Methodological approach of construction businesses failure prediction studies: a review

Alaka, H, Oyedele, LO, Owolabi, HA, Ajayi, SO, Bilal, M & Akinade, OO

Original citation & hyperlink:
Alaka, H, Oyedele, LO, Owolabi, HA, Ajayi, SO, Bilal, M & Akinade, OO 2016, 'Methodological approach of construction businesses failure prediction studies: a review' *Construction Management and Economics*, vol 34, no. 11, pp. 808-842 https://dx.doi.org/10.1080/01446193.2016.1219037

DOI 10.1080/01446193.2016.1219037
ISSN 0144-6193
ESSN 1466-433X

Publisher: Taylor and Francis

This is an Accepted Manuscript of an article published by Taylor & Francis in Construction Management and Economics on 18th August 2016, available online: http://www.tandfonline.com/10.1080/01446193.2016.1219037

Methodological Approach of Construction Businesses Failure Prediction Studies: A review

Abstract

Performance of bankruptcy prediction models (BPM), which partly depends on the methodological approach used to develop it, has virtually stagnated over the years. The methodological positions of BPM studies were thus investigated. Systematic review was used to search and retrieve 70 journal articles and doctoral theses. Their 'general methods' and 'philosophical underpinnings' investigated using summary of findings tables and meta-analysis. 'General methods' results showed positive trends in terms of techniques being used, error cost consideration and model validation, with some use of skewed data being the main drawback. For 'philosophical underpinnings', positivism paradigm was discovered to be at the core of BPM studies. This is deemed inadequate because of the need to consider industries' dynamism, financial variables flaws, and social factors which actually lead to the financial status of firms. The pragmatism paradigm using mixed method is proposed. A research design framework for executing the proposed methodology is presented. This will help BPM developers go through more rigorous and robust methodology to deliver better and more valid models. Limitation of study include not reviewing studies not reported in English language and impact of different countries accounting practices on ratios. Limited availability of theses database resulted in reviewing only 4 theses

Keywords: Insolvency, methodology, research methods, bankruptcy prediction models.

1.0 Introduction

As vital as the business operation is to most countries' economy, firms still fail in large numbers as clearly indicated in the UK government 'The Insolvency Service' (2016) yearly published statistics, with the construction and retail usually sectors toping such list. The negative impact of such failures on economies, owners of failed firms, financiers, clients and other stakeholders can be immense. One of the major ways of preventing firms' failure, and

ensuring financiers and clients give loans and contracts respectively to only healthy firms, is by using bankruptcy prediction models (BPMs) to reveal potential failure so that mitigation steps can be quickly taken as correctly highlighted by Hafiz et al. (2015) in their conference paper. Many studies have thus justifiably attempted to build high performing BPMs. The performance of a BPM is however dependent on, among other factors, the methodological approach used to develop it.

The first work on BPM was done by Beaver (1966), an accountant, who investigated how financial ratios can be used to predict failure of firms. Beaver's aim, which was to test the usefulness of financial ratios, in itself readily side-lined the subjects i.e. firms' employees or owners, and had everything to do with the objects i.e. firms, since financial ratios of firms can be easily sourced from documents/databases. Altman (1968) built on Beaver's work by using a multivariate analysis of ratios to predict failure. The study (i.e. Altman, 1968) proved to be very successful and has more or less become a template for most BPM researchers as most subsequent studies have used a similar approach.

A number of studies have subsequently attempted to develop BPMs. However most, if not all, BPM studies (e.g. Abidali and Harris, 1995; Ng et al., 2011; Chen, 2012; Horta and Camanho, 2013) have also arbitrarily used an approach similar to that of Altman (1968). As a result, no study made mention of its research paradigm or methodological positions let alone justify them. With no BPM gaining wide acceptance, new BPM studies have continued to emerge; the improvement, whenever there is any, has however been insignificant according to Balcaen and Ooghe's (2006) comprehensive review of BPM studies' methods. Balcaen and Ooghe's (2006) review article was able to highlight many challenges with BPMs built with statistical tools and noted that artificial intelligence tools, which are potential solvers of the highlighted problems, were gradually being employed. Since this review, artificial intelligence tools have become gradually more popular with BPM studies but yet only little or no progress has been made in terms of BPM performance. The main aim here is thus to review and critique the methods and methodology used in recent BPM studies so as to explore possible methodological improvements and provide an improved methodology framework. The following are the objectives:

- To carry out a systematic review of recent BPM studies reported in English Language in order to identify their methodological positions

- To discuss and critique the methods and methodological positions taken in BPM studies
- To identify any deficiencies and propose possible improvement(s) to the methods and methodology for developing a valid and robust BPM

As with other studies, there are limitations to this work. The first is that it was impossible to review studies not reported in English language as cost of interpretation could not be met. Also, the British Library e-theses online service (EthOS), which is the only public doctoral theses database, does not host too many theses and thus returned only four doctoral theses for review in this study compared to the 66 journal articles returned by other databases.

The next section makes clear why different researches require different methodology and that systematic review with meta-analysis, which is a statistical approach, indicates that the quantitative, positivism approach is being used for this study. Section 3 explains how the systematic review was used to search four databases for journal articles and the only known public doctoral thesis database, resulting in 70 primary BPM studies. A meta-analysis of the 70 studies was done and two summary of findings tables were provided. The analysis focused on methods (section 4) and methodological trends (section5). The results in section 4 show that the use of advanced artificial intelligence (AI) techniques in building models, the consideration of error cost and model validation have been embraced by a large percentage of the studies though studies that focus on the construction industry have room for improvement on adopting AI techniques. The use of skewed data made up of many existing firms and few failed firms was found to still exist despite its problems. The philosophical underpinnings of the studies were examined (in section 5) using data collection method (external observation), data type (financial variables i.e. quantitative data), objective of studies (to predict bankruptcy) etc. and was identified to be positivism. The narrow methodology used by all the studies was critiqued and it was identified that there needs to be some subjective approach, in conjunction with the objective approach, because of the various limitations of financial variables and the need to talk to social actors involved in firm failure who understand the rudiments of the failure event. A mixed method approach with the pragmatism paradigm, along with methods improvement, were thus proposed in section 6. Section 7 concludes that upcoming studies that adopt the proposal can come up with more robust BPMs.

2.0 Methodology

As noted in Knight and Turnbull' (2008) book on Epistemology, right from the times of the ancient Greeks, researchers never really agreed on the right process of obtaining knowledge. This disagreement is still very much present in today's research world. Taking it as two warring sides, the battle is mainly between qualitative and quantitative approach, which emanate from the interpretivist and positivist research paradigms. More recently approaches like pluralism, or mixed method, have emerged and have been encouraged in research by authors like Raftery et al. (1997). Raftery et al. (1997) in their journal article where they responded robustly to Seymour and Rooke (1995) qualitative research (interpretivism) supporting work, proved the validity of mixed methods. Dainty (2008) supported this idea in his review of methodology of construction management articles. He was able to identify the conscious or unconscious pro-positivism nature of construction management research, the reluctance to adopt pluralism and the potential implications. One of the paradigms taken up by some mixed methods researchers is pragmatism.

Rooke et al. (1997) in opposition to the pluralism or multi-paradigm proposed by Raftery et al. (1997), and in an effort to make clear that paradigms are different ways of thinking that cannot necessarily be combined, made clear in their article that 'different research methods are required for different research purposes' (Rooke et al., 1997, p. 491). The authors of this study believe this argument is valid for this study. For instance, in order to review and critique the methods and methodology used in recent BPM studies, the review method has to be used. Review as research has however been criticized over the years because of its high unreliability (unrepeatability) and sometimes low validity. The systematic review through meta-analysis, which has seen off such criticisms, is thus used to assess recent BPM studies in this research. Tranfield et al. (2003) in their research journal article where they thoroughly assessed review methods were able to conclude that a fundamental feature of systematic review, as against ordinary review, is its ability to produce valid and reliable knowledge as it minimizes bias; this is why it is popular in the all-important medical research world.

The meta-analysis used here, which is a statistical approach, indicates that the quantitative, positivism approach is being adopted. Although qualitative research advocates like Hammersley have expectedly complained in his journal article response to 'on systematic reviews' that "systematic review favours quantitative methods and embodies a scarcely concealed positivism that places qualitative research far down the 'credibility hierarchy"

(2001: 545), such criticism has not in any way proven to reduce the credibility of systematic reviews.

3.0 Research Method

The systematic review method was used to reveal the methods and methodological trends in the bankruptcy prediction models (BPM) study area. The general review of various existing knowledge is a recognised way of contributing to the progression and expansion of knowledge as noted in Aveyard's (2007) book on doing a literature review. This is why it has been widely used as method in various research areas including insolvency prediction studies (e.g. Balcaen and Ooghe 2006) and construction business failure studies (e.g. Edum-Fotwe et al., 1996; Mahamid, 2012).

The single research question here is 'what are the methods and methodology trends in the BPM research area'. To ensure quality and validity, peer reviewed journal articles were considered for their quality and doctoral theses were considered for their thoroughness. The following popularly used robust databases were considered for the journal articles search: Google Scholar (GS), Engineering Village (EV), Wiley Interscience (WI), Science Direct (SD), Web of Science UK (WoS), and Business Source Complete (BSC). Observations revealed that GS, EV, WoS and BSC contained all the journal articles provided in Wiley and Science Direct; this is probably because the later are publishers and their results are limited to articles they have published. The British Library e-theses online service (EthOS) was used for doctoral thesis search since it is the only known public doctoral theses database.

The initial searches in the four databases (GS, WoS BSC and EV) showed that studies tend to use bankruptcy, insolvency and financial distress as synonyms for failure of firms. A search framework which captured all these words was thus designed with the following defined string ("Forecasting" OR "Prediction" OR "Predicting") AND ("Bankruptcy" OR "Insolvency" OR "Distress" OR "Default" OR "Failure"). A process flow of the methodology is presented in Figure 1.

Insert Figure 1 here

To eradicate database bias, ensure high consistency and repeatability, and consequently reliability and quality, all relevant studies that appeared in the four databases were used. These databases contain studies from all over the world hence geographic bias was also eliminated. Considering that Balcaen and Ooghe's (2006) comprehensive BPM review was done in 2006, and to make this review manageable, a start date of 2010 was chosen.

Generally, the topic of the articles that emerged from the search looked okay to determine which ones were fit for the problem being studied. However, this was not the case for all articles. Where otherwise, articles' abstract was read and, if necessary, introduction and/or conclusion were read. In some cases, the complete articles had to be read.

For exclusion, after eliminating unrelated studies that dealt with topics like 'default prediction for surety bonding' (e.g. Awad and Fayek, 2012) and 'contractor default prediction prior to contract award' where focus is on the ability of a contractor to complete a particular type of project (e.g. Russell and Jaselskis, 1992) rather than its insolvency probability, only 70 studies were left: 66 journal articles and four doctoral theses. The 'review studies' in the search results (e.g. Edum-Fotwe et al., 1996) were not considered since the interest is in the methods and methodology of studies that built and/or tested BPM(s).

In the final sample of 70 studies, where the result of more than one technique is presented, only the result of the proposed technique in the study is used to represent accuracy of the study, though all techniques used are presented in the summary of findings table. Where none of the multiple techniques used in a study is proposed by the study, the highest accuracy result is reported. Where the proposed technique is tested on training samples and on test (or validation) sample, the accuracy result for the validation sample is used. Where there is no validation, the given result is used. Where error types are not considered separately from accuracy results and the Receiver Operating Characteristic (ROC) curve is used to measure performance, the area under the curve (AUC) value in its percentage form is taken as the accuracy result. In cases where the results of more than a year of prediction were given, the results of the first year were used to allow for fair comparison since most BPM studies normally present only first year results.

The assessments of the 70 sample studies were carried out under two main themes: 'general methods' and 'philosophical underpinning or paradigm'. A meta-analysis was done with data synthesised through the use of summary of findings tables as required for systematic review

according to various research articles on conducting systematic review (e.g. Khan et al., 2003; Smith et al., 2011) and the highly regarded Cochrane handbook for systematic reviews by Higgins (2008); with key points discussed sometimes backed up with a wider review of literature. Two summary of findings tables were provided.

4.0 General methods of Construction Industry-Bankruptcy Prediction Model Studies

This section deals with the types of techniques, sample characteristics and model testing features of construction industry-bankruptcy prediction model (BPM) studies. Data dispersion is the main data characteristic analysed. In each case, the topic is briefly explained, the trend in BPM is presented through a table, and the effectiveness or poorness of the method/feature and trend are discussed. Table 1 is the first summary of findings table. It contains the 'general methods' features of the primary studies.

Insert Table 1 here

4.1 Techniques Used

The technique used to build a BPM plays a big role in the performance of such model. The two main categories of technique are the statistical and artificial intelligence (AI) techniques, although there is also the uncommon option based model formula.

There has been a surge in the use of AI techniques since 2010 as suggested by Balcaen and Ooghe (2006). Over 80% of the reviewed studies used some form of AI techniques compared to the 67.14% that used statistical techniques. Even so, many studies only employed statistical techniques for comparison purpose (e.g. Heo and Yang, 2014; Cheng and Hoang, 2015 among others). Artificial neural network (ANN) is the most commonly used tool/technique overall (62.86%) while support vector machine (SVM) is the second in the AI techniques queue (38.6). This is not surprising given that many BPM studies that have compared various BPM techniques (e.g. Tseng and Hu 2010; Du Jardin 2010; Yoon and

Kwon 2010; Huang et al. 2012; De Andrés et al. 2012) have identified these two as giving the most accurate results. Their main disadvantage is that provide uninterpretable results.

The most popular statistical techniques as noted by Balcaen and Ooghe (2006) in their comprehensive review of BPMs are multiple discriminant analysis (MDA) and Logistic regression (LR). This trend remains the same as confirmed Table 1, with MDA and LR used by 40% and 55.7% of the reviewed studies respectively. The next most popular statistical technique was probit analysis (PA) which was only used by 4.3% of the studies. The reduction in use of statistical tools represent a positive trend as they have many disadvantages, which have been identified in many BPM research and review studies (e.g. Balcaen and Ooghe 2006; Du Jardin and Séverin 2011; Ng et al., 2011; Tserng et al., 2015 among others) over the years. These disadvantages are normally in form of restrictive assumptions on data including multivariate normality of independent variables, equal group variance-covariance, groups are discrete and non-overlapping etc. All these restrictive assumptions can barely be satisfied together by one data set hence are violated in many studies as identified in the earlier cited studies. Nonetheless LR is deemed relatively less demanding compared to MDA according to Balcaen and Ooghe's, (2006) review and Wood's (2012) PhD thesis on BPMs. The high use of LR as compare MDA and PA hence a positive situation. The barrier option models (5.7%) represent the least used tools.

A survey of the studies that concentrated on the construction industry showed that only 57.9% of them used AI techniques, with the same percentage using statistical tools. This (percentage use of AI techniques) is low and should be improved upon by researchers focusing on the construction industry. The studies that used either or both AI technique and statistical **tools** represent 95%. The only study that used neither of them used a barrier option model

4.2 Data or Sample Characteristics

Data or sample characteristics are very important to the performance BPMs. Data dispersion, defined as the ratio of failing to non-failing firms or vice versa in a sample data, plays a significant role in building a BPM. The relatively high number of existing firms compared to failed firms means that data available to build BPMs are normally highly skewed. This problem has long been identified and a number of techniques have been proposed as solution:

- Technique balancing: the process whereby the technique employed to develop the model has a special feature that is used to balance/equalise the data dispersion.

- Over (under) sampling: is the process whereby the smaller (bigger) group is increased (decreased) until the number of firms in it equal that of the bigger (smaller) group. The increment in oversampling is usually done by using average values of data of the firms in the group to form data of new firms while the decrement in under sampling is done by matching firms with similar properties (size, turnover etc.) from the bigger group to those of the smaller group and the rest discarded.

Of the 70 primary studies reviewed, 68 studies clearly presented the level of dispersion of data. Of this 68, only 38.2% (**or 24 studies**) used equal data dispersion. This is a very poor figure and shows that many authors do not recognise the importance of the data dispersion characteristic. The figure (only 16.7) is even worse concerning the studies that concentrated on the construction industry. This is an area that clearly needs the attention of researchers in the area so that necessary improvements/adjustments can be made to ensure more balanced data is used to develop BPMs.

All studies that used balanced data used the under sampling method (note that random and matching methods are forms of under sampling). The only study that used oversampling (i.e. Cheng and Hoang 2015) eventually used an unbalanced data. Critics of undersampling are sceptical, as mentioned in their research articles (e.g. Cheng and Hoang 2015), that deleted data might be those that are crucial to the learning/development process of any technique. Critics of data balancing in general have also argued in their works (e.g. Tserng et al. 2011b) that it leads to sampling bias and thus the entire population should be used. However it is well established in many studies BPM developing studies (e.g. Neves and Vieira, 2000; Ng et al., 2011; Du Jardin, 2015) that using skewed data generally results in the model being more accurate for predicting the lager group. This means the model will be more likely to incorrectly predict a bankrupt firm as non-bankrupt than vice versa; this is the costlier of the two BPM error types and needs to be well avoided hence equal data dispersion is more appropriate for developing BPMs. Overall it can be concluded that with more than half BPM studies using unequal data dispersion, many BPMs must have been suboptimal.

4.3 Result related trend

The result related trend discussed here has to do with how developed BPMs are tested or validated in the primary studies. It also explores the number of studies that considered error cost when testing models. Error cost is sort of a 'play safe' factor used to improve BPMs.

4.3.1 Validation

In the early days of BPM research like in Altman's (1968) study, it was common to test a developed model with the data used to build the model. Such tests yielded very accurate results. However, with the models not performing as well in practice, further research by Joy and Tollefson (1975) quickly revealed that BPM developers confused ex-post classification results with ex-ante predictive abilities. Many studies (e.g. Joy and Tollefson 1975; Moyer 1977 among others) hence recommended that a built model should be tested on separate data apart from that which was used to build it if there is to be any confidence in the model. This practice has now almost become a norm in the BPM research world with the separate data usually referred to as test or validation data.

Normally the data is pre-divided, usually into a ratio of between 80-20 and 70-30, the bigger portion used for training and the smaller used for validation. The case where the validation samples are removed in batches such that the entire sample, at different times, forms part of the training or validation sample is known as cross validation.

The results reveal a 'just' acceptable trend of BPM studies where less 13% of them did not validate (or failed to report validation of) the developed model. The studies that focused on construction firms had a higher percentage (21%), depicting a less acceptable trend, especially when it is considered that an invalidated model is as bad as any existing model. Although Ng et al. (2011), one of the primary studies reviewed, claimed to have validated their model, that claim is unacceptable because the validation was done using earlier years data of the sample used to build the BPM. Unfortunately, the immediate earlier year's data of a model building sample is not an acceptable replacement for data of firms that were not used in the model building process. No wonder the model misclassified only one firm out of the 32 firms selected from the model building sample to validate the firms. This single misclassification was even put down to unequal data dispersion by Ng et al. (2011) as they tried to explain that the model is very perfect.

Of the nine from the 70 reviewed studies that did not validate their model, only two (80.8% and 79%) reported accuracy figures of below 85%, depicting highly accurate models; such accuracy figures are questionable. In fact, with AI techniques like ANN and SVM, it is very possible to build a model with a 100% prediction accuracy when tested on training (i.e. model building) data. However, such models are not usually very good on test/validation data and are normally condemned for what is known as 'overfitting' to the training data; this makes such models rather poor

4.3.2 Error Cost consideration

There are two types of error in bankruptcy prediction. Type I error where a bankrupt construction firm is wrongly predicted as healthy, and type II error where a healthy construction firm is wrongly predicted as going bankrupt. It is common knowledge that type I error is costlier than type II error. For example, the cost of awarding contracts to an impending contractor who might fail will typically be much larger than the cost of rejecting a healthy contractor.

In order to consider error cost in developing a BPM, it is either a study reports overall accuracy and error values for both type of errors, or simply use sensitivity analysis employing the receiver operating characteristic (ROC) curve where the area under the curve (AUC) represents the accuracy of the model.

Table 1 shows that only 72.9% of the studies considered error type either directly or through the use of ROC; again this gives an acceptable outlook. The studies that focused on construction firms had a higher percentage (84.2%), depicting a more acceptable trend. This trend will bring more confidence to BPM users since it ensures that a failing firm is barely ever mistakenly predicted as a healthy one. This means stakeholders of firms predicted as healthy can be double sure the firm is healthy while stakeholders of firms wrongly predicted as failing will take steps that will ensure their firms become even healthier thereby losing almost nothing. This will also ensure bankers or clients never give loan or contract respectively to a failing firm. The only disadvantage in this case, which is less costly compared to other explained options, is the possibility of a healthy firm missing out on a loan or contract because it is wrongly predicted as failing.

5.0 Paradigm of Construction Industry-Bankruptcy Prediction Model Studies

This section explores the philosophical underpinning or research paradigms of bankruptcy prediction model (BPM) studies. Paradigm, according to Guba's (1990) book on paradigm dialog, is a research culture, a set of assumptions, values and belief, which comprise of but is not limited to epistemology, axiology, ontology, methodology, and aesthetic beliefs. Table 2 presents more summary of findings of the primary studies which will, along with Table 1, be used to analyse the philosophical underpinning(s) of the primary studies

Insert table 2 here

The data collection trend in BPM studies appear to be that of independent observers as most of the primary studies (80%) either used only financial ratios, or financial ratios in combination with other externally observable financial variables like stock market information (14.28%) (Table 2). A few others (2.86%) used credit card information alone while only two (or 2.86%) studies included some form of non-financial variables. Studies that focused on construction firms had similar statistics, with 78.9% using financial ratios alone. All variables are generally used to measure the health of a firm, giving the studies a positivism outlook since the positivism paradigm believes that research can mainly be done by observations and measurements as explained by Levin (1988) in his book on Nihilism and the postmodern situation.

Financial variables, including stock market information, are common in BPM studies mainly for two reasons:

1) Financial ratios are the variables used by the two successful pioneering studies (Beaver 1966; Altman 1968) that most BPM studies are emulating

2) Financial data are usually readily available from third party databases or in publicly available company archives (see Table 2) and thus makes data collection very easy for a researcher.

Of the two primary studies that used non-financial variables (see Table 2), Horta and Camanho, (2013) combined three strategic variables with six financial ratios; their strategic variables chosen from their previous study. The value for each of the three variables

(company main activity, company size and headquarter geographic location) were accessible from archives or financial databases, suggesting a form of independent observation which is a positivist approach. A positivist researcher, is normally independent (of the subject) as an observer, reduces a phenomenon to simpler measurable factors/ elements, explains the elements in terms of how they affect the phenomenon (cause and effect) and usually uses large samples (Burrell and Morgan, 1979; Easterby-Smith et al., 1991). Positivism seeks to explain and predict what happens in the social world by searching for patterns and relationships (Burrell and Morgan, 1979).

In BPM studies, it is the complex failure process (phenomenon) that is reduced to measurable variables, usually financial and/or non-financial variables (simpler elements) measured from databases and/or through Likert scale questionnaire respectively. The relationship between each variable and the failure process is then explained in the studies and the importance of each variable highlighted, usually through a statistical process (see 'variable selection method' column in Table 1), before they are used. An example quote of where primary studies explained a used variable relationship to failure are as follows:

The activity ratio measures "*how well a company has been using its resources*" (Ng et al., 2011, p.601).

Apart from Othman's (2013) doctoral thesis and Stroe and Bărbuță-Mișu (2010) which is a construction industry focused study, that used a sample sizes of 10 and 11 respectively, the least sample size in the primary studies is 35 firms. The mean average firm sample size used by the 70 studies is 7472.7 while that of studies that focused on construction firms in 7506.8. Removing the big numbers outliers (20,000 and above) like 109361 sample firms used by Tsai and Hsu (2013), the mean for the 70 studies and those that focused on construction firms become 1479.8 and 965.59 respectively. Marshall et al. (2013) in their comprehensive review of sample sizes in qualitative research, using 83 qualitative studies from top international journals, clearly proved a sample size of 30 to be high and that saturation is normally reached before reaching this figure (i.e. 30). Majority (97.1%) of the studies used a large sample size, well beyond the 30 figure, which is what is common with quantitative studies. In fact, more than 75% of the studies used above 100 samples. Using large samples, as noted earlier, is a feature of positivism

In positivism, the significance of research is to understand the world that it might be predicted and controlled as explained by Hempel (1965) is his book chapter on aspects of scientific explanation. This is exactly the aim of most BPM studies (see aims of studies in Table 2). In the studies, an attempt is made to understand failure of construction firms and to identify failure indicators; then there is effort to predict potential failure in order to aid control of the situation by owners taking mitigating steps. The aim of BPM studies thus in itself, to an extent, lend them to positivism.

According to Burrell and Morgan (1979) the functionalist/positivist is always seeking to find implementable solutions to real problems and is more concerned with controlling social affairs. This is well in line with the aim of BPM studies which try to provide BPM as a solution to the real problem of either high rate of firm's failure or to the problem of identifying healthy companies for loans or contract. Further, positivist tend to use statistical analysis so as to aid generalization according to Keat's (1979) work titled positivism and statistics in social science; this is typical of BPM studies as shown in Table1. Note that artificial intelligence (AI) techniques are advanced statistical/mathematical methods.

From all the evidences given in this section, it appears that the positivism/functionalism paradigm is predominant in the BPM literature. This is well understandable since prediction, the main aim of the studies, is a main feature of positivism. Although critical realism also supports quantitative data and analysis, and possess some features similar to those of positivism, it is not used mainly for prediction. A critical realist is also not an independent observer, i.e. an objectivist, as is with BPM researchers.

5.1 Implication of the Narrow Methodological Positions in BPM Studies

The restricted use of positivism in BPM studies has led to the continuous exclusive use of the objectivism (Keat 1979), through the use of multivariate analysis of financial ratios. Unfortunately, this singular dimension approach does not fully represent the insolvency situation of many firms including those from the construction industry as highlighted in various studies; and due to the dynamism of the construction industry.

On facts highlighted in various studies, countless number of non-financial indications of insolvency, such as management mistakes, do come up a lot earlier than financial distress as

noted in Abidali and Harris' (1995) BPM study. Financial distress only tends to be noticeable when the failure process is almost complete, around the last two years of failure according to Abidali and Harris (1995). The truth is that it is adverse managerial actions and other social factors that lead to poor financial standings and in turn cause insolvency. Accordingly, many experts, like Argenti (1980) in his popular corporate planning book, have reiterated that financial variables alone are insufficient for early depiction of disastrous factors like shambolic management, acquisition of a failing firm, economic decline etc. (Argenti, 1980). Argenti (1980) and Edmister (1972) in his research on testing financial ratio analysis also stressed that financial ratio models are not enough to predict insolvency of firms until they are used with other economical, managerial and social factors. Further, the tendency of accountants to amend important financial ratios, known as window dressing or creative accounting, reduces the reliability of financial ratios as factors for predicting insolvency according to Degeorge et al., (1999)study on earnings management

In addition, in many countries (including UK, France, etc.) only certain firms that meet some specific criteria like a pre-set minimum asset size, number of employees, etc. are required by law to produce financial statements periodically hence very small companies might not have periodic financial statements (Balcaen and Ooghe, 2006). According to the UK Federation of Small Businesses (2016), in 2015, small businesses accounted for 99.3% of all UK businesses; small or medium-sized (SMEs) account for 99.9%; 76% were 'one-man business'; SMEs represent at least 99% of the businesses in every main industry sector; just under a fifth of all SMEs operate in Construction, compared to just 1% in the Mining, Quarrying and Utilities sector. A lot of SMEs do not have a good accounting system and hence only produce/submit very incomplete and inadequate financial statements; this automatically nullifies the possibility of their involvement in BPM studies since incomplete financial statements are normally discarded of in BPM studies (Balcaen and Ooghe, 2006). This means most BPM studies build models that are not applicable to small firms despite the fact that they (i.e. SMEs) are well known to make up a larger percentage of the failing firms.

Another problem is that some SMEs do outsource their account management to independent accounting firms with the sole intention of ensuring periodic production of their financial statement in order to satisfy legal/government requirements. This sometimes leads to misrepresented financial statements based solely on the amount of information provided to the accounting firm by the construction firm. In a similar fashion, some SME firms simply

produce poor and inaccurate statements themselves simply to fulfil the legal requirements. Any BPM developed from such statements will have limited practical usefulness.

The dynamic nature of many industries with constantly changing trends, as identified for example by Razak Bin Ibrahim et al., (2010) in their study on the dynamics of the global construction industry, means the main causes of failure of firms will vary from time to time. This implies that key players like owners, directors, managers etc. (i.e. subjects) will have to be spoken to in order to understand key reasons behind failure of firms at different times. Ultimately, leaving out the subjects appears not to be a wise choice if a valid BPM is to be built as identified by Alaka et al., (2015) in their conference paper.

The need to involve social factors, which can mainly be considered through a subjective approach, and the need to talk to subjects to understand the timely dynamics of the industries, both call strongly for the adoption of the subjective epistemology in BPM studies. Remenyi et al. (1998, p.35) stressed the importance of investigating "the details of the situation to understand the reality or perhaps a reality working behind them" in their book on doing research in business and management. This is only achievable subjectively since subjectivism emphasizes on seeking explanation to understand a social phenomenon (Burrell and Morgan, 1979). A very good understanding of the failure process of firms by a BPM developer will definitely contribute to an improved model.

6.0 Proposed Improvements for the Methods and Methodology for Developing a Robust BPM

It has been proven in many studies that a hybrid of AI techniques is the best technique for BPMs hence it is proposed here. Effective hybrids perform better than standalone techniques, as proven in many studies that compared hybrid BPMs with standalone technique BPMs (e.g. Tsai, 2014; Zhou et al., 2014; Iturriaga and Sanz, 2015 among others), and generally outperforms even the ANN and SVM which are known to be the most accurate BPM techniques. "However, these hybrid models consume more computational time" (Zhou et al. 2014, p.251).

Equal data dispersion is vital to building a good BPM. Since available financial data is normally skewed, one of the data balancing methods should be used to equalise the dispersion. No method has been clearly proven as being the best. All BPMs should at least go

through some form of model validation using a separate sample if they are to have any practical validity. Error costs also need to be considered if BPMs are to be robust.

Concerning paradigms, having reviewed numerous BPM studies, the authors propose that the best paradigm for developing a BPM is pragmatism. Pragmatism argues that the main determinant of the methodology to be used in a research should be the research questions according to the mixed methodology book by Tashakkori and Teddlie (1998) rather than strictly following a particular paradigm because of a sociological belief, or so as to copy past studies as done in BPM studies. Pragmatists are more concerned with the practical consequences of the research findings and as such believe that one standpoint can never be suitable for answering all types of research questions and there may be multiple realities according to James' (1995, 1907 original) book on pragmatism. This is the maximalist view, noted by Johnson and Onwuegbuzie, (2004) in their mixed methods research study, which argues that nothing in a research phenomenon can escape pragmatics. Pragmatists neither agree with positivists in that demands of a research cannot be fully satisfied by a theory (falsify-ability, objectivity, etc.), nor with interpretivists in that demands of a research can be satisfied (at least partly) by almost any theory (Rorty, 1985). Pragmatism thus allows the use of any, or a mix of multiple methods, approaches, choices, techniques etc. as long as they will help to answer the research questions properly. It allows the researcher to "study what interests you and is of value to you, study in the different ways in which you deem appropriate, and use the results in ways that can bring about positive consequences within your value system" (Tashakkori and Teddlie, 1998, p.30).

The authors believe the relative rigidity of other paradigms as to the methodological positions that fit a research can limit steps needed to be taken to complete quality research, Further, a good BPM study should focus on failure of construction firms (a problem) experienced by construction firms owners (people) and the effect of developing a BPM which will allow timely intervention that can prevent potential failure (consequence of inquiry). Such focus is synonymous with pragmatism which accentuates practical problems experienced by people, research questions in focus and consequences of inquest" (Dewey et al. 1983).

Realism, which is the reality assumed by positivists (Guba and Lincoln, 1994), which is thus used for BPM studies is very appropriate and is consequently proposed here. There is only one reality and it is that 'construction firms do fail and failing construction firms have certain similar attributes'. Finding the most effective attributes to develop a BPM is what is tricky.

This is one of the reasons there are many BPM studies, each trying to prove certain attributes are more effective than others.

Although the objectivism as a feature of positivism (Guba and Lincoln, 1994, p. 109) is suitable for developing a BPM, a combined subjective and objective approach in a facilitation manner is proposed here. While the objective approach will aid the use of existing factors and variables, the subjective approach can be used to identify temporal factors and variables that can be used to develop a timely and robust BPM; this would have taken the dynamism of the construction industry into consideration. The subjective approach can also help identify important social and managerial factors that contribute to insolvency of the construction firms. This has long been advocated by many studies including those in the construction management (CM) area (e.g. Seymour et al., 1997; Dainty, 2008), who queried the focus on objects when at the centre of most CM research is people (subjects), justifying the need for greater emphasis on qualitative enquiry. Management level staff and/or owners of failed and existing firms can use their practical experience to contribute vital information in terms of factors that affect insolvency and survival of firms hence they need to be engaged.

Since both the objectivism and subjectivism are vital for a valid BPM, the integration of quantitative and qualitative research approach is proposed here. This is in line with the much advocated methodological pluralism (Seymour et al., 1997; Mingers and Gill, 1997) which combines methodologies from varying paradigms to provide richer insights into relationships and their interconnectivities (between factors and firm failure in this case); this is the best approach to solving research problems as explained by Mingers and Gill, (1997) in their multimethodology book. The use of the dual (objectivism and subjectivism) approach is only possible with mixed methods which in itself is a feature of pragmatism according to Johnson and Onwuegbuzie's (2004) study. Mixed method can combine the strengths of different methods to provide a more robust approach to answering a research question and avoid preconceived biases which might exist in past studies that are being copied. For example, many BPM studies ended up using unequal data dispersion simply by copying methodologies of past studies

In this vein, the proposed methodology agrees with the popular Seymour and Rooke (1995) work which clearly argued that different researches require different methods and no method should be ruled out a priori. However, it does not support their opposition to the multi-paradigm (see Rooke et al., 1997) approach which pragmatism allows if it is what will bring

about a valid methodology. In fact, such an opposition is tantamount to nullifying some methods a priori since selecting a particular paradigm readily nullifies some methods; an act Rooke et al. (1997) themselves preach against. An improved research methodology framework for developing a BPM is given in Figure 2.

The subjective aspect of the work could be executed with the multiple case study strategy. Case study is defined by Mitchell (1983, p. 192) in his study as a "detailed examination of an event (or series of related events) which the analyst believes exhibits (or exhibit) the operation of some identified general theoretical principles". Yin (1994, p. 13) defined a case study in his popular text on case study researcher as "an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident. Yin (1994) went ahead to explain that case studies usually require more than one source of evidence. The reason for proposing case study is to allow a comprehensive investigation of the failure phenomenon by investigating multiple failed and existing construction firms. The case studies can be accomplished using unstructured interviews and/or maybe focus group discussions. The temporal factors obtained (because of the dynamism of various industries) should be analysed to identify befitting measuring variables which can then be measured with a survey research strategy. This process of identifying new factors and discovering patterns from the field culminates in an inductive research approach (Easterby-Smith et al., 1991)

Insert figure 2 here

The objective aspect should be executed using survey and archival research strategies. Survey can be executed with a Likert scale questionnaire, allowing respondents to rate the extent to which each variable (factor), identified from the case studies, applies to their construction firm. Archival research strategy, which involves collecting financial data from companies' archives, financial databases or stock market is the common process in BPM. The term 'archival' in this strategy does not directly mean 'old' in anyway as pointed out by Bryman (1989) hence using recent financial statements also fall under this category.

The proposed strategy shows the intended mixed method approach (qualitative and quantitative data and analysis) which is a very good approach since it ensures an all-round

effectiveness of research (Creswell and Clark 2007) and is well in line with the proposed pragmatism philosophical stance (Johnson and Onwuegbuzie 2004).

7.0 Conclusion

Judging from the results of this systematic review, there appear to be some methodological issues with bankruptcy prediction model (BPM) studies. While it appears that there is improvement in relation to some methods trend, others appear quite stagnant. The results show that the use of advanced artificial intelligence (AI) techniques in building models and the consideration of error cost in testing models have been highly embraced generally though studies that focus on the construction industry have room for improvement on adopting AI techniques. Also, there exist an acceptable model validation with more than 80% validating their models. The use of skewed data made up of many existing firms and few failed firms still occurs in the BPM world despite the popular knowledge that it leads to skewed accuracy in favour of the larger group.

From the philosophical angle, positivism is about the only paradigm used in BPM studies. This is not surprising since BPMs are usually about large sample size, statistical analysis, operationalization, prediction, generalization, etc. The ease of getting financial ratios or variables through questionnaires explains why objectivism is common with BPM researchers. However, the need to consider the dynamic nature industries/sectors, the flaws of financial variables, and social factors which actually lead to the financial status of firms, make BPM studies' methodology inadequate. The key area of improvement would be to first use a subjective approach, through case studies for example, to identify temporal social factors affecting failure of firms. These factors can capture what happens before the financial state of a firm actually deteriorates and thus help build early warning systems which are more useful as they allow more time for correctional strategies. The factors can subsequently be analysed to establish measurable variables to be used with financial variables, to build a robust BPM. To allow this type of mixed method approach, the pragmatism paradigm has to be used.

The implication to practice here is that BPM developers will be urged to go through a more rigorous and robust methodology in order to deliver better and more valid models. This will in turn result into the availability of better effective models and potential reduction in

bankruptcy of firms as the models can call owners to attention early enough. The implication to study is that an intellectual debate on the validity and robustness of BPM studies' methodology is opened up. Also, a number of studies will now adopt more rigorous methods in building BPMs and there will be a lot of beneficial comparison among studies that have adopted various rigorous methods.

As with other studies, there are limitation to this work. The first is that it was impossible to review studies not reported in English language as cost of interpretation could not be met. The primary studies examination of longitudinal analysis of ratios, impact of different countries accounting practices on ratios, and consequences of consolidated accounts of large construction firms, were not reviewed here. Also, the British Library e-theses online service (EthOS), which is the only public doctoral theses database, does not host too many theses and thus returned only four doctoral theses for review in this study compared to the 66 journal articles returned by other databases. Future review studies should endeavour to take some, if not all, of these limitations into consideration.

Future reviews should also look into BPMs definition and differentiation criteria for failed and non-failed construction firms. There is also need to check the variable selection trend and how studies arrive at an optimized solution when using AI techniques. Upcoming BPM studies should endeavour to combine necessary features as proposed in section six in order to achieve better models.

References

Abellán, J., and Mantas, C. J. (2014). Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, **41** (8), pp. 3825-3830.

Abidali, A.F. and Harris, F. (1995). A methodology for predicting failure in the construction industry. *Construction Management and Economics*, **13** (3), pp. 189–196.

Alaka, H, Oyedele, L, Toriola-Coker, O, Owolabi, H, Akinade, O, Bilal, M and Ajayi, S (2015) Methodological approach of construction businesses failure prediction studies: A review In: Raidén, A B and Aboagye-Nimo, E (Eds) *Procs 31st Annual ARCOM Conference*, 7-9 September 2015, Lincoln, UK, Association of Researchers in Construction Management, pp. 1291-1300.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, **23** (4), pp. 589-609.

Argenti, J. (1980) Practical Corporate Planning. London: Allen and Unwin.

Arieshanti, I., Purwananto, Y., Ramadhani, A., Nuha, M. U., and Ulinnuha, N. (2013). Comparative Study of Bankruptcy Prediction Models. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, **11** (3), pp. 591-596.

Aveyard, H. (2007). *Doing a Literature Review in Health and Social Care*. London: McGraw-Hill

Awad, A., and Fayek, A. R. (2012). Adaptive learning of contractor default prediction model for surety bonding. *Journal of Construction Engineering and Management*, **139** (6), 694-704.

Balcaen and Ooghe (2006). 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, **38** (1) pp. 63–93

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, **4** (1966), pp. 71-111.

Bemš, J., Starý, O., Macaš, M., Žegklitz, J., and Pošík, P. (2015). Innovative default prediction approach. *Expert Systems with Applications*, **42** (17), pp. 6277-6285.

Bryman, A. (1989). *Research Methods and Organisation Studies*. London: Unwin Hyman.

Burrell, G. and Morgan, G. (1979). *Sociological Paradigms and Organisational Analysis (Vol. 248*). London: Heinemann.

Callejón, A. M., Casado, A. M., Fernández, M. A., and Peláez, J. I. (2013). A system of insolvency prediction for industrial companies using a financial alternative model with neural networks. *International Journal of Computational Intelligence Systems*, **6** (1), pp. 29-37.

Chen, M. Y. (2011). A hybrid model for business failure prediction-utilization of particle swarm optimization and support vector machines. *Neural Network World*, **21** (2), pp. 129-152.

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

Chen, N., Ribeiro, B., Vieira, A. S., Duarte, J., and Neves, J. C. (2011a). A genetic algorithm-based approach to cost-sensitive bankruptcy prediction. *Expert Systems with Applications*, **38** (10), pp. 12939-12945.

Chen, H. L., Yang, B., Wang, G., Liu, J., Xu, X., Wang, S. J., and Liu, D. Y. (2011b). A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method. *Knowledge-Based Systems*, **24** (8), pp. 1348-1359.

Cheng, C.B., Chen, C. L. and Fu, C.J. (2006). Financial distress prediction by a radial basis function network with logit analysis learning. *Computers and Mathematics with Applications*, **51** (3–4), pp. 579–588.

Cheng, M. Y., and Hoang, N. D. (2015). Evaluating Contractor Financial Status Using a Hybrid Fuzzy Instance Based Classifier: Case Study in the Construction Industry. *Engineering Management, IEEE Transactions on*, **62** (2), pp. 184-192.

Cheng, M. Y., Hoang, N. D., Limanto, L., and Wu, Y. W. (2014). A novel hybrid intelligent approach for contractor default status prediction. *Knowledge-Based Systems*, **71** (2014), pp. 314-321.

Cho, S., Hong, H., and Ha, B. C. (2010). A hybrid approach based on the combination of variable selection using decision trees and case-based reasoning using the Mahalanobis distance: For bankruptcy prediction. *Expert Systems with Applications*, **37** (4), pp. 3482-3488.

Chuang, C. L. (2013). Application of hybrid case-based reasoning for enhanced performance in bankruptcy prediction. *Information Sciences*, **236**, pp. 174-185.

Creswell, J.W. and Clark, V.L.P. (2007). *Designing and Conducting Mixed Methods Research*. California: SAGE Publications

Dainty, A. (2008). Methodological pluralism in construction management research. *Advanced research methods in the built environment*, **1**, pp. 1-13.

De Andrés, J., Lorca, P., de Cos Juez, F. J., and Sánchez-Lasheras, F. (2011a). Bankruptcy forecasting: A hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS). *Expert Systems with Applications*, **38** (3), pp. 1866-1875.

De Andrés, J., Landajo, M., and Lorca, P. (2012). Bankruptcy prediction models based on multinorm analysis: An alternative to accounting ratios. *Knowledge-Based Systems*, **30,** pp. 67-77.

De Andrés, J., Sánchez-Lasheras, F., Lorca, P., and Juez, F. J. D. C. (2011b). A hybrid device of self organizing maps (SOM) and multivariate adaptive regression splines (mars) for the forecasting of firms'bankruptcy. *Accounting and Management Information Systems/Contabilitate si Informatica de Gestiune*, **10** (3), pp. 351-374.

Degeorge, F., Patel, J., Zeckhauser, R., (1999). Earnings management to exceed thresholds. *Journal of Business*, **72** (1), pp. 1–33.

Dewey, J., Boydston, J. A., and Ross, R. (1983). *The middle works, 1899-1924* (Vol. 13). Illinois: Southern Illinois University Press

Divsalar, M., Firouzabadi, A. K., Sadeghi, M., Behrooz, A. H., and Alavi, A. H. (2011). Towards the prediction of business failure via computational intelligence techniques. *Expert Systems*, **28** (3), pp. 209-226.

Divsalar, M., Roodsaz, H., Vahdatinia, F., Norouzzadeh, G., and Behrooz, A. H. (2012). A Robust Data-Mining Approach to Bankruptcy Prediction. *Journal of Forecasting*, **31** (6), pp.504-523.

Du Jardin, P. (2010). Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy. *Neurocomputing*, **73** (10), pp .2047-2060.

Du Jardin, P., and Séverin, E. (2011). Predicting corporate bankruptcy using a self-organizing map: An empirical study to improve the forecasting horizon of a financial failure model. *Decision Support Systems*, **51** (3), pp. 701-711.

Du Jardin, P., and Séverin, E. (2012). Forecasting financial failure using a Kohonen map: A comparative study to improve model stability over time. *European Journal of Operational Research*, **221** (2), pp. 378-396.

Du Jardin, P. (2015). Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research*, **242** (1), pp. 286-303.

Easterby-Smith, M., Thorpe, R. and Lowe, A. (1991). *Management Research: An Introduction*. London: Sage Publications.

Edmister, R. O. (1972). An empirical test of financial ratio analysis for small business failure prediction. *Journal of Financial and Quantitative Analysis*, **7** (02), pp. 1477-1493.

Edum-Fotwe, F., Price, A. and Thorpe, A. (1996). A review of financial ratio tools for predicting contractor insolvency. *Construction Management and Economics*, **14** (3), pp. 189- 198.

Federation of Small Businesses (2016). *Small Businesses Statistics* [online]. Available from: http://www.fsb.org.uk/media-centre/small-business-statistics [Accessed 10 July 2016].

Gepp, A., Kumar, K., and Bhattacharya, S. (2010). Business failure prediction using decision trees. *Journal of forecasting*, **29** (6), pp. 536-555.

Gordini, N. (2014). A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy. *Expert Systems with Applications*, **41** (14), pp. 6433-6445.

Guba, E. G. (Ed.). (1990). *The paradigm dialog*. Sage Publications.

Guba, E. G., and Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*. Thousand Oaks, CA: Sage.

Hafiz, A., Lukumon, O., Muhammad, B., Olugbenga, A., Hakeem, O., and Saheed, A. (2015, March). Bankruptcy Prediction of Construction Businesses: Towards a Big Data Analytics Approach. In *Big Data Computing Service and Applications (BigDataService), 2015 IEEE First International Conference on*. Redwood City, California, March 30 - April 2 2015. IEEE, pp. 347-352.

Hammersley, M. (2001). On 'systematic' reviews of research literatures: a 'narrative' response to Evans and Benefield. *British educational research journal*, **27** (5), pp. 543-554.

Hempel, C., (1965). Aspects of scientific explanation, in his *Philosophy and Phenomenological Research*. New York: Free Press.

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

Hernandez Tinoco, M. (2013). *Financial distress and bankruptcy prediction using accounting, market and macroeconomic variables*. (Doctoral dissertation, University of Leeds).

Higgins, J. P. (Ed.). (2008). *Cochrane handbook for systematic reviews of interventions* (Vol. 5). Chichester, England: Wiley-Blackwell.

Ho, C. Y., McCarthy, P., Yang, Y., and Ye, X. (2013). Bankruptcy in the pulp and paper industry: market's reaction and prediction. *Empirical Economics*, **45** (3), pp. 1205-1232.

Horta, I.M. and Camanho, A.S. (2013). Company failure prediction in the construction industry. *Expert Systems with Applications*, **40** (16), pp. 6253-6257.

Huang, S. C., Tang, Y. C., Lee, C. W., and Chang, M. J. (2012). Kernel local Fisher discriminant analysis based manifold-regularized SVM model for financial distress predictions. *Expert Systems with Applications*, **39** (3), pp. 3855-3861.

Iturriaga, F. J. L., and Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of US commercial banks. *Expert Systems with Applications*, **42** (6), pp. 2857-2869.

James, W. (1995, 1907). *Pragmatism*. New York: Dover.

Jeong, C., Min, J. H., and Kim, M. S. (2012). A tuning method for the architecture of neural network models incorporating GAM and GA as applied to bankruptcy prediction. *Expert Systems with Applications*, **39** (3), pp. 3650-3658.

Johnson, R. B., and Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher*, **33** (7), pp. 14-26.

Joy, O. M., and Tollefson, J. O. (1975). On the financial applications of discriminant analysis. *Journal of Financial and Quantitative Analysis*, **10** (5), pp. 723-739.

Kasgari, A. A., Divsalar, M., Javid, M. R., and Ebrahimian, S. J. (2013). Prediction of bankruptcy Iranian corporations through artificial neural network and Probit-based analyses. *Neural Computing and Applications*, **23** (3-4), pp. 927-936.

Khademolqorani, S., Zeinal Hamadani, A., and Mokhatab Rafiei, F. (2015). A Hybrid Analysis Approach to Improve Financial Distress Forecasting: Empirical Evidence from Iran. *Mathematical Problems in Engineering*, 2015 (2015), pp. 1-9.

Khan, K. S., Kunz, R., Kleijnen, J., and Antes, G. (2003). Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*, **96** (3), pp. 118-121.

Keat, R. (1979). Positivism and statistics in social science. In: Irvine, J., Miles, I., and Evans, J. (eds). *Demystifying social statistics.* London: Pluto Press (UK).

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

Kim, S. Y. (2011). Prediction of hotel bankruptcy using support vector machine, artificial neural network, logistic regression, and multivariate discriminant analysis. *The Service Industries Journal*, **31** (3), pp. 441-468.

Knight, A.D. and Turnball, N. (2008). *Epistemology*. In: Knight, A.D. and Ruddock, L. (eds.) *Advanced Research Methods in the Built Environment*. Oxford: Wiley Blackwell. pp6. 4-74

Kristóf, T., and Virág, M. (2012). Data reduction and univariate splitting—Do they together provide better corporate bankruptcy prediction?. *Acta Oeconomica*, **62** (2), pp. 205-228.

Lee, S., and Choi, W. S. (2013). A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. *Expert Systems with Applications*, **40** (8), pp. 2941-2946.

Levin, D. M. (1988). *The Opening of Vision: Nihilism and the Postmodern Situation*. London: Routledge

Li, H., Lee, Y. C., Zhou, Y. C., and Sun, J. (2011). The random subspace binary logit (RSBL) model for bankruptcy prediction. *Knowledge-Based Systems*, **24** (8), pp. 1380-1388.

Liang, D., Tsai, C. F., and Wu, H. T. (2015). The effect of feature selection on financial distress prediction. *Knowledge-Based Systems*, **73**, pp. 289-297.

Lin, F., Liang, D., and Chu, W. S. (2010). The role of non-financial features related to corporate governance in business crisis prediction. *Journal of Marine Science and Technology*, **18** (4), pp.504-513.

Mahamid, I. (2012). Factors affecting contractor's business failure: contractors' perspective. *Engineering, Construction and Architectural Management*, **19** (3), pp. 269-285.

Makeeva, E., and Neretina, E. (2013a). A Binary Model versus Discriminant Analysis Relating to Corporate Bankruptcies: The Case of Russian Construction Industry. *Journal of Accounting*, **3** (1), pp. 65-76.

Makeeva, E., and Neretina, E. (2013b). The Prediction of Bankruptcy in a Construction Industry of Russian Federation. *Journal of Modern Accounting and Auditing*, **9** (2), pp. 256-271.

Marshall, B., Cardon, P., Poddar, A., and Fontenot, R. (2013). Does sample size matter in qualitative research?: A review of qualitative interviews in IS research. *Journal of Computer Information Systems*, **54** (1), pp. 11-22.

Mingers, J. and Gill, A. (1997). *Multimethodology: The Theory and Practice of Combining Management Science Methodologies*. Chichester: Wiley.

Mitchell, J. C. 1983. Case and situation analysis. *Sociology Review* **51** (2), pp.187–211.

Moyer, R. C. (1977). Forecasting financial failure: a re-examination. *Financial Management*, **6 (**1) pp. 11-17.

Muscettola, M. (2014). Probability of Default Estimation for Construction Firms. *International Business Research*, **7** (11), pp. 153.

Neves, J. C., and Vieira, A. (2006). Improving bankruptcy prediction with hidden layer learning vector quantization. *European Accounting Review*, **15** (2), pp. 253-271.

Ng, S. T., Wong, J. M., and Zhang, J. (2011). Applying Z-score model to distinguish insolvent construction companies in China. *Habitat International*, **35** (4), pp. 599-607.

Othman, J. (2013). *Analysing Financial Distress in Malaysian Islamic Banks: Exploring Integrative Predictive Methods* (Doctoral dissertation, Durham University).

Raftery, J., McGeorge, D., and Walters, M. (1997). Breaking up methodological monopolies: a multi-paradigm approach to construction management research. *Construction Management and Economics*, **15** (3), pp. 291-297.

Razak Bin Ibrahim, A., Roy, M. H., Ahmed, Z. U., and Imtiaz, G. (2010). Analyzing the dynamics of the global construction industry: past, present and future. Benchmarking: An International Journal, **17** (2), pp. 232-252.

Remenyi, D., Williams, B., Money, A. and Swartz, E. (1998) *Doing Research in Business and Management: An Introduction to Process and Method*. London: Sage.

Rooke, J., Seymour, D., and Crook, D. (1997). Preserving methodological consistency: a reply to Raftery, McGeorge and Walters. *Construction Management and Economics*, **15** (5), pp. 491-494.

Rorty, R. (1985). Texts and lumps. *New Literary History*, **17** (1), pp.1-16

Russell, J. S., and Jaselskis, E. J. (1992). Predicting construction contractor failure prior to contract award. *Journal of construction engineering and management*, **118** (4), pp. 791-811.

Sánchez-Lasheras, F., de Andrés, J., Lorca, P., and de Cos Juez, F. J. (2012). A hybrid device for the solution of sampling bias problems in the forecasting of firms' bankruptcy. *Expert Systems with Applications*, **39** (8), pp. 7512-7523.

Seymour, D., and Rooke, J. (1995). The culture of the industry and the culture of research. *Construction Management and Economics*, **13**(6), pp. 511-523.

Seymour, D., Crook, D., and Rooke, J. (1997). The role of theory in construction management: a call for debate. *Construction Management and Economics*, **15** (1), pp. 117-119.

Shie, F. S., Chen, M. Y., and Liu, Y. S. (2012). Prediction of corporate financial distress: an application of the America banking industry. *Neural Computing and Applications*, **21** (7), pp. 1687-1696.

Smith, V., Devane, D., Begley, C. M., and Clarke, M. (2011). Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC medical research methodology*, **11** (1), 15.

Stroe, R., and Bărbuță-Mișu, N. (2010). Predicting the financial performance of the building sector enterprises–case study of Galati County (Romania). *The Review of Finance and Banking,* **2** (1), pp. 29-39.

Sun, J., Liao, B., and Li, H. (2013). AdaBoost and bagging ensemble approaches with neural network as base learner for financial distress prediction of Chinese construction and real estate companies. *Recent Patents on Computer Science*, **6** (1), pp. 47-59.

Tashakkori, A. and Teddlie, C. (1998). *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. Thousand Oaks, California: Sage Publications.

The Insolvency Service (2016*). Insolvency Statistics – January to March 2016 (Q1 2016)*. London: The Insolvency Service

Tranfield, D. R., Denyer, D., and Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British journal of management*, **14**, pp. 207-222.

Tsai, C. F. (2014). Combining cluster analysis with classifier ensembles to predict financial distress. *Information Fusion*, **16**, pp. 46-58.

Tsai, C. F., and Cheng, K. C. (2012). Simple instance selection for bankruptcy prediction. *Knowledge-Based Systems*, **27** (2012), pp. 333-342.

Tsai, C. F., and Hsu, Y. F. (2013). A Meta-learning Framework for Bankruptcy Prediction. *Journal of Forecasting*, **32** (2), pp. 167-179.

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

Tsai, L. K., Tserng, H. P., Liao, H. H., Chen, P. C., and Wang, W. P. (2012). Integration of accounting-based and option-based models to predict construction contractor default. *Journal of Marine Science and Technology*, **20** (5), pp. 479-484.

Tseng, F. M., and Hu, Y. C. (2010). Comparing four bankruptcy prediction models: logit, quadratic interval logit, neural and fuzzy neural networks. *Expert Systems with Applications*, **37** (3), pp. 1846-1853.

Tserng, H. P., Chen, P. C., Huang, W. H., Lei, M. C., and Tran, Q. H. (2014). Prediction of default probability for construction firms using the logit model. *Journal of Civil Engineering and Management*, **20** (2), pp. 247-255.

Tserng, H. P., Liao, H. H., Tsai, L. K., and Chen, P. C. (2011a). Predicting construction contractor default with option-based credit models—Models' performance and comparison with financial ratio models. *Journal of Construction Engineering and Management*, **137** (6), pp. 412-420.

Tserng, H. P., Liao, H. H., Jaselskis, E. J., Tsai, L. K., and Chen, P. C. (2011b). Predicting construction contractor default with barrier option model. *Journal of Construction Engineering and Management*, **138** (5), pp.621-630.

Tserng, H. P., Lin, G. F., Tsai, L. K., and Chen, P. C. (2011c). An enforced support vector machine model for construction contractor default prediction. *Automation in Construction*, **20** (8), pp. 1242-1249.

Tserng, H. P., Ngo, T. L., Chen, P. C., and Quyen Tran, L. (2015). A Grey System Theory-Based Default Prediction Model for Construction Firms. *Computer-Aided Civil and Infrastructure Engineering*, **30** (2), 120-134.

Virág, M., and Nyitrai, T. (2014). Is there a trade-off between the predictive power and the interpretability of bankruptcy models? The case of the first Hungarian bankruptcy prediction model. *Acta Oeconomica*, **64** (4), pp. 419-440.

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

Wood, A. P. (2012). *The performance of insolvency prediction and credit risk models in the UK: A comparative study, development and wider application* (Doctoral dissertation, University of Exeter).

Xiong, T., Wang, S., Mayers, A., and Monga, E. (2013). Personal bankruptcy prediction by mining credit card data. *Expert Systems with Applications*, **40** (2), pp. 665-676.

Yang, Z., You, W., and Ji, G. (2011). Using partial least squares and support vector machines for bankruptcy prediction. *Expert Systems with Applications*, **38** (7), pp. 8336-8342.

Yeh, C. C., Chi, D. J., and Lin, Y. R. (2014). Going-concern prediction using hybrid random forests and rough set approach. *Information Sciences*, **254**, pp. 98-110.

Yin, R. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks: Sage Publishing

Yoon, J. S., and Kwon, Y. S. (2010). A practical approach to bankruptcy prediction for small businesses: Substituting the unavailable financial data for credit card sales information. *Expert systems with Applications*, **37** (5), pp. 3624-3629.

Yu, Q., Miche, Y., Séverin, E., and Lendasse, A. (2014). Bankruptcy prediction using extreme learning machine and financial expertise. *Neurocomputing*, **128**, pp. 296-302.

Zhiyong, L. (2014). *Predicting financial distress using corporate efficiency and corporate governance measures* (Doctoral dissertation, University of Edinburgh).

Zhou, L., Lai, K. K., and Yen, J. (2012). Empirical models based on features ranking techniques for corporate financial distress prediction. *Computers and Mathematics with Applications*, **64** (8), pp. 2484-2496.

Zhou, L., Lai, K. K., and Yen, J. (2014). Bankruptcy prediction using SVM models with a new approach to combine features selection and parameter optimisation. *International Journal of Systems Science*, **45** (3), pp. 241-253

Table 1: Summary of finding showing data, techniques and model testing related features of the primary studies

	Author (Year)	Techniques used	Class balancing method	Variable selection method	Data Country	Accuracy	Error type considered	Total sample	Existing firms (%)	Failed firms (%)	Validation	Sample % for validation	Var. type & no.	Period
1.	Cho et al. (2010)	ANN; DT; CRB; LR	USP	t-test& decision tree	Korea	73.7		1000	50	50	CV	20	7 (max) FR	00–02
2.	Du Jardin (2010)	ANN; MDA; LR	USP	Error backward-order (stepwise)	France	94.03	Y	1020	50	50	CV	51	7FR	01-03
3.	Gepp et al. (2010)	DT; MDA	UB	Lit. rev (stepwise)	US	87.6	Y	200	71	29	CV	20	20FR	71-81
4.	Kim & Kang (2010)	ANN	USP	cumulative accuracy profiles	Korea	71.02	Y	1458	50	50	CV	10	7 FR	02–05
5.	Lin et al. (2010)	SVM; MDA	MT/USP	Stepwise regression	Taiwan	94.44	Y	108	50	50	CV	0.93	4FR & 6 non-fin	01-05
6.	Stroe & Bărbuță-Mișu (2010)	MDA	-	-	Romania	77.8	-	11	-	-	10 firms	-	5 FR	01-06
7.	Tseng and Hu (2010)	**ANN; LR**	**UB**	Literature rev. (stepwise)	England	93.75	-	77	58.4	41.6	HOV	20	5FR & 1 market var	85 - 94
8.	Yoon & Kwon (2010)	SVM; ANN; MDA; LR	USP	t-test & chi-square	Korea	74.2	Y	10000	50	50	HOV	30	13CCV	00-02
9.	Chen (2011)	SVM	MT	Lit. rev. (stepwise), GA	Taiwan		N	80	50	50	CV	20	13FR & 1 bus. var	00-10
10.	Chen et al. (2011a)	SVM; ANN; GA	Random		France	90.6	Y	1200	50	50	CV	20	30FR	06-07
11.	Chen et al. (2011b)	SVM; ANN ;GA	UB		Poland &Australia	90.6	ROC	240	53.3	46.7	CV	20	30FR	97-01
12.	De Andrés et al. (2011a)	ANN; MDA; LR	UB	Altman's ratios (stepwise)	Spain	92.38	N	59474	99.77	0.23	CV	20	5FR	07
13.	De Andrés et al. (2011b)	SOM & MARS hybrid.		SPDA	Spain	88.72	Y	63107	99.6	0.4	CV	20	5FR	07-08
14.	Divsalar et al. (2011)	ANN; GA; LR	UB	SFS	Iran	95	ROC	140	51.4	48.6	HOV	28.6	4FR	99-06

	Author (Year)	Techniques used	Class balancing method	Variable selection method	Data Country	Accuracy	Error type considered	Total sample	Existing firms (%)	Failed firms (%)	Validation	Sample % for validation	Var. type & no.	Period
15.	Du Jardin & Séverin (2011)	ANN; MDA; LR	Random	Error backward-order (stepwise)	France	82.61	Y	2360	50	50	HOV	37.3	10FR	95-04
16.	Kim (2011)	SVM; ANN; MDA; LR	MT	Stepwise	Korea	95.95	Y	66	50	50	-	-	7FR	95– 02
17.	Li et al. (2011)	RS; MDA; LR	Random	Stepwise & t-test	China	88.46		370	50	50	MCV	30	23FR	
18.	Ng et al. (2011)	MDA	UB	SP	China	96.9	N	35	88.5	11.5	N	-	7 FR	
19.	Tserng et al. (2011a)	BSM; CB; BS	USP	OMV	US	90%	ROC	87	66.7	33.3	Y	-	SM	70-06
20.	Tserng et al. (2011b)	Barrier option model; MDA	UB	OMV	US	84.5	ROC	121	76	24	CV	-	6OMV& 4FR	70-06
21.	Tserng et al. (2011c)	ESVM; LR	OSP	SPDA	US	80.31	ROC	168	69.6	30.4			7 FR	
22.	Yang et al. (2011)	SVM; ANN	UB	Pearson cor. & PLS	-	79	Y	120	53.3	46.7	-	100	22FR	
23.	Chen (2012)	Hybrid SFNN	UB	All in FS	Taiwan	85.1	-	42	35	65	CV	10	25 FR	98-08
24.	De Andrés et al. (2012)	ANN; MDA	MT	automatic backward elimination	Spain	76.03	Y	122	50	50	CV	19.6	8FR	09-11
25.	Divsalar et al. (2012)	ANN; GA; LR	UB	SFS	Iran	91.18	ROC	136	52.2	47.8	HOV	25	4FR	99-06
26.	Du Jardin & Séverin (2012)	ANN; MDA; LR	Random	Multiple	France	81.6	Y	17840	50	50	HOV	68.4	Chosen from 41 FR	98-04
27.	Huang et al. (2012)	SVM; ANN; DT; LR	MT but UB	Selected from Taiwan economic journal	Taiwan	86.61	N	150	66.67	33.33	CV	10	48FR	00-07
28.	Jeong et al. (2012).	SVM; ANN; DT; CBR;MDA; LR	Random	GAM	Korea	81	N	2542	50	50	CV	20	9FR	01-04
29.	Kristóf & Virág (2012)	ANN; DT; LR	UB	PCA	Hungary	88.8	ROC	504	86.3	13.7	HOV	25	19FR; 6 PCA factors	04

	Author (Year)	Techniques used	Class balancing method	Variable selection method	Data Country	Accuracy	Error type considered	Total sample	Existing firms (%)	Failed firms (%)	Validation	Sample % for validation	Var. type & no.	Period
30.	Sánchez-Lasheras et al. (2012)	SOM & MARS hybrid.	UB	SPDA	Spain	89.58	Y	63107	99.6	0.4	CV	20	5 FR	
31.	Shie et al. (2012)	SVM; ANN; DT; LR	MT but UB	Factor analysis & PCA	USA	81.82	Y	54	55	44.4	MCV		17 FR	06-09
32.	Tsai & Cheng (2012)	SVM; ANN; DT; LR	UB		Australia, Japan, Germany	86.06	Y	653	45.3	54.7	CV	10	15FR (for Autralia)	-
33.	Tsai et al. (2012)	BSM & LR hybrid	UB	SPR	US	87.32		121	76	24	CV		4FR & 1OMV	70-06
34.	Wood (2012)	Univariate; MDA;LR; ANN; Hillgeist et al. (2004) market model; BS; Barrier option model; Barrier option as Down-and-Out-Call ; New naïve down-and-out call	UB		UK	83	ROC	2422	92.7	7.3	HOV	50	17FR	00-09
35.	Zhou et al. (2012)	SVM; ANN; DT; MDA; LR	Random	Spearman correlation	USA	75.6	ROC	834	50	50	HOV	10.8	26FR	93-08
36.	Arieshanti et al. (2013)	SVM; ANN;KNN	UB	-		77.5	-	240	53.3	46.7	CV	20	30FR	-
37.	Callejón et al (2013)	ANN	Random	self-organizing map	multiple	92.11	-	1140	50	50	HOV	12.28	6FR	07- 09
38.	Chuang (2013)	CBR	UB	Multiple	-	90.1	-	321	86.9	13.1	CV	20	8FR	99-06
39.	Hernandez Tinoco, (2013)	LR	UB	reported results, theoretical propositions & empirical assessment	UK	86.85	ROC	3,020	87.25	12.74	HOV		4FR, 4MV, 2 MaI	80-11
40.	Ho et al. (2013)	LR	UB	Lit rev (stepwise)	US	93	ROC	122	90.2	9.8	HOV	20	8FR	99-09
41.	Horta & Camanho (2013)	SVM	USP & OSP	Lit rev	Portugal	97.6	ROC	10559	85	15	HOV	20	6FR &	08-10

	Author (Year)	Techniques used	Class balancing method	Variable selection method	Data Country	Accuracy	Error type considered	Total sample	Existing firms (%)	Failed firms (%)	Validation	Sample % for validation	Var. type & no.	Period
													3Strat. var	
42.	Kasgari et al. (2013)	ANN; GA; LR; Probit	UB	Lit rev	Iran	94.11	ROC	136	52.2	47.8	HOV	23.5	4FR	99-06
43.	Lee & Choi (2013)	ANN; MDA	MT but UB	t-test &correlation analysis	Korea	92	Y	75	6.67	33.33	HOV	4.0	6FR	00-09
44.	Makeeva & Neretina (2013a)	MDA; LR; PA	BA	FA& SPR	Russia	86.44	Y	120	50	50	-	-	6FR	02-10
45.	Makeeva & Neretina (2013b)	MDA; LR; PA	USP	FA& SPR	Russia	86.44	Y	120	50	50	-	-	22 FR	02-10
46.	Othman (2013)	MDA; LR,	UB	FA	Malaysia	90		10	40	60	-	-	13FR	05-10
47.	Sun et al. (2013)	ANN-AB hybrid; ANN bagging hybrid; ANN	USP	t-test, CA & SPDA	China	93.07	Y	85	61.2	38.8	CV	33.3	9 FR	01-10
48.	Tsai & Hsu (2013)	ANN; DT; LR	UB	MC	Multiple (use given average)	79.11	Y	109361	50.1	49.9	CV	20	40 variables	-
49.	Xiong et al. (2013)	SVM; Credit scoring	Random (training data was balanced)	sequence clustering algorithm	Canada	84.46	ROC	17670	57.59	42.41	HOV	77.36	Credit card details	05-06
50.	Abellán & Mantas (2014)	RS	-	Lit. rev. (stepwise)	Multiple (Australia data used)	93.64	ROC	690	-	-	MCV	30	14FR	-
51.	Cheng et al. (2014)	LS-SVM and DE hybrid; SVM; ANN	OSP by SMOTE	SPDA	-	92.13	ROC	76	82.9	17.1	CV	20	7 FR	70-11
52.	Gordini (2014)	SVM; GA; LR	UB	VIF & stepwise	Italy	71.5	Y	3100	51.6	48.4	HOV	30	8FR	09-12
53.	Heo & Yang (2014)	DT; ANN; SVM; MDA; AdaBoost	Random	Calculated based on financial data of sample firms	Korea	78.5	Y	2762	50	50	HOV	20	5FR	08-12
54.	Muscettola (2014)	LR	UB	SPR	Italy	80.94	Y	1338	87.2	12.8	HOV	-	9FR	07-11
55.	Tsai (2014)	ANN; DT; LR; K-means; SOM	UB	SOM	Multiple (Australia	91.61	Y	690	44.5	55.5	CV	20	14FR	-

	Author (Year)	Techniques used	Class balancing method	Variable selection method	Data Country	Accuracy	Error type considered	Total sample	Existing firms (%)	Failed firms (%)	Validation	Sample % for validation	Var. type & no.	Period
					value used in this table)									
56.	Tsai et al. (2014)	SVM; ANN; DT	UB	-	Multiple (Japan value used in this table)	86.37	N	690	44.5	55.5	CV	10	15 FR	-
57.	Tserng et al. (2014)	LR	UB	LR; AUC	US	79.18	ROC	87	66.7	33.3	CV	1.15	6FR & 1 market variable	70-06
58.	Virág & Nyitrai (2014)	SVM; ANN; RS	Random	Modelling	Hungary	89.32	N	156	50	50	CV	25	10FR	
59.	Wang et al. (2014)	SVM; ANN; DT; LR; FS-Boosting; bagging; boosting	Random	Computed from Compustat and from the Moody's Industrial Manual	US	79.99	Y	132	50	50	CV	10	24FR	70-82
60.	Yu et al. (2014)	SVM; MDA	Random	Forward variable selection technique	France	93.2	N	500	50	50	CV	33.3	9FR	02-05
61.	Yeh et al. (2014)	SVM; ANN; DT; RS; RF	UB	RF	Taiwan	96.99	Y	220	75	25	MCV	33	12FR, 1 non-FR & 2 human capital	04-08
62.	Zhiyong (2014)	DEA; LR	UB		China	93.4	ROC	2,104	89.54	10.64	HOV	About 50	Many FR, gov. measures & MaI var	98-10
63.	Zhou et al. (2014)	ANN; DT; MDA; LR; KNN	Random	GA	US	75.6	ROC	2010	50	50	HOV	-	-FR	80-06
64.	Bemš et al. (2015).	ANN; CBR; MDA;	UB	Gini index	-	0.301	ROC	459	67	33	HOV	Another	5FR	-

	Author (Year)	Techniques used	Class balancing method	Variable selection method	Data Country	Accuracy	Error type considered	Total sample	Existing firms (%)	Failed firms (%)	Validation	Sample % for validation	Var. type & no.	Period
		LR; Gini index										2 661 observations		
65.	Cheng & Hoang (2015)	KNN& FFA hybrid, SVM; MDA; LR	OSP	Lit rev	-	96.0	ROC	76	82.9	17.1	CV	20	20 FR	70–11
66.	Du Jardin (2015)	ANN; MDA; LR; SOM; survival analysis	Matching	multiple	France	80.8	Y	125860	50	50	-	50	-	05-11
67.	Iturriaga & Sanz (2015)	SVM; ANN; MDA; LR	Random	Mann-Whintney test & Gini index	US	93.27	ROC	876	50	50	HOV	13.5	7FR	02-12
68.	Khademolqorani et al. (2015)	ANN; DT; MDA; LR; K-means	UB	Factor analysis	Iran	94	ROC	180	67.8	32.2	HOV	44.4	14FR	11
69.	Liang et al. (2015)	SVM; ANN; DT; KNN; Naïve Bayes	Stratified sampling	GA	China	92.98	Y	688	50	50	CV	10	45FR	
70.	Tserng et al. (2015)	GST	UB	GST	US	84.8	-	92	73.9	26.1	-	-	8 FR	72-08

Statistical tools: Discriminant function, FA, LR, MARS, MCDM, MDA, ordinary least-squares, PA.

Artificial intelligence tools: AB, ANN, bagging, DE, DT, ESVM, FFA, KNN, SFNN, SOM, SVM.

Option based models: Barrier option model, BS, BSM, CB.

-: Not stated or not clear or not done or not applicable DT: Decision tree ANN: Artificial neural network BSM: Black-Scholes-Merton BA: balanced or equally dispersed data BS: Bharath & Shumway naïve model CA: Correlation analysis CB: Crosbie & Bohn model CCV: variables from credit card data Cor.: correlation CV: Cross validation DE: Differential Evolution algorithms DEA–DA: Data Envelopment Analysis–Discriminant Analysis ESSVM: enforced support vector machine FA: Factor analysis FFA: firefly algorithm FR: financial ratio FS: financial statements GAM: generalized additive model Gov.: governance GST: Grey system theory HOV: Hold out validation KNN: K-nearest neighbour Lit rev: Literature review LR: Logistic regression LS-SVM: Least square- support vector machine MaI: Macroeconomic indicators MARS: Multivariate Adaptive Regression Splines. MC: meta classifier MDA: Multiple discriminant analysis MCDM: Multiple criteria decision-making MCV: Multiple cross-validation MT: Matching MVR: market value ratio N: No OMV: Option model variables OSP: Oversampling PA: Probit analysis PCA: principal component analysis PLS: partial least squares RF: random forest ROC: Error cost considered using Receiver Operating Characteristic Curve SFNN: self-organizing feature map optimization, fuzzy, and hyper-rectangular composite Neural Networks SFS: sequential feature selectionSM: Stock market variables SMOTE: Synthetic Minority Over-sampling Technique SOM: Self Organizing Maps Neural Networks SP: Stepwise SPDA: discriminant analysis SPR: Stepwise regression Strat. Var: Strategic Variables SVM: Support vector machine UB: unbalanced data USP: Under sampling VIF: variance Inflation Factor Y: Yes

Table 2: Summary of finding showing features of the primary studies used to determine their philosophical underpinnings

	Author (Year)	IRSM	Variables employed	Aim of study	Journal	Data source	Authors background	Sector
1.	Cho et al. (2010)	Sampl e data	Depreciation ratio; Depreciation ratio ; Net profit to sales; Operating income to sales; Accounts receivables/accounts payables; Current liabilities/total assets; Cash flow to total debt; Cash flow/(loan + interest expense); Fixed assets turnover; Interest cost– interest income; (Current assets–current liabilities)/total equity; Break even point sales/sales	Propose new CBR approach	ESA		Business School	Manufactu ring
2.	Du Jardin (2010)		Cash/Total Assets; Total Debt/Shareholder Funds; Cash/Total Debt; (Cash + Mark. Sec.)/Total Assets; EBIT/Total Assets; EBITDA/Total Assets; Shareholder Funds/Total Assets	To Check variable selection methods effect	NC	Diane	Business School	Retail
3.	Gepp et al. (2010)		-	Compare DT & MDA	JF	Compustat	Business School	Manufactu ring & retail
4.	Kim & Kang (2010)	-	Ordinary income to total assets; EBITDA to Interest expenses; Total debt to total assets; Retained earning to total assets; Cash ratio; Inventory to sales; Total assets	Check enhanced ANN against ord. ANN	ESA	commercial bank	Business Administratio n; Computer. Engineering	Manufactu ring
5.	Lin et al. (2010)	Lit rev.	Debt Ratio; Accounts Receivable Turnover Ratio; Assets Turnover; Operating Income to Capital; Shareholding of Board Members - Current vs. Prior Year; Ratio of Pledged Shares of Board Members; Shareholding of Board Members; Necessary Controlling Shares; Other Investment Assets; Board Member Bonus to Pretax Income	Use SVM with ratios & non-financial variables	JJMST	TEJ	Business Management; Computer Science.	-
6.	Stroe & Bărbuță-Mişu (2010)		Gross surplus of exploitation/ TD; own capital/total liabilities; CA/stocks; financial expenses/turnover; personnel costs/added value.	To predict the financial performance of CCs.	TRFB	-	Business management	Constructi on
7.	Tseng and Hu (2010)	-	After-tax profit/total assets; Retained earnings/total assets; Cash/current liabilities; Cash/total liabilities; Working capital/operating expenditure; Working capital/total assets;	**Comparing models**	ESA	Datastream & FT Extel Company Research	Business Administratio n	General industry
8.	Yoon & Kwon (2010)	Credit card data & bus. type	The months suspended during 6 months; The term between the month of max sales and min sales; The term of continual trade until the point of the time of basis; Months happened card transactions; Average of sales in a month; Average of sales in 6 months; Amount of difference between max sales and min sales; Minimum amount of sales in 6 months; Variance of sales in 6 months; Average of sale per a transactions; Average of transactions in 6 month; Maximum	Use credit card data for small bus. & compare techniques	ESA	K-VAN (a credit card) service company & Korea Federation of Banks	Business; Ind. Systems Engineering	Merchant

	Author (Year)	IRSM	Variables employed	Aim of study	Journal	Data source	Authors background	Sector
			transactions in a month; Minimum transactions in a month					
9.	Chen (2011)	Lit rev	Business Cycle Index; Debt to Equity Ratio; Gearing Ratio; Debt/Equity (DE); Return on Asset (ROA); Earnings per Share (EPS); Return on Equity (ROE); Current Ratio; Acid-Test Ratio; Current Assets to Total Assets; Cash Flow to Total Debt Ratio; Cash Flow Ratio; Inventory to Total Assets Ratio; Inventory to Sales Ratio	Use PSO with SVM	NNW	TSEC	IT	Electronic
10.	Chen et al. (2011a)		Too many (30 financial ratios)	Integrate error cost into model	KBS	Diane	IT; economics	All
11.	Chen et al. (2011b)		Too many (30 financial ratios)	Propose FKNN	ESA	-	IT	All
12.	De Andrés et al. (2011a)	-	working capital/total assets; retained earnings/total assets; earnings before interest and taxes (EBIT)/total assets; market value of equity/book value of total debt; sales/total assets	Propose hybrid model (C-means & MARS)	ESA	SABI database	Accounting.; EEM	Manufactu ring
13.	De Andrés et al. (2011b)		WC/TA; RE/TA; EBIT/TA; market value of equity/book value of TD; S/TA.	Proposes a hybrid approach to the forecasting of bankruptcy of Spanish CCs	AMIS	BVD & SABI	AEF & ME	Constructi on
14.	Divsalar et al. (2011)	Lit rev	sales to current assets ratio; operational income to sales ratio; quick assets to total assets ratio; total liability to total assets ratio	Use GA & NN	ES	ThSE	Management & Accounting	Multiple
15.	Du Jardin & Séverin (2011)	-	Shareholder funds/Total assets; Total debt/Shareholder funds; (Cash + Marketable securities)/Total assets; Cash/Current liabilities; Cash/Total debt; EBITDA/Total assets; EBIT/Total assets; Change in shareholders' equity; Cash/Total sales; EBIT/Total sales	Use self-organizing map	DSS	Diane	Business School	Retail
16.	Kim (2011)	Lit rev & NICE	ROE; current ratio; fixed asset turnover; fixed asset to long-term capital ratio; ordinary income to owners' equity ratio; growth in owners' equity; growth in asset	Compare techniques	SIJ	NICE & FSS	HTM	Hotels
17.	Li et al. (2011)	-	Gross income/sales; Net income/sales; Ebit/total asset; Net profit/total assets; Net profit/current assets; Profit margin; Net profit/equity; Total assets turnover; Current assets turnover; Fixed assets turnover; Current ratio; Cash/current liability; Asset-liability ratios; Equity/debt ratio; Liability/tangible net asset; Liability/equity market value; Interest coverage ratio; Growth rate of total assets; Current assets/total assets; Fixed assets/total assets; Current liability/total liability; Earning per share; Net assets per share; Cash flow per share	Propose Random subspace LR (RSBL)	KBS	SaSE & SeSE	Economics; Business	All
18.	Ng et al. (2011)		Interest cover ratio; Total cash flow ratio; Operating margin; CRt; ROA; Operation asset / TA; Current asset turnover; Operation index.	To develop a BPM to evaluate the financial performance of construction contractors in China.	HI	GTIOne	CE	Constructi on
19.	Tserng et		sum of short-term liabilities & one-half of long-term liabilities; total	To empirically explore	JCEM	CIF, WRDS,	CE & AEF	Constructi

	Author (Year)	IRSM	Variables employed	Aim of study	Journal	Data source	Authors background	Sector
	al. (2011a)		liabilities; market value of equity; contractor's assets value; equity volatility; asset volatility; one-year Treasury bill rate; actual ROA.	whether the OBMs have an advantage in predicting construction contractor default		CRSP		on
20.	Tserng et al. (2011b)	Lit rev	Stock price × Shares outsourcing; Book value of TD at the end of December; Annualized percent standard deviation of asset returns; One-year Treasury bill rate. (6 stock variables calculate from this four)	Investigates how well the BOM can measure the contractor default risk using a large sample	JCEM	CIF,WRDS & CRSP	CE & AEF	Constructi on
21.	Tserng et al. (2011c)		ROA; Net WC /TA; Revenues /FA; Turnover of TA; S /NW; ROE; Profits /net WC	Proposes ESVM for the default prediction in the CI	AC	CIF & CRSP	CE	Constructi on
22.	Yang et al. (2011)	-	Cash/current liabilities; Cash/total assets; Current assets/current liabilities; Current assets/total assets; Working capital/total assets; Working capital/sales; Sales/inventory Sales/receivables; Net profit/total assets; Net profit/current assets; Gross profit/sales; Net profit/liabilities; Net profit/equity; Sales/total assets; Sales/current assets; Liabilities/total income; Receivables/liabilities; Net profit/sales; Liabilities/total assets; Long term liabilities/equity; Current assets/sales	Propose hybrid model (PLS & SVM)	ESA	Referred to a thesis	IT	
23.	Chen (2012)	All ratios in fin statem ents	Profit margin; ROA; After-tax rate of return; Operating profit /paid-in capital; Pre-tax net profit /paid-in capital; Earnings per share; Operating margin; Operating profit; Growth rate; After-tax net profit growth rate; Revenue growth rate; Growth rate of TA; Growth in the total ROA; Equity ratio; Debt /assets; Long-term funds /FA; Dependence on borrowing; Inventory turnover ratio; Receivable turnover ratio; TA turnover ratio; FA turnover ratio; NW turnover ratio; CRt; Acid-test ratio Times interest earned ratio	To use SFNN to provide a new method for forecasting corporate financial distress	ESA	TSE	CE	Constructi on
24.	De Andrés et al. (2012)	Lit rev	Current Assets/ Current Liabilities; Equity/ Long-term Liabilities; Operating Revenue/ Total Employees; Staff Costs/ Operating Revenue; Net Financial Income/ Total Revenue; Net Income/Total Assets; Operating Revenue/Total Assets; Equity/Total Liabilities	To improve performance of classifiers	KBS	SABI database	Accounting	Manufactu ring
25.	Divsalar et al. (2012)	Lit rev	sales to current assets ratio; operational income to sales ratio; quick assets to total assets ratio; total liability to total assets ratio	Use new type of GA called GEP	JF	ThSE	Management & Accounting; Construction	Multiple
26.	Du Jardin & Séverin (2012)	Financ ial statem ents	Chosen from 41 financial ratios covering solvability, liquidity, profitability, financial structure, activity and rotation	To use Kohonen map to stabilize temporal accuracy	EJOR	Diane	Business School	Retail
27.	Huang et al. (2012)	TEJ	48 ratios covering profitability index; per share rates index; growth rates index; debt-paying ability index; management ability index.	Propose hybrid KLFDA & MR-SVM	ESA	TSE	Business; Accounting; Finance.;	Any

	Author (Year)	IRSM	Variables employed	Aim of study	Journal	Data source	Authors background	Sector
							Economics	
28.	Jeong et al. (2012).	Lit rev.	Gross value added to total assets; Growth rate of total assets; Ordinary income to total assets; Rate of earnings on total capital; Net working capital to total assets; Depreciation expense; Operating assets turnover; Net interest expenses; Employment costs	To fine-tune ANN factors	ESA	-	Management; Business; IT	Manufactu ring
29.	Kristóf & Virág (2012)	-	The initial list of 31 ratios was given without specifying the 19 used in developing models	To compare bankruptcy prediction method	AO	Hungarian corporate database	Finance	Multiple
30.	Sánchez-Lasheras et al. (2012)		WC/TA, RE/TA; EBIT (EBIT)/TA; market value of equity/book value of TD; S/TA.	Propose a hybrid using SOM for clustering sound firms and ach cluster is replaced by a director vector which summarizes all of them.	ESA	BVD & SABI	CE, ME & AEF	Constructi on
31.	Shie et al. (2012)	From financ ial statem ents	Given in undefined abbreviations	Proposed enhanced PSO-SVM	NCA	NYSE	Finance; Information management; Human resource	Banking
32.	Tsai et al. (2012)	Lit rev	Default Probability from the Option-based Model; ROA; FA to NW; Dr; Accounts Receivable Turnover	To validate a hybrid model for predicting construction contractor default	JMST	CIF	CE & AEF	Constructi on
33.	Tsai & Cheng (2012)	-	-	Check effect of outlier on BPMs	KBS	UCI machine learning repository	Information management	-
34.	Wood (2012)		total assets; total debt; current assets; current liabilities; cash; retained earnings; EBIT; sales; inventories; operating expenses; depreciation; net income; funds from operations; ROE; dividends; market capital; and price volatility	To provide empirical analysis of extant corporate insolvency prediction technologies, in an attempt to discover if there is a single technology, methodology, or set of variables which outperforms all of its counterparts	Thesis	London Share Price Database and Thomson One Banker	Accounting	All Non-finance
35.	Zhou et al. (2012)	Lit rev	Current ratio; Net income/total assets (ROA); Retained earnings/total assets; Working capital/total assets; EBIT/total assets; Sales/total assets; Sales/total assets; Quick ratio; Cash flow/total debt; Current assets/total assets; Quick assets/current liabilities; Total debt/total assets; Cash/current liabilities; Cash/total assets; Quick assets/total	To find the best variables for accuracy	C&MA	Compustat	Management; Business	Non financial firms

	Author (Year)	IRSM	Variables employed	Aim of study	Journal	Data source	Authors background	Sector
			assets; Total assets; Net income/stockholders' equity; Current liabilities/total assets; Net income/sales; Inventories turnover; Quick assets/sales; Sales/cash; Working capital/sales; Dividend; Fixed assets/ (stockholder's equity + long-term liabilities); Ordinary income/total assets; Stock holders' equity/total assets					
36.	Arieshanti et al. (2013)	-	Too many (30 financial ratios)	To compare techniques	TCEC		Informatics Engineering	
37.	Callejón et al (2013)	Lit rev	EBIT / Current liabilities; Equity / Non-current liabilities; (Net income + Depreciation amortization and write-offs) / Current financial liabilities; EBIT / Total assets; Net profit / Total assets; log Total Assets	To increase predictive power of ANN	IJCIS	AMADEUS	Accounting & finance; Computer science	Any
38.	Chuang (2013)	Lit rev	Accounts receivable turnover ratio; Sales Growth ratio; Gross Margin Growth ratio; Total equity growth ratio; Return on total assets growth ratio; Operating Income/Capital; Liabilities ratio; Debt/Equity %	To Hybridise CBR	IS	TEJ; TSEC	Information management	IT
39.	Hernandez Tinoco, (2013)		Total Funds from Operations to Total Liabilities; Total Liabilities to Total Assets; the No Credit Interval; and Interest Coverage; firm's equity price; lagged cumulative security residual return; Size of the company measured by its market capitalisation relative to the total size of the FTSE All Share Index; Market Capitalisation to Total Debt; the Retail Price Index (RPI); and the United Kingdom Short Term (3-month) Treasury Bill Rate Deflated	investigates the information content of different types of variables in the field of financial distress/default prediction	Thesis	Datastream and Thomson One Banker (Worldscope) & LSDP	Accounting & Finance	All Non-finance
40.	Ho et al. (2013)	-	Total Assets/GNP Price-level Index; Total Liabilities/Total Assets; Current Liabilities/Current Assets; Net Liability Working Capital/Total Assets; (Pretax Income + (Depreciation and Amortization))/Total Liabilities; Net Income/Total Assets; Income Negative for Two Years	Develop BPM for US paper companies	EE	Compustat & others	Economics & management; Paper business	Pulp and paper
41.	Horta & Camanho (2013)	Lit rev	**Financial variables:** ROS ratio; ROA ratio; ROE ratio; CRt; WC/TA; e net value of sale/average CA **Non- financial variables:** company main activity; company size & headquarter geographic location	To use SVM to predict company failure in the CI	ESA	INCI	Engineering	Constructi on
42.	Kasgari et al. (2013)	-	sales to current assets ratio; operational income to sales ratio; quick assets to total assets ratio; total liability to total assets ratio	Compare ANN to other techniques	NCA	ThSE	Management & accounting	Any
43.	Lee & Choi (2013)	Financ ial statem ents	Total asset growth rate; Net profit ratio before income tax expense Net profit ratio before income tax expense per capita; Retained earnings to total asset; Operating cash flow to total liabilities; Receivables turn-over rate	To do multi industry investigation	ESA	KSE	Business	Constructi on
44.	Makeeva &	Lit rev	EBIT* /TA; Cash/current liabilities; Account receivable/S; Interest	To reveal the key	JAFE	BVD Russia	AEF	Constructi

	Author (Year)	IRSM	Variables employed	Aim of study	Journal	Data source	Authors background	Sector
	Neretina (2013a)		coverage; S/TA;	determinants, which cause bankruptcy of Russian CCs				on
45.	Makeeva & Neretina (2013b)	Lit rev	profitability, liquidity, & turnover measures, size & interest coverage coefficients	To use models appropriate for predicting bankruptcy in the CI	JMAA	BVD Russia	AEF	Constructi on
46.	Othman (2013)	Lit rev	(Shareholders' Equity + Total Income)/Total Assets; (Shareholders' Equity + Total Income)/(Deposits and nondeposit Funds); Net Working Capital/Total Assets; (Shareholders' Equity + Total Income)/(Total Assets + Contingencies and Commitments); Capital/Assets; Total Financing/Total Assets; Non-performing Financing/Total Financing; Permanent Assets/Total Assets; Liquid Assets/Total Assets; Liquid Assets/(Deposits and Non-deposit Funds); Interest Expenses/Total Expenses; Interest Income/Total Income; Total Income/Total Expenses; Interest Income/Interest Expenses; Income-Expenditure Structured Ratio; Provision for financing Losses/Total Assets; Provision for financing Losses/Total financing; Net Income Before Tax/Average Total Assets; Net Income (Loss)/(Common Stocks and Preferred Stocks/4); Net Income (Loss)/Average Shareholders' Equity; Net Income (Loss)/Average Total Assets	To empirically explore, examine and analyse the financial distress of the Malaysian Islamic banks	Thesis	Annual reports	Business school	Islamic Banking
47.	Sun et al. (2013)	Lit rev	Cash ratio; Equity & long-term debt /FA & longterm investment ratio; Net profit /TA ratio; Profit margin; Account receivable turnover; Fixed asset turnover; Net assets per share; Equity /TA ratio; Growth rate of prime operating revenue	Use ANN & ensemble learning approaches to develop CI-BPM	RPCS	Chinese listed firms	AEF	Constructi on
48.	Tsai & Hsu (2013)	CCI& O	Not listed	Present met-learning framework (hybrid)	JF	Various database websites	Information management	Any
49.	Xiong et al. (2013)	-	Credit card details	Use sequence on credit card data	ESA	A Canadian Bank	Computer Science; Mathematics	Any
50.	Abellán & Mantas (2014)	-	Not given	To correctly use bagging scheme	ESA	Referedn to another paper	Computing	Manufactu ring
51.	Cheng et al. (2014)	Lit rev	Net WC /TA; S /NW; Turnover of TA; Revenues /FA; ROA (ROA); ROE (ROE); Profits /Net WC.	Use SMOTE, LS-SVM, & DE hybrid to develop construction contractor BPM	KBS	WRDS	CE	Constructi on
52.	Gordini (2014)	Lit rev	Return on equity; Return on investment; Ebitda/turnover; Interest charges/Ebitda; Cash flow/total debts; Financial debts/equity; Total debts/Ebitda; Current ratio	Test GA accuracy & compare to other techniques	ESA	CERVED	Business management	Manufactu ring

	Author (Year)	IRSM	Variables employed	Aim of study	Journal	Data source	Authors background	Sector
53.	Heo & Yang (2014)	Fin reports	Earnings before interest and taxes/Total assets; Earnings before taxes/Capital; Working capital/Total assets; Working capital/Sales; Current assets/Total assets; Current assets/Current liabilities; Cash/Total assets; Cash/Current liabilities; Value of natural logarithm of the total assets; Sales/Capital; Sales/Current assets; Sales/Total assets	To prove AdaBoost is right for Korean construction firms	ASC	NICE DnB	Computing	Construction
54.	Muscettola (2014)		Intangibles / TA; Return on Debt; Investment turnover; Gross profit / S; NW / TA. Current liabilities / TD; Current liabilities / S; Ebitda / Interest expense; Constant	To create a new predictive model of CCs	IBR		AEF	Construction
55.	Tsai (2014)	CCI&O	Not given	To compare hybrids	IF	Various database websites	Infornmation Management	Any
56.	Tsai et al. (2014)	-	Not given	To compare classifier ensembles	ASC	Various database websites	Information management; Economics & business	Any
57.	Tserng et al. (2014)		current ratio; et working capital to total asset; debt ratio; accounts payable turnover ratio); the turnover of the total assets; return on asset-ROA; book to ma	To use LR to predict contractors default	JCiEM	Compustat & CRSP	Construction	Construction
58.	Virág & Nyitrai (2014)	Calculated from sample data	Quick liquidity ratio; Liquidity ratio; Cash funds ratio; Cash flow and total debts ratio; Assets turnover ratio; Debt ratio; Solvency ratio; Short-term-loans-covered Current Assets; Return on sales; Return on assets	To show RS accuracy is competitive with SVM & ANN	AO	first Hungarian bankruptcy model's database	Finance	Any
59.	Wang et al. (2014)	Compustat and from the Moody's Industrial Manual	Cash/Current debt; Cash/Sales; Cash/Total assets; Cash/Total debt; Cash flow from operations/Sales; Cash flow from operations/Total assets; Cash flow from operations/Total assets; Cost of goods sold/Inventory; Current assets/Current debt; Current assets/Sales; Current assets /Total assets; Current debt/Total debt; Income/Sales; Income/Total assets; Income/Total debt; Income plus depreciation/Sales; Income plus depreciation/Total assets; Income plus depreciation/Total debt; Sales/Receivables; Sales/Total assets; Total assets/Total debt; Working capital from operations/Sales; Working capital from operations/Total assets; Working capital from operations/Total debt	Inject feature selection into boosting	ESA	Compustat & MIM	Management, information system	Generally industry
60.	Yeh et al. (2014)	Lit rev & TEJ	Current assets/current liabilities; Current assets/total assets; Total assets (log or natural log); Total liabilities/total assets; Change in total liabilities/total assets; Net income before tax/net sales; Net income	To increase accuracy using RF&RS	IS	TEJ	Business; accounting	

	Author (Year)	IRSM	Variables employed	Aim of study	Journal	Data source	Authors background	Sector
			before tax/total assets; Retained earnings/total assets; Net worth divided/total liabilities; Atlman's Z-score; 1 if negative net income, 0 otherwise; 1 if a big 4 auditor performs the audit, 0 otherwise; Net income/workforce; Net income before tax/workforce; R&D expense/total assets					
61.	Yu et al. (2014)	From Philip Du Jardin 's data	Profit before tax/shareholders' funds; Net income /shareholders' funds; EBITDA/total assets; EBITDA/ permanent assets; EBIT/total assets; Net income/total assets; Shareholders' funds/total assets; Total debt/ shareholders' funds; Total debt/total assets	Produce BPM using ELM	NC	DIANE	Computing	Trade
62.	Zhiyong (2014)		Profitability; Liquidity and liability; Capital composition; Cash flow; Operation capacity; Growth rates; Capitals; Costs; Assets; Debts; Profits; GDP growth; Inflation, consumer prices; Unemployment; Lending interest rate; Board composition; Ownership structure; Management compensation; Director and manager characteristic	To estimate cross-sectional models that incorporate corporate efficiency measures and its components, dynamic efficiency & corporate governance measures to predict financial distress	Thesis	Wind database	Management science and Business economics	All
63.	Zhou et al. (2014)	Lit rev	Net income/total assets; current ratio; fixed assets/(stockholder's equity + long-term liabilities); stock holders' equity/total assets; Working capital/total asset; Cash flow/total debt; Quick assets/current liabilities; Cash/total asset; Quick assets/total asset; Total assets; Cash flow/total assets; Net income/stockholders' equity; Current liabilities/total assets; Net income/sales; total liabilities/total assets; Size; Net income/net worth; Quick assets/sales; Sales/cash; Working capital/sales; Dividend	Propose new feature selection method	IJSS	Compustat	Management	Any
64.	Bemš et al. (2015).	-	debt ratio; debt to equity; return on costs; current ratio; payables turnover	Introduce new scoring method called Gini index	ESA	-	Management; Computing	
65.	Cheng & Hoang (2015)	Lit rev	Quick Ratio; Net WC to TA; CA /Net Assets; Total Liabilities /NW; RE /S; Dr; Times Interest Earned; Revenues /Net WC; Accounts Receivable Turnover; Accounts Payable; S /NW; Quality of Inventory; FA /NW; Turnover of TA; Revenues /FA; ROA; ROE; ROS; Profits /Net Working Capita	To use FIC hybrid f for contractor default prediction	IEEETEM	WRDS	CE	Construction
66.	Du Jardin (2015)	From theory	-	To improve BPM accuracy beyond one year	EJOR	Diane	Business	Services; Construction; Retail
67.	Iturria1 & Sanz (2015)	Taken from	Net income/average total equity capital; Number of times net loans and lease-financing receivables exceed equity capital; Construction, land	To develop ANN BPM for US banks	ESA	FDIC	Economics	Banking

	Author (Year)	IRSM	Variables employed	Aim of study	Journal	Data source	Authors background	Sector
		financial reports	development and other land loans plus closed end loans secured by family residential properties first liens plus revolving open-end loans plus loans secured by farmland plus secured by nonfarm non-residential properties as a percentage of total capital; Efficiency ratio (average total costs/total assets); Average real estate owned other than bank premises/average total assets; Provision for loans and leases receivables losses/average assets; Credit card plans in domestic offices plus other revolving credit plans in domestic offices plus other consumer loans in domestic offices as a percentage of total capital					
68.	Khademolqorani et al. (2015)	-	total asset; ratio of the current assets to the current liabilities; ratio of the amount of cash and equivalents, short, and accounts receivable, term investments to the current liabilities; ratio of the total liabilities to the total assets; ratio of the total liabilities to the total owner's equity; ratio of the sales to the number of the inventories; ratio of the total liabilities to the daily sales; ratio of the sales to the fixed assets; ratio of the sales to the total assets; ratio of the net income to the sales; ratio of the current liabilities to the owner's equity; ratio of the total liabilities to the owner's equity; ratio of the working capital to the total assets	To develop a novel hybrid	MPE	ThSE	Industrial & systems engineering	Manufacturing
69.	Liang et al. (2015)	-	Not given	To compare feature selections	KBS	TEJ	Computer science; Information management	Manufacturing & retail
70.	Tserng et al. (2015)	Lit rev	ROE; ROS; profitability including ROA; profits /net WC; Dr; net WC /TA; RE /S; CA /net assets.	To use GST to develop CI-BPM	CACIE	CIF, WRDS & CRSP	CE	Construction

Note: where ratio number goes beyond 25, the group ratio names given in the paper are presented

AC: Automation in Construction AEC: Associated General Contractors AEF: Accounting &/or Finance/ &/or Economics AMADEUS: Analyse Major Databases from European Sources ANN: Artificial neural network ASC: Applied Soft Computing AMIS: Accounting and Management Information Systems AO: Acta Oeconomica BD: Building Department BOM: barrier option model BVD: Bureau van Dijk C&MA: Computers & mathematics with applications CA: Current assets CACIE: Computer-Aided Civil and Infrastructure Engineering CC: construction companies or construction firms CCI&O: Credit card information & others CE: Civil engineering or Construction engineering or civil & architecture engineering CI: construction industry CIF: Compustat Industrial file CM: Construction management CME: Construction Management and Economics CoE: Computer engineering CR: Corporate Records CRt: current ratio CRSP: Centre for Research in Securities Prices Differential Evolution: algorithms DR: Debt ratio DSMEAT: Department of Statistics at the Ministry of Economic Affairs of Taiwan DSS: Decision support systems EE: Empirical economics EBIR: Earnings/profit before interest & tax EEM: Exploitation and Exploration of Mines

(Department of) EJOR: European Journal of Operational Research ELM: extreme learning machine ES: Expert systems ESA: Expert Systems with Application ESVM mode: Enforced support vector machine-based model FA: Fixed asset FDIC: Federal Deposit Insurance Corporation FIC: fuzzy instance based classifier FKNN: fuzzy k-nearest neighbour FRBIC: Federal Reserve Bulletin, insurance companies GEP: gene expression programming GST: grey system theory HI: Habitat International HTM: Hotel & tourism management (Department of) IBR: International Business Research IEEETEM: IEEE transactions on engineering management IF: Information fusion IJCIS: International journal of computational intelligence systems IM: Information Management IN: Industry Norms INCI: Instituto da Construção e do Imobiliário [The Institute of Construction & Real Estate] IRSM Initial ratio selection method IS: Information sciences IJSS: International journal of systems science IT: Information technology JAFE: Journal of Accounting, Finance and Economics JBFA: Journal of Business Finance & Accounting JCEM: Journal of Construction Engineering and Management JCiEM: Journal of civil engineering and management JMAA: Journal of Modern Accounting and Auditing JF: Journal of forecasting JMST: Journal of Marine Science and Technology KBS: Knowledge-Based Systems KLFDA: kernel local fisher discriminant analysis KSE: Korea Stock Exchange Lit rev: Literature review LN: Lexis nexis LS: Least Squares MARS: multivariate adaptive regression splines ME: Manufacturing engineering MIM: Moody's Instruction Manual MPE: Mathematical problems in engineering MR: manifold-regularized NC: Neurocomputing NCA: Neural computing and applications NNW: Neural network world NICE: National Information and Credit Evaluation Inc. NW: Net worth OBM: option-based models NYSE: New York Stock Exchange PICE: Proceedings of the Institution of Civil Engineers PLS: partial least squares PSO: particle swarm optimization RPCS: Recent Patents on Computer Science RE: retained earnings ROA: Return of assets ROE: Return on Equity ROS: Return on sales RSBL: random subspace binary logit SaSE: Shanghai Stock Exchange SAUS: Statistical abstract of the United States SeSE: Shenzhen Stock Exchange SFNN: integrate the concepts of self-organizing feature map optimization, fuzzy, & hyper-rectangular composite Neural Networks SIJ: The Service Industries Journal SOM: Self Organized Maps SSE: Singapore Stock Exchange SVM: Support Vector Machine SB: Source Book TA: Total assets TCB: Taiwan Central Bank TCEC: TELKOMNIKA (Telecommunication computing electronics and control) TCPA: Taiwan Construction & Planning Agency TD: Total debt TEJ: Taiwan Economic Journal TEJD: Taiwan Economic Journal Database TRFB: The Review of Finance and Banking TSE: Taiwan Stock Exchange TSEC: Taiwan Stock Exchange Corporation ThSE: Tehran stock exchange USDLS: U.S. Department of Labour Statistics WC: Working capital WRDS: Wharton Research Data Services