Predicting Bank Failure: Statistical Technique versus Intelligent Technique

Hong Hanh LE, Jean-Laurent VIVIANI

Abstract—This research measures the prediction accuracy of 5 methods to bank failure through a sample of 658 American banks (398 distressed and 260 comparable healthy banks). Two traditional statistical techniques (linear regression analysis and multinomial logistic regression analysis) are compared to three machine learning approaches (artificial neural network, support vector machine and k-nearest neighbor). The dataset of 5 years before bankruptcy was selected to demonstrate the efficacy of these mentioned methods. Based on fundamental ratios extracted from financial statements, 5 methods offer a highly accurate prediction result. The empirical result reveals that k nearest neighbor is the most precise method. The study suggests insightful methods for bankruptcy prediction issues.

KEYWORDS*:* bankruptcy prediction, statistical analysis, intelligent techniques, neural network, support vector machine, K-nearest neighbor, US banks

1. INTRODUCTION

Some people believe that failure sometimes is the opportunity for more efficient reconstruction. However, not every downfall is welcomed as a precious chance to open the new door. The collapse of several banks, especially these days, heralds serious damaged financial systems worldwide. According to the statistics by Federal Deposit Insurance Corporation (FDIC), more than 500 banks declared failure in the United State of America during the period from 2008 to 2014. This series of distressed banks heavily influenced the financial system and caused the worst financial crisis since the Great Depression of 1930s.

With the purpose of preventing banks from bankruptcy, Basel Committee on Banking Supervision provides regulators to improve the quality of banking supervision. Regarding the protection of financial system through banking activities, Basel III was introduced. To enhance Basel II (2004), which aims to ensure minimum capital regulation, supervisory review process, disclosure and market discipline, Basel III (2010) was proposed to ensure that banks do not take on excessive debt, and do not rely too much on short term funds. The guideline promotes a more resilient banking system by focusing on four vital banking parameters: capital, leverage, funding and liquidity. The target of these extensions, undeniably, is to protect the banking system from crisis.

As a consequence, the study of bankruptcy prediction of banks is becoming essential for everyone including investors/creditors, borrowing firms, and governments [1]. Firstly, allowing financial market regulators to predict potential distress, enable them to control the operation of banks more efficiently. Secondly, protecting customers' bank accounts from loss, this is one of the main goals of banks' function. Thirdly, supporting banks to have the right strategy and protecting themselves from breakdown risks. Finally, predicting accuracy helps reduce the effect of systematic risk when the distress of one bank can cause a "Domino effect" that can weaken the macro economy.

Since 1968, thanks to the accomplishment of E.I Altman from New York University, multivariate analysis in forecasting the bankruptcy was implemented worldwide. Over the years, various remarkable researches consolidate the important role of this approach, and contribute appreciably for the economic systems. In parallel with the development of computational sciences, other interesting approaches were explored to promote the power of technology. One of these intelligent applications, which is called 'Machine learning', becomes more popular among researchers and practitioners [2]. A commonly cited formal definition of machine learning, proposed by computer scientist [3]¸explained that a machine is said to learn if it is able to take experience and utilize it such that its performance improves up on a similar experience in the future. We, in this study, also join the stream of researchers to examine the effectiveness of this intelligent method.

Lots of researches today use details information to forecast the status of banks and provide more adequate points of view. Gang, Yang [4] chose 30 ratios which cover various particular sides of firm. Melek [5] selected 20 financial ratios with six feature groups including capital adequacy, asset quality, management quality, earnings, liquidity and sensitivity to market risk. However, in this research, we select five 'classical' ratios to simplify the implementation process but still obtain a high accuracy to forecast the probability of failing before it happens. 398 failed United State banks are investigated during 5 years before going bankruptcy and 260 healthy US banks are used to exam the accuracy of the model objectively.

Two approaches are executed: statistical technique (Linear regression analysis (LRA) and Multinomial logistic regression (MLR)); intelligent technique (k- Nearest neighbor (kNN), artificial neural networks (ANNs) and Support Vector Machines (SVMs))

Our research answers the following questions:

- Could the 'classical' ratios make a good prediction?
- Could we make an improvement by using machine learning in comparison with multivariate approaches?
- Which methods of machine learning will bring out the best performance?
- What features of bank's indicators are the most powerful to predict the bankruptcy?

The brief history of bank's failure prediction is introduced in section 2. Section 3 is explanation on variables. Methodologies are described in section 4. Section 5 introduces the data and software. The result is summarized in section 6. Conclusion and discussion are included in section 7.

2. LITERATURE REVIEWS

To review the history of bankruptcy, we should not ignore the important contribution of Professor *E.I.Altman* [6] in the late 1960's. Thanks to his foundation, multivariate analysis has extended worldwide. The original Z-score (1968) shows its advantage by analyzing five main sides of a firm: the liquidity and size dimensions; the true productivity or profitability of the assets; measure financial leverage as well as consider the capability of management in dealing with competitive conditions. John, Robert [7] however suggests that the accuracy of Altman's model declined from 83.5% to 57.8% when applied to their samples. Laura, Ettore Croc, Federica [8] experimented on 3242 banks across 12 European countries and revealed that Z-score is a good predictive model to identify banks in distress and also has the advantages of simple calculation.

However, in some standpoints, statistical techniques are no longer preferred in view of their low accuracy [2].The attention to and confidence in intelligent techniques has increased enormously during the past 5–10 years. A great number of studies suggest that intelligent techniques perform more effectively than traditional ones. The main difference between intelligent and statistical techniques is that statistical techniques usually require researchers to consider structures to different models, and to construct the model by estimating parameters to fit the data or observation, while intelligent techniques allow learning the particular structure of the model directly from the data [4]. Moreover, the statistical analysis depends on lots of strict hypothesis such as the normal distribution, correlations, which can cause the difficulty in translating forecast results.

Among several methods of Machine learning, neural networks seem to be the most popular tool in prediction issues [2]. Ky Tam [9] implemented neural network on 118 banks (59 failed and 59 non-failed banks) of Texas during 1985-1987. He indicated that neural network has performed more effectively than other methods (Discriminant analysis, factor-logistic, kNN and Decision tree). Félix, Iván [10] employed data from the Federal Deposit Insurance Corporation between 2002 and 2012, the results state that failed banks are more concentrated in real estate loans and have more provisions. The developed model of neural networks can detect 96.15% of the failures in this period and outperforms traditional models of bankruptcy prediction. However, by taking the sample of 1180 firms and selecting 13 accounting-based ratios Esteban [11] recommended that Boosting method achieved improvement against neural network in prediction.

Melek [5] examined ANNs, SVMs and multivariate statistical methods to predict the failure of 65 Turkey financial banks. 20 financial ratios belonging to 6 main groups were chosen: Capital adequacy, Asset quality, Management, Earning, Liquidity and the sensitivity to the market risk. Overall, the result proved that SVMs had achieved the highest accuracy among others. Mapping input vectors into a high-dimensional feature space, SVMs transforms complex problems (with complex decision surfaces) into simpler problems that can use linear discriminant functions, and it has been successfully introduced in several financial applications recently. SVM was also proved to work better than back-propagation neural networks through the research of Laura [8] for a sample of 3242 EU banks.

Unlike Neural network and Support Vector Machine, k-nearest neighbor algorithm is not implemented widely in finance field. This model is implemented widely in biological and transportation field. This method, however, can function appreciatively and obtain the high accuracy in predicting.

Despite the fact that there is numerous researches on bankruptcy prediction, researchers are not satisfied yet. Recently, instead of using single intelligent technique, "ensemble techniques" was applied extensively with the expectation to reach higher accuracy. Ensemble is a machine learning paradigm where multiple learners are trained to solve the same problem. In contrast to ordinary machine learning approaches that attempt to learn one hypothesis from the training data, ensemble methods try to construct a set of hypotheses and integrate them [12]. Gang Wang [4] collected 2 datasets including 240 and 132 companies from 1970 to 2001 with 30 and 24 financial variables respectively. They found that Feature-Selection Boosting method can predict with highest accuracy rate (87%) in comparison with single ANNs, Decision Tree SVMs, Bagging, Boosting or LRA. Deligianni and Kotsiantis [13] observed that "ensemble method" can foresee longer period before the bankruptcy through the dataset of 150 Greek firms.

In brief conclusion, there are several studies on predicting the bankruptcy of banks (Appendix 1). However, the history recorded that intelligent techniques seems to work more effectively than statistical techniques. This study will execute both techniques in different methods and anew, attempt to make a comparison on two aspects: The accuracy and the importance of each ratio.

3. VARIABLES

Altman [6] proposed a Z-score formula based on discriminant analysis that helps forecast the bankruptcy of firms within two years, carried out on 66 manufacturers and small firms who own assets of less than 1 million US dollar. His study investigated 5 main aspects: liquidity of assets, profitability, and operation efficiency, influence of market price and total sales. These ratios were used in lots of studies [14], [5] etc. His model is to objectively forecast for firms during 2 years before collapsing.

$$Z = 1.2Z_1 + 1.4Z_2 + 3.3Z_3 + 0.6Z_4 + 0.99Z_5 \quad (1)$$

Where

Z_1 = Working Capital / Total Assets; Z_2 = Retained Earnings / Total Assets; Z_3 = Earnings before Interest and Taxes / Total Assets; Z_4 = Market Value of Equity / Book Value of Total Liabilities; Z_5 = Sales / Total Assets. He found that the ratio profile for the bankrupt group fell at -0.25, and for the non-bankrupt group at +4.48.

This formula becomes practical because it requires basic accounting information and it is easy to calculate. Taking this advantage, we create 5 new ratios based on Z-score formula. Our adjusted ratios are appropriate for the operation of the banking system and help measure 5 main features of banks. (Table 1).

TABLE 1
RATIOS EXPLANATION

Z_1	Capital/ Total assets	Measure the financial ability of bank.
Z_2	Retained earnings/ Total assets	Implies profitability that reflects the company's age and earning power.
Z_3	(Total of Interest income and Non-interest income)/Total assets	Represents the productivity of earning power compared with the value of assets operating efficiency apart from tax and leveraging factors.
Z_4	Total equity capital/ Total liabilities	Measure the independence of company's financial controller
Z_5	Net income (loss)/Total asset	Measure how profitable the company is over the total asset

We expect that these 5 ratios can forecast the bankruptcy of bank effectively. The idea is to investigate these ratios by implementing: LRA, MLR, ANNs, SVMs and kNN. Table 2 introduces the brief comparison among methods.

TABLE 2
COMPARISION OF 5 METHODS

	LRA	**MLR**	**kNN**	**ANNs**	**SVMs**
Linear classify	Yes	Non linear	Non linear	Non linear	The both
Normality assumption	Yes	No	No	No	No
Result function	Create a function	Calculate probability	No function	No function	No function
Time to train	Quick	Quick	Quick	Slow	Slow
Outcome target	Numeric	Nominal	Nominal	Nominal	Nominal

4. METHODOLOGIES

4.1. Statistical techniques

4.1.1. Multinomial Logistic regression

Multinomial Logistic Regression (MLR) is the linear regression analysis to conduct when the dependent variable is nominal with more than two levels. MLR is an extension version of logistic regression, which analyzes binary dependents. MLR is used to explain the relationship between one dependent nominal variable with one or more independent variables. MLR compares multiple groups through a combination of binary logistic regressions. This method is often preferred to discriminant analysis as it is more flexible in assumptions and types of data that can be analyzed. In this study, each bank is assigned only one of three statuses: healthy - failed or weak.

The MLR provides us the linear function:

f(x)= Log$\frac{\rho_i}{1-\rho_i}$= $\alpha_i + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots . + \varepsilon_i$ (2)

For each 'i', the probability of occurrence is:

$\rho_i = \frac{e^{\alpha_i + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots . + \varepsilon_i}}{1 + e^{\alpha_i + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots . + \varepsilon_i}}$ (3)

Which:

ρ_i And ρ_j are the probabilities of an event

x_{i1}, x_{i2}: Dependent variables

α : the constant

B_1, β_2: Regression coefficients

Since then we have the probability of being healthy, failed and weak of bank as function (4), (5), (6).

$p_{\text{healthy}} = \frac{e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots \varepsilon_1}}{1 + e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots \varepsilon_1}}$ (4)

$p_{\text{failed}} = \frac{e^{\alpha + \beta 1 x 1 + \beta 2 x 2 + \ldots \varepsilon_{10}}}{1 + e^{\alpha + \beta 1 x 1 + \beta 2 x 2 + \ldots \varepsilon_{10}}}$ (5)

$p_{\text{weak}} = \frac{e^{\alpha + \beta 1 x 1 + \beta 2 x 2 + \ldots \varepsilon_{10}}}{1 + e^{\alpha + \beta 1 x 1 + \beta 2 x 2 + \ldots \varepsilon_{10}}}$ (6)

The sum of p_{healthy}, p_{failed} and p_{weak} is always equal 1

4.1.2. Linear regression analysis

Linear Regression Analysis (LRA) is a powerful tool for predicting or explaining a nonmetric dependent variable. This is concerned with the classification of distinct sets of observations and it tries to find the combination of variables that predicts the group to which an observation belongs. The combination of predictor variables is called a linear discriminant function, and this function can then be used to classify new observations whose group membership is unknown [5]. LRA is a classification method which assumes that data in each class are Gaussian-distributed and that there is a unique covariance matrix for each class [14]. The expected linear function is as follows:

$Z = a + bZ_1 + c\, Z_2 + dZ_3 + eZ_4 + fZ_5$ (7)

Subject to:

Z is the predictive score

a is constant

b, c, d, e, f: are coefficients

Z_1, Z_2, Z_3, Z_4, Z_5: are independent variables

We define 1 as the status of healthy bank, 0.5 is the status of weak bank and 0 is for failed bank.

For $Z > 0.6$: the bank is predicted as healthy

$0.3 \leq Z \leq 0.6$: the bank is listed as weak

$Z \leq 0.3$: the bank is considered to be failed

4.2. Intelligent techniques

Although the discriminant analysis and linear regression model have become the most commonly used in bankruptcy prediction, their inherent drawbacks of statistical assumptions such as linearity, normality and independence among variables have constrained both applications [1]. To solve the limitation of linear problems, intelligent techniques in this study will be considered as Machine learning approaches achieving a forward movement by introducing nonlinear separation. Several methods are implemented to classify, predict and contribute a great help to financial analysis. The basic concept showed that a machine is said to learn if it is able to take experience and utilize it such that its performance improves up on similar experiences in the future [3]

In this paper, 2 machine learning approaches are implemented: Lazy learning and Black Box Methods through 3 algorithms: K-nearest neighbor, Artificial neural network and Support Vector Machine.

Neural network and support vector machine are extremely well-nown models and are appraised as one of the most powerful tools in prediction even when their conceptions are not easy to be translated. These models are referred to as black box processed because the mechanism that transforms the input into the output is obfuscated by a figurative box.

On the contrary, K-nearest neighbor is regarded as a lazy learning, an intelligibility model by their characteristics of classifying unlabeled examples by assigning them the class of the most similar labeled examples.

The pros and cons of each method are explained as follows (table 3)

4.2.1. Lazy learning – K nearest neighbor

In a brief description, nearest neighbor classify by mapping the different characteristics of dataset closely to different label groups, since then, the given data with common features will be placed in the same group.

For example, a bank holds low equity/total assets ratio and low liquidity asset/total assets ratio could be judged susceptible to bankruptcy, because most of collapsed banks own the same characteristics. If all banks are placed on one plane, each bank with their characteristics is represented by one point. Uniform banks tend to be placed closely together. It is required to calculate the distance between each point to each group in order to find out which group (healthy – failed-weak banks) that the banks belong to.

There are many ways to calculate this distance. Traditionally, the K nearest neighbor (kNN) algorithm deploys Euclidean distance (Equation 8). *p* and *q* are two random points, each of them has *n* features. The distance between p and q can be calculated as:

$$\text{Dist (p,q)}=\sqrt{(p_1-q_1)^2+\ \dots+(p_n-q_n)^2}\ (8)$$

To classify the bank as healthy-weak-failed, we should begin by assigning the number of *k*. With *k* being the number of nearest points, we will choose the nearest neighbor to consider the status of bank.

For example (Figure 1), a random bank A (in Red box) is considered among 3 groups: Healthy- Normal-Fail banks which is Blue-Green and Black boxes respectively. There are 5 other nearest banks: B, C, D, E, and F. Distance from each bank to bank A is mentioned in table 4.

TABLE 3
K-NEAREST NEIGHBOR EXPLANATION

Bank name	Status of bank	Distance to A
B	Fail	0.5
C	Fail	0.7
D	Normal	0.6
E	Fail	0.4
F	Healthy	1

If we select *k*=1, the closest bank with A is bank E, so A belongs to failed groups.

If we select *k*=3, the closest banks with A are: B, D, E. Among them, B and E are failed banks, D is normal bank, so A belongs to failed groups.

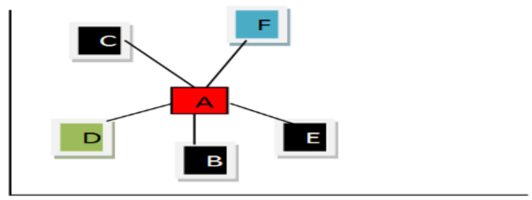

Figure 1: K-nearest neighbor distance example

We can select any value of *k* to find the best grouping way. There are some divergent hypotheses on selecting the ‘best’ *k*. Some researchers supposed that *k* should be the square root of number of features. However, others assume that *k* performs the best if it is between (2, 10). In this research, we experiment with various value of *‘k’* to find the optimal one.

4.2.2 Artificial neural networks

Taking advantages of computers’ potentials, Artificial Neural Networks (ANNs) was designed to simulate humans’ brain. The idea is to learn from examples using several constructs and algorithms just similar a human being learns new things [2].The advantages of ANNs are their flexible nonlinear modeling capability, strong adaptability, as well as their learning and massive parallel computing abilities [15]. However, they cannot explain the causal relationship among variables, which constrains its application to managerial problems [1]. Neural networks have many different topologies for problem dissimilarities. Among them, back-propagation is the most well-known and commonly used, categorized as one of the supervised learning models [1]. It provides the mapping function between the input (x) and output (y) from the database. The purpose of back-propagation training is to minimize an error backward while information is transmitted forward [16] Depending on the important role of each x-variable, the weight (w values) will be calculated ignoring how these weight is determined. The higher weight of neuron is, the stronger the input multiplied with is. y(x)= f($\sum_{i=1}^{n} w_i x_i$)

(y) Will be compared with a threshold of neuron. For example, the healthy bank has value =1 and failed bank has value= -1, hence, if (y) is greater than the threshold α=0, the bank is healthy and vice versa.

$$Output = \begin{cases} 1, \sum_{i=1}^{n} w_i x_i > threshold = 0 \\ -1, \sum_{i=1}^{n} w_i x_i > threshold = 0 \end{cases}$$

The process is described in figure 2

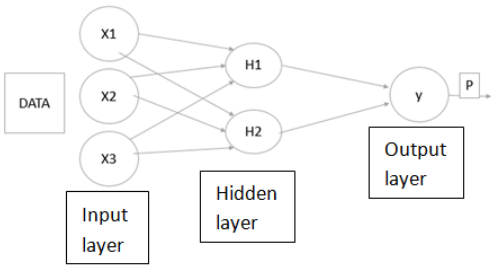

Figure 2: Artificial neural network process

In this research, a single hidden layer BPNN is used for one-step-ahead forecasting. An aggregate ***I*** of **Input** receive unprocessed signals directly from the database. There are x_{t-n}, x_{t-n+1} ..... x_2, x_1. Each input node is responsible for processing a single feature in the dataset; the feature's value will be transformed by the node's activation function ***f*** to obtain the target output y_t.
The value of input and output of hidden layer are:
$I_j = \sum_{t-n}^{t-1} w_{ji} * x_i + \beta_0$ (j=1….h)
$y_t = f_h(I_0)$
The input and output value of the output layer:
$I_0 = \sum_{j=1}^{h} w_{0j} * x_j + \alpha_0$ (0=1)
$y_t = f_0(I_0)$
Where:
I: set of input; y: target output; y_t is the forecasted value of point t
n and h: the number of input layer nodes and hidden layer nodes
w_{ji}: the weights of the input and hidden layers;
w_{oj}: the connection weights of the hidden and output layers;
β_j and α_o: the threshold values of the hidden and output layers.
β_j and $\alpha_o \in (-1,1)$
f_h and f_o : the activation functions of the hidden and output layers.
In this study, the database will be split randomly into 2 groups: training dataset and test dataset. For training set, we selected randomly 5000 observations and the rest 2260 banks is put in test set. We implement ANNs on the training set, since then we will compare the result with test set.

4.2.3. Support Vector Machine

Unlike lot of other methods that focus on whole training data, Support Vector Machine pays great attention to the most 'difficult to recognize data point' based on the idea: if SVM can point out the toughest points, the others will be seen easily. The Support Vector Machine is a non- linear generalized portrait algorithm developed in Russia in the 60s [17] SVM create a linear model to implement nonlinear class boundaries through some nonlinear mapping input vectors into a high-dimensional feature space [18] This is a supervised learning model that owns learning algorithms to analyze data and recognized patterns. SVM can perform data in a higher dimension, since then figure out the best hyperplane that distinguish them into 3 classes.
Suppose that we have L training point, where input x_i has D attributes. Basically, there exists a linear $y = w.x + b$ that separate the 2 classes when D=2 and a hyperplane on graphs of $x1, x2, \dots xD \; for \; D > 2$
For the 2 class classification, we have 2 functions:

$$\begin{cases} xi.w + b \geq 1 \; if \; yi = 1 \\ wi.x + b \leq -1 \; If \; yi = -1 \end{cases}$$

To combine, we have a new equation:

$$y_i(x_{i.}.w + b) - 1 \geq 0 \; \forall_i$$

The role of SVMs is to find out the points that lay the closet to the separating hyperplanes. Call H1 and H2 are 2 planes that these points lie on. We have:

$$(H1): xi.w + b = +1$$
$$(H2): xi.w + b = -1$$

With d1 and d2 are the distances from H1 and H2 respectively. The hyperplane's equidistance between H1 and H2 means that d1=d2, is defined as the SVM's margin. The purpose of SVMs is to maximize this margin.
The vector geometry shows that the margin is equal to $\frac{1}{\|w\|}$, so we need to minimize $\|w\|$ or, as Quadratic Programming optimization, we need to find: min $\frac{1}{2}\|w\|^2$ subject to $yi(xi.w + b) - 1 \geq 0$.
By using Lagrange multiplier to find minima, we add α, where ∀α:

$L_P = \frac{1}{2}\|w\|^2 - \alpha(y_i(x_{i.}.w + b) - 1 \geq 0 \; \forall_1)$

$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{L} \alpha_i((y_i(x_{i.}.w + b)) - 1$

$= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{L} \alpha_i((y_i(x_{i.}.w + b)) + \sum_{i=1}^{L} \alpha_i$ (9)

By setting derivatives to zero, we obtain:

$\frac{\sigma L}{\sigma w} = 0 \Rightarrow w = \sum_{i=1}^{L} \alpha_i y_i \; x_i$ (10)

$\frac{\sigma L}{\sigma b} = 0 \Rightarrow \sum_{i=1}^{L} \alpha_i \, y_i = 0$ (11)

Substituting (10) and (11) into (9), we obtain new formulation L_D which is dependent on α. LD is referred to as the Dual form of the Primary Lp.

$L_D = \sum_{i=1}^{L} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j x_i x_j$ s.t $\alpha_i \geq 0, \sum_{i=1}^{L} \alpha_i \, y_i = 0$
$= \sum_{i=1}^{L} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i H_{i,j} \, \alpha_j$ where $H_{i,j} = y_i y_j x_i x_j$

$= \sum_{i=1}^{L} \alpha_i - \frac{1}{2}\alpha^T H_\alpha$ s.t $\alpha_i \geq 0, \sum_{i=1}^{L} \alpha_i \, y_i = 0$

Any point that satisfies $\sum_{i=1}^{L} \alpha_i \, y_i = 0$ which is a support vector x_s will have the form: $y_s(x_s.w + b) = 1$ or,

$y_s\left(\sum_{m \in S} x_m . \alpha_m y_m + b\right) = 1$ Where S is the set of indices of the Support Vector, defined by finding the indices i where $\alpha_i \geq 0$

Multiplying through by y_s, we obtain:

b= $y_s - \sum_{m \in S} x_m . \alpha_m y_m = \frac{1}{Ns}\sum_{s \in S}(y_s - \sum_{m \in S} x_m . \alpha_m y_m . x_s)$

Each new point x' is classified by evaluating

y'=$sgn(w.x'+b)$

For regression that is not linearly separate, we need to choose a kernel and relevant parameters which would help mapping data from non-linearly separable into a feature space where it is linearly separable. For the first step, choosing the appropriate kernel and map $x \rightarrow \emptyset(x)$

- Create hyperplane H, where $H_{ij} = y_i y_j \emptyset(x_i)\emptyset(x_j)$
- Choose how significant misclassification should be treated, by selecting a suitable value for the parameter C.
- Choose how significant misclassifications should be treated and how large the insensitive loss region should be, by selecting suitable values for the parameters C and ϵ
- Find $\alpha^+$and α^- so that:

$$\sum_{i=1}^{L}((\alpha_i{}^+ - \alpha_i{}^-)t_i - \epsilon \sum_{i=1}^{L}(\alpha_i{}^+ - \alpha_i{}^-) - \frac{1}{2}\sum_{i,j}(\alpha_i{}^+ - \alpha_i{}^-)((\alpha_j{}^+ - \alpha_j{}^-)\emptyset(x_i)\emptyset(x_j)$$

Is maximized subject to the constraints:

$0 \leq \alpha_i{}^+, \alpha_i{}^- \leq C$ And $\sum_{i=1}^{L} \alpha_i{}^+ - \alpha_i{}^- = 0\ \forall i$

This is done using a QP solver

- Calculate w=$\sum_{i=1}^{L}(\alpha_i{}^+ - \alpha_i{}^-)\ \emptyset\, x_i$
- Determine the set of Support Vector S by finding the indices *I* where 0 <α< C and $\varepsilon_i = 0$
- Calculate b:

  b= $\frac{1}{Ns}\sum_{s\in S}(t_i - \epsilon - \sum_{m=1}^{L}(\alpha_i{}^+ - \alpha_i{}^-)\emptyset(x_i)\emptyset(x_m)$
- Each new point x' is determined by evaluating:
- y'=$\sum_{i=1}^{L}(\alpha_i{}^+ - \alpha_i{}^-)\emptyset(x_i)\emptyset(x') + b$

Find α so that: $\sum_{i=1}^{L}\alpha_i - \frac{1}{2}\alpha^T H\alpha$ is maximized, subject to the constraints: $0 \leq \alpha_i \leq C\ \forall i$ and $\sum_{i=1}^{L}\alpha_i y_i = 0$

- After solving by QP solver, we need to calculate $w = \sum_{i=1}^{L}\alpha_i y_i\ \emptyset x_i$
- Determine the set of Support Vector S by finding the indices such that $0 \leq \alpha_i \leq C\ \forall i$
- Calculate b=$\frac{1}{Ns}\sum_{s\in S}(y_s - \sum_{m\in S}\emptyset x_m.\alpha_m y_m.\emptyset x_s)$

Each new point x' is classified by evaluating

y'= $sgn(w.\emptyset(x)' + b)$

TABLE 4:
COMPARISION OF 3 METHODS

	Strengths	**Weaknesses**
Artificial neural network	- Can be adapted to classification or numeric prediction problems - Among the most accurate modeling approaches - Makes few assumptions about the data's underlying relationships	- Reputation of being computationally intensive and slow to train, particularly if the network topology is complex - Easy to overfit or underfit training data - Results in a complex black box model that is difficult if not impossible to interpret
Support Vector Machine	- Can be used for classification or numeric prediction problems - Not overly influenced by noisy data and not very prone to overfitting - High accuracy	- Can be slow to train, particularly if the input dataset has a large number of features or examples - Results in a complex black box model that is difficult if not impossible to interpret
k-nearest neighbor	- Simple and effective - Makes no assumptions about the underlying data distribution - Fast training phase	- Does not produce a model, which limits the ability to find novel insights in relationships among features - Require a large amount of memory - Finding the best value of 'k' require testing various times.

5. DATA AND SOFTWARES

The accountant information of 658 American banks were collected, including 398 collapse banks and 260 healthy banks, through Bloomberg database during the period from 2003 to the first half of 2015. Each bank is investigated during a 5 –year period. In total, 3288 observations are collected. The failed banks were collected from the Federal Deposit Insurance Corporation (FDIC) failed banks list update at May 2015. Healthy banks collected as a comparable data must meet the requirement of having an asset of smaller than 1000 million US Dollar.

We categorize banks into 3 groups: Healthy, Failed banks and Weak. Healthy bank include 1299 observations collected from 260 institutions which are not bankrupt through 5 years. Failed banks include 398 observations of 398 collapsed banks and weak banks contain 1591 observations which are extracted from failed banks during the period from 2 to 5 years before being distress.

To follow up the assumption of linear regression analysis, the normality test is implemented. The visualization of data distribution is shown in figure 3. The dark blue, red and light blue are colors of healthy, failed and weak banks respectively. We use Shapiro Wilk test (table 5), the result rejected the Null hypothesis that the population is normally distributed. This result can lead to the inefficiency of linear regression prediction model.

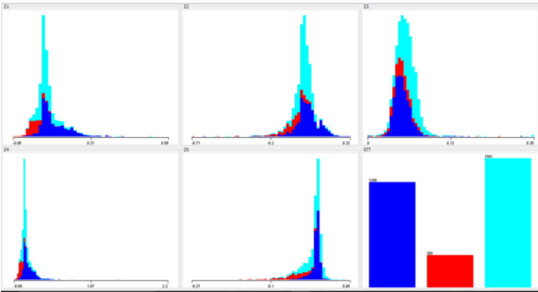

Figure 3: Data distribution

TABLE 3
SHAPIRO WILK TEST RESULT

	W	p-value	Reject Null hypothesis
Z1	0.83	<2.2E-16	Yes
Z2	0.88	<2.2E-16	Yes
Z3	0.79	<2.2E-16	Yes
Z4	0.61	<2.2E-16	Yes
Z5	0.74	<2.2E-16	Yes
STT	0.77	<2.2E-16	Yes

Figure 4 provides the descriptive statistics of the whole data set and the sub-sample data sets. The 3 groups are represented under 3 numbers: 0, 0.5 and 1, which means failure, weak and healthy banks respectively. Figure 5 suggests the Levene's test which is used to assess the equality of variances for a variable calculated for our 3 groups. The sig. (Figure 5) rejects the null hypothesis of equal variances and it is concluded that there is a difference between the variances in the population.

Descriptives

		N	Mean	Std. Deviation	Std. Error
Z1	.0	398	.0388974887	.0300037904	.0015039541
	.5	1591	.0969665581	.0529191677	.0013267158
	1.0	1299	.1343501736	.0697715844	.0019358603
	Total	3288	.1047067731	.0654704266	.0011417717
Z2	.0	398	-.063453427	.0907890030	.0045508415
	.5	1591	.0127034015	.0552375162	.0013848383
	1.0	1299	.0430590622	.0769845459	.0021359888
	Total	3288	.0154776063	.0766194957	.0013362059
Z3	.0	398	.0545986372	.0198958101	.0009972868
	.5	1591	.0656187715	.0200765413	.0005033311
	1.0	1299	.0549529808	.0139234506	.0003863156
	Total	3288	.0600710599	.0186604408	.0003254288
Z4	.0	398	.0414982749	.0329359888	.0016509319
	.5	1591	.1133670207	.1091620794	.0027367600
	1.0	1299	.1518231086	.0851946538	.0023637839
	Total	3288	.1198605418	.0996447125	.0017377542
Z5	.0	398	-.046027003	.0370542124	.0018573598
	.5	1591	-.002055206	.0228394074	.0005725979
	1.0	1299	.0032193408	.0114824462	.0003185883
	Total	3288	-.005293995	.0265488765	.0004629992

Figure 4: Descriptive statistics of the whole dataset

Test of Homogeneity of Variances

	Levene Statistic	df1	df2	Sig.
Z1	159.258	2	3285	.000
Z2	78.008	2	3285	.000
Z3	10.031	2	3285	.000
Z4	48.192	2	3285	.000
Z5	243.130	2	3285	.000

Figure 5: Levene's test for equality of variances

We examine the correlation among the 5 attributes (figure 6). The significant correlations between all independent variables and dependent variable demonstrate that the status of bank is significantly promoted by 5 ratios. The correlation results show that each ratio has influence on the status of bank. All ratios except Z3 have positive direction with status. The p-value is significant, so it rejected the null hypothesis that the independent variables (Z1, Z2, Z3, Z4, and Z5) have no influence on dependent variable (Z). However, it also showed that the correlation among ratios is remarkable. The correlation between (Z1, Z2), (Z2, Z4) and (Z2, Z5) are arranged as having 'medium correlation' which could cause 'Multicollinearity'. This can make the result of multinomial logistic regression analysis or linear regression become less precise, on the contrary with assumption of having no or little multicollinearity

Correlations

		Z1	Z2	Z3	Z4	Z5	STT
Z1	Pearson Correlation	1	,465**	-,068**	,934**	,314**	,453**
	Sig. (2-tailed)		,000	,000	,000	,000	,000
	N	3288	3288	3288	3288	3288	3288
Z2	Pearson Correlation	,465**	1	,155**	,347**	,539**	,402**
	Sig. (2-tailed)	,000		,000	,000	,000	,000
	N	3288	3288	3288	3288	3288	3288
Z3	Pearson Correlation	-,068**	,155**	1	-,084**	,244**	-,110**
	Sig. (2-tailed)	,000	,000		,000	,000	,000
	N	3288	3288	3288	3288	3288	3288
Z4	Pearson Correlation	,934**	,347**	-,084**	1	,210**	,334**
	Sig. (2-tailed)	,000	,000	,000		,000	,000
	N	3288	3288	3288	3288	3288	3288
Z5	Pearson Correlation	,314**	,539**	,244**	,210**	1	,470**
	Sig. (2-tailed)	,000	,000	,000	,000		,000
	N	3288	3288	3288	3288	3288	3288
STT	Pearson Correlation	,453**	,402**	-,110**	,334**	,470**	1
	Sig. (2-tailed)	,000	,000	,000	,000	,000	
	N	3288	3288	3288	3288	3288	3288

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 6: Correlations among variables

In this study, SPSS 20 and Weka version 3.6.13 is used for the empirical test.

6. EMPIRICAL RESULTS

It is obvious that different models had different accuracies, as be expected. Our result presents that k-nearest neighbor performs better than the other methods. Among those, LRA obtain the lowest prediction accuracy.

The statistical analysis, with the advantage of giving clear function, quick training data, gives us the high percentage of accuracy. Figure 7 presents the confusion matrix that announces the ability to predict of each model. However, both of them have problem with data assumption that lead to less trust worthy result. In detail, Linear Regression Analysis provides the function:

$$Z = 0.66 + 3.67*Z_1+0.27*Z_2-3.455*Z_3-1.51*Z_4+4.4*Z_5$$

This function reveals that ratio Z_1 (Capital / Total Assets), Z_3 (Earnings before Interest and Taxes) and Z_5 (Sales / Total Assets) have typical influence on the status of banks. Moreover, Z3 (Earnings before Interest and Taxes) and Z4 (Total Equity Capital/ Total liabilities) have a negative direction with the Z score. Through this method, 68.9% banks are placed correctly. However, the model is not precise perfectly because, as mentioned in data part, data is not normal distributed (Table 5)

Multinomial Logistic Regression obtains 72.7% of accuracy which is even better than artificial neural network and Support vector machine. MLR shows three functions that calculate the probability of being healthy:

$$p_{healthy} = \frac{e^{0.99 + 130*Z1+ 2.38*Z2 -64.2*Z3 - 94*Z4+25*Z5}}{1+e^{0.99 + 130*Z1+ 2.38*Z2 -64.2*Z3 - 94*Z4+25*Z5}}$$

$$p_{failed} = \frac{e^{0.91 - 61.23*Z1+1.34*Z2-0.59*Z3 + 15.8*Z4-32.56*Z5}}{1+e^{0.91 - 61.23*Z1+1.34*Z2-0.59*Z3 + 15.8*Z4-32.56*Z5}}$$

$$\rho_{\text{weak}} = 1 - (\rho_{\text{healthy}} + \rho_{\text{fail}})$$

However, this result is not totally trust worthy because of the multicollinearity (Figure 6) which can lead to problems with understanding which variable contributes to the explanation of the dependent variable and technical issues in calculating a multinomial logistic regression.

	REAL RESULT					
	LRA			MLR		
PREDICTED RESULT	Healthy	Failure	Weak	Healthy	Failure	Weak
Healthy	1147	25	25	828	26	445
Failure	7	149	224	11	251	136
Weak	146	224	149	214	63	1314
Accuracy rate	69%			73%		

Figure 7: Confusion matrix of statistical techniques

To assess the performance of intelligent techniques as banking failure prediction methods, we compare the discriminatory power with the output of two statistical techniques. The intelligent techniques perform very well. Figure 8 shows the correct classification of support vector machine, artificial neural network and K-nearest neighbor. The result present that support vector machine methods obtain 70% of accuracy, artificial neural network obtain 71%. Remarkably, k nearest neighbor obtained up to 99% of accuracy.

	REAL STATUS								
	SVM			ANN			K-nearest neighbor		
PREDICTED RESULT	Healthy	Failure	Weak	Healthy	Failure	Weak	Healthy	Failure	Weak
Healthy	682	10	607	703	39	557	1299	0	0
Failure	3	231	164	5	307	86	3	395	0
Weak	149	66	1376	135	132	1324	16	0	1575
Accuracy rate	70%			71%			99%		

Figure 8: Confusion matrix of intelligent techniques

Generally, intelligent techniques perform better and more trust worthy than statistical techniques because it requires less statistical data assumption. However, intelligent techniques are frequently criticized for being like black boxes because the manner in which they receive the inputs and provide the output is not very transparent. Figure 9 shows the proportion of banks that were inaccurately recognized among 1988 failed and weak observations by using intelligent techniques. The results reveal that, our methods predict very well through 5 years and perform the best since 3 years before being bankruptcy of bank. Figure 10, 11, and 12 showed the detailed accuracy by class of 3 intelligent methods. The ROC Area of 3 methods are more than 76%, especially, kNN has ROC Area is 100%. This result is good and excellent at separating.

Period before bankruptcy	ANN	SVM	KNN
1	5.33%	4.23%	0.15%
2	2.52%	2.36%	0.20%
3	1.96%	2.77%	0.30%
4	3.67%	4.73%	0.20%
5	4.53%	6.19%	0.15%

Figure 9: Inaccurate classification during 5 years period

```
=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.564     0.097      0.792     0.564     0.659      0.774    healthy
                 0.688     0.042      0.695     0.688     0.692      0.883    fail
                 0.825     0.388      0.666     0.825     0.737      0.718    weak
Weighted Avg.    0.705     0.231      0.719     0.705     0.7        0.76
```

Figure 10: SVM detailed accuracy by class

```
=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         0.01       0.986     1         0.993      1        healthy
                 0.992     0          1         0.992     0.996      1        fail
                 0.99      0          1         0.99      0.995      1        weak
Weighted Avg.    0.994     0.004      0.994     0.994     0.994      1
```

Figure 11: Knn detailed accuracy by class

```
=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.541     0.07       0.834     0.541     0.656      0.834    healthy
                 0.771     0.059      0.642     0.771     0.701      0.943    fail
                 0.832     0.379      0.673     0.832     0.744      0.795    weak
Weighted Avg.    0.71      0.218      0.733     0.71      0.704      0.828
```

Figure 12: Ann detailed accuracy by class

7. CONCLUSION AND DISCUSSION

This study provides empirical experiment on predicting the bankruptcy of banks through 2 techniques: statistical analysis and intelligent analysis. Among 5 methods, kNN records the highest prediction accuracy with 99% for the period of 5 years. Other methods LRA, MLR, SVM, ANN obtain from 69% to 73%. This result convinced that kNN method is more successful than others.

Another important conclusion, MLR obtained the higher accuracy than two out of three intelligent techniques, however, because of multicollinearity assumption; the MLR prediction is not as precise as SVM and ANN. On the other hand, intelligent techniques result is not easy to translate and difficult to figure out as clearly as statistical analysis.

In brief, we cannot deny that using ‘classical’ accounting ratios still work effectively. A part of the study proves that the ratios Z_1, Z_3, Z_5 are more significant in predicting the bank’s bankruptcy than the others; which indicates that the measures of net liquidity, the productivity of earning power and the capacity of gaining profit compared with the value of total assets are extremely important for the banks and play a significant role in prediction. However, by using only kNN, we will not able to find out the influence of each ratio to the model.

The limitation of this study is that we only focus on accounting information and do not pay attention to market activities of banks. Moreover, we could not figure out the role of each ratio in intelligent techniques. However, this research still helps regulators have an overview of the system of small and medium banks. It is also interesting for very first approach on analyzing the performance of banks.

We suggest that in future research; some other ratios can be added to analyze more details about the operation of banks. Moreover, recent trend using ‘ensemble method’ which is the combination of several machine learning algorithms can be implemented for upcoming studies and help improve the limitations of separate learning algorithm.

Appendix 1: Brief summary on the history of bankruptcy prediction

Name of authors	**Year published**	**Size of data**	**Methods**	**Result**
Altman [6]	1968	66 US firms	Multivariable discriminant analysis	95% for 1-year period. For overall 5-year period: 79%
E.I. Altman, R.G. Haldeman, P. Narayanan [20]	1977	111 firms	Zeta-analysis	96% for the first period and 70% for 5-year periods
D.Martin [22]	1977	5700 commercial banks	Logit model, linear discriminant, quadratic regression	Logit model performed the best with 87% of accuracy
J.A Ohlson [23]	1980	2163 Firms	Logit	96% for 1 year period, 95% for 2 year period
J.R.Dietrich, R.S. Kaplan [21]	1982	150 firms	Loan classification model from OLS and MDA	67% right classification
KY TAM [9]	1990	118 Banks	Discriminant analysis, Neural network,	Neural network performs the best
I. Olmeda, E. Fernandez [25]	1997	66 banks	Neural network, logit, Discriminant analysis	Neural network perform better than the others
John Stephen Grice, Robert W. Ingram [19]	2001	1974 firms	Z-score	For the first period: 57.8% correct replacement. The second period: 88.1%
J.H. Min, Y.-C. Lee [24]	2005	Korean firms	SVM,LDA, logit, Neural network	SVM obtained the best performance (83%)
Gang Wang, Jian Ma, Shanlin Yang [4]	2014	372 firms	SVM, NN, Boosting,	FS-Boosting gets the highest average accuracy, 81.5%
Laura Chiaramonte, Ettore Croci, Federica Poli [8]	2015	3242 banks	Z-score and the CAMELS related covariates	Natural logarithm of the Z-score show better predictive power

BIBLIOGRAPHY

I.

[1] *A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. Expert Systems with Applications.* **Sangjae Lee, Wu Sung Choi.** 2013, Vol. 40, pp. 2941–2946.

[2] *Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review.* **P. Ravi Kumar, V. Ravi.** s.l. : European Journal of Operational Research, 2007, Vol. 180, pp. 1-28.

[3] **Lantz, Brett.** *Machine learning with R.* s.l. : Packt Publishing Ltd, 2013. p. 10.

[4] *An improved boosting based on feature selection for corporate bankruptcy prediction .* **Gang Wang, Jian Ma, Shanlin Yang.** s.l. : . Expert Systems with Applications,, 2015, Vols. 41,, pp. 2353–2361.

[5] *Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey.* **Melek Acar Boyacioglu, Yakup Kara , O¨ mer Kaan Baykan.** s.l. : Expert Systems with Applications, 2009, Vol. 36, pp. 3355-3366.

[6] *Financial ratios, discriminant analysis and the prediction of corporate bankruptcy.* **Altman, E.I.** s.l. : Journal of finance, 1968, Vol. 23, pp. 589-609.

[7] *Tests of the generalizability of Altman's bankruptcy prediction model .* **John Stephen Grice, Robert W. Ingram.** s.l. : . Journal of Business Research,, 2001, Vol. 54, pp. 53– 61.

[8] *Should we trust the Z-score? Evidence from the European Banking Industry.* **Laura Chiaramonte, Ettore Croci, Federica Poli.** s.l. : Global Finance Journal, 2015, Vols. 28,, pp. 111-131.

[9] *Neural Network Models and the Prediction of Bank Bankruptcy.* **Tam, Ky.** s.l. : . OMEGA, 1991, Vol. 19 no.5, pp. 429--445.

[10] *Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks.* **Félix J. López Iturriaga, Iván Pastor Sanz.** s.l. : Expert Systems with Applications, 2015, Vol. 42, pp. 2857–2869.

[11] *Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks.* **Esteban Alfaro, Noelia García, Matías Gámez, David Elizondo.** s.l. : Decision Support Systems, 2008, Vol. 45, pp. 110–122.

[12] *Popular Ensemble Methods: An Empirical Study.* **David Opitz, Richard Maclin.** s.l. : Journal of Artificial Intelligence Research, 1999, Vol. 11, pp. 169-198.

[13] *Forecasting Corporate Bankruptcy with an Ensemble of Classifiers.* **Despina Deligianni, Sotiris Kotsiantis.** s.l. : Lecture Notes in Computer Science, 2012, Vol. 7297, pp. 65-72.

[14] *A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method.* **Chen, Hui-Ling, et al.** s.l. : Knowledge-Based Systems, 2011, Vol. 24.8 , pp. 1348-1359.

[15] *A Bayesian regularized artificial neural network for stock market forecasting.* **Ticknor, Jonathan L.** s.l. : Expert Systems with Applications, 2013, Vol. 40.14, pp. 5501-5506.

[16] *Back propagation neural network with adaptive differential evolution algorithm for time series forecasting.* **Wang, Lin, Yi Zeng, and Tao Chen.** s.l. : Expert Systems with Applications, 2015, Vol. 42.2, pp. 855-863.

[17] *Pattern recognition using generalized portrait method.* **Vapnik V., Lerner A.** s.l. : Automation and Remote Control, 1963, Vol. 24, pp. 774–780.

[18] *Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. ,,.* **Jae H. Mina, Young-Chan L Maclin ee.** s.l. : Expert Systems with Applications, 2005, Vol. 28, pp. 603–614.

[19] *Empirical Analysis of the Commercial loan classification decision.* **J. Richard and Dietrich, Robert S.Kaplan.** s.l. : The accounting review, 1982, Vol. 57, pp. 18-38.

[20] *ZETA ANALYSIS, a new model to identify bankruptcy risk of corporations.* **E.I Altman, R.G Halderman, P.Narayanan.** s.l. : Journal of Banking and Finance, 1977, Vol. 1, pp. 29-54.

[21] *Empirical Analysis of the Commercial loan classification decision.* **J. Richard and Dietrich, Robert S.Kaplan.** s.l. : The accounting review, 1982, Vol. 57, pp. 18-38.

[22] *Early warning of bank failure: A logit regression approach.* **D.Martin.** s.l. : Journal of Banking and Finance, 1977, Vol. 1, pp. 249-276.

[23] *Financial ratios and the probabilistic prediction of bankruptcy.* **Ohslon, J.A.** s.l. : Journal of Accounting research, 1980, Vol. 18, pp. 109-131.

[24] *Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters.* **Jae H. Mina, Young-Chan Lee.** s.l. : Expert Systems with Applications , 2005, Vol. 28, pp. 603–614.

[25] *Hybrid Classifiers for Financial Multicriteria Decision Making: The Case of Bankruptcy Prediction.* **Ignacio Olmeda, Eugenio Fernandez.** s.l. : Computational Economics, 1997, Vol. 10, pp. 317–335.

[1]