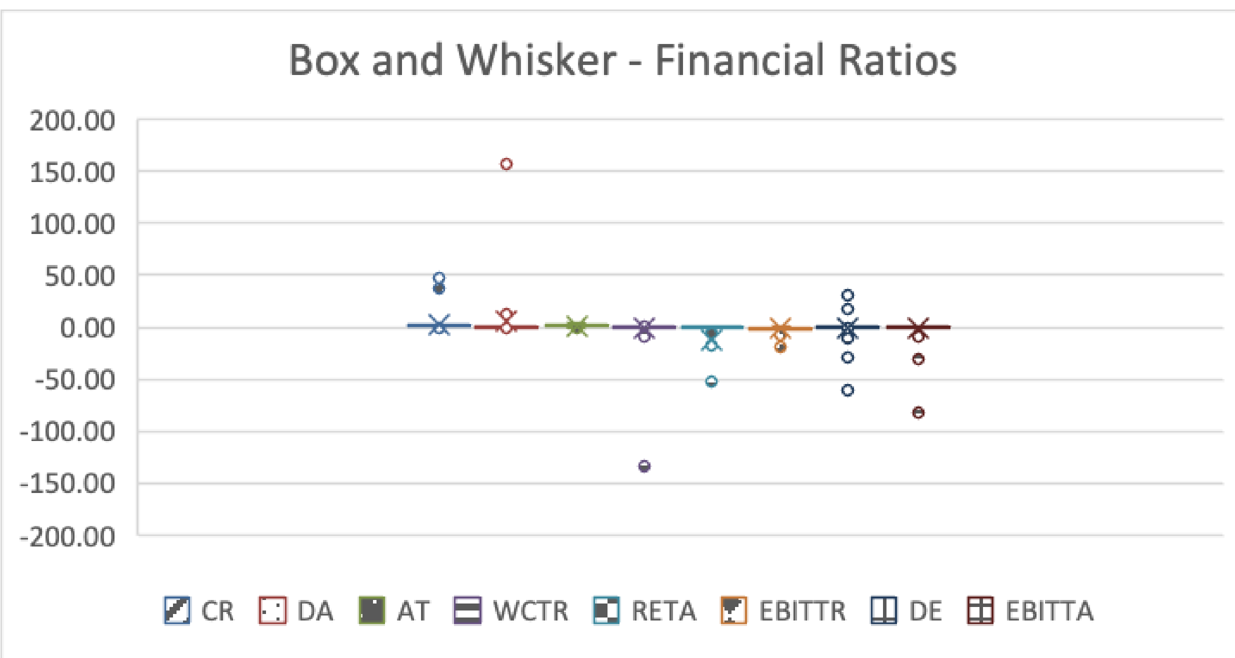


Figure 13 Box & whisker diagram for financial ratios

3.8 Data scaling

Since input variables have significant differences in the value ranges; scaling were required to correctly input data to the model. In other words, setting each variable's mean to a zero value and a unit variance. The following equation shown in eq. 2 was used for scaling (Clavel, 2019).

$$x' = \frac{x - \bar{x}}{\sigma} \quad (2)$$

3.9 Correlation between variables:

Since this research employs a logistic regression model as one of the predictive models, logistic regression models are susceptible to overfitting problems from multicollinear variables. The input variables must be independent of each other, which is calculated through multicollinearity that results in values ranging from -1 to 1; where values approaching an absolute value of 1 or -1 show high correlation while values approaching 0 show minimal correlation. Correlation tests were performed for financial ratios and macroeconomic variables to identify highly correlated variables and remove them from the model inputs. Table 18 shows a high correlation between GNP with CPI and GNP with TEC, therefore it is recommended to remove GNP from the input variables which shall improve the model's prediction efficiency.

Table 18 Correlation between macroeconomic variables

	IR	*CPI*	*GDPC*	*GNP*	*TBF*	*TE*
IR	1.000					
CPI	-0.751	1.000				
GDPC	0.368	-0.221	1.000			
GNP	-0.694	0.992	-0.149	1.000		
TBF	0.093	-0.014	0.349	-0.010	1.000	
TE	-0.489	0.875	0.205	0.909	0.054	1.000

Table 19 is another correlation matrix between financial ratios which shows that EBIT to total assets ratio is highly collinear with total debt to total assets and retained earnings to total assets; this might be due to the shared denominator between the three ratios; accordingly, to reduce collinearity, it is recommended to remove EBIT to total assets ratio from the input variables to improve the validation process and increase the prediction accuracy.

Table 19 Correlation between financial ratios

	CR	*DA*	*AT*	*WCTR*	*RETA*	*EBITTR*	*DE*	*EBITTA*
CR	1							
DA	-0.043	1						
AT	0.189	0.133	1					
WCTR	0.069	-0.954	-0.171	1				
RETA	0.047	-0.965	-0.129	0.857	1			
EBITTR	-0.111	-0.506	0.007	0.425	0.457	1		
DE	0.018	-0.013	-0.049	0.013	0.017	0.033	1	
EBITTA	0.060	-0.919	-0.240	0.850	0.934	0.499	0.022	1

3.10 Significance of input variables

Linear regression determines the significance of multiple or single variables with the output; this method was used to calculate the significance value through the F-test; which concludes that the lower the significance F; the greater the significance of the variable. Table 20 contemplates that the consumer price index is a significant factor to determine the status of the contracting company based on the value of the significance F which is lower than 5% and the plot shows the minimal error between the predicted values and the actual values.

Table 20 ANOVA table for CPI variable

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1.0E+00	1.1E+02	1.1E+02	4.5E+02	1.8E-54
Residual	2.2E+02	5.4E+01	2.5E-01		
Total	2.2E+02	1.7E+02			

Table 21 also shows the ANOVA analysis for asset turnover ratio (AT) where the significance F value is less than 5% meaning that this variable is highly significant to the output.

Table 21 ANOVA table for AT variable

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	87.5237827	87.5237827	237.871627	1.2389E-36
Residual	216	79.4762173	0.36794545		
Total	217	167			

3.11 Summary

This chapter provides the approach followed to prepare data for model input, starting by collection from various sources, then the collected data were filtered for outliers such as companies with major income not from construction activities. The representative sample was selected through normal distributions Z-score. Afterward, pre-processing the data occurred by making sure that both datasets (operational and distressed) are spanning over the same period and no remaining outliers from the previous stages are left transforming the collected raw data into data points. The following step was to calculate the descriptive statistics for each class of input variables showing the minimum, maximum, mean, and standard deviation which were illustrated in tables 17 and 18. Since the macroeconomic variables are different in units, feature scaling was required to make sure that all input variables are equally important to the model before initializing the model. Finally, multicollinearity and significance tests are performed to indicate which input variables are highly correlated with other input variables and which variables are not significant towards the output; those variables are removed from the model to improve the training of the model and prevent over-fitting.

Chapter 4: Model Development

This chapter shows the model's development phase and illustrates the approach followed to generate the results represented. Several models are developed to compare and select the highest performance measures, the models include (ANN, SVM, and logistic regression). The ANN is developed using python and NeuralTools while the SVM and logistic regression are performed using python. In terms of the datasets, two datasets were used to determine the effect of macroeconomic variables on the model, the first dataset included financial ratios and macroeconomic variables (FR & MV) while the second dataset included the financial ratios only.

4.1 Artificial Neural Networks (ANN)

4.1.1 Training Data Set and Testing Data Set

Each dataset containing input variables and actual output is divided into a ratio of 4:1 training to testing respectively, splitting a total of 217 points into 162 training data points and 55 testing data points.

4.1.2 ANN Architecture

Figure 14 displays one of the early attempts to run the model with 12 input variables including financial ratios and macroeconomic variables with two hidden layers with 8 and 4 neurons respectively and an output layer which consists of only 1 neuron which corresponds to the status (whether 1 or 0). In later trials, the number of neurons in the hidden layers were changed while the optimum number of hidden layers turned out to be 2 layers as the optimum scenario. This was achieved by trial and error through changing the numbers with different runs to the model and selecting the model parameters that generate the highest accuracy.

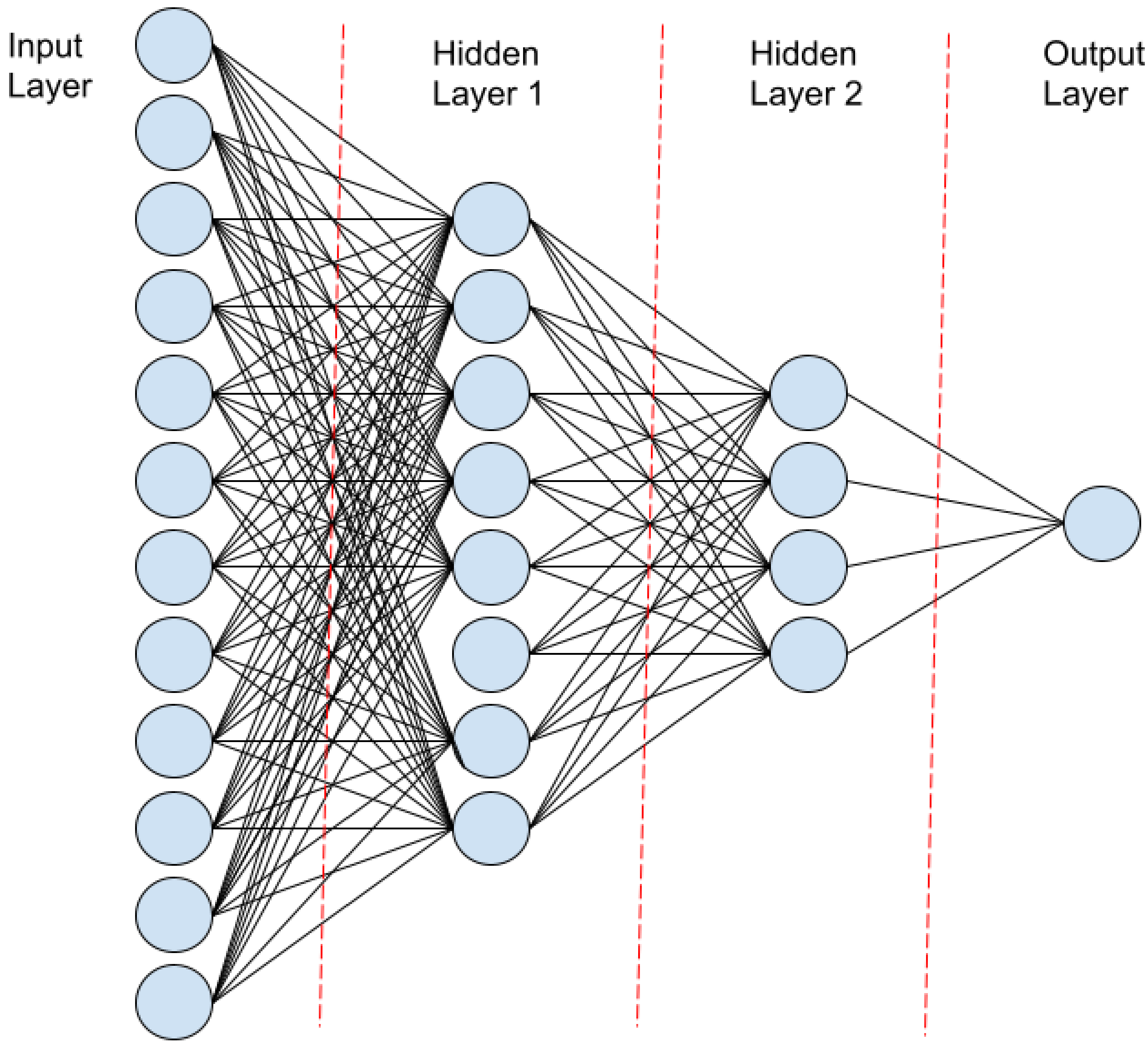


Figure 14 Sample ANN Architecture

4.1.3 Activation Function

Since the model consists of a categoric output, an activation function of threshold or sigmoid between the last hidden layer and the output layer would be preferred to generate a prediction of a 0 or 1 denoting the status of the construction company (1 being operational and 0 being distressed). The Softmax functions were tested but they yielded lower accuracy than the sigmoid activation function. While for the activation function utilized for transferring the values from the input layer to the 1st hidden layer and the following hidden layers till the output layer. Any activation function could be used throughout the network such as: rectifier, tanh, sigmoid and linear. The most efficient function that resulted in the highest accuracy was the rectifier function.

4.1.4 Error Adjustment

Neural networks calculate the cost function according to the error between the predicted and the actual values; to update the model and improve the prediction accuracy; the weights need to be updated through backpropagation. Backpropagation is selected for its highly efficient training in complex multi-layered neural networks (Sapna et al, 2012). As for the selected minimization method; the most commonly used is the gradient descent function which calculates the slope within the cost function and chooses to update the weights according to the largest value of negative slope. A drawback of this method is that it is susceptible to being trapped in a local minimum. Other minimization functions may avoid this issue based on stochastic nature measures such as the stochastic gradient descent function which generates random values for the cost function and updates based on the error calculated between the predicted and the actual output.

4.1.5 Confusion Matrix

The output of the model generated by python is a print for the predicted test set and the actual values of the test set. To determine the prediction accuracy of the model; a confusion matrix is produced as the following output shown in figure 15 which is a sample of an earlier developed stage of the model. Based on the results provided in figure 15 the first line represents the distressed group including a total of 10 cases tested where 6 cases were predicted correctly while the remaining cases were incorrectly predicted. The second line represents the non-distressed cases summed up to a total of 118 and all of these cases were correctly predicted by the model.

[[6 4]

[0 118]]

Figure 15 ANN confusion matrix extract

4.2 Support Vector Machines

In this research the author investigated employing RBF and linear kernel models in predicting financial distress and compared the results against the ANN and logistic regression models. Figure 16 shows a form of confusion matrix as a chart where the top left and bottom right are the correct predictions while the other two cells represent the misclassification either false positive (bottom left cell) or false negative (top right cell).

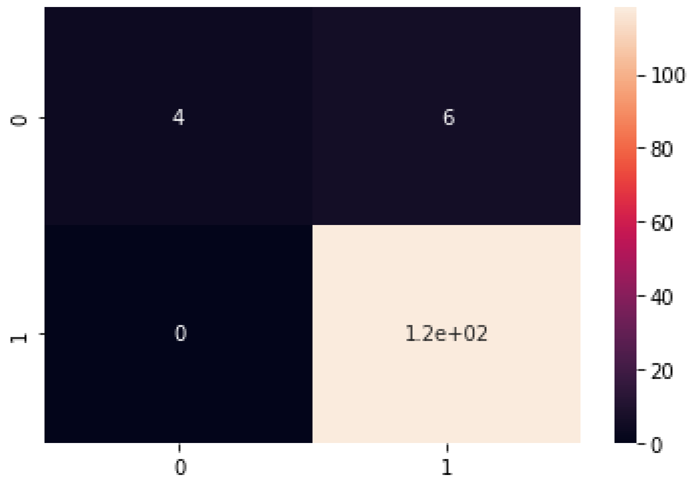


Figure 16 heat map confusion matrix

4.3 Logistic Regression

Logistic regression is also compared to ANN & SVM. The model is developed by selecting the same training to testing ratio used in ANN & SVM.

The code shown in figure 17 shows a sample of the code used in the development phase of the model to run the logistic algorithm model and print the predicted values next to the actual values.

```
from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression(random_state = 0)
classifier.fit(X_train, y_train)
y_pred = classifier.predict(X_test)
print(np.concatenate((y_pred.reshape(len(y_pred),1),
y_test.reshape(len(y_test),1)),1))
```

Figure 17 Sample of python code

4.4 ANN using NeuralTools

NeuralTools is an excel add-in developed by Palisade (Palisade, 2022). NeuralTools utilizes artificial neural networks to predict dependent variables. The user starts by inputting independent and dependent variables. Then, training and testing parameters are selected by the user and training to testing ratio and the training algorithm is set. The user then selects the stoppage criteria based on the desired error percentage achieved, model time or number of trials. Afterwards, the model starts training and testing producing statistical reports and confusion matrices for the training and testing data.

Chapter 5: Discussion and Analysis

This chapter represents the results produced from this research and compares the outputs generated by the developed models (ANN, SVM & LR). The effect of macroeconomic variables is also tested by applying two datasets on the model dataset 1 (FR & MV) and dataset 2 (FR only).

5.1 Research Overview

5.1.1 Python models

Based on the proposed literature, neural networks and SVM are the most common machine learning techniques used to obtain high prediction accuracy. This work represents developing ANN, SVM and LR models to identify the construction companies that might be subject to distress and those which will remain operational the techniques used employ deep learning, artificial, machine learning, and basic linear methods. Eq. 3 shows the equation utilized by Jang (2019) to calculate accuracy and Eq. 4 displays the F1 score, which is the harmonic mean of precision and recall which measures the effect of false predictions with respect to the true positive values (Jang, 2019).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \qquad (3)$$

$$\text{F1} - \text{score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \qquad (4)$$

TP = True positive - FP= False positive
TN= True negative - FP= False positive

The confusion matrices for the python developed models are represented in table 22 for dataset 1 (FR & MV). ANN produced the highest accuracy (96.88%) exceeding all other models. Followed by the linear SVM kernel function and the linear regression which both generated (95.3%). The RBF SVM kernel function yielded the lowest accuracy (94.53%). Table 23 shows the developed models on dataset 2 (FR only) showing similar trends in contrast to dataset 1 (FR & MV) but the performance measures are lower with respect to dataset 1.

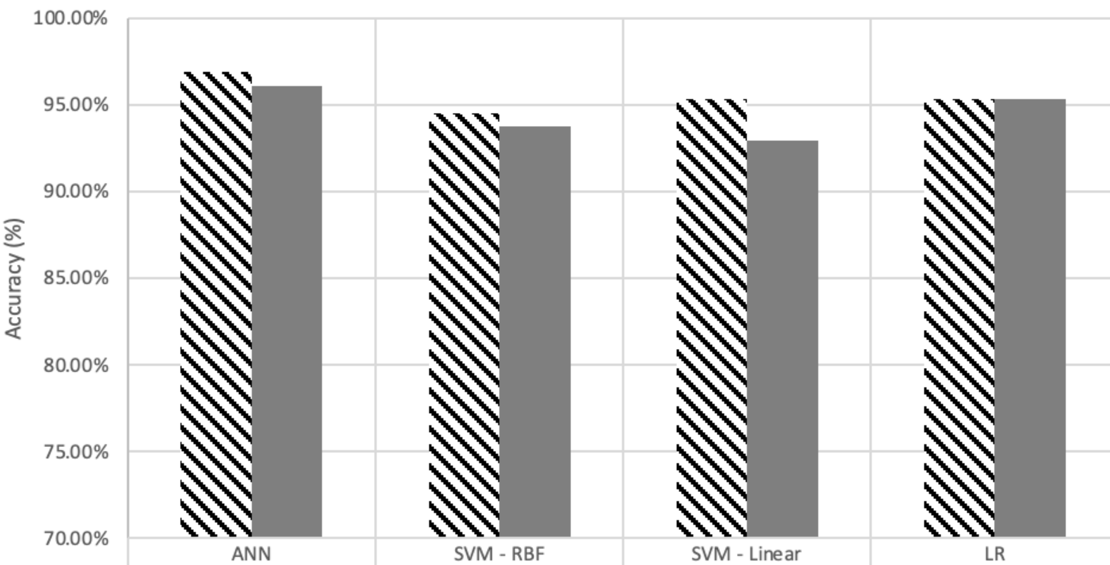
Table 22 Confusion Matrix for Python models using FR & MV (Dataset 1)

ANN		**SVM - RBF**		**SVM - Linear**		**LR**	
6	4	3	7	4	6	4	6
0	118	0	118	0	118	0	118
Total	128	Total	128	Total	128	Total	128
Accuracy	**96.88%**	Accuracy	**94.53%**	Accuracy	**95.31%**	Accuracy	**95.31%**
F1 Score	**98.33%**	F1 Score	**97.12%**	F1 Score	**97.52%**	F1 Score	**97.52%**

Table 23 Confusion Matrix for Python models using FR (Dataset 2)

ANN		**SVM - RBF**		**SVM - Linear**		**LR**	
5	5	2	8	1	9	4	6
0	118	0	118	0	118	0	118
Total	128	Total	128	Total	128	Total	128
Accuracy	**96.09%**	Accuracy	**93.75%**	Accuracy	**92.97%**	Accuracy	**95.31%**
F1 Score	**97.93%**	F1 Score	**96.72%**	F1 Score	**96.33%**	F1 Score	**97.52%**

It is evident from the decrease in accuracy for all models except the LR between table 22 and table 23 that macroeconomic variables improve prediction accuracy and their significance is highly correlated to the status of the contracting companies. This observation shall alarm stakeholders with speculation of an approaching macroeconomic crisis which is common for the construction markets across all countries. The ANN model yielded the highest accuracy and F1-score throughout both datasets 1 & 2. Followed by SVM-linear and LR models which generated similar results in both datasets and the least performing model SVM-rbf in terms of accuracy and F1-score. Figure 18 displays this difference between all models with respect to the datasets that illustrate the significance of the macroeconomic variables and the differences between model outputs.



	ANN	SVM - RBF	SVM - Linear	LR
Dataset 1	96.88%	94.53%	95.31%	95.31%
Dataset 2	96.09%	93.75%	92.97%	95.31%

Figure 18 Accuracy Comparison between Python models for both datasets

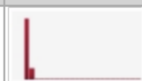
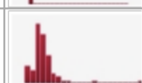
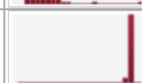
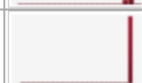
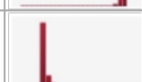
5.1.2 NeuralTools ANN model

This section investigates the performance of NeuralTools program to predict financial distress using both datasets 1 & 2. The program generates descriptive statistics and variable impacts before training and testing the model.

5.1.2.1 Statistical Analysis

NeuralTools implements statistical analysis on the data sets. Table 24 below represents the descriptive statistics developed in the data collection chapter as well as a 95% confidence interval for the values of each variable. Figure 19 illustrates the statistical data for the macroeconomic variables.

Table 24 NeuralTools Descriptive Statistics output (FR)

Name	Graph	Minimum	Maximum	Mean	Std. Deviation	5%	95%	Count
CR		0.000	48.006	1.937	4.062	0.133	2.794	217
DA		0.00	599.51	4.00	41.99	0.00	1.31	217
AT		-0.433	13.774	1.619	1.306	0.000	3.295	217
WCTR		-133.44	19.80	-0.398	9.21	-0.232	0.538	217
RETA		-927.12	0.816	-7.80	72.26	-5.69	0.439	217
EBITTR		-23.213	1.000	-0.342	2.344	-0.689	0.169	217
DE		-59.95	292.62	2.54	22.86	-1.19	5.86	217
EBITTA		-81.559	0.502	-0.845	6.447	-1.051	0.204	217

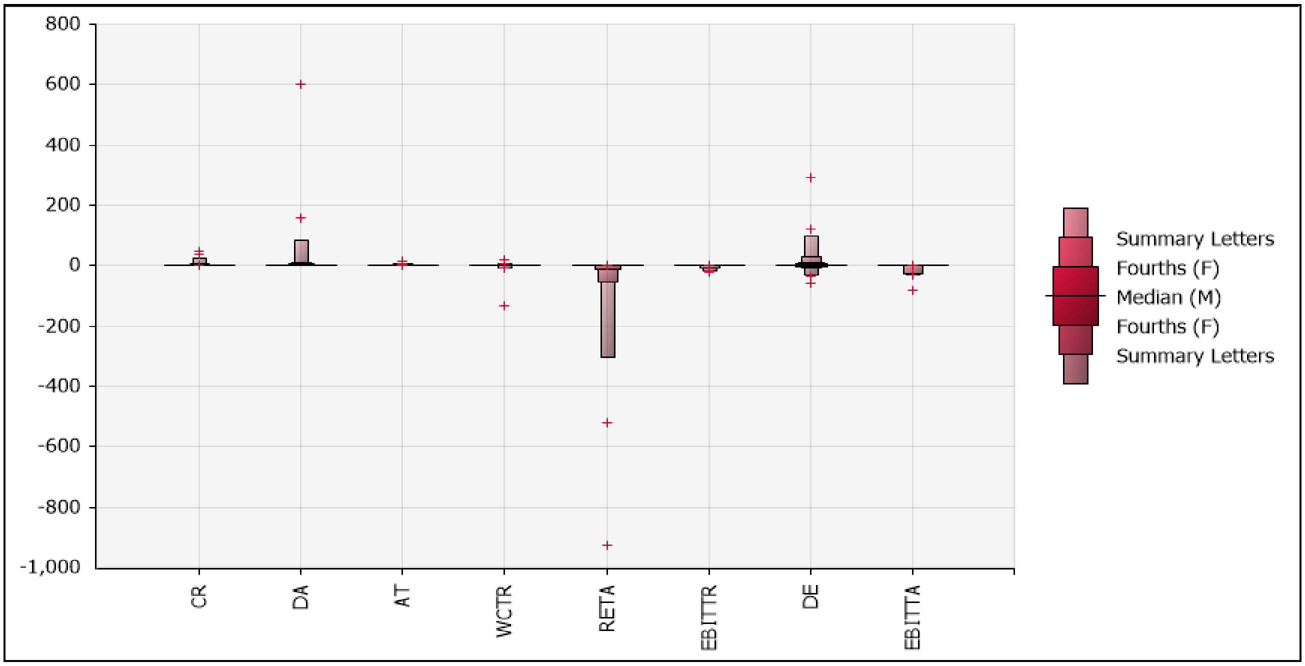


Figure 19 NeuralTools Bar & Whisker diagram

The results shown in figure 19 illustrate high variances in the DA, RETA, and DE variables which are mainly attributable to the distressed data set containing high debt values with respect to equity and assets. Where DA, DE, and RETA reached a value of 599.5, 292.62, and -917.12 respectively as shown in table 24.

5.1.2.2 Variable Impact

figure 20 tackles an important aspect which is the cause and effect relationship between the independent variables and the dependent variable. The figure shows that EBITTA, which is the EBIT to total assets ratio, has the greatest relative impact on the output followed by the debt to equity ratio. As for the macroeconomic variables, the total number of employees has the 5th highest impact followed by GDP in construction as the 7th highest impact.

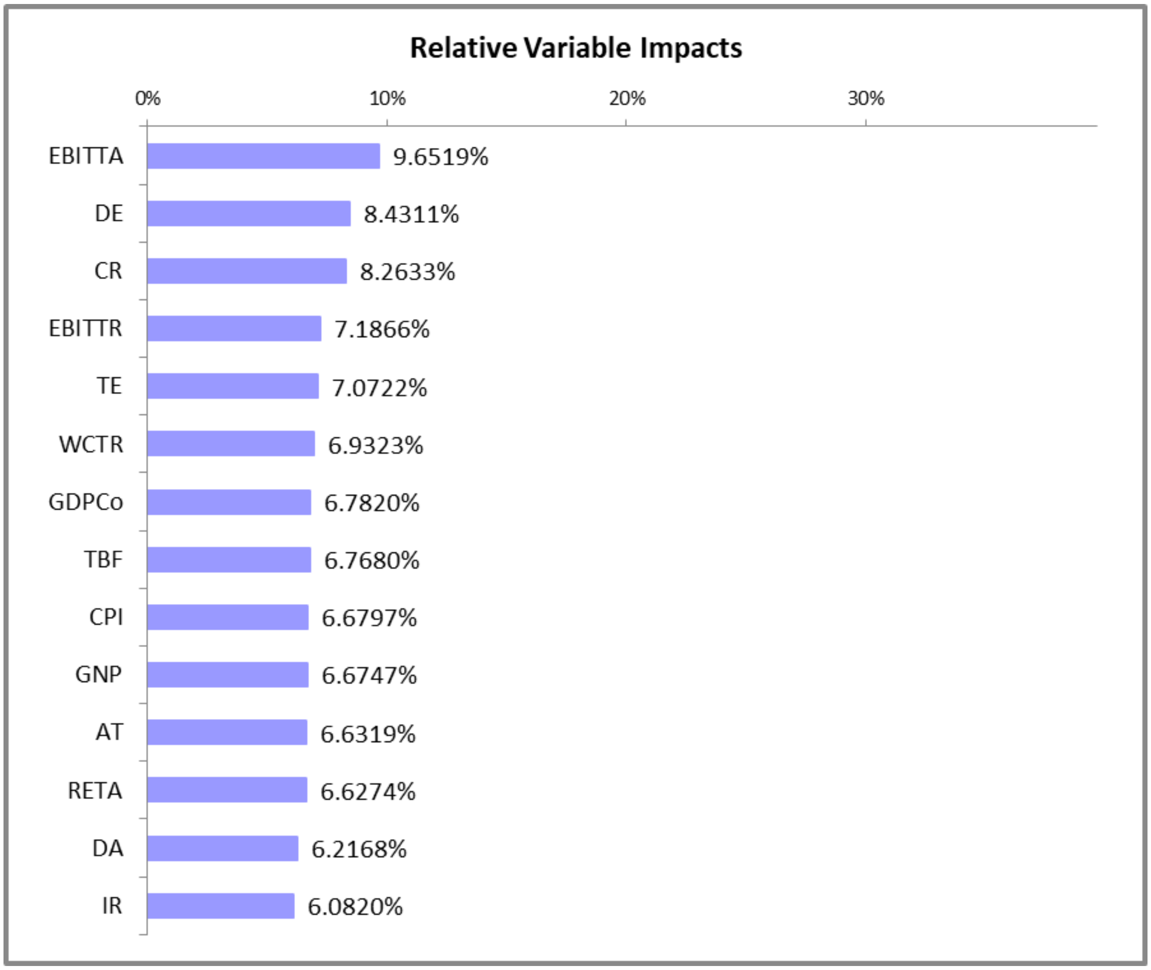


Figure 20 Relative Variable Impacts

5.1.2.3 ANN Model Output

Table 25 shows the performance measures of the model based on the confusion matrix provided accuracy of the prediction model resulted in 90.55% and F1-score of 94.95%

Table 25 Performance measures for testing cases for dataset 1

Classification Performance Measures (for testing cases)	
	Performance Measure
Accuracy	90.5512%
Precision	91.8699%
Recall	98.2609%
F1 Score	94.9580%

Table 26 shows the testing confusion matrix produced by the program using dataset 1, with a total of 43 testing cases, 12 cases were predicted to be distressed, and 2 of them were actually distressed, while for the non-distressed category, 115 cases were predicted to be non-distressed, and 113 cases were actually non-distressed.

Table 26 Confusion Matrix for testing cases for dataset 1

Classification Matrix (for testing cases)		
	0	**1**
0	2	10
1	2	113

Table 27 shows the performance measures generated using dataset 2, it is evident that dataset 1 has generated higher accuracy, precision, recall, and F1-score compared to dataset 2, this trend also follows the same output generated by the ANN developed using python.

Table 27 Performance measures for testing cases for dataset 2

Classification Performance Measures (for testing cases)	
	Performance Measure
Accuracy	90.5512%
Precision	91.0569%
Recall	99.1150%
F1 Score	94.9153%

Table 28 displays the testing confusion matrix using dataset 2, the output shows a higher percentage of bad predictions since out of 14 distressed cases only 3 predictions were correctly predicted and for the non-distressed cases, 26 out of 30 cases are correctly predicted.

Table 28 Confusion Matrix for testing cases for dataset 2

Classification Matrix (for testing cases)		
	0	**1**
0	3	11
1	1	112

Figure 21 shows the difference between the accuracy prediction for dataset 1 (including FR & MV) and dataset 2 (including FR only). The results provided show high correlation to the python developed models whereas adding macroeconomic variables increases the prediction performance of the model.

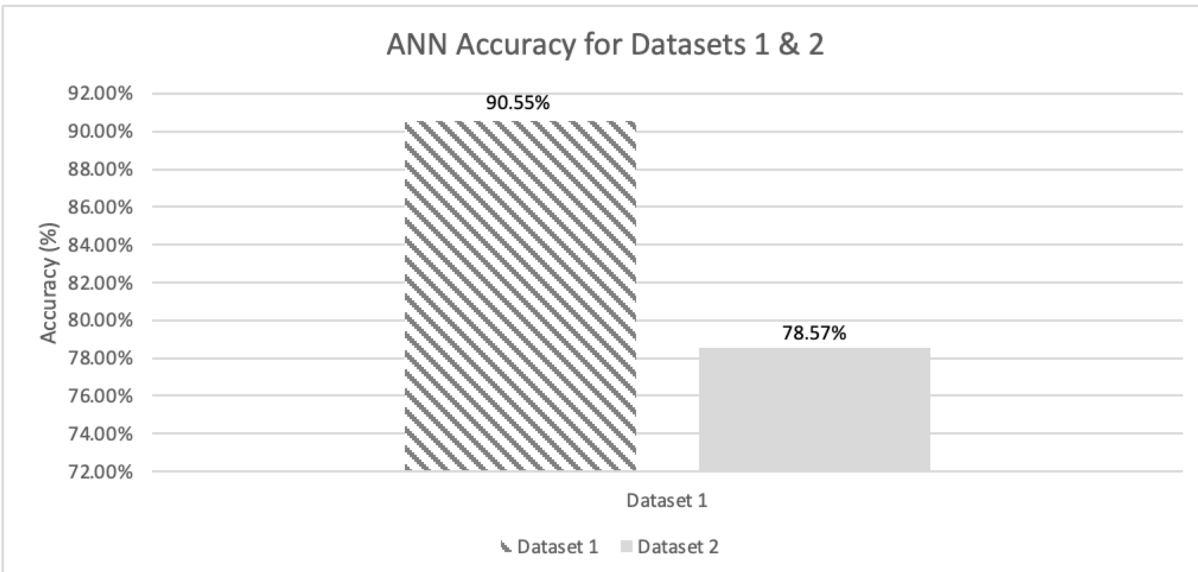


Figure 21 ANN NeuralTools Dataset comparison

5.1.3 Validation Dataset

According to the results provided by the techniques developed, the highest performing technique (ANN) was validated using a different dataset. Table 29 summarizes the results provided by the validation dataset where the output is represented by the probability that the companies are non-distressed.

Table 29 Validation dataset results

	Company Name	Year	Actual Status[a]	ANN Output
1	Granite Construction Inc.	2019	1	0.995
2	Granite Construction Inc.	2020	1	0.992
3	Granite Construction Inc.	2021	1	0.996
4	Tutor Perini Corp	2019	1	0.953
5	Tutor Perini Corp	2020	1	0.740
6	Tutor Perini Corp	2021	1	0.844
7	EMCOR Group Inc	2019	1	0.987
8	EMCOR Group Inc	2020	1	0.974
9	EMCOR Group Inc	2021	1	0.945
10	Argan Inc	2019	1	0.811
11	Argan Inc	2020	1	0.945
12	Kingfish Holding	2022	0	0 (approx.)
13	McDERMOTT INTERNATIONAL, INC.	2020	0	$6.89 * 10^{-7}$
14	SPORTS FIELD HOLDINGS, INC.	2019	0	$1.95 * 10^{-23}$

[a] 1 denotes the company is non-distressed and 0 denotes the company is distressed

The results presented in table 29 show an accuracy of 100% as all eleven non-distressed companies produced an output exceeding 50% with a minimum value of 74%. While for the distressed companies, all three companies produced a probability below 50% with a maximum value below 1%.

Chapter 6: Conclusion and Recommendations

6.1 Conclusion

In this research, a contractor prequalification model was developed that predicts company distress using multiple models. Eight financial ratios and six macroeconomic variables were considered that are significant to the model's output. Several techniques were investigated, and the ANN model proved to be the highest accuracy and F1-score. The research employed the techniques on two datasets: a dataset containing financial ratios and macroeconomic variables (dataset 1) and another dataset consisting of financial ratios only (dataset 2).

The results also show that the ANN model performed using python achieved a better prediction in both accuracy and F1-score with respect to the ANN performed using NeuralTools.cWith regards to datasets, the performance measures (accuracy & F1-score) generated by dataset 1 achieve better accuracy and F1- score in comparison to model trials using dataset 2 only, which indicate the significance and impact of macroeconomic variables on the status of the contracting company.

6.2 Research Contribution

This work contributes to the available body of knowledge in the prequalification area by implementing and comparing different techniques and different types of input datasets. The analysis provided in the research tackles an important aspect by employing different datasets. The model could be utilized to predict the future performance of a construction company and whether it is going to be operational or subject to financial distress.

Through multiple trials, the optimum architecture utilized to produce the highest accuracy was achieved using 2 hidden layers with rectifier activation functions in the hidden layers and a sigmoid activation function for the output layer.

The model also incorporates a new input variable which is the GDP for the construction sector which improves the predictive accuracy of the model since it is significant towards the status of the contractor.

This work also compares the accuracy produced by developed software such as NeuralTools vs developed python codes and the chosen parameters in each model. The results presented in this matter conclude that using the python developed code is more efficient as it generates a higher accuracy and F1-score with respect to the NeuralTools model.

6.3 Recommendations for future research

This work provides a framework for determining an important parameter in the construction industry; pre-qualification is vital for construction stakeholders to anticipate and avoid any predicted problems that might arise from unfit contractors.

Adding more input cases as ANN models require a large sample of data points to improve accuracy. Where the larger number of samples increases the accuracy prediction of neural networks models (W.Ng, 2019).

Investigating the use of other classification techniques to check whether they would result in improved performance measures such as random forest models, K nearest neighbor (KNN), Naive Bayes, and Decision Trees.

Further study for the variables that impact the status of construction companies is also crucial to improve the modeling accuracy. By introducing more significant variables the confidence in the models' outputs increases.

References

Adeleye, T., Huang, M., Huang, Z., and Sun, L. (2013). “Predicting Loss for Large Construction Companies.”Journal of Construction Engineering and Management, Vol.139 No.9, pp.1224–1236.

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy. The Journal of Finance, 23(4), 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x

Aggrawal, Charu (2018). Neural Networks and Deep Learning. Springer. ISBN 978-3-319-94463-0 (eBook) https://doi.org/10.1007/978-3-319-94463-0.

Altman, E. I., Haldeman, R. G., and Narayanan, P. (1977). "Zeta analysis: a new model to identify bankruptcy risk of corporations." J. of Banking and Finance, (June), 29-54.

Arditi, D., Koksal, A., and Kale, S. (2000). “Business failures in the construction industry.” Engineering, Construction and Architectural Management, 7(2), 120–132.

Al-Sobiei, O. S., D. Arditi, and G. Polat. 2005. “Predicting the risk of contractor default in Saudi Arabia utilizing artificial neural network (ANN) and genetic algorithm (GA) techniques.” Constr. Manage. Econ. 23 (4): 423–430. https://doi.org/10.1080/01446190500041578.

Beaver, W. H. (1966). "Financial ratios as predictors of failure." Empirical Res. in Accounting: Selected Studies, Univ. of Chicago, Ill., 77- Ill.

Bureau, U. S. (2013, 12 30). US Census Data. Retrieved from US Census Data: https://www.census.gov/

Chen, J. H. (2012). “Developing SFNN models to predict financial distress of construction companies.” Expert Systems with Applications, 39(1), 823–827.

Cheng, M. Y., N. D. Hoang, L. Limanto, and Y. W. Wu. 2014. “A noble hybrid intelligent approach for contractor default status prediction.”Knowl.-Based Syst. 71 (Nov): 314–321. https://doi.org/10.1016/j.knosys.2014.08.009.

Chung, Chang-Chih et al (2016). “Bankruptcy Prediction using Cerebellar Model Neural Networks”. Springer. https://doi.org/10.1007/ s40815-015-0121-5

Clavel, Fred (2019). “Basics: Standardization and the Z score” .https://fredclavel.org/2019/03/18/basics-standardization-and-the-z-score/

Doumpos, M., & Zopoudinis, C. (1999). “A multicriteria discriminant method for the prediction of financial distress: the case of Greece”. Multinational Finance Journal.

Holt, G. D., et al. 1994. “Evaluating prequalification criteria in contractor selection.” Build. Environ., 294, 437–448.

“How SVM works”. IBM, 17 Aug 2021. https://www.ibm.com/docs/it/spss-modeler/SaaS?topic=models-how-svm-works

Hosny, Ossama et al (2013). Prequalification of Egyptian construction contractors using fuzzy-AHP models”. Engineering, Construction, and Architectural Management. Emerald. DOI 10.1108/ECAM-09-2011-0088

Ioffe, Sergey; Christian Szegedy (2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". arXiv:1502.03167 [cs.LG].

Ismail, Ihab. “Financial Cash Flow Determinants of Company Failure in the Construction Industry”. PhD, The University of Michigan, 2014.

Jang, Y., et al. 2019. “Predicting Business Failure of Construction Contractors Using Long Short-Term Memory Recurrent Neural Network”. ASCE. https://doi.org/10.1061/j.1943-7862.0001709

Kolmar, Chris. “These Are the 10 Biggest Construction Companies in America.” *Zippia*, https://www.zippia.com/advice/biggest-construction-companies-in-america

Lam, K. C., E. Palaneeswaran, and C. Yu. 2009. “A support vector machine model for contractor prequalification.” Automat. Constr. 18 (3): 321–329. https://doi.org/10.1016/j.autcon.2008.09.007.

LeCun, Yann, et al. Ecient BackProp. 1998. http://yann.lecun.com/exdb/publis/pdf/lecun-98b.pdf

Nguyen, V. U. 1985. Tender evaluation by fuzzy sets, Journal of Construction Engineering and Management 111: 231– 243.

doi:10.1061/(ASCE)0733-9364(1985)111:3(231)
Ohlson, J. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. Journal of Accounting Research, 109‑131.

Palisade (2022) – the Leader in Risk and Decision Analysis Software.” *Palisade*,

https://www.palisade.com/company/

Plebankiewicz, E. “Contractor prequalification model using fuzzy sets”. Journal of Civil Engineering and Management 15 (4) (2009) 377–385.

Project Management Institute. ”A Guide to the Project Management Body of Knowledge (PMBOK Guide)”. 6th ed., Project Management Institute, 2017.

Russell, J. S., and H. Zhai. 1996. “Predicting contractor failure using stochastic dynamics of economic and financial variables.” J. Const. Eng. Manage. 122 (2): 183–191. https://doi.org/10.1061/(ASCE)0733-9364 (1996)122:2(183).

Ren, Jinchang. “ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging”. ELSEVIER. 2012.

Sang, J., Ham, N.-H., Kim, J.-H., and Kim, J.-J. (2014). “Impacts of Macroeconomic Fluctuations on Insolvency: Case of Korean Construction Companies.” Journal of Management in Engineering, 30(5), 05014009.

S.Sapna et al (2012). BACKPROPAGATION LEARNING ALGORITHM BASED ON LEVENBERG MARQUARDT ALGORITHM. Computer Science & Information Technology (CS & IT) DOI : 10.5121/csit.2012.2438

Thomson Reuters Eikon. (2018). [Teck Resources Ltd income statement, annual standardized in millions of US dollars, 2000-2016] [Data set]. https://eikon.thomsonreuters.com/index.html

U.S. Department of Commerce, Bureau of Economic Analysis, *Industry Data, GDP-by-Industry*, Annual Tables, available at https://apps.bea.gov/iTable/index_industry_gdpIndy.cfm as of Nov. 3, 2021.

U.S. Securities and Exchange Commission (SEC). (2022). EDGAR ‑ Company Filings. Retrieved from the U.S. Securities and Exchange Commission: http://www.sec.gov/edgar/searchedgar/companysearch.html

W. Ng et al. (2019). "The influence of training sample size on the accuracy of deep learning models for the prediction of soil properties with near-infrared spectroscopy data". SOIL https://d-nb.info/1221721836/34

Zhang, G., Hu, M.Y., Patuwo, B.E. and Indro, D.C. (1999), ''Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis'', European Journal of Operational Research, Vol. 116, pp. 16-32.