Computational Economics
https://doi.org/10.1007/s10614-021-10135-4

The Impact of Financial Enterprises' Excessive Financialization Risk Assessment for Risk Control based on Data Mining and Machine Learning

Yuegang Song[1] · Ruibing Wu[2]

Accepted: 31 May 2021

Abstract
The purpose is to make full use of data mining and machine learning technology under big data to improve the ability of trade financial enterprises to cope with the risk of excessive financialization. In view of the above needs, based on previous studies, genetic algorithm (GA), neural network and principal component analysis (PCA) methods are used to collect and process the data, and build a risk assessment model of excessive financialization of financial enterprises. The performance of the model is analyzed through the data of specific cases. The results suggest that the data mining technology based on back propagation neural network (BPNN) can optimize the input variables and effectively extract the hidden information from the data. The specific examples show that most of the current enterprises do not have greater financial risk. However, most of the financial enterprise indexes show that the actual enterprise assets are gradually financialized. The total accuracy rate of financial risk assessment model based on deep belief network (DBN) is over 91%, and the accuracy of the model can reach 80% even if the sample size is small. Therefore, the financial risk assessment model proposed can effectively analyze the relevant financial data, and provide reference for the financial decision-making research of financial enterprises.

Keywords Artificial neural networks · Risk assessment · Genetic algorithm · Financialization · Data mining

✉ Yuegang Song
2016028@htu.edu.cn

Ruibing Wu
57378436@qq.com

[1] Business School, Henan Normal University, Xinxiang 453007, China

[2] School of Economics, Sichuan University, Chengdu 610065, China

Published online: 10 June 2021

1 Introduction

With the continuous development of the world economy in recent years, more and more enterprises are active in the financial market, which promotes the sustainable development of the national economy. However, the following problems are the fracture of many enterprise capital chains due to the poor risk awareness, financial market turbulence, commercial fraud or mismanagement. Small enterprises are on the verge of bankruptcy because of their poor ability to resist risks, which is mainly because more enterprises do not invest in the real economy but pay more attention to virtual financial markets (Anginer et al., 2018). With the advent of big data era, although the risk control model of enterprises plays a certain role in credit management, due to the lack of deep understanding of data, the corresponding model accuracy is low, and it is unable to evaluate the borrower comprehensively (Florio & Leoni, 2017). Due to the more in-depth research on big data tools by many scholars, more and more methods are applied to the financial field (Trelewicz, 2017). Traditional risk control model has problems such as single dimension and limited assessment ability. However, using data mining method can analyze the data of financial industry in depth, involving many dimensions and more comprehensive assessment (Moradi & Mokhatab Rafiei, 2019). With the establishment of distributed database and the gradual maturity of various data platform architectures, more data are stored and calculated, and the dimensions of data analysis are also wider (Medvedev et al., 2017). Especially with the construction of Hadoop cloud computing platform, the distributed computing ability of big data has been greatly improved, and the data foundation and computing ability are more efficient. Compared with the traditional risk control model, the model based on big data has great changes in the amount of information and accuracy (Ding et al., 2019; Ivanov et al., 2019). Big data can improve the credit system and help trade enterprises reduce credit risk (Saura et al., 2019; Urbinati et al., 2019). Data mining can use the network data to improve the scoring model and optimize the approval process through the integration of algorithms, so as to form a sustainable closed-loop management mode (Aldridge, 2019). Therefore, the use of data mining and its application in financial risk assessment has become one of the hot spots in this field.

Machine learning is a branch of artificial intelligence, which enables the system to learn automatically and improve from experience without explicit programming (Gu et al., 2020). The use of data mining and machine learning technology can be a good solution to the risk of trade enterprises. Among them, Zhu et al. (2019) proposed an enhanced hybrid integration algorithm by combining the two classical integration algorithms of random space and MultiBoosting, which can significantly improve the accuracy of predicting the credit risk of small and medium-sized enterprises (Zhu et al., 2019a); based on the relevant algorithms of big data machine learning, Yang (2020) constructed a variability model to adapt to early warning and control of financial risks, which minimizes the risks of financial enterprises and maximizes the interests of enterprises (Yang, 2020); Kim et al. (2019) proposed a machine learning financial risk detection and classification method based on feature selection, which can reduce enterprise financial risk by 10% (Kim et al., 2019); Gulsoy and Kulluk (2019) proposed an objective risk measurement method for the loan process of small and medium-sized business companies, and performed the classification task of data mining by collecting the data of current customers in the process of bank credit assessment (Gulsoy & Kulluk, 2019). It suggests that the application of data mining and machine learning in the financial field, especially in the financial risk management and control, is the trend of the times. However, in the related studies, different machine algorithms are not compared and analyzed, and the prediction accuracy of risk control is low, which seriously restricts the development of related enterprises.

Therefore, based on the above problems, machine learning and data mining algorithms are introduced into the risk assessment of financial enterprises. Back propagation neural network (BPNN) algorithm has strong learning and nonlinear mapping ability, data mining algorithm can effectively deal with complex and changeable financial data, deep belief network (DBN) can efficiently process and classify data, so it is conducive to deal with the problems faced by financial enterprises, and has certain feasibility. The above methods are used to build a suitable data processing and risk assessment model, which has a very important practical value for the assessment of financial enterprises and the processing of enterprise data.

This exploration is divided into five parts. The first part is the introduction, which is to put forward the importance of studying enterprise financial risk assessment, introduce data mining and machine learning theory, and determine the main research ideas. The second part is literature review. Through the analysis of the related research of data mining and machine learning in the field of financial risk control, the existing problems in the current research are clarified. The third part is to introduce the research methods, put forward the financial enterprise data processing and risk assessment model, and describe the details of the specific modeling, parameters and datasets. The fourth part is to discuss the research results, analyze the examples of the proposed model, obtain the performance and advantages of the model, and compare the proposed model with other algorithms. The fifth part is the conclusion, including the conclusion, actual contribution, limitations and future prospects.

2 Recent Related Work

2.1 Financial Risk Assessment Based on Data Mining

There are many studies about the application of data in financial risk assessment. However, most of the studies tend to separate the data from the model when discussing the relationship between the data and the model, ignoring the influence of the data characteristics on the model selection. When the model is used, some assumptions usually need to be made. If these assumptions are ignored, the effectiveness of the model may be greatly reduced. Jin et al. (2018) found that data mining technology has a good performance in the research of uncertainty theory. Through three improved association rule algorithms, the efficiency of data mining was greatly improved. Meanwhile, the concept hierarchy tree model of enterprise financial risk and the financial crisis early warning model of dynamic maintenance of time series were proposed. It was verified that the algorithm is suitable for enterprise financial risk analysis and crisis warning (Jin et al., 2018); Tavana et al. (2018) proposed a model using artificial neural network and Bayesian network, used the model to analyze the related finance, and found that the data mining method is applicable and has high efficiency, accuracy and flexibility (Tavana et al., 2018); Kara et al. (2020) developed a financial risk processing framework based on data mining to identify and evaluate the risks of different types of enterprises in the supply chain. The results reveal that the model can find hidden and useful information from unstructured risk data, so as to make wise risk management decisions (Kara et al., 2020); Pejić et al. (2019) found that with the support of big data technology, information stored in various sources of semi-structured and unstructured data can be collected, which is a better way to deal with financial enterprise risk assessment (Pejić Bach et al., 2019).

2.2 Financial Risk Assessment Based on Machine Learning

Many scholars have reported the research of machine learning in the field of financial risk assessment. Chen et al. (2017) studied the difference between the risk control of the Internet financial industry in the aspect of loan default and that of the traditional financial industry in the aspect of loan default, and found that the Internet financial industry has learned from many methods of the traditional financial industry in the aspect of risk control (Chen et al., 2017); Weng et al. (2018) predicted the Internet financial data by constructing a regression model based on kernel function, and verified that the model is superior to the existing methods in risk control (Weng et al., 2018). Sun et al. (2018) used logistic regression model to calculate loan default probability of people with different credit, which provided ideas for the further improvement of credit scoring card model (Sun & Vasarhelyi, 2018); in order to solve the problem of financial risk assessment more effectively, Gigović (2019) trained more than 300,000 pieces of data by using logistic regression model, support vector machine (SVM) and ensemble learning model random forest, and realized the automatic system of data processing and data modeling (Gigović et al., 2019); Zhu et al. (2019) predicted the default of online loan users by using the ensemble learning model random forest, so that the probability of default of users can be quantified, and a better credit rating system has been established (Zhu et al., 2019b); Thackham and Ma (2020) took a vehicle loan platform as the research object, constructed logistic regression model and Cox model, and studied the factors that affect users' default after borrowing. The results reveal that the factors positively correlated with the default rate of loan are: registered residence in the field and historical default records, while the factors negatively correlated with the loan violation rate are that relatives know the high credit rating (Thackham & Ma, 2020); Gao (2021) studied the experience of traditional micro lending in risk control and the role of bank's credit assessment system in risk control. On this basis, the risk factors in online lending were re added, and a new risk control model was constructed by using BPNN (Gao, 2021).

2.3 Related Research Summary

To sum up, more and more models are used in the field of credit risk classification. With the development of technology, the performance of the model is also improving. However, there is still a lack of research on the relationship between data features and model selection. In the field of credit risk, so far, there is no research on the data characteristics of credit data to make clear guidance for model selection. Therefore, the introduction of data-driven in credit classification modeling will significantly improve the classification effect, and then make the credit classification results more reliable. Literature research on risk control model mainly focuses on the identification of loan default in traditional financial industry. The research method is mainly to use single or individual algorithm to build the model, and the research on the comparison of multiple models is relatively less. In the application of Internet financial risk control model, the comparative study of many different types of machine learning algorithms can provide reference and basis for algorithm selection.

3 Method

3.1 Data Mining Technology Based on Artificial Neutral Network (ANN)

The core of finance is risk control. With the development of Internet technology, financial risk prediction is combined with big data analysis to establish a sound financial risk control system and promote the development of financial enterprises. By collecting all kinds of internal and external risk data, big data platform enriches the dimensions of risk data assessment and provides strong support for risk control. Data mining technology relies on application scenarios to mine risk attributes in risk data, build a complete data analysis system, and form a three-dimensional big data risk control system. The big data risk control system covering the whole life cycle of financial enterprises has been established to conduct integrated analysis on the breadth, depth and freshness of the risk data existing in the enterprise finance, covering the risk control requirements of all links (Abdulameer, 2018).

ANN is to solve problems intelligently by studying the structure and thinking processing mode of human brain physiological structure. ANN consists of a large number of simple neurons. The interconnection between neurons is essentially a mathematical model. By adjusting the connection weights between neurons, the input and output sample information is learned, the information hidden in the sample data is mined, and the ability of new samples to predict the results is obtained (Mrówczyńska et al., 2020). When BPNN carries out data training, the signal propagates forward, and when the output value is different from the ideal result, the network will adjust the parameters among the neuron layers through the error back-propagation, so as to minimize the prediction error square of the model (Sadiq et al., 2020). The structure of BPNN includes input layer, hidden layer and output layer. The neurons between the layers are connected, while the neurons in the layers are not connected. BPNN with hidden layer can fit the nonlinear function at any progress, effectively deal with the prediction problem with complex internal relationship, and do not need to describe the mapping relationship (Das et al., 2018; Khan et al., 2017). Therefore, in the data mining and prediction of financial enterprise financialization, BPNN has a good processing effect, so the data mining technology of enterprise financialization is established based on BPNN.

Principal component analysis (PCA) is a process of computing the original data matrix and generating a new linearly independent combination by means of dimensionality reduction. By combining variables, the data information in the original data is retained, and the combined variables with linearly independent are obtained (Asghari & Nematzadeh, 2016). Genetic algorithm (GA) is a complex evolutionary process from low level to high level by simulating the selection, inheritance and variation of organisms in nature. Figure 1 shows the algorithm flow. Therefore, the natural selection process is combined with natural genetic method in the biological world, and the optimal solution is iteratively searched in the sample selection process through the process

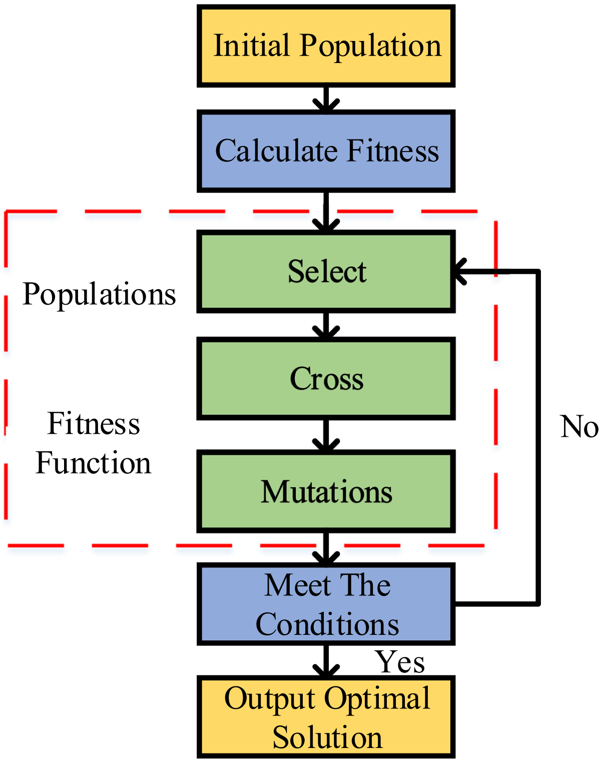

Fig. 1 GA flow chart

of chromosome replication, crossover and variation (Chen et al., 2020a; Chiba et al., 2018; Tang et al., 2020). GA has the following advantages: (1) it can search and solve non numerical problems by chromosome coding, and has good parallel ability; (2) it can improve the flexibility of the algorithm by using objective function as search information, and the genetic operator used has good scalability; (3) it has good search ability and prediction ability (Senthil & Ayshwarya, 2018; Sornam & Devi, 2016). Therefore, GA has a good application effect in processing engineering optimization and financial analysis and prediction.

PCA is used to reduce the dimension of data, project the data from the high-dimensional space to the low-dimensional space, re-express the data, and reduce the impact of data with increasing differences in classification. Moreover, the dimensionality reduction processing can re-select the data and select the data which has a greater effect on network classification from the original data combined with the relevant performance indicators (Alharbi & Alghahtani, 2019). Then, combined with the process of replication, crossover and variation of GA, the suitable fitness function is selected to guide the evolution process, and the high-quality genes are inherited to the next generation. Finally, the individuals with good quality are obtained and survive with a high probability. Therefore, GA is used to select the initial variables.

Based on this idea, GA-BPNN model is proposed. Figure 2 shows the structure of GA-BPNN. In this model, the fitness function represents the reciprocal of the sum of the squared errors between the output value and the actual value of the neural network.

$$f(x) = \frac{1}{mse} \tag{1}$$

$$mse = \sum_{i=1}^{n} (p_i - t_i)^2 \tag{2}$$

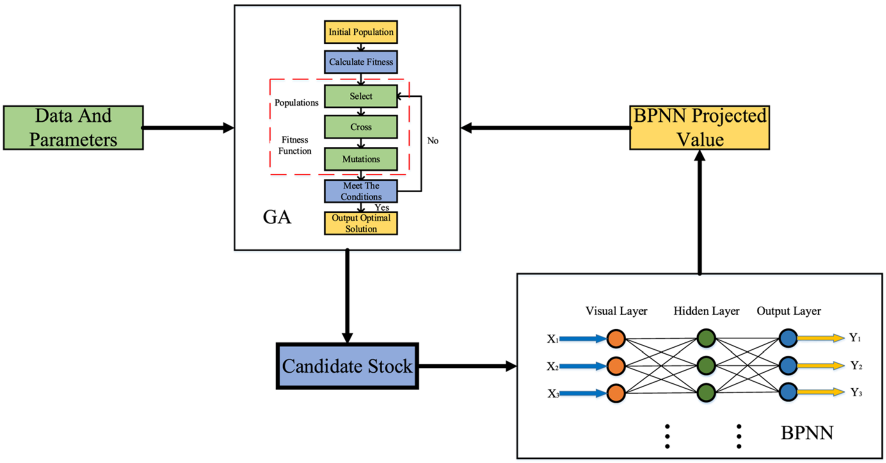

Fig. 2 GA-BPNN combination model

x represents chromosome, p represents actual value of target variable, t represents predicted value, and *mse* represents sum of the squared errors. The fitness function is used to get the combination of variables with small prediction error and eliminate the variables with poor prediction results.

In order to accurately describe the performance of GA-BPNN (Genetic Algorithm Back Propagation Neural Network) combination model, mean absolute percentage error (MARE), mean absolute percent error (MAPE), root mean square relative error (RMSRE), root mean square percentage error (RMSPE) and mean square percent error (MSPE) are used for comparison and verification. The calculation equations of MARE, MAPE, RMSRE, RMSPE and MSPE are as follows.

$$MARE = T^{-1} \sum_{t=1}^{T} \left| \frac{A_{\mathrm{t}} - P_{\mathrm{t}}}{A_{\mathrm{t}}} \right| \tag{3}$$

$$MAPE = T^{-1} \sum_{t=1}^{T} \left| \frac{A_{\mathrm{t}} - P_{\mathrm{t}}}{A_{\mathrm{t}}} \right| \times 100 \tag{4}$$

$$RMSRE = \left[T^{-1} \sum_{t=1}^{T} \left(\frac{A_{\mathrm{t}} - P_{\mathrm{t}}}{A_{\mathrm{t}}} \right)^2 \right]^{\frac{1}{2}} \tag{5}$$

$$RMSPE = \left[T^{-1} \sum_{t=1}^{T} \left(\frac{A_{\mathrm{t}} - P_{\mathrm{t}}}{A_{\mathrm{t}}} \right)^2 \times 100 \right]^{\frac{1}{2}} \tag{6}$$

$$MSPE = T^{-1} \sum_{t=1}^{T} \left(\frac{A_{\mathrm{t}} - P_{\mathrm{t}}}{A_{\mathrm{t}}} \right)^2 \times 100 \tag{7}$$

A_{t} and P_{t} represent the actual output value and theoretical output value in period t, and T represents the sample size. These indexes represent the difference between the actual value and the ideal value of the model output. Therefore, the smaller the index value is, the better the output of the model is (Aljarah et al., 2018; Ijjina & Chalavadi, 2016). The financial reports of 1000 financial enterprises are selected as experimental data to train and test the model. First, the combination model is used to reduce the dimension of the obtained data, the iterations of GA are 50, and the number of neurons in BPNN input layer is 30. The output optimization model is studied systematically. Then, the BPNN is used to mine the financial data of financial enterprises.

3.2 Risk of Excessive Financialization of Financial Enterprises

The development situation of the world financial enterprise organization is constantly changing. Enterprises turn to group development, business tends to be diversified, industrial boundaries are gradually blurred, cross industries are formed, and the development income is increasing. With the development of financial liberalization, driven by interests, entity enterprises are keen to invest in financial assets, which is called enterprise financialization (Bonizzi et al., 2020; Pariboni et al., 2020). The financial risk control model based on big data is used to analyze the risk data and to analyze and evaluate the current degree of enterprise financialization. According to the assessment results, scientific risk control should be implemented. Since there is no unified definition of enterprise financialization risk, the measurement of enterprise financialization is not uniform. At present, the research includes: the ratio of financial asset holding to total asset, the ratio of financial channel income to annual total income, the ratio of financial activity investment payment to enterprise investment activity total expenditure are used to quantify the degree of enterprise financialization. The financial level of enterprises will be quantified and analyzed from three aspects: financial assets and asset financialization, financial income and income financialization, financial investment and investment financialization.

(1) One aspect of enterprise financialization is that the enterprise transfers part of its assets from the production and operation field to the financial field, which is finally reflected in the financial assets of the enterprise's balance sheet. The degree of enterprise financialization can be described by the scale of financial assets and the proportion of financial assets. According to the regulations of the *Accounting Standards for Business Enterprises*, monetary funds, long-term receivables and interest receivable, investment real estate, loans and advance in cash, are eliminated (Davis, 2017). The equations of enterprise financial assets *Fass* and asset financialization *F_ass* are as follows.

$$Fass = FD+TFA+AFA+BFA+HMI+LAA \tag{8}$$

$$F\_ass=\frac{Fass}{TA} \tag{9}$$

Fass is financial assets; *FD* is financial derivatives; *TFA* is trading financial assets; *AFA* is available-for-sale financial assets; *BFA* is buying back the sale of financial assets; *HMI* is held-to-maturity investment; *LAA* is issuing loans and advance in cash; *F_ass* is asset financialization; and *TA* is total assets.

(2) Another manifestation of enterprise financialization is to use the financial assets held by the enterprise for financial investment, and to maximize the value of shareholders by the income from investment. The degree of enterprise financialization can be reflected by the status of enterprise financial income and the ratio of financial income to operating profit. Financial income includes investment income and profit and loss from fair value changes, but there are differences in

interest income and exchange earning (Bortz & Kaltenbrunner, 2018). Therefore, only investment income and profit and loss from fair value changes are regarded as financial income, and financial channel profit is analyzed from both broad and narrow sense. The equations of financial income and income financialization are as follows.

$$Fben = LI + GLC - GIAJV \tag{10}$$

$$F\_ben = \frac{Fben}{OP} \tag{11}$$

LI is investment income; *GLC* is profit and loss from fair value changes; *GIAJV* is investment income of associated enterprises and cooperative enterprises; *Fben* is financial income; *F_ben* is income financialization; *OP* is operating profit.

(3) For the quantification of financial assets in the previous sections, the cash in bank and cash in stock of financial enterprises are not considered, but these funds are used for the operation of production costs and financial investment. Therefore, the level of enterprise financialization can be also reflected through the financial investment status of financial enterprises and the proportion of financial investment in the cash outflow of investment activities. Financial investments are represented by cash paid for investments in the statement of cash flows. From the static point of view, the higher the ratio is, the higher the level of enterprise financialization is; from the dynamic point of view, the increase of the proportion indicates that the degree of enterprise financialization is deepening year by year (Cupertino et al., 2019; Grossule, 2019). Therefore, the equations of financial investment *Finv* and investment financialization *F_inv* of financial enterprises are as follows.

$$Finv = DCPI \tag{12}$$

$$F\_inv = \frac{Finv}{DFI} \tag{13}$$

DCPI represents cash paid for investment; *Finv* represents financial investment; *DFI* represents cash outflow from investment activities; *F_inv* represents investment financialization.

Through the extraction of data from the balance sheet, profit statement and statement of cash flows of financial enterprises, the indexes related to the financialization of financial enterprises are established. The absolute quantitative indexes include the amount of financial assets, financial income and financial investment; and the relative quantitative indexes are asset financialization, income financialization and investment financialization (Pernell, 2020). Through different quantitative indexes, the degree and trend of financial enterprise financialization are analyzed.

3.3 Financial Enterprises Financialization Risk Assessment Model based on DBN

In order to evaluate the risk of excessive financialization of financial enterprises, relevant financial risk measurement and early warning models are used for reference and combined with machine learning algorithm to establish a risk assessment model suitable for financial enterprise financialization degree assessment. Through the study of DBN, the improved DBN of classification partitioning restricted Boltzmann machine (CPRBM) is proposed, which can fully mine the characteristics of labelled data in financial enterprise data, and effectively monitor the financialization degree of enterprises.

Neural network is a multi-level network established by simulating the information transmission between neurons in human brain, including input layer, hidden layer and output layer. The neural nodes between adjacent layers are connected. The neurons in the layer are not connected. Traditional neural network training uses the iterative optimization method of weight adjustment to find the optimal solution. However, it is easy to produce local optimal solution, and the efficiency is low (Alameen & Gupta, 2020; Lim et al., 2018). Therefore, the deep learning method is used to train the neural network layer by layer, and multi-layer neural network is established for unlabelled data. Through massive data training, the network extracts and classifies the features of relevant data to form a new feature neural network finally. DBN is a multi-level neural network which integrates deep learning and feature learning. DBN adopts unsupervised learning method for layer by layer training, so its structure consists of multi-layer unsupervised RBM and one layer of supervised ANN (Alameen & Gupta, 2019). Figure 3 shows its structure.

DBN is composed of many RBM units. RBM is an ANN composed of random neurons. The neurons form a full connection, with a hidden layer and a visible layer. RBM neurons are random neurons, the value of state is determined by probability and statistics, and the output state is a binary parameter (Huda et al., 2018). Boltzman machine (BM) model has strong unsupervised learning ability, which can extract data features from complex data. However, the training speed is slow and the calculation accuracy is low, so it is difficult to get the required samples. Compared

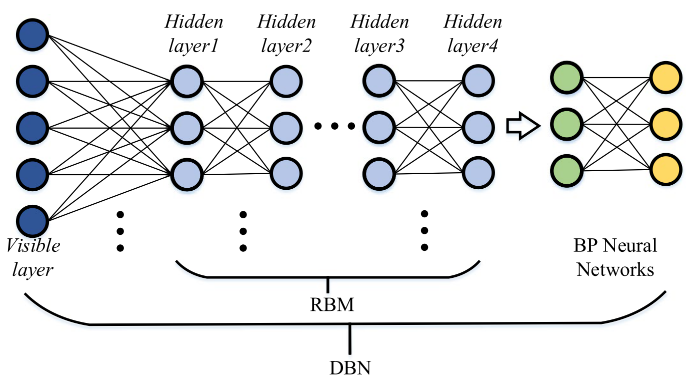

Fig. 3 DBN structure diagram

with BM, RBM is an undirected graphical model, which adopts intra layer connectionless. The structure consists of visible nodes and hidden nodes. The visible layer is the bottom layer of the network, and the hidden layer is the top layer of the network. The two layers are connected by weight. However, if the RBM visible layer unit state is given, the activation condition of hidden layer unit is independent (Ali et al., 2020; Chemweno et al., 2018). Through the obtained internal random samples, with the increase of the hidden layer in the network, the calculated results can approximate the discrete distribution arbitrarily.

First, the learning process of DBN is bottom-up unsupervised learning. The network is trained layer by layer using unlabelled training data to extract features from the data (Harshvardhan et al., 2020). Second, the last layer of BPNN receives the feature information from RBM (Restricted Boltzmann machine) network and conducts supervised learning to train entity relation classifier. According to the data characteristics of financial enterprise financialization, DBN based on classification learning RBM is proposed. The weight of RBM is increased for label related penalty terms, and the penalty term obeys Gaussian distribution. Third, RBM uses the label information to punish the weights in training, which can stimulate the learning ability of the relevant weights on the label information, and effectively improve the extraction of label features by the network. Finally, in the DBN network fine tuning stage, it is necessary to stop the penalty term and get the weight value highly related to the label information (Vidhya & Shanmugalakshmi, 2020).

In order to improve the learning ability of DBM, partition function is added to RBM and label is introduced into training to establish the CPRBM (Chrome Plated-Restricted Boltzmann machine). The model has a visible layer, two hidden layers, and an output layer (Fig. 4). Where, x_{label} represents sample data, y_{label} represents sample label, w_1, w_2, w_3 represent connection weight between two layers, and *Penalty* represents penalty term of classification partitioning. Red line represents the

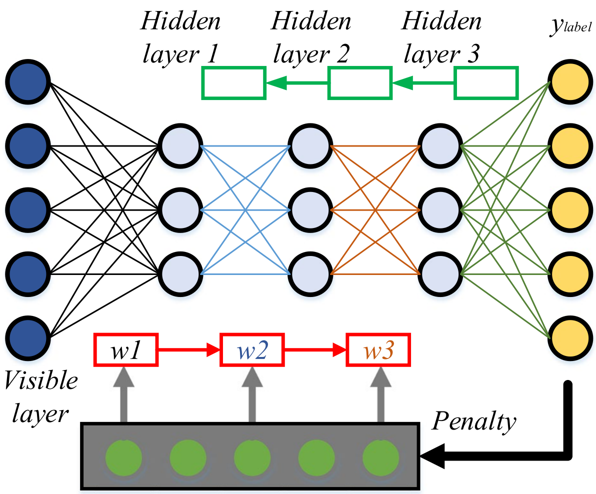

Fig. 4 RBM structure chart of classification partitioning

bottom-up generation process of RBM data in the first stage; the green line represents the second stage fine tuning data generation process.

First, the v and h_1 layers of RBM are trained, and then the h_1 and h_2 layers are trained. The generation probability equation is obtained as follows.

$$p(h_1 = 1|v) = sigmrnd(vw^{1T}) \tag{14}$$

$$p(v_1' = 1|h_1) = sigmrnd(h_1 w^1) \tag{15}$$

$$p(h_1' = 1|v_1') = sigm(v_1' w^{1T}) \tag{16}$$

Parameters such as weight and bias are adjusted according to the reconstruction error principle. The penalty item of classification partitioning is generated by label data. It acts on the connection weight and bias between the visible layer and the hidden layer, and plays a regulating role in the training network model (Ajesh et al., 2020; Harrou et al., 2018). Classification partitioning RBM needs to set the parameters of Gaussian classification penalty vector for each hidden layer. The classification partitioning matrix q is as follows.

$$Q = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1Snum} \\ q_{21} & q_{22} & \cdots & q_{2Snum} \\ \vdots & \vdots & \ddots & \vdots \\ q_{m1} & q_{m2} & \cdots & q_{mSnum} \end{bmatrix} \tag{17}$$

For the sample of each hidden layer, only the multiplication vector is used to process the sample, and then the penalty vector of this layer is as follows.

$$q = yQ \tag{18}$$

The purpose of classification partitioning is to increase the uncertainty of network training, and at the same time give different penalty weights to different training datasets, resulting in different effects of weights.

3.4 Model Training and Parameter Setting

(1) *Index selection* literature analysis shows that risk analysis based on financial indicators alone cannot meet the market demand, so the data will be input from the perspective of financial indicators and non-financial indicators. Regarding financial indicators, debt-paying ability indicators should be more concerned in the early warning of the company's credit risk, because it can more directly affect the credit risk of listed companies. Once the debt-paying ability indicators fluctuate abnormally, the enterprise is not far away from the occurrence of credit risk. Hence, five representative financial indicators are selected, which are current ratio, quick ratio, cash flow interest protection ratio, debt to asset ratio and cash ratio. Profitability also exerts a very crucial impact on credit risk early warning, reflecting the size of the profitability of enterprises in the market. Hence, four

profitability indicators are selected, which are sales gross profit margin, return on equity, return on total assets and sales cash ratio. The operation ability affects the management level of the enterprise, so it exerts an impact on the credit risk of the listed enterprises. Operation ability is the guarantee for an enterprise to maintain normal and reasonable operations. Three representative operational capability indicators are selected: receivables turnover ratio, inventory turnover ratio, total assets turnover. Development ability reflects the future development level of the enterprise and also exerts a direct or indirect impact on the credit risk of listed companies. Enterprises with strong development ability have better market prospects, and their ability to bear credit risk will become increasingly stronger. Thereby, the increase rate of main business revenue, growth rate of operating profit, net profit growth rate, and net assets growth rate are selected as four financial indicators to represent the development ability of the enterprise (Shen et al., 2019).

The indicator of the enterprise scale is selected from the perspective of the enterprise scale. Generally, the larger scale indicates the stronger the ability to resist economic risks and the greater the credibility in the hearts of investors. The employee quantity is employed as an indicator since it is a crucial factor to measure the scale of an enterprise (Yan et al., 2019). The industry cognition indicator is selected from the perspective of industry cognition. Generally, the longer the enterprise is established, the more experience it has in the industry. Hence, it can better grasp the development law of the industry and the impact of other economic fluctuations on the industry. Therefore, the industry cognition index is measured by the time of establishment. Considering that the operation of enterprises may be affected by external legal proceedings, the number of legal proceedings also reflects the size of the judicial risk of the normal operation of enterprises to a certain extent. Therefore, the judicial risk indicator is also included as a non-financial indicator to evaluate the credit status of enterprises, and is measured by the number of legal cases of enterprises (Yan et al., 2019).

(2) *Data source* most of the operation mode of financial companies differs from that of other types of companies. For example, most of them do not have inventory in their operation. Hence, there is no accounting account for inventory, and financial indicators such as inventory turnover rate cannot be obtained. Consequently, the non-financial listed companies listed in Shanghai and Shenzhen stock exchanges are selected. Obviously, the annual reports of enterprises audited by accounting firms can reflect the financial situation of enterprises more objectively and fairly, so the financial data collected come from the annual reports of listed companies. The collected non-financial index data comes from the industry-recognized data collection platform Tianyancha and Qichacha. All the listed companies in Shanghai and Shenzhen Stock Exchange are selected as the sample of the credit crisis to ensure the comprehensiveness of the sample. The remaining sample is 56 in total by excluding the financial listed companies and some samples with missing data; after the financial listed companies and missing samples are eliminated from the 200 component stocks of CSI Midcap 200 index, 170 non-financial listed companies are obtained as normal credit samples without default risk, a total of 226 samples. Finally, the ratio of normal

credit sample companies to abnormal credit sample companies is about 3:1. The financial data and non-financial data of 226 sample companies in 2016 are counted. The 226 sample companies are randomly divided into two groups. The first group is a training group that consists of 158 samples of 39 ST companies and 119 normal credit companies; the second group is the test group, including 17 samples of ST companies and 51 samples of normal credit companies, a total of 68 samples. The ratio of training group to test group is about 7:3.

(3) *Model parameters* regarding BPNN, the maximum training times are set to 100 times, the trainlm function is used for training, the mean square error is set to le-7, the learning rate is set to 0.01, and the momentum factor is set to 0.9. For SVM, σ is set to 1.5. In PCA, mtry is set to 2, ntree is set to 500, and other parameters are set to default. For the LSA model, refresh time is set to 30 s, N-bit is set to 1, E-bit is set to 0, and the range of group step timer is from 10 to 1800s. The parameters of the DTA and RFA models are consistent with that of the LSA model. For the LGBM model and DBN model, the learning rates of retraining and fine-tuning are both 0.05, and the mini-batch size is 100. Since the classification accuracy of samples tends to be stable after 20 iterations, the number of iterations (epoch) of samples is set to 20. For different combinations of layers and hidden layer nodes, the performance of the model is the best when the number of layers of DBN is 3 and the number of hidden layer nodes is 64.

4 Results and Discussion

4.1 GA-BPNN Performance

Figure 5 shows the MARE, MAPE, RMSRE, RMSPE and MSPE results of GA-BPNN training and prediction errors.

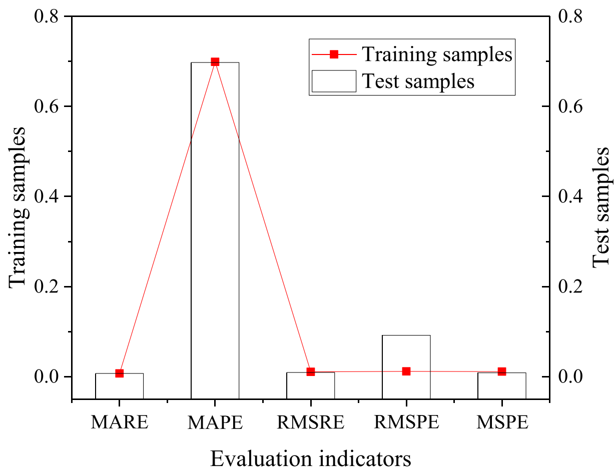

Fig. 5 Performance evaluation of GA-BPNN model

Figure 5 shows that the application of GA makes up for the problem that BPNN cannot select suitable variables. GA-BPNN has good effect in optimizing input variables, and can effectively extract hidden information from data. Therefore, GA combined with PCA method can effectively improve the extraction effect of BPNN.

4.2 An Analysis of the Financialization of China's Financial Enterprises

The degree of financialization in the life cycle stage of Chinese financial enterprises from 2015 to 2019 obtained from data mining is analyzed. Figure 6 shows the result.

Figure 6 shows that the absolute index of enterprise financialization is on the rise. The amount of financial assets and financial investment of the entity companies are increasing year by year, but the financial income is fluctuating. Among the relative indexes, asset financialization showed an upward trend after 2015; although income financialization was unstable in those years, the overall proportion was as high as 25%; investment financialization continued to increase year by year after 2015. Generally, the problem of over financialization has not yet occurred in enterprises. However, the financial indexes show that the real enterprises are changing from real to virtual and huge amount of enterprise wealth is transferred to financial industry. Therefore, it is necessary to prevent the financial crisis caused by the bubble of the financial industry, and assess and warn the possible phenomenon of enterprise financialization.

4.3 Performance Analysis of DBN

The accuracy of the algorithm is verified in three cases: (1) the training dataset of medium scale sample includes 4100 sample data. 60% are randomly selected for training, and 40% are randomly selected for testing. Figure 7 shows the classification

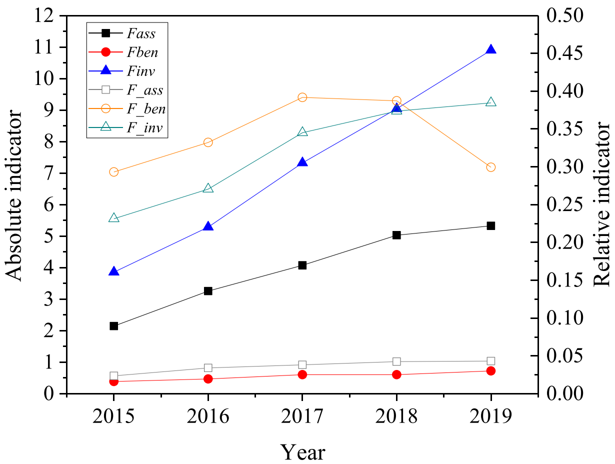

Fig. 6 Analysis of the degree of financialization of China' financial enterprises from 2015 to 2019

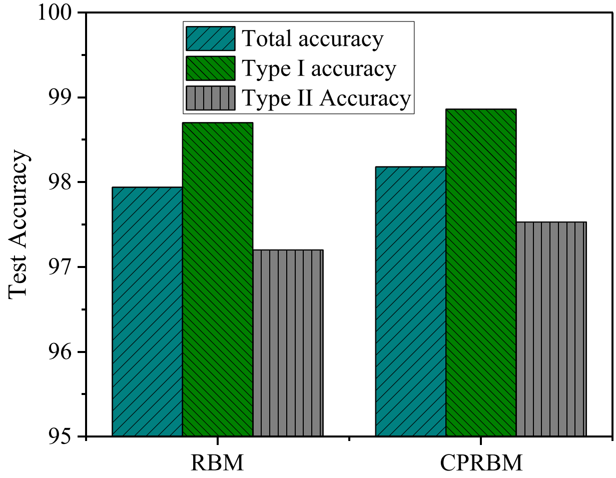

Fig. 7 Test results of medium scale testing dataset

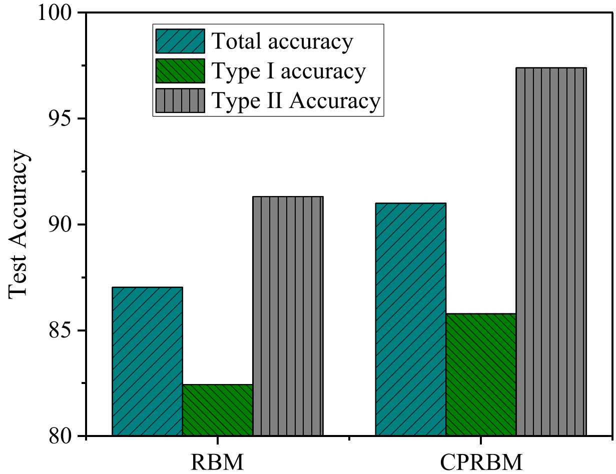

Fig. 8 Test results of small scale data sample set

results. Among them, the first type of accuracy represents the sample results with classification of 1, and the second type of accuracy represents the sample result with classification result of 0; (2) in the testing dataset of small scale sample, the total number of samples is 2000, and the proportion of data classified as 0 and 1 is 1:1; there are 1600 training data and 400 testing data, and Fig. 8 shows the classification results; (3) for the testing dataset of small scale, 60% labelled samples are deleted, so that the ratio of the number of samples classified as 1 to that classified as 0 is 1:3; Fig. 9 shows the classification results.

Figure 7 shows that RBM and CPRBM have excellent classification training probability. Moreover, the performance of the optimized CPRBM algorithm is

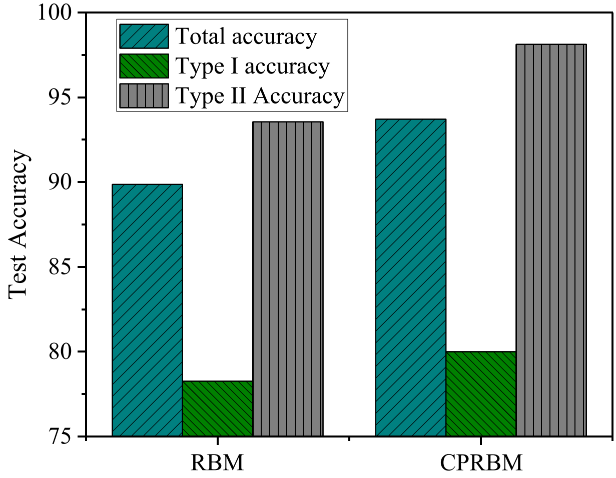

Fig. 9 Small scale data unbalanced sample set test results

better than that of the RBM algorithm. The total accuracy, Type I accuracy and Type II accuracy are better than those of RBM, which shows that the penalty vector used in the algorithm can effectively motivate the model to learn classification features.

Figure 8 shows that the accuracy of CPRBM is better than that of RBM, and the label data features learned by CPRBM are better than those of RBM, which can effectively prevent the occurrence of over fitting.

Figure 9 shows that, compared with RBM, CPRBM has higher total accuracy and two separate accuracy rates, and has better recognition effect for samples classified as 0; due to the small number of samples classified as 1 in the sample data, the recognition accuracy rate is low.

The test results of three groups of models with different feature data show that the optimized CPRBM algorithm has better classification performance than RBM algorithm. In the three groups of CPRBM experiments, the total accuracy rate is higher than 91%. Even in the case of a small number of samples, CPRBM still has a good accuracy rate of 80%. The test of small scale samples shows that due to the reduction of the number of samples classified as 1, the learning ability of the system for multiple samples is enhanced, and it can identify more accurately. Therefore, for a certain type of feature samples, by increasing the number of samples and using different penalty matrix in training, the recognition accuracy of the algorithm can be improved, and the system can be prevented from over fitting and falling into the local optimal solution.

In conclusion, based on the DBN financialization risk assessment model, CPRBM is used to verify the data of financial enterprises. The test results show that CPRBM can effectively improve the accuracy of training and test results compared with RBM algorithm. Even if no restrictive conditions are set on the sample, CPRBM can still effectively collect the hidden information of financial enterprises, and has a high accuracy rate in predicting the degree of enterprise financialization.

Table 1 Comprehensive comparison results of different data mining and machine learning risk control models

Performance	SVM	LSA	DTA	RFA	LGBM	DBN
AUC	0.7831	0.779	0.6974	0.7746	0.7746	0.7935
F1-Score	0.4055	0.3198	0.2335	0.3376	0.4336	0.4764
Accuracy	0.7884	0.7765	0.7561	0.7828	0.7877	0.7828
Precision	0.6913	6818	0.5579	0.7248	0.6591	0.6059
Recall	0.2869	0.2089	0.1476	0.2201	0.3231	0.3928
Overview	0.59104	0.5532	0.4785	0.56798	0.59562	0.61028

4.4 Comparative Analysis of Different Data Mining and Machine Learning Risk Control Models

In Table 1, all the optimized risk control models are randomly incorporated into the real financial data of an enterprise and the results of each index are scored comprehensively. Table 1 is the result. In a single model, the AUC (Area under the ROC curve) results show that the authenticity of the fusion model is 79.35%, followed by the SVM model, which is 78.31%. Most of the data mining and machine learning models are around 0.7, which suggests that these models can be used in the risk control prediction of trade-oriented enterprises. In the comprehensive assessment of accuracy and recall rate, the value of the fusion model is 0.4764, followed by the Light Gradient Boosting Machine (LGBM) model, which reaches 0.4336. It can be seen that the model proposed in this paper has obvious advantages in performance. In terms of accuracy, SVM is the best, the ratio is as high as 78.84%, followed by LGBM model, and the values of other models are higher than 75%. Therefore, in terms of accuracy, all models can meet the requirements. In terms of precision, random forest algorithm (RFA) is the highest, which is 72.48%, followed by SVM (69.13%). The highest recall rate is the fusion model (39.28%), followed by LGBM (32.31%). In a word, each model has its own unique advantages in different aspects of performance. The authenticity, recall and accuracy of the fusion model are high, and the precision of RFA is high.

5 Conclusion

In the big data environment, through the prediction and analysis of the existing financial risks, the financial risk assessment model of financial enterprises is established and optimized based on the original model. Through the analysis of data mining algorithm, the differences of different algorithm models in practical application and performance are comprehensively evaluated. The accuracy of the model is higher than 81%. When the dataset is small, the algorithm can get the optimal results under different datasets. When the dataset is large, DBN neural network model consumes less time; each model has its own unique advantages in different performance

(Chen et al., 2020b). Although the performance of all models is analyzed as comprehensively as possible in the research process, due to the limitation of the ability and the research funds, it is necessary to improve the research in the following aspects. (1) In the model performance assessment, only common indexes are used to judge the quality of the model, there is no explanation of the internal mechanism, and the internal relationship among these indexes is not analyzed from a whole perspective; (2) the data constructed mainly comes from the financial data of the enterprise, and the time span is small. It takes a long time to try different data mining and machine learning algorithms to train models. Therefore, when a large amount of data is analyzed, how to reduce the running time and improve the efficiency of algorithm learning is a problem that needs to be studied. In the future, in-depth research will be conducted in these directions.

Acknowledgements This research was funded by financial support of the National Social Science Foundation (The impact of heterogeneity of regional trade in services agreements on the reconstruction of global value chain of China's manufacturing industry Project no:20BJY091) and The Innovation Team Project in Philosophy and Social Sciences for Colleges and Universities of Henan Province (Coordinated development of urban and rural areas and Rural Revitalization; Project No: 2021-CXTD-04).

Author Contributions Yuegang Song: writing—original draft preparation; formal analysis, data curation; Conceptualization, methodology; Ruibing Wu: writing—review and editing, visualization, supervision. All authors have read and agreed to the published version of the manuscript.

Declarations

Conflict of interest All Authors declare that they have no conflict of interest.

Human and animal rights This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Abdulameer, A. T. (2018). An improvement of MRI brain images classification using Dragonfly algorithm as trainer of artificial neural network. *Ibn AL-Haitham Journal for Pure and Applied Science, 31*(1), 268–276.

Ajesh, F., Ravi, R., & Rajakumar, G. (2020). Early diagnosis of glaucoma using multi-feature analysis and DBN based classification. *Journal of Ambient Intelligence and Humanized Computing, 1*, 10–16.

Alameen, A., & Gupta, A. (2019). Clustering and Classification based real time analysis of health monitoring and risk assessment in Wireless Body Sensor Networks. *Bio-Algorithms and Med-Systems, 15*(4), 78–86.

Alameen, A., & Gupta, A. (2020). Optimization driven deep learning approach for health monitoring and risk assessment in wireless body sensor networks. *International Journal of Business Data Communications and Networking (IJBDCN), 16*(1), 70–93.

Aldridge, I. (2019). Big Data in Portfolio allocation: A new approach to successful Portfolio optimization. *The Journal of Financial Data Science., 1*(1), 45–63.

Alharbi, A., & Alghahtani, M. (2019). Using genetic algorithm and ELM neural networks for feature extraction and classification of type 2-diabetes mellitus. *Applied Artificial Intelligence, 33*(4), 311–328.

Ali, S. A., Raza, B., Malik, A. K., et al. (2020). An Optimally Configured and Improved Deep Belief Network (OCI-DBN) Approach for Heart Disease Prediction Based on Ruzzo-Tompa and Stacked Genetic Algorithm. *IEEE Access, 8*, 65947–65958.

Aljarah, I., Faris, H., & Mirjalili, S. (2018). Optimizing connection weights in neural networks using the whale optimization algorithm. *Soft Computing, 22*(1), 1–15.

Anginer, D., Demirgüç-Kunt, A., & Mare, D. S. (2018). Bank capital, institutional environment and systemic stability. *Journal of Financial Stability., 37*, 97–106.

Asghari, M., & Nematzadeh, H. (2016). Predicting air pollution in Tehran: Genetic algorithm and back propagation neural network. *Journal of AI and Data Mining, 4*(1), 49–54.

Bonizzi, B., Kaltenbrunner, A., & Powell, J. (2020). Subordinate financialization in emerging capitalist economies. *The Routledge International Handbook of Financialization, 9*, 177–187.

Bortz, P. G., & Kaltenbrunner, A. (2018). The international dimension of financialization in developing and emerging economies. *Development and Change, 49*(2), 375–393.

Chemweno, P., Pintelon, L., Muchiri, P. N., et al. (2018). Risk assessment methodologies in maintenance decision making: A review of dependability modelling approaches. *Reliability Engineering & System Safety, 173*, 64–77.

Chen, M. R., Chen, B. P., Zeng, G. Q., et al. (2020a). An adaptive fractional-order BP neural network based on extremal optimization for handwritten digits recognition. *Neurocomputing, 391*, 260–272.

Chen, M., Liu, Q., Huang, S., & Dang, C. (2020b). Environmental cost control system of manufacturing enterprises using artificial intelligence based on value chain of circular economy. *Enterprise Information Systems*. https://doi.org/10.1080/17517575.2020.1856422

Chen, Z., Li, Y., Wu, Y., & Luo, J. (2017). The transition from traditional banking to mobile internet finance: An organizational innovation perspective-a comparative study of Citibank and ICBC. *Financial Innovation., 3*(1), 1–16.

Chiba, Z., Abghour, N., Moussaid, K., et al. (2018). A novel architecture combined with optimal parameters for back propagation neural networks applied to anomaly network intrusion detection. *Computers & Security, 75*, 36–58.

Cupertino, S., Consolandi, C., & Vercelli, A. (2019). Corporate social performance, financialization, and real investment in US manufacturing firms. *Sustainability, 11*(7), 1836–1846.

Das, D., Pratihar, D. K., Roy, G. G., et al. (2018). Phenomenological model-based study on electron beam welding process, and input-output modeling using neural networks trained by back-propagation algorithm, genetic algorithms, particle swarm optimization algorithm and bat algorithm. *Applied Intelligence, 48*(9), 2698–2718.

Davis, L. E. (2017). Financialization and investment: A survey of the empirical literature. *Journal of Economic Surveys, 31*(5), 1332–1358.

Ding, H., Peng, C., Tian, Y., & Xiang, S. (2019). A risk adaptive access control model based on Markov for big data in the cloud. *International Journal of High-Performance Computing and Networking., 13*(4), 464–475.

Florio, C., & Leoni, G. (2017). Enterprise risk management and firm performance: The Italian case. *The British Accounting Review., 49*(1), 56–74.

Gao, J. (2021). Performance evaluation of manufacturing collaborative logistics based on BP neural network and rough set. *Neural Computing and Applications., 33*(2), 739–754.

Gigović, L., Pourghasemi, H. R., Drobnjak, S., & Bai, S. (2019). Testing a new ensemble model based on SVM and random forest in forest fire susceptibility assessment and its mapping in Serbia's Tara National Park. *Forests, 10*(5), 408–412.

Grossule, E. (2019). Regulatory strategies towards the commodity market financialization risk: Position limits' regime, transparency and enforcement tools. *European Business Law Review, 30*(2), 90–96.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies., 33*(5), 2223–2273.

Gulsoy, N., & Kulluk, S. (2019). A data mining application in credit scoring processes of small and medium enterprises commercial corporate customers. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery., 9*(3), e1299–e1306.

Harrou, F., Dairi, A., Sun, Y., et al. (2018). Statistical monitoring of a wastewater treatment plant: A case study. *Journal of Environmental Management, 223*, 807–814.

Harshvardhan, G. M., Gourisaria, M. K., Pandey, M., et al. (2020). A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review, 38*, 100285–100296.

Huda, S., Yearwood, J., Hassan, M. M., et al. (2018). Securing the operations in SCADA-IoT platform based industrial control system using ensemble of deep belief networks. *Applied Soft Computing, 71*, 66–77.

Ijjina, E. P., & Chalavadi, K. M. (2016). Human action recognition using genetic algorithms and convolutional neural networks. *Pattern Recognition, 59*, 199–212.

Ivanov, D., Dolgui, A., & Sokolov, B. (2019). The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics. *International Journal of Production Research., 57*(3), 829–846.

Jin, M., Wang, Y., & Zeng, Y. (2018). Application of data mining technology in financial risk analysis. *Wireless Personal Communications., 102*(4), 3699–3713.

Kara, M. E., Fırat, S. Ü. O., & Ghadge, A. (2020). A data mining-based framework for supply chain risk management. *Computers & Industrial Engineering., 139*, 105570–105578.

Khan, R. A., Suleman, T., Farooq, M. S., et al. (2017). Data mining algorithms for classification of diagnostic cancer using genetic optimization algorithms. *IJCSNS, 17*(12), 207–215.

Kim, H., Kim, J., Kim, Y., Kim, I., & Kim, K. J. (2019). Design of network threat detection and classification based on machine learning on cloud computing. *Cluster Computing., 22*(1), 2341–2350.

Lim, K., Lee, B. M., Kang, U., et al. (2018). An optimized DBN-based coronary heart disease risk prediction. *International Journal of Computers Communications & Control, 13*(4), 492–502.

Medvedev, V., Kurasova, O., Bernatavičienė, J., Treigys, P., Marcinkevičius, V., & Dzemyda, G. (2017). A new web-based solution for modelling data mining processes. *Simulation Modelling Practice and Theory., 76*, 34–46.

Moradi, S., & Mokhatab Rafiei, F. (2019). A dynamic credit risk assessment model with data mining techniques evidence from Iranian banks. *Financial Innovation., 5*(1), 15–42.

Mrówczyńska, M., Sztubecki, J., & Greinert, A. (2020). Compression of results of geodetic displacement measurements using the PCA method and neural networks. *Measurement, 8*, 107693–107703.

Pariboni, R., Paternesi Meloni, W., & Tridico, P. (2020). When melius abundare is no longer true: Excessive financialization and inequality as drivers of stagnation. *Review of Political Economy, 32*(2), 216–242.

Pejić Bach, M., Krstić, Ž, Seljan, S., & Turulja, L. (2019). Text mining for big data analysis in financial sector. *A Literature Review. Sustainability., 11*(5), 1277–1283.

Pernell, K. (2020). Market governance, financial innovation, and financial instability: Lessons from banks' adoption of shareholder value management. *Theory and Society, 7*, 30–35.

Sadiq, M. T., Yu, X., & Yuan, Z. (2020). Exploiting dimensionality reduction and neural network techniques for the development of expert brain-computer interfaces. *Expert Systems with Applications, 9*, 114031–114036.

Saura, J. R., Herráez, B. R., & Reyes-Menendez, A. (2019). Comparing a traditional approach for financial brand communication analysis with a Big Data analytics technique. *IEEE Access., 7*, 37100–37108.

Senthil, S., & Ayshwarya, B. (2018). Lung cancer prediction using feed forward back propagation neural networks with optimal features. *International Journal of Applied Engineering Research, 13*(1), 318–325.

Shen, C.-W., Min, C., & Wang, C.-C. (2019). Analyzing the trend of O2O commerce by bilingual text mining on social media. *Computers in Human Behavior, 101*, 474–483.

Sornam, M., & Devi, M. P. (2016). A Survey on Back Propagation Neural Network. *International Journal of Communication and Networking System, 5*(1), 70–74.

Sun, T., & Vasarhelyi, M. A. (2018). Predicting credit card delinquencies: An application of deep neural networks. *Intelligent Systems in Accounting, Finance and Management., 25*(4), 174–189.

Tang, S. Z., Li, M. J., Wang, F. L., et al. (2020). Fouling potential prediction and multi-objective optimization of a flue gas heat exchanger using neural networks and genetic algorithms. *International Journal of Heat and Mass Transfer, 152*, 119488–119493.

Tavana, M., Abtahi, A.-R., Di Caprio, D., & Poortarigh, M. (2018). An artificial neural network and Bayesian Network model for liquidity risk assessment in banking. *Neurocomputing, 275*, 2525–2554.

Thackham, M., & Ma, J. (2020). On maximum likelihood estimation of competing risks using the cause-specific semi-parametric Cox model with time-varying covariates–An application to credit risk. *Journal of the Operational Research Society*,. https://doi.org/10.1080/01605682.2020.1800418.

Trelewicz, J. Q. (2017). Big data and big money: The role of data in the financial sector. *IT Professional., 19*(3), 8–10.

Urbinati, A., Bogers, M., Chiesa, V., & Frattini, F. (2019). Creating and capturing value from Big Data: A multiple-case study analysis of provider companies. *Technovation, 84*, 21–36.

Vidhya, K., & Shanmugalakshmi, R. (2020). Deep learning based big medical data analytic model for diabetes complication prediction. *Journal of Ambient Intelligence and Humanized Computing, 6*, 12–19.

Weng, B., Lu, L., Wang, X., Megahed, F. M., & Martinez, W. (2018). Predicting short-term stock prices using ensemble methods and online data sources. *Expert Systems with Applications., 112*, 258–273.

Yan, X., Chen, M., & Chen, M.-Y. (2019). Coupling and coordination development of Australian energy, economy, and ecological environment systems from 2007 to 2016. *Sustainability, 11*, 6568.

Yang, B. (2020). Construction of logistics financial security risk ontology model based on risk association and machine learning. *Safety Science, 123*, 104–123.

Zhu, L., Qiu, D., Ergu, D., Ying, C., & Liu, K. (2019b). A study on predicting loan default based on the random forest algorithm. *Procedia Computer Science., 162*, 503–513.

Zhu, Y., Zhou, L., Xie, C., Wang, G. J., & Nguyen, T. V. (2019a). Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach. *International Journal of Production Economics., 211*, 22–33.