Predicting Corporate Investment/Non-investment Grade by using Interval-Valued Fuzzy Rule-Based Systems – A Cross-Region Analysis

Petr Hájek
Institute of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice, Studentská 84, Pardubice, Czech Republic
e-mail: petr.hajek@upce.cz, tel.: +420 466 036 147, fax: +420 466 036 010

Abstract

Systems for predicting corporate rating have attracted considerable interest in soft computing research due to the requirements for both accuracy and interpretability. In addition, the high uncertainty associated primarily with linguistic uncertainties and disagreement among experts is another challenging problem. To overcome these problems, this study proposes a hybrid evolutionary interval-valued fuzzy rule-based system, namely IVTURS, combined with evolutionary feature selection component. This model is used to predict the investment/non-investment grades of companies from four regions, namely Emerging countries, the EU, the United States, and other developed countries. To evaluate prediction performance, a yield measure is used that combines the return and default rates of companies. Here, we show that using interval-valued fuzzy sets leads to higher accuracy, particularly with the growing granularity at the fuzzy partition level. The proposed prediction model is then compared with several state-of-the-art evolutionary fuzzy rule-based systems. The obtained results show that the proposed model is especially suitable for high-dimensional problems, without facing rule base interpretability issues. This finding indicates that the model is preferable for investors oriented toward developed markets such as the EU and the United States.

Keywords: interval-valued fuzzy rule-based systems; evolutionary algorithms; financial distress; credit rating.

Introduction

Credit ratings are increasingly considered to be an important decision-making support to financial market participants such as investors, financial institutions and regulators. Credit ratings play an important role in the financial system by reducing information asymmetry between investors and borrowers. Credit ratings are assigned to issuers, as well as to specific debt issues, such as bonds, notes, and other debt securities. In issuer's credit ratings, an issuer's overall capacity and willingness to meet its financial obligations is addressed. This evaluation is based on a complex analysis performed by professionals who interpret both financial and non-financial information from multiple sources. Since this evaluation can be slow and costly, automatic credit rating prediction has become a central problem in artificial intelligence (AI) research [1]. Prediction models have been extensively developed to replicate and explain the credit rating processes performed by credit rating agencies [2].

A wide range of AI methods have been applied to predict credit ratings, including statistical classifiers [3], decision trees [1], neural networks (NNs) [4], [5], support vector machines (SVMs) with both supervised [6], [7] and semi-supervised learning [8], case-based reasoning [9], artificial immune systems [10], rough sets [11], fuzzy rule-based systems (FRBSs) [12], and ensemble approaches [10], [13]. Recent efforts have also indicated that AI methods should be integrated into the feature selection process to improve prediction accuracy [1]. Finally, the paradigm of soft computing [14] has recently provided the most encouraging results in related problems such as financial failure prediction [15], [16] and a consumer's credit scoring [17]. This refers to the integration of different, seemingly unrelated, AI methods such as FRBSs, NNs, evolutionary algorithms (EAs), rough set theory and probabilistic reasoning in various combinations to exploit their strengths. However, significantly insufficient attention has been paid to its application in corporate credit rating prediction.

Most importantly, the hybridization between FRBSs and EAs provides advantages that are desirable in imitating the credit rating process. First, evolutionary FRBSs provide good interpretability in terms of fuzzy if-then rules (in contrast to non-if-then fuzzy classifiers) and, thus, simulate the credit rating decision-making process of financial experts. Note that FRBSs, similarly as other possibilistic classifiers, may assign a soft class label with degrees of membership in each class [18]. This is similar to probabilistic classifiers that usually have the posterior probabilities for the classes as output. However, in the credit rating process, experts use linguistic labels to represent the partial truth of their opinions, rather than partial knowledge. Thus, the experts are able to verify the classification paradigm, for example the consistency and completeness of the rule base. Second, EAs are employed to learn or tune different components of FRBSs such as rule bases, the antecedents and consequents of if-then rules, parameters of membership functions (MFs), and so on. Thus, EAs enable the automatic design of FRBSs through their capability to encode and evolve rule antecedent aggregation operators, different rule semantics, rule base aggregation operators and defuzzification methods [19]. In addition, decision makers can fix some components of FRBSs in order to improve interpretability and accuracy. Owing to these qualities, evolutionary FRBSs represent one of the most popular approaches in the soft computing literature [20]. However, determining the precise values of MFs can be problematic in many application domains due to the uncertainties associated with dynamic unstructured environments, linguistic uncertainties, disagreement among experts, and noise in the data [21]. Therefore, several generalizations of FRBSs have been developed to design MFs effectively. Interval-valued FRBSs (IVFRBSs) [22] are widely considered to be the most important representatives of these generalizations. Recently, methods have been developed to optimize the design of IVFRBSs that, compared with traditional FRBSs, provide an additional degree of freedom and flexibility in handling uncertainty [23], [24].

This study evaluates IVFRBSs optimized by EAs, namely IVTURS (interval-valued fuzzy rule-based system with tuning and rule selection) [25], to predict corporate investment/non-investment grade. First, the genetic feature selection process is carried out to obtain IVFRBSs with both comprehensible rule bases and high prediction accuracy. Second, we employ a range of state-of-the-art evolutionary FRBSs to compare their performance with IVFRBSs when predicting the grades of companies in four regions, namely the United States, the EU, other developed and Emerging countries. To measure prediction performance, we use a yield measure that is suitable for investors' decision making because it combines return and default rates. We also measure the interpretability of FRBSs at both the rule base and the fuzzy partition level.

The proposed system is aimed to provide an accurate and interpretable decision-support tool mainly to investors. Investors may use the system to match the relative credit risk of an issuer with their own risk tolerance in making investment decisions and portfolio management. However, the system may also be useful for other market participants, such as companies to make financing decisions (on the cost of capital and capital structure, for example) or financial institutions to assess counterparty risk.

As opposed to previous studies using AI to predict corporate rating grades, the proposed methodology integrates: (1) interval-valued fuzzy sets to address the issue of high linguistic uncertainty of expert opinions; (2) genetic feature selection to achieve high interpretability and verifiability of the rule base; and (3) evolutionary optimization of FRBSs to guarantee a high accuracy of the prediction system. In addition, this is, to our best knowledge, the first study comparing the performance of rating prediction models across multiple regions.The remainder of this paper has been organized in the following way. Section 2 provides a brief overview of soft computing applications in financial distress prediction. Section 3 describes the research methodology, including missing data treatment and the feature selection process. Section 3 also introduces the design of IVFRBSs for corporate investment/non-investment grade prediction. Section 4 presents the datasets used in this study. Section 5 examines the performance of IVFRBSs, mainly in terms of the yield obtained. Section 6 concludes and discusses both the results and the possible future research directions.

1. Soft Computing in Corporate Financial Distress Prediction

Financial distress can be defined as a situation that clearly shows an enterprise's financial difficulty, such as statutory bankruptcy and credit default [26]. Credit default is usually estimated by rating grades. In the domain of financial distress prediction using soft computing, the research to date has tended to focus on bankruptcy rather than rating grades prediction. Here, we review a large and growing body of the literature that has investigated various combinations of AI methods in financial distress prediction.

Several papers have systematically reviewed recent research in this field. Kumar and Ravi [27] conducted a comprehensive review of the work on bankruptcy prediction during the 1968–2005 period. Based on this overview, the authors claimed that the most successful prediction models, rather than using a single method, are based on hybrid soft computing systems. These were categorized into (1) ensemble classifiers [28], (2) intelligent feature selection combined with classification [29], and (3) tightly integrated hybrid systems (evolutionary NNs, fuzzy NNs, etc.) [30]. However, later review studies have regarded only the latter two as soft computing approaches [31]. In their review, Verikas et al. [31] identified the following soft computing approaches for financial distress prediction: (1) genetic algorithms (GAs) in hybrid techniques (to select a subset of input features, to find the appropriate hyper-parameter values of a predictor, or to determine predictor parameters) [32], [33]; (2) rough sets in hybrid techniques (to select a subset of input features) [34]; (3) fuzzy set theory-based techniques (to increase transparency) [12]; (4) self-organizing maps in hybrid systems (for data exploration and visualization) [35]; and (5) combining traditional and soft computing techniques [36]. Lin et al. [37] categorize four soft computing approaches, GAs, Group method of data handling, rough sets, and fuzzy sets.

Sun et al. [26] also categorized soft computing approaches. In the first type, one algorithm (usually GAs or rough sets) is applied to choose the features of another classification algorithm. Second, one algorithm is applied to optimize the parameters for another classification algorithm. Finally, a new classification algorithm is produced by integrating two or more algorithms. For example, Cheng et al. [38] embedded logit analysis into the output layer of radial basis function NNs, whereas Chaudhuri and De [39] developed fuzzy SVMs to handle uncertainty and impreciseness in corporate data.

Other hybrid systems combine multiple criteria decision-making methods with soft computing methods. Wu and Hsu [40] employed TOPSIS to determine the optimal classifier and subsequently extracted knowledge from the classifier by using decision trees. Shen and Tzeng [41] combined feature selection using rough sets with multiple criteria decision-making methods to collect the knowledge of domain experts.

Taken together, previous studies of financial distress prediction have reported that (1) the feature selection process improves prediction accuracy, and (2) hybrid systems improve both accuracy and transparency. Although extensive research has been carried out on financial distress prediction, to our best knowledge, no single study exists that adequately covers the advantages of evolutionary IVFRBSs in a financial distress prediction model. However, evolutionary IVFRBSs have recently been employed in related financial applications, namely credit scoring, fraud detection and stock market trend prediction [42].

2. Research Methodology

3.1 Rating Grades and Datasets

A credit rating (rating grade) is an expert evaluation of the general creditworthiness of an obligor. This evaluation is conducted by experts from a rating agency, and it is based on a variety of financial and non-financial criteria. For companies, key financial indicators usually include profitability, leverage, cash flow, liquidity, and financial flexibility, whereas non-financial indicators include country risk, industry determinants, business risk, and competitive position. The rating grade is assigned to a company from a predefined rating scale. Higher rating grades represent a lower probability of default. For example, the Standard & Poor's rating agency uses 13 long-term rating grades (from AAA to D, and R, SD, and NR), which can be further categorized into two rating grades, namely IG (investment grade, from AAA to BBB) and NG (non-investment grade, from BB to D, R, SD, and NR). Whereas AAA represents the highest rating grade, D stands for default on all obligations, SD for default on selected obligations, NR for not rated, and R denotes under regulatory supervision.In addition, Standard & Poor's uses modifiers for long-term rating grades AA to CCC by the addition of a plus or minus sign to show relative standing within the major rating grades. The latter categorization is particularly important for investment decision making due to the restrictions imposed on investment instruments. Therefore, this two-class categorization is used in this study, although previous studies have examined both two-class and multi-class problems [5].

Prior studies have also focused on U.S. companies due to better data availability, although several country-specific studies have examined Taiwanese [43], Korean [6], and European companies [2]. Huang et al. [4] compared prediction models for U.S. and Taiwanese corporate datasets, whereas Hajek and Michalak [1] performed a comparative study of U.S. and European companies. Both these studies suggest that specific market-related financial ratios and their weights, respectively, are used as determinants in corporate rating evaluation. However, no studies have thus far compared the performance of rating prediction models across multiple countries or regions. Here, we compare the performance of IVFRBSs for four datasets, namely the United States, the EU, Other developed and Emerging countries. To provide consistent and comparable results, this categorization was adopted from the Standard & Poor’s rating agency. It is based on the fact that these regions differ in particular in financial reporting and asset valuation practices and accounting techniques. The emerging market dataset covers the following subgroups: Latin America & the Caribbean, China, India, Eastern Europe & Russia, Small Asia, and Africa & the Middle East. Australia, New Zealand and Canada are included in the dataset of other developed countries.

The assigned corporate rating grades are comparable across these regions because a single scale is used by Standard & Poor’s. As regards the industries, mining companies and financial institutions (banks and non-bank financial institutions) were excluded from the datasets because they require specific input variables.

3.2 Input Variables

As shown in Table 1, we examined several subgroups of financial indicators: valuation ratios (market value ratios); dividends; growth potential; financial strength (leverage and liquidity ratios); and profitability ratios (management effectiveness indicators). This list of input variables is based on both the methodological reports released by leading rating agencies and the recent literature on variable selection in rating grades prediction (see, for example, [1] for an overview). Taken together, previous studies suggest that financial strength, profitability and valuation ratios represent the most important categories of the determinants of rating grades.

Table 1

3.3 Data Preprocessing

About 6% of the data were missing due to incomplete records. In a comparative study by [44], the following imputation methods performed best for the rule induction learning classifiers: fuzzy *k*-means clustering, ε-SVR (support vector regression), and event covering. Using the IVTURS with 5 membership functions as the benchmark classifier, we compared the 5×2 cross-validation performance of the three imputation methods in terms of classification accuracy. On average, the accuracy was increased by 0.45±4.63%, 6.84±6.64%, and 1.96±5.40% for fuzzy *k*-means clustering, ε-SVR, and event covering, respectively, compared with the base learner (ignoring the missing data). Therefore, ε-SVR was used for their imputation, although we recognize that the performance improvement was achieved at the expense of hyper-parameter optimization. In the ε-SVR model, all variables except the missing one are used to estimate the missing value as an output variable. The RBF kernel function with radius γ=0.4, penalty parameter C=23, and epsilon in loss function ε=1.0 were used as the parameters of the ε-SVR imputation model.

The data were divided into training and testing sets by using 5×2 cross-validation (two-fold cross-validation repeated five times), which is recommended because it directly measures variation and it has been found to be more powerful for comparing classifiers than 10-fold cross-validation in terms of statistical testing [45].

As indicated above, financial ratios are rather specific for companies in different regions. Moreover, recent studies have demonstrated that the performance of financial distress prediction models can be significantly improved by using a feature selection/extraction scheme [29], [46]. Wrapper feature selection methods have provided particularly promising results compared with filter approaches [1]. Therefore, we applied a genetic feature selection process developed specifically for FRBSs by [47]. This approach is called steady-state EA for feature selection and it represents a wrapper algorithm that uses accuracy provided by the *k*-nearest neighbors algorithm on training data as an evaluation function. For the experiments, the parameters of the algorithm were defined as follows: k=1, generations=5000, population=100, and the number of features to be selected={3, 5, 7, … , 15}.

3.4 Interval-Valued Fuzzy Rule-Based Systems

IVFRBSs utilize interval-valued fuzzy sets (IVFSs) to generalize FRBSs. IVFSs, also known as grey or vague sets, represent a special case of both L-fuzzy sets and type-2 fuzzy sets [48]. An IVFS A on a non-empty set X is an object having the form $A=\{\langle x, M_A(x)\rangle | x\in X\}$, where the function $M_A:X\rightarrow D[0,1]$ such that $x\rightarrow M_A(x)=[M_{AL}(x), M_{AU}(x)]$ defines the lower extreme and the upper extreme, respectively, of the interval $M_A(x)$ [49]. Hereinafter, we denote the lower bounds by x^L and upper bounds by x^U; this is $x=[x^L, x^U]$. The length of the interval $x=[x^L, x^U]$ is called the degree of uncertainty of x and is defined as $\pi(\mathrm{x})= x^U - x^L$.

The interval-valued restricted equivalence function (IV-REF): $[0,1]^2 \rightarrow [0,1]$ was introduced to measure the degree of equivalence between two intervals [49]. Bustince et al. [50] proposed several methods for the construction of IV-REFs from automorphisms and implication operators. Most importantly, for any t-norm T and any t-conorm S in [0, 1], $\text{IV-REF}(x, y)=[T(\text{REF}(x^L, y^L),\text{REF}(x^U, y^U)), S(\text{REF}(x^L, y^L),\text{REF}(x^U, y^U))]$ is an IV-REF.

The IVFRBS used in this study is called IVTURS [25]. This system is composed of three stages: (1) initialization, (2) extension of the fuzzy reasoning method on IVFSs, and (3) tuning the parameters of the IVFRBS.

The initial IVFRBS was generated by using a fuzzy association rule-based classification (FARC) method [51]. This approach uses a search tree to find the most interesting base of rules. In the initial IVFSs, lower bounds are obtained from the initial MFs and upper bounds are initially set to have a 50% larger amplitude than that of lower bounds (ignorance parameter W=0.25). In other words, the support of upper bounds is 25% larger than that of lower bounds. Initial IV-REFs are defined by using the identity function as automorphism $\varphi(x)=x$. The fuzzy reasoning process of the IVFRBS starts by using IV-REFs to compute the interval matching degree between the patterns and the antecedent of the rules. This is done by applying an interval-valued t-norm ($\mathbf{T}_{T_a,T_b}$) to the equivalence degree between the interval membership degree and ideal interval membership degree 1_L in the antecedents of the rules can be quantified as follows

$$[A_j^L(x^p), A_j^U(x^p)] = \mathbf{T}_{T_a,T_b}\begin{pmatrix} IV-REF([A_{j1}^L(x^{p1}), A_{j1}^U(x^{p1})],\ [1,\ 1]),\ \dots\ , \\ IV-REF([A_{jn}^L(x^{pn}), A_{jn}^U(x^{pn})],\ [1,\ 1]) \end{pmatrix}, \tag{1}$$

where x^p are input patterns, p=1, 2, … , P, n is the number of input variables, and $A_j=[A_j^L, A_j^U]$ are IVFSs, and j=1, 2, … , R are the indexes of the rules.

Next, interval association degrees are calculated from the interval matching degrees $[A_j^L(x^p), A_j^U(x^p)]$ and rule weights $w_j=[w_j^L, w_j^U]$. The rule weights are determined as follows

$$[w_j^L, w_j^U] = \frac{\sum x^p \in C[A_j^L(x^p), A_j^U(x^p)]}{\sum_{p=1}^{P}[A_j^L(x^p), A_j^U(x^p)]}, \tag{2}$$

where C is class corresponding to the j-th rule, and P is the number of examples of the training set. A combination operator is used to combine the interval matching degrees with the rule weights. The interval aggregation function is then used to aggregate the positive interval association degrees into the interval soundness degree. In the last step, the maximum interval soundness degree is selected to determine the predicted class.

The evolutionary method CHC [52] is applied to find the most appropriate sets of both IV-REFs and fuzzy rules. IV-REFs are optimized by using two automorphisms $\varphi_1(x)=x^a$ and $\varphi_2(x)=x^b$, where the values of parameters a and b are chosen from the interval [0.01, 100]. A subset of the initial rule base is selected by using a binary codification that expresses whether the fuzzy rule belongs to the rule base. Classification accuracy was used as the fitness function for chromosome evaluation. Note that the ignorance parameter W of IVFSs is not evolutionary tuned in the IVTURS used in this study, originally called IVTURS-FARC. In fact, this method has shown to be superior to the approach that simultaneously tuned the ignorance parameter W and IV-REFs [25].

3.5 Performance Evaluation

An important consideration in evaluating financial distress prediction models is to estimate their ability to predict each class value. Among the standard performance measures applied in classification tasks, the false positive (FP) rate has been reported as being particularly important in financial distress prediction studies owing to its possible serious financial consequences. Specifically, the wrong prediction of a company that becomes financially distressed may eventually lead to the loss of investment. On the contrary, predicting financial distress for a healthy company may lead to the loss of potential return. However, this problem of cost-sensitive classification has not been adequately addressed in previous studies.

In this study, we use a performance measure that is suitable for the classification of IG/NG because it maximizes an investor's yield. This yield measure combines the return and default rates of the IG and NG classes, respectively. Standard & Poor's reported a 0.03% default rate (the probability of default) for IG and 1.71% for NG in 2011. Consequently, the return rates for NG are higher due to the higher credit risk. In this case, Standard & Poor's reported an average return (between 1983 and 2013) of 8.1% for IG and 9.4% for NG. By combining these default and return rates, we use a yield measure that takes into consideration the probabilities of both returns and losses as follows

$$\text{Yield [\%]}=(TP\times(8.1\times(1-0.0003)-100\times0.0003)+TN\times(9.4\times(1-0.0171)-100\times0.0171)+FN\times0.0+FP\times(8.1\times(1-0.0171)-100\times0.0171) / (TP+FP+FN+TN), \quad (3)$$

where TP is the true positive, FP is the false positive, FN is the false negative and TN is the true negative rate (see Table 2).

Table 2

In other words, we calculate the yield as the expected rate of return (return rate × survival rate – loss in case of default × default rate) for TP (return rate=8.1%, survival rate=99.97%, loss in case of default=100%, and default rate=0.03%) and TN (return rate=9.4%, survival rate=98.29%, loss in case of default=100%, and default rate=1.71%), whereas it is assumed that the investment would not be realized at all (actual return would be lower than expected return) for FN (return rate=0.0%) and as (expected return rate for IG × actual survival rate for NG – expected loss in case of IG default × actual default rate for NG) for FP, respectively. Note that for simplification, we assume 100% loss in case of IG/NG default. Hence, the objective of the classification is to maximize the yield matrix defined in Table 3. Hereinafter, we mainly report the yield in terms of percentages because of its easy interpretability to investors. However, to achieve easy comparability of the results, it is also possible to express this yield metric in the [0,1] range as the normalized yield:

$$\text{Yield}_{norm}=(\text{Yield} - \text{Yield}_{min})/(\text{Yield}_{max} - \text{Yield}_{min}), \quad (4)$$

where Yield_{min} is the yield for the classification accuracy of 0%, and Yield_{max} is that for 100% accurate classifier, respectively.

Table 3

Interpretability measures represent other important criteria of an FRBS's performance. However, there is an inverse relation between the accuracy and interpretability of FRBSs. The optimal trade-off between accuracy and interpretability largely depends on a user's requirements. Gacto et al. [53] presented an overview of interpretability measures, categorizing them according to knowledge-based components (rule base/fuzzy partition) and complexity/semantic interpretability. Here, we focus on traditional complexity measures, mainly due to the lack of semantic measures developed specifically for IVFSs. At the rule base level, we use two traditional measures, namely the number of rules and the number of conditions in the antecedent of a rule. The number of MFs is also used to control the complexity (granularity) at the fuzzy partition level.

3. Datasets

Since rating grades are assigned by the agency based on financial ratios covering a longer time period, we collected data from the publicly available Reuters Global Market Data for three consecutive years (2008–2010) and calculated the average values to overcome the business cycle effect (see Table 4). The three-year perspective was adopted to mimic the "rating through the cycle" process used by rating agencies to achieve rating stability [54].

The companies selected in the datasets were required to have a rating grade assigned by the Standard & Poor's rating agency in 2011 and, at the same time, they had financial statements available at the Reuters Global Market Data for 2008–2010. All the companies from the United States were listed on the New York Stock Exchange or Nasdaq. Most of the EU companies were from the United Kingdom (47), Germany (36) and France (29). Canada (50) and Australia & New Zealand (25) were included in the other developed countries. Emerging countries were mostly represented by the following sub-regions: Latin America & Caribbean (30), Small Asia (22), Africa and Middle East (8) and China (8). Regarding the industry breakdown, the following industries were most frequent in the datasets: retail (the U.S. dataset), telecom. services (the EU dataset), paper/forest products (Other developed), and construction (Emerging).

The rating categories (IG/NG classes) were obtained from Standard & Poor's for 2011. The collected rating grades can be classified as long-term issuer credit ratings (company's long-term overall creditworthiness expressed in foreign currency). Therefore, issue-specific characteristics such as maturity or seniority were excluded. Consistently for all datasets (the United States, the EU, Other developed and Emerging), the IG class was associated with higher market capitalization, a higher stock price to sales ratio and a higher price to book value ratio, but lower market risk Beta and leverage ratios. Regarding the frequencies of rating grades, 46.3% of all companies were classified into the IG class, although this proportion varied from 39.4% (U.S. dataset) to 68.7% (EU dataset) (see Table 4). These differences may be partially explained by the size of the companies in the datasets. Even smaller U.S. companies have rating grades assigned by rating agencies, whereas it is mostly large companies in other regions.

Table 4

Fig. 1 shows the results of the genetic feature selection process. Since 5×2 cross-validation led to 10 pairs of training/testing sets, the steady-state EA was employed for the feature selection in the 10 training datasets. The evaluation function in Fig. 1 is represented by the classification accuracy Acc [%] of the *k*-NN algorithm for the training data. For further experiments, those sets of variables were chosen for which the highest accuracies were achieved. On average, the number of variables varied: 4.2 for the Emerging, 5.4 for the Other developed, 8.2 for the EU and 9.4 for the U.S.

dataset.

Fig. 1: The accuracy of wrapper feature selection for different numbers of selected variables.

Only those input variables that appeared in at least 60% of the 10 datasets are presented in Table 5. Interestingly, the profitability ratios were the least important determinants despite the fact that ROA (or alternatively ROE) has been considered to be a significant determinant of rating grades in most prior studies [4]. This inconsistency may be due to the high variation in profitability in the monitored period 2008–2010, which was strongly affected by the financial crisis. This explanation is also supported by the higher importance of the determinants of resilience to the financial crisis such as company size (market capitalization) and financial market volatility (beta regression coefficient). Specifically, the volatility was reported to have critical implications for long-term resilience, whereas ROA is no match for assessing the resilience [55]. In most studies, market capitalization is regarded to be a significant determinant of firm survival during the financial crisis [56]. Market capitalization was selected for all the datasets, which corresponds to the significantly lower values observed for NG class in Table 4. Similarly, higher financial market volatility was observed for NG class across regions, indicating a high importance of Beta variable in feature selection. Overall, the informative value of the selected variables was in agreement with the differences between IG and NG classes detected in Table 4.

Table 5

The behaviour of prediction models strongly depends on the complexity of the datasets. To control the problem's complexity, we used several of the measures suggested by [57] and performed the complexity analysis of the datasets using the Keel software [58]. Specifically, these measures included (1) Fisher's discriminant ratio (F1) to measure overlaps in the feature values from different classes; (2) linear separability (L1); (3) the fraction of the points lying next to the class boundary (N1) to measure the separability of classes' distributions; and (4) the average number of samples per dimension (T2) to evaluate the curse of dimensionality. The high value of Fisher's discriminant ratio (F1) in Fig. 2 suggests that the EU dataset represents a more linear problem compared with the other three. By contrast, the lower value of F1 for the Emerging dataset indicates a strongly nonlinear problem, probably attributed to the greater variety of countries (markets) included in this dataset. The values of the L1 and N1 measures suggest that the EU dataset is more linearly separable and has larger margins between classes, respectively. Finally, the curse of dimensionality is the most severe for the U.S. dataset.

Fig. 2: Complexity measures of datasets, F1 (overlap of individual feature values), N1 and L1 (separability of classes), and T2 (curse of dimensionality).

4. Experimental Results

In this section, we compare the performance of IVTURS, a representative of IVFRBSs, with state-of-the-art evolutionary FRBSs. All experiments were performed by using the KEEL software [58]. In this study, IVTURS was trained by using the following set of parameters (these values were adopted from [25] and [51]: triangular MFs; the support of upper bounds 25% greater than lower bounds; *t*-norm minimum and *t*-conorm maximum; initial parameters of automorphisms a=1 and

b=1; population size=50; number of evaluations=20000; and bits per gene=30. An example of the triangular MFs (MFs=5) generated for the Div input variable is presented in Fig. 3.

Fig. 3: An example of the IVFSs for the Div input variable. The solid lines are the lower bounds, dashed lines are the upper bounds, and values of the linguistic variable Div are denoted by VS (very small), S (small), M (medium), L (large) and VL (very large).

In the first set of experiments, we examined the effect of granularity at the fuzzy partition level on both prediction performance (represented by the Acc [%] and Yield [%] measures) and the interpretability of the rule base (the number of rules R and number of conditions in the antecedent of a rule Ant). Specifically, we tested the performance of IVTURS for MFs={3, 5, 7}. The results in Table 6 suggest that a low number of MFs (3) was not sufficient for any of the datasets in terms of Yield. In fact, a substantial increase in Yield (and Acc) was achieved for MFs=5 (Other developed and EU datasets) and MFs=7 (Emerging and U.S. datasets), respectively. Regarding the relationship between fuzzy partition and rule base interpretability, a higher number of rules were required for higher granularity. On the other hand, the number of conditions in the antecedent remained low across all datasets irrespective of granularity. The higher dimensionality of the U.S. datasets accounts for the lower rule base interpretability. The sample of the rule base for the U.S. dataset for MFs=5 is presented in Appendix 2. Nevertheless, not even this case exceeds the recommended limit values of Ant=7±2 [50].

Table 6

We further compared IVTURS with other state-of-the art evolutionary FRBSs to demonstrate that IVFSs present a more efficient tool to handle the strong uncertainty in the evaluation process of rating grades. In FRBSs, EAs can be used to tune [21]: (1) fuzzy sets represented by MFs; (2) rule base; or (3) both these components simultaneously. In the first case, EAs mostly modify the shapes of MFs. Three strategies have been developed to optimize the rule base, namely the Michigan approach, Pittsburgh approach and iterative rule learning. The Michigan approach utilizes concept classifier systems, where each individual represents one coded rule. In the Pittsburgh approach, on the other hand, the whole rule base is coded in one chromosome. Iterative rule learning combines both approaches, coding one rule in one chromosome and generating the whole rule base gradually. In the comparative analysis, we used representatives of all three approaches:

The Michigan approach:

- An FRBS based on genetic cooperative-competitive learning (GCCL) [59], where the MFs for each input variable in the antecedents are fixed. The GCCL method was examined by using the following setting: MFs={3, 5, 7}, population=100, generations=10000, individuals replaced in the population=20, crossover probability p_c={0.8, 0.9, 1.0}, mutation probability p_m=0.1.
- Genetic Programming-based learning of COmpact and ACcurate fuzzy rule-based classification systems for High-dimensional problems (GP-COACH) [60]. The GP-COACH method was trained by using the following values of parameters: MFs={3, 5, 7}, population (initial number of rules R)=200, generations=20000, crossover probability p_c=0.5, mutation probability p_m=0.2.

The Pittsburgh approach:

- Fuzzy learning based on genetic programming grammar operators and simulated annealing (SP) [61], where the number of rules R must be given. Parameters for execution: MFs={3, 5, 7}, R={4, 8, 16, 32}, iterations=10000, mutation probability p_m=0.5, subpopulations=10.

Iterative rule learning:

- Fuzzy genetics-based machine learning (FHGBML) [62]. Parameters for execution: R={4, 8, 16, 32}, rule bases=200, generations=1000, crossover probability p_c=0.9, probability for a Michigan iteration=0.5, "do not care" probability for inactive variables=0.5.
- Structural learning algorithm in a vague environment with feature selection (SLAVE) [63] which extracts the best rule in each iteration. Parameters for execution: population=100, iterations without change=500, mutation probability p_m=0.5, crossover probability p_c=0.1.
- Improved SLAVE (NSVL) [64] extends the iterative model of SLAVE for learning both the antecedent and the consequent of the rule in each iteration. Parameters: population=100, iterations without change=500, mutation probability p_m=1.0, crossover probability p_c={0.8, 0.9, 1.0}.
- Steady-state genetic algorithm for extracting fuzzy classification rules (SGERD) [65]. Parameters: the number of rules per class in the final population Q=min[(14×n)/(2×M), 20] from 14×n candidate rules, where n is the number of input variables, and M is the number of classes.

Regarding the granularity of the above-mentioned iterative rule learning approaches, multiple fuzzy partitions with different granularities (MFs={2,3,4,5}) were simultaneously applied to generate fuzzy rules. Thus, the appropriate granularity of the fuzzy partition can be obtained for each attribute (see e.g. [62]). To compare the complexity of the models at the fuzzy partition level, only the Michigan and Pittsburg approaches were therefore employed (Table 7).

Table 7

Table 8 shows average performance across all datasets. Weighted average of values from Table 6 and Table 7 was used in the case of Acc and Yield, where the weights corresponded to the frequencies of companies in the datasets. Note, however, that overall performance generally depends on the structure of an investor's portfolio.

Table 8

Although the GCCL method performed worst in terms of both Acc and Yield, the experimental results for the individual datasets showed that it performed well for the Other developed dataset (Table 7), indicating that this algorithm is suitable for low-dimensional financial distress prediction problems.

The GP-COACH method performed particularly well for the higher granularity of the fuzzy partition with a prediction performance close to that of IVTURS. In fact, GP-COACH with MFs=7 achieved Acc=84.21% and Yield=6.32% for the Other developed dataset and Acc=90.91% and Yield=7.48% for the EU dataset (Table 7). However, in the case of the U.S. dataset, this was at the expense of rule base interpretability (R=17.60 and Ant=11.56).

The SP method also provided good performance for both the Other developed (Acc=84.74% and Yield=6.89%) and the EU datasets (Acc=89.29% and Yield=7.52%) (Table 7). However, these results were achieved for lower complexity at the fuzzy partition level (MFs=3). Interestingly, the increasing granularity did not lead to better prediction performance. As indicated above, the prediction performance of IVTURS increased with growing complexity at the fuzzy partition level. The advantages of lower and upper MFs are thus apparent for higher numbers of MFs.

In a further set of experiments, the granularity of the fuzzy partition was fixed at MFs=5 for the purpose of comparability with the iterative rule learning approaches (Table 9). In addition, five linguistic labels (from very small/very weak to very large/very strong) are also used by experts from leading rating agencies such as Moody's or Standard & Poor's in their rating methodologies (available at their web pages). It is therefore appropriate to set this level of fuzzy partition to enable the verification of the FRBSs by the experts. As indicated above, the GCCL method performed well for the Other developed dataset with Acc=85.26% and Yield=7.06% for MFs=5; however, it resulted in both low prediction performance and low interpretability for the U.S. dataset. On the other hand, the FHGBML method performed particularly well for the Emerging and U.S. datasets in terms of prediction performance, indicating good data performance with narrow margins between the classes. However, these results were achieved at the expense of the interpretability measures. Consistent with the findings of [58], we observed that the NSVL method dominated the original SLAVE method in terms of both prediction performance and interpretability. Although NSVL did not perform the best in terms of prediction performance, it provided highly interpretable FRBSs for all the datasets. Similarly, the SGERD method resulted in highly interpretable rule bases but low prediction performance in contrast to NSVL. Finally, IVTURS performed best for the EU and U.S. datasets, suggesting that it is particularly suitable for more linearly separable datasets.

Table 9

Again, we compared average performance across all datasets in Table 10. Similarly as in Table 8, weighted average of values from Table 9 was used in the case of Acc and Yield. An average prediction performance close to that of IVTURS was achieved by both the Pittsburgh approach (SP) and iterative rule learning (FHGBML and NSVL). In contrast to FHGBML, NSVL also performed well in terms of interpretability.

To compare the performance of FRBSs in terms of Yield, we employed the Friedman test [66], which is the most common nonparametric procedure for performing multiple statistical comparisons between more than two algorithms. This procedure first ranks the algorithms for each problem separately and then it tests the equality of the average ranks of the algorithms. In addition to evolutionary FRBSs, we also provide the performance for (1) FURIA [67] (the state-of-the-art FRBS extending the well-known RIPPER algorithm, the rule optimization process was carried out two times), (2) multilayer perceptron (MLP) (trained using one hidden layer with the number of neurons={5, 10, 15, 20} and 1000 iterations of backpropagation algorithm [68]), (3) SVM (Sequential Minimal Optimization algorithm [69] with polynomial kernel function and complexity parameter C={2^0, 2^1, … , 2^8}), and (4) probabilistic SVM (PSVM) [70] (trained with Gaussian RBF kernel function, the width of the Gaussian kernel θ={0.1,0.3, … ,10} and complexity parameter C={2^0, 2^1, … , 2^8}).

Table 11 shows that the existence of statistical differences among the tested methods was confirmed for multiple datasets with p=0.001. To detect pairwise differences between two algorithms within the Friedman test, we used the post-hoc Li procedure, which has proved superior in comparative empirical studies [70]. IVTURS performed the same as the SP, FHGBML, NSVL, FURIA, MLP, SVM and PSVM statistically in terms of pairwise differences.

Table 11

To compare the classification accuracy of evolutionary FRBSs and other methods with that of IVTURS on individual datasets, we used McNemar χ^2 test, which has proved reliable for comparing different classifiers based on the contingency table of misclassified testing data [45]. The results of the test indicate that the FHGBML, SVM and PSVM methods performed the same as IVTURS statistically for all the datasets. However, regarding interpretability at the rule base level, FHGBML had the disadvantage of high numbers of conditions in the antecedent of a rule, which exceeded the recommended limit value of 7±2 for the EU and U.S. datasets.

To test for the robustness of the results, we first examined the performance of the methods on more recent data. Considering both the stability of rating grades in time and the temporal sequence of the patterns, we verified the learnt models on data from 2011-2013 to predict IG/NG classes from 2014 (see Table 12). Since the stability of rating grades leads to high autocorrelation of the predicted classes, we have to overcome the problem of sample selection bias. Therefore, we used only the data from the original testing samples for this validation procedure.

The results in Table 13 summarize the performance of the methods in terms of average Acc, Yield and $\text{Yield}_{\text{norm}}$ for both periods. Although the performance of Yield was statistically comparable for the SP and FHGBML methods, even small differences in Yield may be critical from a financial point of view. Considering the weighted average Yield for classes from 2014, IVTURS achieved a mean value of 7.02%, the highest along with that of FURIA, but 0.15% greater than that of FHGBML and 0.78% greater than that of NSVL (Table 13). These differences would be even greater in the case of an investor oriented toward developed markets owing to the dominance of IVTURS for the EU and U.S. datasets. On the other hand, IVTURS provided the lowest Yield for the Emerging dataset compared with the other two methods. This finding may be attributed to the fact that the Emerging data represent the most nonlinear problem with the lowest number of input variables (4.2 on average), indicating that (1) the number of rules generated by IVTURS was not sufficient (compared with FHGBML, for example) and (2) the advantages of IVFSs become evident for prediction problems with higher dimensionality.

Table 12

Table 13

We further examined the performance of IVTURS using Yield as the fitness function. The effect of using this fitness function was twofold. First, as expected, Yield was increased to 8.10% on average across regions, while accuracy decreased to Acc=80.37%. Second, the rule base complexity increased to the number of rules R=9.88 on average. Moreover, these effects became stronger with growing complexity of the datasets, with the strongest effect on the U.S. dataset (Yield=8.01%, Acc=78.94%, and R=11.91).

5. Conclusion

The present study was designed to determine the effect of using IVFSs to predict corporate rating grades. Our work allows us to conclude that additional freedom and design flexibility in determining MFs can be used to achieve higher accuracy in financial distress prediction problems. The findings of this study indicate that this advantage increases with growing complexity at the fuzzy partition level.

To take advantage of hybrid systems, that is combining EAs and FRBSs, we employed evolutionary IVFRBSs. Thus, we were able to obtain highly interpretable and accurate IVFRBSs. To demonstrate the advantage of using IVFSs, we compared them with state-of-the-art evolutionary FRBSs. To measure prediction performance, we used a performance measure that maximizes an investor's yield by using historical data on return and default rates. Thus, the direct financial interpretability of prediction performance was provided. In addition to prediction performance, we measured the interpretability of the FRBSs at the fuzzy partition and rule base levels.

Taken together, our findings suggest that IVTURS is especially suitable for high-dimensional problems with sufficient granularity at the fuzzy partition level. By contrast, GCCL performed best for low-dimensional problems, while facing rule base interpretability issues in the case of higher dimensionality, similar to the GP-COACH algorithm. The SP algorithm provided the most promising results in terms of fuzzy partition interpretability, although it was outperformed by IVTURS with higher granularity in terms of prediction performance. Several state-of-the-art evolutionary FRBSs performed statistically the same as IVTURS in terms of the Yield obtained, namely the SP and FHGBML methods. However, SP and FHGBML suffered from rule base interpretability issues.

Finally, a number of important limitations need to be considered. First, the datasets apply only to non-financial companies. However, financial companies are also attracting considerable interest in the financial distress prediction literature. Specific input variables must be designed for financial companies and, therefore, further investigation into IVFRBSs in the financial distress prediction of these companies is strongly recommended. Second, although country risk was considered by using the datasets from different regions, industry risk was not considered in this study. Industry risk is thought to be an important component of business risk profile. Moreover, different industry breakdown may also affect feature selection and classification performance. As demonstrated in [12], a higher accuracy can be achieved for some industries and, financial strength variables are more important for manufacturing companies, while sales and growth rates are critical for retail industry. Future research should therefore consider additional variables, such as industry-specific growth trends or market structure. Third, the system was proposed for binary class prediction, ignoring the ordinal nature of rating grades. Indeed, multiple class prediction is a challenging problem, requiring a more complex model to distinguish the minor differences of determinants' values between adjacent rating grades. Fourth, IVTURS performed effectively on the tested datasets mainly due to the initial fuzzy association rule-based classification method. Thus, the most interesting rules were preselected to decrease the computational cost. Although IVTURS (with an average computational time of 105 s) was more effective than the SP (168 s) and FHGBML (223 s) algorithms, the NSVL algorithm performed significantly more effectively (12 s). In addition to the more effective modification of IVFRBSs, further work needs to be carried out to develop IVFRBSs for strong nonlinear problems. It would, for example, be interesting to assess the effects of using ensembles of IVFRBSs in technology credit scoring [72] or financial distress prediction [73]. Another future direction is the use of intervals' numbers to design arbitrary MFs because (1) good generalizability and interpretability can be obtained and (2) this approach is straightforward applicable to IVFSs [74], [75]. Finally, the performance evaluation of FRBSs in our study was limited to several (mostly conflicting) objectives. Therefore, a future study investigating multiobjective evolutionary FRBSs [76], [77] would also be interesting. Moreover, the results in terms of yield were strongly affected by the balance of IG/NG classes in the data. As a result, a higher yield was achieved for the datasets with a higher proportion of IG companies (the EU dataset with 68.7%), as opposed to the U.S. dataset with 39.4% of IG companies. The normalized version of the yield measure overcomes this problem, but the study of the effect of imbalanced datasets remains an open question for future research.

Acknowledgements

This work was supported by the scientific research project of the Czech Sciences Foundation Grant No: 13-10331S. We gratefully acknowledge the help provided by constructive comments of the anonymous referees.

References

[1] P. Hajek, K. Michalak, Feature selection in corporate credit rating prediction, Knowl.-Based Syst. 51 (2013) 72–84.

[2] M. Doumpos, D. Niklis, C. Zopounidis, K. Andriosopoulos, Combining accounting data and a structural model for predicting credit ratings: Empirical evidence from European listed firms, J. Bank. Financ. 50 (2015) 599–607.

[3] K. Hung, H. W. Cheng, S. S. Chen, Y. C. Huang, Factors that affect credit rating: An application of ordered probit models, Romanian J. Econ. Forecasting 16(4) (2013) 94–108.

[4] Z. Huang, H. Chen, C. J. Hsu, W. H. Chen, S. Wu, Credit rating analysis with support vector machines and neural networks: a market comparative study, Decis. Support Syst. 37(4) (2004) 543–558.

[5] P. Hajek, Municipal credit rating modelling by neural networks, Decis. Support Syst. 51(1) (2011) 108–118.

[6] Y. C. Lee, Application of support vector machines to corporate credit rating prediction, Expert Syst. Appl. 33(1) (2007) 67–74.

[7] J. Sánchez-Monedero, P. Campoy-Muñoz, P. A. Gutiérrez, C. Hervás-Martínez, A guided data projection technique for classification of sovereign ratings: The case of European Union 27, Appl. Soft Comput. 22 (2014) 339–350.

[8] P. Hajek, V. Olej, Credit rating modelling by kernel-based approaches with supervised and semi-supervised learning, Neural Comput. Appl. 20(6) (2011) 761–773.

[9] K. S. Shin, I. Han, A case-based approach using inductive indexing for corporate bond rating, Decis. Support Syst. 32(1) (2001) 41–52.

[10] P. Hajek, V. Olej, Predicting firms' credit ratings using ensembles of artificial immune systems and machine learning – An over-sampling approach, in Artificial Intelligence Applications and Innovations, L. Iliadis, I. Maglogiannis, and H. Papadopoulos, Eds., Berlin Heidelberg: Springer (2014) 29–38.

[11] C. C. Yeh, F. Lin, C. Y. Hsu, A hybrid KMV model, random forests and rough set theory approach for credit rating, Knowl.-Based Syst. 33 (2012) 166–172.

[12] P. Hajek, Credit rating analysis using adaptive fuzzy rule-based systems: an industry-specific approach, Cent. Eur. J. Oper. Res. 20(3) (2012) 421–434.

[13] H. C. Wu, Y. H. Hu, Y. H. Huang, Two-stage credit rating prediction using machine learning techniques, Kybernetes 43(7) (2014) 1098–1113.

[14] L. A. Zadeh, Soft computing and fuzzy logic, IEEE Software 11(6) (1994) 48–56.

[15] P. Ravisankar, V. Ravi, I. Bose, Failure prediction of dotcom companies using neural network–genetic programming hybrids, Inform. Sciences 180(8) (2010) 1257–1267.

[16] D. Wang, C. Quek, G. S. Ng, Bank failure prediction using an accurate and interpretable neural fuzzy inference system, AI Communications 29(4) (2016) 477–495.

[17] C. F. Tsai, M. L. Chen, Credit rating by hybrid machine learning techniques, Appl. Soft Comput. 10(2) (2010) 374–380.

[18] L. Kuncheva, Fuzzy Classifier Design. Berlin: Springer Verlag, 2000.

[19] F. Herrera, Genetic fuzzy systems: taxonomy, current research trends and prospects, Evol. Intel. 1(1) (2008) 27–46.

[20] O. Cordón, F. Herrera, F. Hoffmann, L. Magdalena, Genetic Fuzzy Systems. Singapore: World Scientific Publishing Company, 2001.

[21] H. Hagras, C. Wagner, Towards the wide spread use of type-2 fuzzy logic systems in real world applications, IEEE Comput. Intel. Mag. 7(3) (2012) 14–24.

[22] Q. Liang, J. M. Mendel, Interval type-2 fuzzy logic systems: theory and design, IEEE T. Fuzzy Syst. 8(5) (2000) 535–550.

[23] O. Castillo, P. Melin, A review on the design and optimization of interval type-2 fuzzy controllers, Appl. Soft Comput.12(4) (2012) 1267–1278.

[24] P. Melin, O. Castillo, A review on type-2 fuzzy logic applications in clustering, classification and pattern recognition, Appl. Soft Comput. 21 (2014) 568–577.

[25] J. A. Sanz, A. Fernández, H. Bustince, F. Herrera, IVTURS: A linguistic fuzzy rule-based classification system based on a new Interval-valued fuzzy reasoning method with tuning and rule selection, IEEE T. Fuzzy Syst. 21(3) (2013) 399–411.

[26] J. Sun, H. Li, Q. H. Huang, K. Y. He, Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches, Knowl.-Based Syst. 57 (2014) 41–56.

[27] P. R. Kumar, V. Ravi, Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review, Eur. J. Oper. Res. 180(1) (2007) 1–28.

[28] J. Heo, J. Y. Yang, AdaBoost based bankruptcy forecasting of Korean construction companies, Appl. Soft Comput. 24 (2014) 494–499.

[29] D. Liang, C. F. Tsai, H. T. Wu, The effect of feature selection on financial distress prediction, Knowl.-Based Syst. 73 (2015) 289–297.

[30] M. B. Gorzalczany, Z. Piasta, Neuro-fuzzy approach versus rough-set inspired methodology for intelligent decision support, Inform. Sciences 120(1) (1999) 45–68.

[31] A. Verikas, Z. Kalsyte, M. Bacauskiene, A. Gelzinis, Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey, Soft Comput. 14(9) (2010) 995–1010.

[32] C. H. Wu, G. H. Tzeng, Y. J. Goo, W. C. Fang, A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy, Expert Syst. Appl. 32(2) (2007) 397–408.

[33] H. Ahn, K. J. Kim, Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach, Appl. Soft Comput. 9(2) (2009) 599–607.

[34] C. C. Yeh, D. J. Chi, M. F. Hsu, A hybrid approach of DEA, rough set and support vector machines for business failure prediction, Expert Syst. Appl. 37(2) (2010) 1535–1541.

[35] J. Huysmans, B. Baesens, J. Vanthienen, T. Van Gestel, Failure prediction with self-organizing maps, Expert Syst. Appl. 30(3) (2006) 479–487.

[36] F. M. Tseng, L. Lin, A quadratic interval logit model for forecasting bankruptcy, Omega 33(1) (2005) 85–91.

[37] W. Y. Lin, Y. H. Hu, C. F. Tsai, Machine learning in financial crisis prediction: a survey. IEEE T. Syst., Man, and Cyber. 42(4) (2012) 421–436.

[38] C. B. Cheng, C. L. Chen, C. J. Fu, Financial distress prediction by a radial basis function network with logit analysis learning, Comput. Math. Appl. 51(3) (2006) 579–588.

[39] A. Chaudhuri, K. De, Fuzzy support vector machine for bankruptcy prediction, Appl. Soft Comput. 11(2) (2011) 2472–2486.

[40] T. C. Wu, M. F. Hsu, Credit risk assessment and decision making by a fusion approach, Knowl.-Based Syst. 35 (2012) 102–110.

[41] K. Y. Shen, G. H. Tzeng, A decision rule-based soft computing model for supporting financial performance improvement of the banking industry, Soft Comput. 19 (2015) 859–874.

[42] J. Sanz, D. Bernardo, F. Herrera, H. Bustince Sola, H. Hagras, (2014). A compact evolutionary interval-valued fuzzy rule-based classification system for the modeling and prediction of real-world financial applications with imbalanced data, IEEE T. Fuzzy Syst. 23(4) (2015) 973–990.

[43] S. C. Huang, Integrating nonlinear graph based dimensionality reduction schemes with SVMs for credit rating forecasting, Expert Syst. Appl. 36(4) (2009) 7515–7518.

[44] J. Luengo, S. García, F. Herrera, On the choice of the best imputation methods for missing values considering three groups of classification methods. Knowl. Inform. Syst. 32(1) (2002) 77–108.

[45] T. G. Dietterich, Approximate statistical tests for comparing supervised classification learning algorithms, Neural Comput. 10(7) (1998) 1895–1923.

[46] C. Orsenigo, C. Vercellis, Linear versus nonlinear dimensionality reduction for banks' credit rating prediction, Knowl.-Based Syst. 47 (2013) 14–22.

[47] J. Casillas, O. Cordón, M. J. Del Jesus, F. Herrera, Genetic feature selection in a fuzzy rule-based classification system learning process for high-dimensional problems, Inform. Sciences 136(1) (2001) 135–157.

[48] G. Deschrijver, E. E. Kerre, On the relationship between some extensions of fuzzy set theory, Fuzzy Sets Syst. 133(2) (2003) 227–235.

[49] L. A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning - I, Inform. Sciences 8(4) (1975) 199–245.

[50] H. Bustince, E. Barrenechea, M. Pagola, Restricted equivalence functions, Fuzzy Sets Syst. 157(17) (2006) 2333–2346.

[51] J. Alcala-Fdez, R. Alcala, F. Herrera, A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning, IEEE T. Fuzzy Syst. 19(5) (2011) 857–872.

[52] L. J. Eshelman, The CHC adaptive search algorithm: How to have safe search when engaging, in Foundations of Genetic Algorithms, G. J. E. Rawlins, Ed., San Mateo, CA: Morgan Kaufman (1991) pp. 265–283.

[53] M. J. Gacto, R. Alcalá, F. Herrera, Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures, Inform. Sciences 181(20) (2011) 4340–4360.

[54] F. Fei, A. M. Fuertes, E. Kalotychou, Credit rating migration risk and business cycles. J. Bus. Financ. Account. 39(1-2) 229–263.

[55] G. M. Markman, M. Venzin, Resilience: Lessons from banks that have braved the economic crisis - And from those that have not. Int. Bus. Rev. 23(6) (2014) 1096–1107.

[56] S. Claessens, H. Tong, S. J. Wei, From the financial crisis to the real economy: Using firm-level data to identify transmission channels. J. Int. Econ. 88(2) (2012) 375–387.

[57] T. K. Ho, M. Basu, Complexity measures of supervised classification problems, IEEE T. Pattern Anal. 24(3) (2002) 289–300.

[58] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, F. Herrera, Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework, J. of Multiple-Valued Logic Soft Comput. 17 (2010) 255–287.

[59] H. Ishibuchi, T. Nakashima, T. Murata, Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems, IEEE T. Syst. Man Cy. B 29(5) (1999) 601–618.

[60] F. J. Berlanga, A. J. Rivera, M. J., del Jesús, F. Herrera, GP-COACH: Genetic Programming-based learning of COmpact and ACcurate fuzzy rule-based classification systems for High-dimensional problems, Inform. Sciences 180(8) (2010) 1183–1200.

[61] L. Sánchez, I. Couso, J. A. Corrales, Combining GP operators with SA search to evolve fuzzy rule based classifiers, Inform. Sciences 136(1) (2001) 175–191.

[62] H. Ishibuchi, T. Yamamoto, T. Nakashima, Hybridization of fuzzy GBML approaches for pattern classification problems, IEEE T. Syst. Man Cy B 35(2) (2005) 359–365.

[63] A. Gonzalez, R. Perez, SLAVE: A genetic learning system based on an iterative approach, IEEE T. Fuzzy Syst. 7(2) (1999) 176–191.

[64] A. Gonzalez, R. Perez, Improving the genetic algorithm of SLAVE, Mathw. Soft Comput. 16(1) (2009) 59–70.

[65] E. G. Mansoori, M. J. Zolghadri, S. D. Katebi, SGERD: A steady-state genetic algorithm for extracting fuzzy classification rules from data, IEEE T. Fuzzy Syst. 16(4) (2008) 1061–1071.

[66] M. Friedman, The use of ranks to avoid the assumption of normality implicit in the analysis of variance, J. Amer. Statist. Assoc. 32(200) (1937) 675–701.

[67] J. Hühn, E. Hüllermeier, FURIA: An algorithm for unordered fuzzy rule induction, Data Mining and Knowl. Discov.19(3) (2009) 293–319.

[68] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning representations by back-propagating errors, in Cognitive Modeling, T. A. Polk and C. M. Seifert, Eds., Cambridge, MA: MIT Press (1988) pp. 213–220.

[69] J. C. Platt, Fast training of support vector machines using sequential minimal optimization, in Advances in Kernel Methods - Support Vector Learning, B. Schoelkopf, C. Burges, and A. Smola, A., Eds., Cambridge, MA: MIT Press (1998) pp. 41–65.

[70] H. Chen, P. Tino, X. Yao, Probabilistic classification vector machines. IEEE T. Neural Networks 20(6) (2009) 901–914.

[71] S. García, A. Fernández, J. Luengo, F. Herrera, Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power, Inform. Sciences 180(10) (2010) 2044–2064.

[72] S. Y. Sohn, D. H. Kim, J. H. Yoon, Technology credit scoring model with fuzzy logistic regression, Appl. Soft Comput. 43 (2016) 150–158.

[73] V. Georgescu, V. Georgescu, Using genetic algorithms to evolve type-2 fuzzy logic systems for predicting bankruptcy, Kybernetes 46(1) (2017) 142–156.

[74] V. G. Kaburlasos, A. Kehagias, Fuzzy inference system (FIS) extensions based on the lattice theory, IEEE T. Fuzzy Syst. 22(3) (2014) 531–546.

[75] V. G. Kaburlasos, G. Papakostas, Learning distributions of image features by interactive fuzzy lattice reasoning in pattern recognition applications, IEEE Comput. Intell. Mag. 10(3) (2015) 42–51.

[76] M. Fazzolari, R. Alcala, Y. Nojima, H. Ishibuchi, F. Herrera, A review of the application of multiobjective evolutionary fuzzy systems: Current status and further directions, IEEE T. Fuzzy Syst. 21(1) (2013) 45–65.

[77] M. B Gorzałczany, F. Rudziński, A multi-objective genetic optimization for fast, fuzzy rule-based credit classification with balanced accuracy and interpretability, Appl. Soft Comput. 40 (2016) 206–220.

Appendix 1: List of abbreviations and acronyms

Acc	accuracy	IVFRBS	interval-valued fuzzy rule-based system
AI	artificial intelligence	IVFS	interval-valued fuzzy sets
EA	evolutionary algorithm	IV-REF	interval-valued restricted equivalence function
FARC	fuzzy association rule-based classification	IVTURS	interval-valued FRBS with tuning and rule selection
FHGBML	fuzzy genetics-based machine learning	MF	membership function
		NN	neural network
FRBS	fuzzy rule-based system	PSVM	probabilistic support vector machine
GA	genetic algorithm	SGERD	steady-state genetic algorithm
GCCL	genetic cooperative-competitive learning	SLAVE	structural learning algorithm in a vague environment
GP-COACH	genetic programming-based learning of compact and accurate FRBS	SVM	support vector machine
IVFRBS	interval-valued fuzzy rule-based system	SVR	support vector regression

Appendix 2: Sample of IVTURS' if-then rules for U.S. region

Symbol	Rule
R_1	If Beta is M Then Class is NG with CF=[0.89,0.94]
R_2	If TD/E is L Then Class is NG with CF=[0.88,0.90]
R_3	If MC is VS and Beta is S and IH is M Then Class is NG with CF=[0.64,0.77]
R_4	If QR is VS and Beta is S and IH is VL Then Class is NG with CF=[0.69,0.77]
R_5	If QR is VS and PS is S and ROA is S Then Class is NG with CF=[0.46,0.55]
R_6	If TD/E is M and Div is VS Then Class is NG with CF=[0.78,0.87]
R_7	If Div is M Then Class is IG with CF=[0.85,0.97]
R_8	If PR is M Then Class is IG with CF=[0.49,0.86]
R_9	If MC is S Then Class is IG with CF=[0.49,0.86]
R_{10}	If QR is M and Beta is VS and Div is S Then Class is IG with CF=[0.55,0.73]
R_{11}	If Beta is VS and Div is S and IH is L Then Class is IG with CF=[0.51,0.78]
R_{12}	If Beta is S and PS is S and IH is L Then Class is IG with CF=[0.44,0.51]

VS is very small, S is small, M is medium, L is large, VL is very large, and CF denotes certainty factor's interval

Table 1: Input variables for IG/NG classes' prediction

Valuation ratios	
Beta	beta regression coefficient
PS	stock price to sales
PBV	price to book value ratio
MC	market capitalization
IH	shares held by institutional holders
Dividends	
Div	dividend yield
PR	payout ratio
Growth rates	
EPS	expected growth in earnings per share
PEG	(stock price / earnings) / EPS growth
S	growth in sales last year
Financial strength	
QR	quick ratio
TD/E	total debt to equity
TD/EBITD	market debt to EBITD
Profitability ratios	
ETR	effective tax rate
ROA	return on assets

Table 2: Confusion matrix for the prediction of IG/NG classes

Prediction/Target	IG	NG
IG	TP	FP
NG	FN	TN

Table 3: Yield matrix for the prediction of IG/NG classes

Prediction/Target	IG	NG
IG	8.07	6.25
NG	0	7.53

Table 4: Mean values ± St. Dev. of the input variables for the IG and NG classes

	Other developed		Emerging		U.S.		EU	
Variable	IG	NG	IG	NG	IG	NG	IG	NG
Beta	0.75	1.36	0.81	1.13	1.04	1.36	0.98	1.38
	±0.47	±0.65	±0.46	±0.55	±0.29	±0.39	±0.38	±0.59
PS	3.53	1.35	2.88	2.48	1.55	1.06	1.40	0.92
	±4.55	±1.61	±2.71	±2.83	±1.12	±1.66	±1.20	±1.08
PBV	2.81	1.66	3.00	2.71	3.52	2.89	3.34	2.45
	±1.84	±1.18	±2.93	±2.28	±3.72	±4.10	±5.43	±3.60
MC[a]	23,226	1,469	17,355	2,961	21,744	2,030	31,267	4,088
	±40,705	±3,360	±15,943	±4,742	±41,265	±3,280	±47,374	±9,985
IH [%]	41.0	45.7	35.5	34.4	71.6	73.5	39.3	52.3
	±18.2	±23.2	±21.5	±20.5	±22.2	±24.0	±16.2	±19.0
Div [%]	3.30	1.26	1.64	1.10	8.25	2.42	2.86	2.16
	±2.27	±2.11	±1.87	±2.20	±14.21	±9.01	±2.35	±2.91
PR	0.7	1.9	0.3	0.2	1.5	0.8	0.8	2.5
	±0.5	±1.1	±0.2	±0.2	±4.8	±2.4	±1.0	±2.5
EPS [%]	6.4	14.0	16.9	28.5	6.6	9.9	6.2	17.5
	±7.07	±10.4	±13.5	±17.6	±7.7	±11.6	±9.8	±48.0
PEG	2.96	7.53	2.39	1.34	3.40	2.61	2.92	1.57
	±1.63	±28.6	±2.59	±1.98	±4.43	±3.61	±2.02	±1.06
S [%]	8.0	1.9	11.0	22.9	9.9	8.2	1.4	5.5
	±9.3	±3.2	±22.1	±71.8	±20.2	±19.7	±3.9	±14.1
QR	8.11	3.94	8.59	4.22	8.56	7.33	9.89	9.02
	±11.30	±5.63	±9.51	±5.58	±12.39	±13.37	±8.62	±15.97
TD/E	0.44	0.76	0.23	0.47	0.30	0.77	0.46	0.90
	±0.21	±0.41	±0.13	±0.14	±0.23	±0.39	±0.22	±0.48
TD/EBITD	15.5	13.9	13.8	22.3	1.08	2.38	24.3	5.51
	±81.9	±84.3	±75.9	±85.5	±7.95	±3.59	±102.7	±21.60
ETR	17.7	9.3	18.9	18.3	26.2	21.3	24.7	14.4
	±15.1	±16.0	±14.0	±17.6	±13.8	±19.2	±15.4	±18.8
ROA [%]	11.7	8.0	20.6	15.1	24.3	22.0	22.2	24.9
	±16.5	±18.2	±15.7	±15.2	±22.0	±44.1	±22.0	±99.6
N	37	38	43	59	239	368	136	62
N [%]	49.3	50.7	42.2	57.8	39.4	60.6	68.7	31.3

[a] in mil. US dollars

Table 5: Most frequently selected input variables

Region	Valuation ratios	Dividends	Growth	Financial strength
Emerging	Beta, PBV, MC		PEG	TD/E, TD/EBITD
EU	Beta, PS, MC	PR	PEG	
U.S.	Beta, MC	Div	EPS, S	TD/E, TD/EBITD
Other developed	PS, PBV, MC	Div, PR	S	QR

Table 6: Prediction Performance of IVTURS

MFs=3	Emerging	EU	U.S.	Other developed
Acc [%]	80.39±5.82	86.87±4.29	83.16±0.76	76.32±1.66
Yield [%]	6.76±0.36	7.52±0.15	6.90±0.11	6.66±0.38
R	4.80±1.60	5.00±1.90	7.00±2.10	6.00±1.67
Ant	1.93±0.29	2.13±0.44	2.19±0.51	1.92±0.32
MFs=5				
Acc [%]	79.61±5.35	88.69±1.74	83.29±1.39	82.11±4.21
Yield [%]	6.79±0.39	7.59±0.23	6.94±0.09	6.87±0.49
R	6.80±2.04	9.80±1.60	9.80±2.79	8.00±2.45
Ant	1.72±0.23	1.93±0.25	2.11±0.48	1.38±0.15
MFs=7				
Acc [%]	83.14±4.22	89.49±1.45	85.13±1.45	82.63±9.93
Yield [%]	6.85±0.30	7.63±0.21	7.02±0.06	6.64±0.68
R	8.20±2.48	11.20±3.82	13.80±5.08	10.20±3.06
Ant	1.78±0.14	1.71±0.29	2.24±0.21	1.37±0.17

MFs is the number of membership functions, Acc is accuracy, *R* is the number of rules, and Ant is the number conditions in the antecedent of a rule.

Table 7: Prediction performance of IVTURS compared with evolutionary FRBSs (Michigan and Pittsburgh approaches) – Average across regions

MFs=3	GCCL	GP-COACH	SP	IVTURS
Acc [%]	68.25±5.44	75.32±4.50	82.28±3.00	83.09±2.06
Yield [%]	5.65±1.08	6.26±1.06	6.84±0.50	6.96±0.44
R	12.30±6.65	5.75±1.73	7.00±1.38	5.70±1.82
Ant	1.61±0.47	2.76±1.03	6.80±2.68	2.04±0.39
MFs=5				
Acc [%]	72.76±4.95	78.94±5.00	82.72±3.70	83.90±2.09
Yield [%]	6.16±0.97	6.69±0.77	6.98±0.59	7.05±0.46
R	23.60±8.74	6.80±2.87	8.60±2.18	8.60±2.22
Ant	1.61±0.33	4.71±1.87	6.80±2.68	1.78±0.28
MFs=7				
Acc [%]	76.83±4.55	83.82±3.13	82.56±2.63	85.61±2.82
Yield [%]	6.44±0.99	6.82±0.61	6.80±0.46	7.04±0.54
R	22.80±12.27	8.70±2.23	5.20±1.78	10.85±3.61
Ant	1.46±0.37	7.68±2.65	6.80±2.68	1.77±0.20

Table 8: Prediction performance of evolutionary FRBS (Michigan and Pittsburgh approaches)

		GCCL		
MFs=3	Emerging	EU	U.S.	Other developed
Acc [%]	69.80±7.08	72.73±8.74	66.25±4.01	70.53±6.09
Yield [%]	5.57±0.73	7.57±0.12	4.99±0.24	5.99±0.59
R	6.60±2.15	9.80±5.56	22.20±15.07	10.60±3.83
Ant	1.40±0.37	1.32±0.21	2.12±1.04	1.61±0.26
MFs=5				
Acc [%]	69.80±4.57	79.80±7.90	69.41±4.13	85.26±4.28
Yield [%]	5.73±0.54	7.60±0.15	5.35±0.37	7.06±0.30
R	10.20±3.54	21.60±8.14	52.20±20.86	10.40±2.42
Ant	1.46±0.39	1.43±0.29	2.46±0.53	1.10±0.10
MFs=7				
Acc [%]	69.41±5.90	89.09±2.34	73.42±4.46	82.11±9.18
Yield [%]	4.13±0.58	7.61±0.14	5.64±0.34	5.74±0.69
R	13.40±4.50	26.00±10.99	38.40±31.27	13.40±2.33
Ant	1.55±0.39	1.35±0.20	1.59±0.64	1.34±0.24
		GP-COACH		
MFs=3	Emerging	EU	U.S.	Other developed
Acc [%]	73.33±5.76	87.07±3.69	72.76±4.30	67.89±6.53
Yield [%]	5.73±0.60	7.77±0.09	5.94±0.75	5.62±0.71
R	5.00±1.55	4.20±0.75	8.60±2.58	5.20±2.04
Ant	1.84±0.33	2.64±0.71	3.55±1.94	2.99±1.14
MFs=5				
Acc [%]	78.43±4.11	87.68±1.96	75.86±6.01	81.58±6.00
Yield [%]	6.22±0.48	7.60±0.13	6.28±0.80	6.65±0.52
R	5.20±1.60	5.20±2.40	11.20±7.00	5.60±0.49
Ant	2.62±0.66	5.07±1.56	6.58±3.79	4.57±1.47
MFs=7				
Acc [%]	78.04±3.80	90.91±1.43	82.43±3.10	84.21±6.86
Yield [%]	5.08±0.30	7.48±0.08	6.68±0.38	6.32±0.46
R	5.80±1.83	6.60±1.50	17.60±4.41	4.80±1.17
Ant	6.38±2.16	7.00±2.30	11.56±4.56	5.77±1.59
		SP		
MFs=3	Emerging	EU	U.S.	Other developed
Acc [%]	80.78±4.37	89.29±3.04	79.93±2.13	84.74±8.05
Yield [%]	6.46±0.31	7.52±0.29	6.68±0.19	6.89±0.39
R	6.40±1.96	6.40±1.96	8.00±0.00	7.20±1.60
Ant	4.20±1.60	8.20±3.49	9.40±3.67	5.40±1.96
MFs=5				
Acc [%]	81.18±4.22	88.48±2.44	81.12±3.74	82.63±5.91
Yield [%]	6.67±0.35	7.64±0.08	6.65±0.39	6.97±0.67
R	12.80±3.92	7.20±1.60	7.20±1.60	7.20±1.60
Ant	4.20±1.60	8.20±3.49	9.40±3.67	5.40±1.96
MFs=7				
Acc [%]	81.96±5.87	85.86±5.42	81.78±0.94	81.05±4.53
Yield [%]	6.69±0.38	7.39±0.26	6.66±0.21	6.50±0.35
R	5.60±1.96	4.80±1.60	5.60±1.96	4.80±1.60
Ant	4.20±1.60	8.20±3.49	9.40±3.67	5.40±1.96

Table 9: Prediction performance of IVTURS compared with evolutionary FRBSs for MFs=5

Dataset	GCCL	GP-COACH	SP	FHGBML
Emerging				
Acc [%]	69.80±4.57	78.43±4.11	81.18±4.22	**81.57±2.66**
Yield [%]	5.73±0.54	6.22±0.48	6.67±0.35	**6.99±0.38**
R	10.20±3.54	5.20±1.60	12.80±3.92	22.00±5.97
Ant	1.46±0.39	2.62±0.66	4.20±1.60	4.20±1.60
EU				
Acc [%]	79.80±7.90	87.68±1.96	88.48±2.44	87.27±4.81
Yield [%]	7.60±0.15	7.60±0.13	7.64±0.08	7.57±0.19
R	21.60±8.14	5.20±2.40	7.20±1.60	18.80±4.45
Ant	1.43±0.29	5.07±1.56	8.20±3.49	8.20±3.49
U.S.				
Acc [%]	69.41±4.13	75.86±6.01	81.12±3.74	81.78±3.21
Yield [%]	5.35±0.37	6.28±0.80	6.65±0.39	6.86±0.24
R	52.20±20.86	11.20±6.70	7.20±1.60	17.60±4.50
Ant	2.46±0.53	6.58±3.79	9.40±3.67	9.40±3.67
Other developed				
Acc [%]	**85.26±4.28**	81.58±6.00	82.63±5.91	80.00±5.67
Yield [%]	**7.06±0.30**	6.65±0.52	6.97±0.67	6.81±0.57
R	10.40±2.42	5.60±0.49	7.20±1.60	22.20±5.31
Ant	1.10±0.10	4.57±1.47	5.40±1.96	5.40±1.96
	SLAVE	NSVL	SGERD	IVTURS
Emerging				
Acc [%]	78.82±4.87	79.61±2.66	74.51±6.32	79.61±5.35
Yield [%]	6.55±0.42	6.55±0.20	5.72±1.08	6.79±0.39
R	3.80±1.60	2.00±0.00	2.40±0.80	6.80±2.04
Ant	3.68±1.80	1.70±0.51	1.90±0.20	1.72±0.23
EU				
Acc [%]	87.47±2.68	87.27±2.18	84.24±4.31	**88.69±1.74**
Yield [%]	7.57±0.10	7.41±0.32	7.11±0.95	**7.59±0.23**
R	5.60±2.06	3.60±0.80	2.60±0.80	9.80±1.60
Ant	6.96±3.12	2.72±0.77	1.95±0.10	1.93±0.25
U.S.				
Acc [%]	78.36±2.60	79.87±4.06	69.93±3.76	**83.29±1.39**
Yield [%]	6.51±0.16	6.59±0.31	5.42±0.32	**6.94±0.09**
R	9.60±3.07	6.80±2.48	2.80±0.75	9.80±2.79
Ant	8.37±3.40	3.47±0.62	2.00±0.00	2.11±0.48
Other developed				
Acc [%]	75.26±3.57	80.00±5.91	76.32±9.42	82.11±4.21
Yield [%]	6.39±0.44	6.77±0.31	6.44±0.88	6.87±0.49
R	5.20±0.75	3.00±0.89	2.60±0.49	8.00±2.45
Ant	4.85±1.82	1.68±0.43	1.93±0.13	1.38±0.15

Table 10: Prediction performance of IVTURS compared with evolutionary FRBSs for MFs=5 – Average across regions

	GCCL	GP-COACH	SP	FHGBML
Acc [%]	76.76±4.95	78.94±5.00	82.72±3.70	82.72±3.67
Yield [%]	5.98±0.99	6.57±0.77	6.88±0.59	7.03±0.48
R	23.60±8.74	6.80±2.87	8.60±2.18	20.15±5.05
Ant	1.61±0.033	4.71±1.87	6.80±2.68	6.80±2.68
	SLAVE	NSVL	SGERD	IVTURS
Acc [%]	80.00±2.93	81.34±3.68	73.78±4.57	83.90±2.09
Yield [%]	6.72±0.58	6.76±0.45	5.87±1.08	7.05±0.54
R	6.05±1.87	3.85±1.04	2.60±0.71	8.60±2.22
Ant	5.97±2.53	2.39±0.58	1.95±0.11	1.78±0.28

Table 11: Statistical comparison of accuracy (McNemar χ^2 test) and yield performance (Friedman and Li test)

	Accuracy				Yield	
	Emerging	EU	U.S.	Other developed	Multiple	
	χ^2	χ^2	χ^2	χ^2	aver. rank	*p*-value
GCCL	9.763***	24.329***	106.265***	2.750#	8.3	0.0044***
GP-COACH	0.082	0.390	43.107***	0.029	7.8	0.0159**
SP	0.070	0.000	4.614**	1.531	5.5	0.6610
FHGBML	0.552	0.837	2.294	1.114	5.8	0.5107
SLAVE	0.000	1.841	17.579***	8.446***	7.7	0.0201**
NSVL	0.000	1.029	10.404***	1.021	7.0	0.0872*
SGERD	6.961***	22.012***	100.765***	5.959**	9.4	0.0001***
FURIA	1.114	14.792***	1.853	2.914#	4.8	0.8094
MLP	1.161	20.779***	6.500**	0.000	6.3	0.2826
SVM	0.265	0.028	0.403	0.500	5.4	0.7422
PSVM	0.250	0.028	0.388	0.180	5.2	0.8608
IVTURS					5.0	
Friedman *p*-value					0.0001	

McNemar χ^2 test and post-hoc Li procedure was performed vs. IVTURS, *** IVTURS performed significantly better at p=0.01, ** at p=0.05, * at p=0.1 and # significantly worse at p=0.1.

Table 12: Frequencies of companies in the IG/NG classes from years 2011 and 2014

		Emerging		EU		U.S.		Other developed	
Year	Frequency	IG	NG	IG	NG	IG	NG	IG	NG
2011	*N*	43	59	136	62	239	368	37	38
	N [%]	42.2	57.8	68.7	31.3	39.4	60.6	49.3	50.7
2014	*N*	42	58	135	62	245	344	36	36
	N [%]	42.0	58.0	68.5	31.5	41.6	58.4	50.0	50.0

Table 13: Prediction performance of IVTURS with evolutionary FRBSs and other methods on IG/NG from 2011 and 2014 – Average across regions

	IG/NG from 2011			IG/NG from 2014		
	Acc [%]	Yield [%]	$Yield_{norm}$	Acc [%]	Yield [%]	$Yield_{norm}$
GCCL	76.76±4.95	6.44±0.99	0.79±0.22	68.29±3.49	6.04±0.26	0.68±0.23
GP-COACH	78.94±5.00	6.69±0.77	0.73±0.11	72.19±4.60	6.26±0.57	0.74±0.13
SP	82.72±3.70	6.98±0.59	0.83±0.09	76.90±4.48	6.73±0.22	0.81±0.09
FHGBML	82.72±3.67	7.04±0.48	0.84±0.09	75.09±2.91	6.87±0.19	0.83±0.07
SLAVE	80.00±2.93	6.76±0.58	0.75±0.21	65.38±2.42	6.00±0.57	0.76±0.11
NSVL	81.34±3.68	6.83±0.45	0.75±0.16	70.81±3.55	6.24±0.52	0.78±0.09
SGERD	73.78±4.57	6.17±1.08	0.71±0.20	67.97±14.01	5.88±0.50	0.63±0.18
FURIA	84.02±3.45	7.06±0.38	0.86±0.05	80.99±4.20	7.02±0.16	0.84±0.05
MLP	80.89±5.93	6.96±0.52	0.80±0.04	70.86±2.37	6.13±0.14	0.72±0.15
SVM	83.23±3.08	7.05±0.82	0.83±0.07	74.54±10.40	6.62±0.14	0.80±0.22
PSVM	84.64±4.62	7.07±0.55	0.87±0.09	79.03±2.21	6.93±0.12	0.83±0.09
IVTURS	83.90±2.09	7.05±0.54	0.86±0.07	78.68±2.79	7.02±0.15	0.83±0.07