**International Journal of Information Systems and Supply Chain Management**
Volume 17 • Issue 1

# Application of Artificial Intelligence Data Mining Algorithm in Enterprise Management Risk Assessment

Juntao Zhu, Zhengzhou Railway Vocational and Technical College, China*

## ABSTRACT

For governmental and non-governmental enterprises to tackle risk management with conviction, enterprise management risk assessment (EMRA) is required. This work proposes a research methodology based on an AI-based data mining algorithm (MSVM+EFCNN) for evaluating enterprise-related risks. Initially, all the possible risk assessment indexes of the enterprise are established using a large variety of identification parameters. Then, the data mining algorithms are trained by considering the previous data for building an EMRA model. At last, the current conditions are analyzed using a cluster of risk indicators, and the risk index is identified via the EMRA model. The support vector machine is used for classification purposes, and the fuzzy-based convolutional neural network is enhanced with a genetic algorithm for creating the enterprise risk assessment. The results obtained after keen analysis and experimentation indicate that the data mining algorithms used in this work can evaluate the enterprise-related risks effectively.

## KEYWORDS

Convolutional Neural Networks, EMRA, Genetic Algorithm, Support Vector Machine

## INTRODUCTION

In recent years, more attention has been paid to the role played by artificial intelligence applications in numerous industries. Based on data from the Gravity Recovery and Climate Experiment (GRACE), this study uses LSTM networks to track and predict TWSC and GWSC during the period from 2003 to 2025 for five basins in Saudi Arabia (Haq, Azam, et al., 2021). By analyzing large amounts of remotely sensed Moderate Resolution Imaging Spectroradiometer (MODIS) data, meteorological records, and simulated global climate data, this study provides a comprehensive overview of changes in vegetation, snow cover, and temperature patterns in Uttarakhand State, India. To further investigate the potential for predicting these environmental variables, we performed regression using machine learning (ML) methods, such as support vector machines (SVMs) and long short-term memory (LSTM) networks. In the present study, we used high-resolution PS nanosatellite data to evaluate agricultural activity in the Al-Qassim region of Saudi Arabia (KSA) across time. To assess the impact of time on the vegetation pattern, we generated an NDVI time series. Sparse foliage and brilliant exposed soil

DOI: 10.4018/IJISSCM.342119

*Corresponding Author

due to poor soil moisture limit NDVI in the current study area. Therefore, to evaluate the relative computing cost of NDVI and the area of vegetation, we employed an ML-based random forest (RF) classification model (Haq, Baral, et al., 2021). The study's goals include filling in these blanks and providing useful information for classifying and forecasting air pollution.

To classify air pollution, we created five ML models, one of which was new and was given the term synthetic minority oversampling technique with deep neural network (SMOTEDNN). Hyperparameter tuning and effective data pre-processing were used across all five models. We created three statistical autoregression-based models for air pollution forecasting. The accuracy of all the models developed in this study improved overall. The unique SMOTEDNN model significantly outperformed the other models used in this study and earlier research in terms of accuracy (99.90%). The regionally distributed likelihood of permafrost occurrence might be reliably estimated using the results of logistic regression models. However, the results from these models varied depending on which topoclimatic and topographic variables were employed as predictors during the model's computation (Haq & Baral, 2019). This was especially true for the location points chosen for rock glaciers' initiation lines. This study focuses on creating a novel automated weed-identifying method for a real-world dataset consisting of 4,400 UAV images and 153,360 discrete weed features using convolutional neural network (CNN) classification. The best parameters for the proposed CNN LVQ model were determined with the help of snapshots. After extensive hyperparameter tuning, the generated CNN LVQ model achieved an overall accuracy of 99.44 percent in weed detection, a substantial improvement over previously reported studies.

We built and fine-tuned the climate deep long short-term memory (CDLSTM) model to make accurate temperature and precipitation predictions for all Himalayan states. To make predictions and evaluate the established CDLSTM model's efficacy, we implemented the Facebook Prophet (FB-Prophet) model. Both models were assessed using a wide range of performance indicators, and the results showed that they were very accurate and had low error rates. We developed the models using GridsearchCV for precise hyperparameter tuning. Overfitting and underfitting were prevented in both DNN models by using early stopping. They were among the best-performing models in terms of accuracy and efficiency after extensive testing on the produced suite of tools. The study's novelty lies in the generated models' ability to fill these gaps using a real-world dataset while maintaining a low false alarm rate. The ANN model estimates align well with ground-penetrating radar measurements of ice thickness for five transverse cross-sections of Chhota Shigri Glacier.

## LITERATURE REVIEW

Global economic expansion has increased business participation in the financial market, encouraging long-term economic growth. Many enterprises' capital chains have collapsed due to a lack of risk awareness, financial market volatility, commercial fraud, and poor management (Luo et al., 2023). Small businesses are at risk of bankruptcy due to a lack of risk tolerance and a focus on virtual financial markets rather than the real economy (Haq, 2022a). A lack of understanding of data and a comprehensive evaluation of the borrower limits the accuracy of business risk control models. As a result of extensive academic research, big data tools are becoming increasingly popular in the financial industry. Traditionally, risk control models have only a single dimension and a limited ability to assess risk (Yi, 2023). Data mining can assist in conducting a more thorough analysis of financial data (Zhao & Sun, 2021). Increased data storage and calculation occur as distributed databases and platform architectures mature, increasing the number of data analysis dimensions (Chen et al., 2022). With the advent of the Hadoop cloud computing platform, large-scale data processing has become possible (Haq, 2022b). Compared to the traditional risk management model, big data provides more information and accuracy than the latter. Big data can assist businesses in reducing credit risk and improving the credit system, among other things (Haq, 2022c). With the help of data mining, it is possible to maintain a long-term closed-loop management mode while improving the scoring model and the approval process (Pan & Zhang, 2021). As a result, professionals in the industry are becoming increasingly interested in data mining and financial risk assessment techniques (Shen et al., 2021).

Machine learning is a subset of artificial intelligence that allows systems to improve independently without the need for explicit programming (Omidipoor et al., 2020). Using data mining and machine learning, businesses can reduce their risk exposure (Haq, 2022d). This paper proposes an improved hybrid integration algorithm that combines two classical integration algorithms, random space and MultiBoosting, with the assistance of big data machine learning algorithms. In their paper, Guo et al. (2023) proposed a machine-learning technique for financial risk detection, and Shah et al. (2023) developed a variability model for early warning and control. The following sections provide an overview of the five sections of this study. The first section discusses the research objectives and the significance of studying enterprise financial risk assessment in more detail. This work investigates related data mining and machine learning research to clarify current financial risk control issues (Wang et al., 2021).

Using machine learning to assess risk has become a popular research topic, as has artificial intelligence, and big data has become more prevalent. Institutions' long-term stability requires effective enterprise risk management (Regin et al., 2023). Traditional methods for determining default users are no longer adequate due to modern data types, large user populations, and high-risk prediction accuracy requirements. Numerous academics use machine learning (Owonifari et al., 2023). Overall, the method is accurate in terms of prediction and generalization, as discussed in detail in the research findings. Initially, scientists used statistical learning to assess risk (Hu, 2023). Credit risk analysis paved the way for techniques like regression. A mathematical statistics-based model was created to investigate the credit risk evaluation problem (Zheng et al., 2019). These techniques, however, have some drawbacks. The final classification effect is weak because the sample classification is based on variance rather than mean. Lenders' credit ratings were generated using an algorithm based on their credit history and current situation. It uses existing credit data to predict default risk for users who do not yet have a credit history (Thayyib et al., 2023). However, linear regression has drawbacks. The possible value range is plus or minus infinity. Recently, logistic regression was used to solve this problem. Sharma & Mukhopadhyay(2023) invented logistic regression to score creditworthiness (Sharma & Mukhopadhyay, 2023). This problem is solved by converting the linear regression value to a probability and setting an empirical threshold between zero and one. The sigmoid function is used to achieve this.

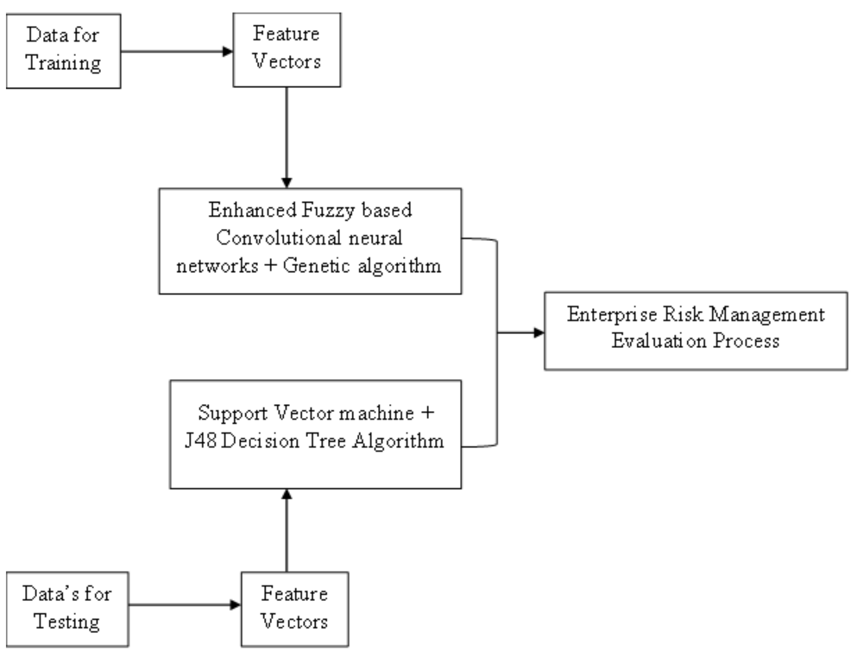
Recent developments in automated risk assessment models have surpassed traditional methods in almost all cases. Machine learning techniques like BP neural networks, KNN, and SVM are prevalent (Chen & Zhang, 2021). KNN outperformed the other two algorithms on two-class classification problems. User credit data was used to create scoring using an artificial neural network model. Studies have shown that ensemble models, like the RF, can accurately assess risk (Zeng et al., 2021). Using supervised learning, researchers claim that historical risk data is represented. The model predicts user behavior and characteristics (Kunduru, 2023). This study employs machine-learning algorithms to assess enterprise risk. A company's risk is assessed using three representative machine algorithms: RF, SVM, and AdaBoost (representative machine algorithms). Three machine learning algorithms train and test an evaluation model based on historical corporate data. The company's current state and risk assessment results are evaluated (Lin et al., 2021). The experiment uses a real-world dataset to evaluate and compare three machine learning algorithms.

Big data analysis has been studied since the 1980s. Analytics includes data mining and machine learning. Data mining involves looking for patterns in large databases, data marts, or data warehouses. Both data marts and data warehouses manage business data (Alotaibi, 2023). This information is vital to a specific organization, even though all departments can access it. Each data mart contains information about the entire company. This document can describe a company's daily operations. A data mart can generate transaction reports and conduct statistical analysis for specific departments or business operations (Dahooie et al., 2021). In a paper on database and data warehouse technology, Wang and Yu (2023) cover pattern recognition, high-performance computing (HPC), machine learning (such as neural networks and fuzzy systems), and spatial or temporal data analysis. The new data management technologies of data mining and knowledge discovery from databases (KDD) have emerged in recent decades (Prokofieva, 2023). In KDD, large datasets are mined for knowledge. As a result, data mining is useful in the finance industry. Large data sets must be searched for hidden patterns by banks. Client financial and behavioral data may be collected before and after the credit is granted (Huang et al., 2021). Most banks and financial institutions go above and beyond regarding customer service, such as providing information for opening an enterprise savings account. The schedule of credit terms covers mortgages, auto loans, insurance, and stock investments. Using data mining and machine learning to forecast future financial events, such as bankruptcies, customer credit ratings, future trades, and financial risk assessments, offers an alternative to stock market forecasting (Fares et al., 2023). Financial institutions can now use artificial intelligence in fund management and asset allocation (Milana & Ashta, 2021). Machine learning algorithms isolate and analyze data in large databases. This tool can spot trends and predict outcomes. However, many banking studies have used machine learning to predict future events and help make decisions. Financial institutions and banks increasingly use data mining and artificial intelligence techniques to compete effectively in today's market.

## PROPOSED METHOD

The enterprise risk management model is developed based on a deep learning algorithm. Methods like K nearest neighbors (KNN) and SVM have already been used. Decision tree models and random forest algorithms are also used in risk assessment. Here, the modified SVM is combined with an enhanced fuzzy-based convolutional neural network (EFCNN) to train the enterprise risk assessment models. The proposed method for managing the enterprise risk assessment is shown in Figure 1.

**Figure 1. Proposed architecture for enterprise risk management assessment**



## MODIFIED SUPPORT VECTOR MACHINE (M-SVM) FOR ENTERPRISE MANAGEMENT RISK ASSESSMENT

The SVM combined with the J48 decision tree classifier is used to determine the decision boundary locations and works on the principle of statistical models. The main task involves mapping the collected data to the high-dimensional space feature. The transformation of linear regression is enhanced with the availability of non-linear functions with estimation problems. The taken samples for training are symbolized as $(\alpha_i, \beta_i)$, where $i = 1, 2, \ldots M$, $\alpha_i \in T^n$ is considered as the input vector, $\beta_i \in T$ is marked as the equivalent output value, and M is the total count of the training samples taken. Then, the linear model corresponding to the space with high dimensions is denoted as:

$$\delta(\alpha, \omega) = \sum_{j=1}^{k} \omega_j \Phi(\alpha)_j + \gamma \tag{1}$$

where $\alpha$ is the input vector, $\omega$ is the coefficient vector of feature space, $\Phi(\alpha)_j$, where j = 1, 2,….k, is considered the non-linear transfer function, and ω is the coefficient of the correlation corresponding to the $\Phi(\alpha)_j$ feature model. Here, γ is the high dimensional space's derivative term. The identified risk function is mentioned as:

$$T(\omega) = \frac{1}{4}\omega^3 + \delta H_2 (Q \sum_{i=1}^{M} D_e(\beta_i, \delta(\alpha_i, \omega)) \tag{2}$$

where $\omega$ is the Euclidean distance, $Q$ represents the penalty coefficient, and $D_e(\beta_i, \delta(\alpha_i, \omega))$ is the loss function, where the value of $\beta$ can range from 1, 2,….M. The output value is mentioned as $\beta$. $\delta(\alpha_i, \omega)$ is the obtained output value of the high dimensional space. The loss function corresponding to the coefficient is given as:

$$D_e(\beta_i, \delta(\alpha_i, \omega)) = \begin{cases} 0, & \left|\delta(\alpha, \omega) - \beta\right| < e \\ \left|\delta(\alpha, \omega) - \beta\right| - e, & \left|\delta(\alpha, \omega) - \beta\right| > e \end{cases} \tag{3}$$

For minimizing the enterprise risk function $T(\omega)$, the obtained equation can be written as:

$$\delta(\alpha) = \sum_{i=1}^{k} (x_i - x_i^*) L(\alpha_i, \alpha) + \gamma \tag{4}$$

Here, the values of $x_i$ and $x_i^*$ can be formulated using the optimization algorithm, and this value corresponds to the eigenvalues with nonlinear transformation. The calculation done here is based on a statistical model, and the J48 decision tree classifier used here, in addition to the normal SVM, provides better results in the classification. It also identifies the features that best classify the base materials for the given training data set. The basic idea behind the use of M-SVM is the minimization of the squares of the constraints, which is considered from the risk functions mentioned in the enterprise's management as shown below:

$$\delta(\alpha) = \sum_{i=1}^{k} (x_i - x_i^*) L(\alpha_i, \alpha) + \gamma \sum_{i=1}^{L} (X_0 + (\chi_i \delta_j - \beta_i) \tag{5}$$

The algorithm is enhanced by using a gradient boosting algorithm, especially for solving problems due to regression. The decision tree using the J48 algorithm can be defined using some basic rules. Initially, the inequality conditions are measured by setting the value of *T*, and the predicted value is:

$$x = \frac{2}{3}\prod\left[\frac{E+G}{E-G}\right] \tag{6}$$

The preconditions behind the prediction values that are taken by classifying the functions using the SVM+DT combination are given by:

$$\delta(\alpha) = 1 + \frac{\sum_{i=1}^{k}(x_i - x_i^*)L(\alpha_i,\alpha) + \gamma\sum_{i=1}^{L}(X_0 + (\chi_i\delta_j - \beta_i)}{\sum_{i=1}^{k}(x_i - x_i^*)L(\alpha_i,\alpha) - \gamma\sum_{i=1}^{L}(X_0 + (\chi_i\delta_j - \beta_i)} \tag{7}$$

The weight distribution from the data is used to train the data points that produce the hypothesis pertaining to the classification-related process. The classification error calculation is performed by using the formula:

$$\eta_t : \zeta_t = \sum_{i:\eta_t(\alpha_i)\neq\beta_i}(\xi_v + \lambda_r) + \sum_{i=1}^{L}\mu_r{}^m \tag{8}$$

The weights correspond to the above equation, which has a weak hypothesis that is generated according to the error rate that occurred due to classification and the data training points, which is updated by considering the weights shown below as:

$$\lambda_{t+1}(i) = \frac{\lambda_t(i)}{S_t} \times \begin{cases}\gamma_t\chi_t(\alpha_i) = \beta_i \\ 0 \quad Others\end{cases} + \delta\sum_{i=1}^{L}\mu_r{}^m \tag{9}$$

The weak hypothesis is determined using the corresponding weights incorporating the loss functions along with the weak hypothesis that is joined along with the prediction function, calculated using the formula:

$$\lambda_{loss}(\alpha) = \frac{\arg\max}{\beta\in\gamma}\sum_{i=1}^{L}\ln\left(\frac{\kappa}{\chi_t}\right)\left[\lambda_t(\alpha)\right] + \frac{\arg\min}{\beta\in\gamma}\sum_{i=1}^{L}\ln\left(\frac{Z}{\chi_t}\right)\left[\lambda_t(\alpha)\right] \tag{10}$$

The already existing and pre-recorded data are taken for the benchmarking process, and the feature vector is created using the modified SVM method using the decision tree algorithm. The evaluation model is generated for testing the data, acquiring the data, and constructing the feature vector model. The model for evaluating the training section is given as input to get the value of recent risk evaluation results for the enterprise.

### Enhanced Fuzzy-Based Convolutional Neural Networks (EFCNN) for Enterprise Management Risk Assessment

Fuzzy-based convolutional neural networks (FCNN) are enhanced with a genetic algorithm (GA). Typically, the FCNN contains two sections: the extraction of features where the neuron inputs are connected to the surface of the previous layer, and the total vector of the feature is obtained by extracting the relevant features of individual blocks. Next, the layer is used to map the feature maps. To ensure the feature for the correct displacement, the activation function is added to the mapping structure of the feature maps of the convolutional networks. The weights for sharing the mapping layer are used to reduce the parameters of the risk model.

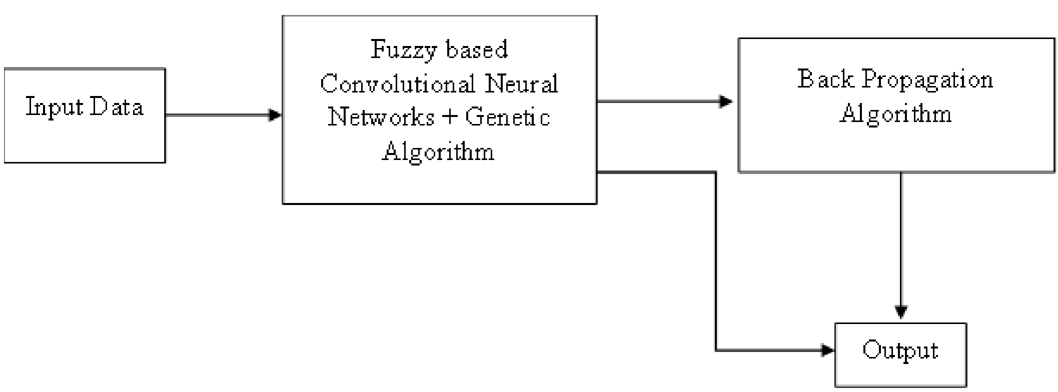
The FCNN is used to extract the features of the input data. Then, the risk assessment-based information regarding the fuzzy inference system is created. Finally, the FCNN model is used to train the information about the internal and external processes. The adjustment is made to achieve the risk assessment model by maintaining optimal performance. The enhancement, made with the help of a genetic algorithm, uses principal component analysis to deal with the information gathered from the enterprise, and the sequence of information is formulated using the process of dimensionality reduction. The genetic algorithm is used here to combine the genetic method with the human perception of handling the information. As humans are involved in the decision-making process, the genetic algorithm is used to combine the complex world with the biological world. Non-numerical problems can be easily solved using the genetic algorithm, which can better handle parallel problems. The algorithm flexibility can be compromised by using the information inputs and thus can obtain scalability. Using genetic algorithms in combination with FCNNs enhances the ability to make predictions to a certain extent. The backpropagation algorithm is used to improve the prediction results. The overall process involved in this EFCNN is shown in Figure 2.

Let us consider the dimension of the input vector as “k”, “τ” represents the $i^{th}$ value of the matrix, $\varepsilon i$ is the convolution kernel, and $\lambda_{\upsilon}(\alpha)$ is the corresponding value of the output whose value is given as:

$$\lambda_{\upsilon}(\alpha) = \Gamma\left[\sum_{\tau=1}^{k}(\omega_{\tau}\alpha_{\tau} + Y)\right] \tag{11}$$

Based on the data used for training and testing, the feature vector is developed using the methods mentioned above, the SVM combined with the decision tree algorithm is used for classification purposes, and the feature formulation module is performed using FCNNs in combination with the

**Figure 2. Proposed architecture EFCNN**



genetic algorithm. The statistical analysis is performed effectively by using this method. In the test phase, the data is acquired, and the feature vector for that value is again remolded, thus obtaining the enterprise's risk evaluation-based output.

## EXPERIMENTS AND ANALYSIS

The procedures used for the validation come from the risk assessment methods for the financial statements considered by banking sectors that serve customers by issuing credit cards and loans. The data taken for the research include one set for the training section (loan repayment, credit card dues) and another set for the testing section.

### Indicators Used for the Evaluation

The risk analysis based on statistical or financial indicators meets the demands of today's market. Hence, the data are taken from financial regulators and their alternatives. The debt being paid is indicated along with the status to create a warning process for the company. This analysis is made by considering a large number of enterprises. An enterprise's ability to pay fluctuates in abnormal situations. Here, indicators like current account, cash flow, interest rate, asset value, and seller response are considered for the analysis. The size of the market reflects the enterprise's profitability. This is based on the indicators of the selected probability. The ability to be concerned with the operation typically affects the enterprise's management levels.

The impact on operational ability might guarantee that the enterprise can maintain its operations reasonably. Each indicator taken so far might correspond to the development of the enterprise's expertise level and the ability to retain better prospects in the market by considering the market value. The enterprise might have a strong development ability to handle various risks to increase profitability. Hence, the rate of the main business will increase to obtain better revenue with a reasonable profit rate.

### Datasets Considered

The datasets used for training purposes are taken from the enterprise's loan section, and the total process is divided into two categories: companies based on performance (+1) and regular companies (–1) based on financial status. Initially, all the data taken for training is processed despite its efficiency and sturdiness. The data are considered large numbers, as the number of categories is taken as input, and the different processes are considered more effective. The deviation test is performed for more than five iterations to eliminate the abnormal data. The total samples considered here are from more than 1,120 loan holders from 345 branches associated with enterprise-related research. The various parameters are considered within these limitations based on the sectors considered. The credit card repayment procedures are one of the indicators, and cardholders who cannot pay their dues on time are considered another parameter.

To analyze the performance of the proposed method, the convergence curve's accuracy and region are taken as performance evaluation indicators. Based on the accuracy of the information, the selected index is simple, and the evaluation samples used for the evaluation are based on the total samples considered. The area of convergence can be measured by considering the probability of occurrence, its performance, and the ranking process using the data mining algorithms. Hence, it can be used in the data mining algorithms by considering the false-positive rates as the basis of the horizontal axis, taking into consideration the class rate, and the vertical axis, which is taken by adjusting the threshold values according to the training samples. The points connected in accordance with the region of the convergence curve can be connected directly by using the index. It is evaluated based on a classifier. In this work, the n-fold cross-validation is taken by considering n = 20 as it has a relatively low variance and means. Here, more than 1,200 risk data points for corporate sectors are considered and divided into 20 equal sections. It is represented as $(R_1, R_2, R_3, \ldots R_{20})$ by considering $\{R\}$ as the dataset used for testing, and the other data are considered for training sets, thus generating

**Table 1. The performance variation of algorithms used**

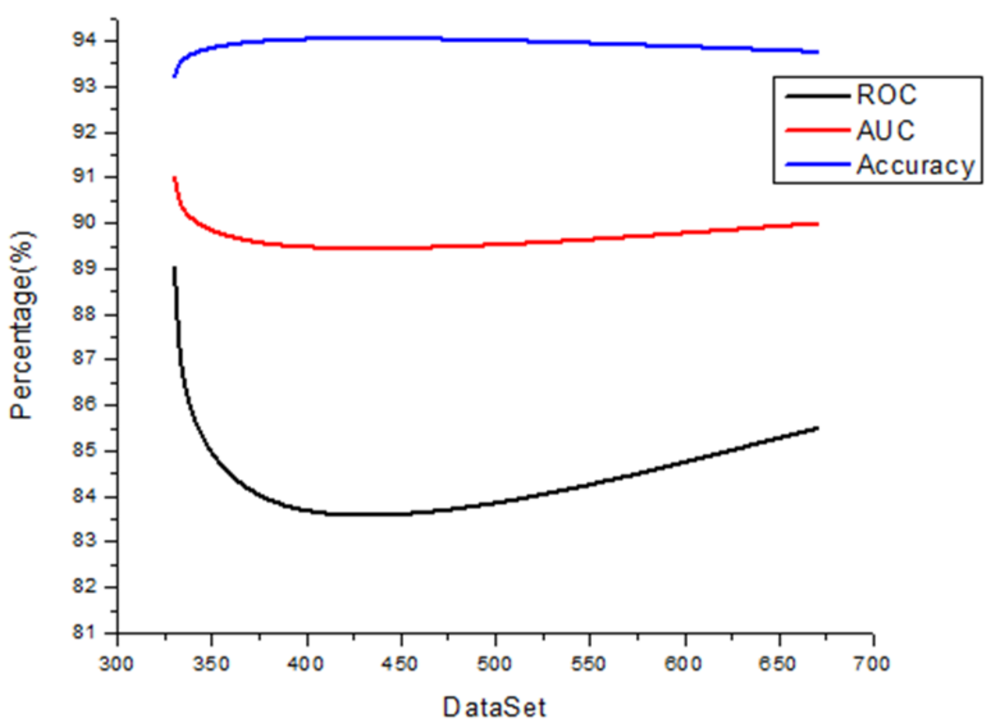
| Algorithm | Dataset | | Accuracy | ROC | AOC |
|---|---|---|---|---|---|
| | Testing | Training | | | |
| MSVM | 330 | 226 | 93.23 | 0.89 | 0.91 |
| EFCNN | 340 | 304 | 94.31 | 0.82 | 0.89 |
| Combined | 670 | 530 | 93.77 | 0.855 | 0.90 |

the group of training sets $\{Test_i, Train_i\}$ where $i = 1, 2, \ldots 20$. The accuracy value average is calculated and is shown in Table 1.

By combining the evaluation and classification methods, the MSVMs and the J48 decision tree algorithm for classification purposes are selected for accuracy, region of convergence, and area of convergence value. The combination of MSVM and EFCNN algorithms provides better results in evaluating indications in enterprise risk assessment. The performance indication of the combined algorithms is illustrated in Figure 3.

The genetic algorithm uses the principal component analysis to process the collected data from the enterprise sectors. The training samples are reconstructed using the information content to handle the wide range of applications. It is also used to optimize the obtained features. The primitive variables are selected according to the horizontal coordinates, and the numerical values corresponding to the Eigenvalues are accumulated using the coefficient matrix. From the collected primitive samples,

**Figure 3. Performance indicators for various parameters**



the contribution rate in the form of percentage is computed, and the accumulation rate is calculated. This is illustrated in Table 2.

The value of *i* can be from (1, 2,…n), but we only consider sample instances here. The highest score of any instance is 100, and the various scores in Table 2 are based on the weights provided. After getting the primitive samples, they must be preprocessed. After preprocessing, the parameters for the genetic algorithm are computed by calculating the eigenvalues and the corresponding side information in the form of a percentage. The results obtained after the conversion using the principal component analysis are shown in Table 3.

The changes in the values taken for training purposes are standardized. The computation uses the genetic algorithm to extract the needed features for enterprise risk management, so the information is extracted by considering the security constraints. The eigenvalues are taken in association with the number of samples and are shown in Figure 4.

Figure 5 shows the representation of the ordinates considering the involvement rate and the growth that occurs due to the involvement. The training values taken from the primitive aspects are combined with the eigenvector to analyze the security risks in the financial management sector.

**Table 2. Collected primitive samples for training and testing purposes**

| Value of *i* | $\propto_1$ | $\propto_2$ | $\propto_3$ | $\propto_4$ | $\propto_5$ | $\propto_6$ | $\propto_7$ | $\propto_8$ | $\propto_9$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 95 | 97 | 91 | 93 | 87 | 89 | 83 | 85 | 79 |
| 2 | 96 | 98 | 92 | 94 | 88 | 90 | 84 | 86 | 80 |
| 3 | 94 | 96 | 90 | 92 | 86 | 88 | 82 | 84 | 78 |
| 4 | 85 | 87 | 81 | 83 | 77 | 79 | 73 | 75 | 69 |
| 5 | 98 | 100 | 94 | 96 | 90 | 92 | 86 | 88 | 82 |
| 6 | 92 | 94 | 88 | 90 | 84 | 86 | 80 | 82 | 76 |
| 7 | 95 | 97 | 91 | 93 | 87 | 89 | 83 | 85 | 79 |
| 8 | 96 | 98 | 92 | 94 | 88 | 90 | 84 | 86 | 80 |

**Table 3. Eigen values vs. involvement rate vs. growth rate**

| Number of samples | Change in values | Eigen values | Rate of involvement in % | Growth rate in % | Security constraints | |
|---|---|---|---|---|---|---|
| | | | | | $R_1$ | $R_2$ |
| 1 | $\propto_{1,} \propto_5$ | 1.185 | 27.56 | 55.35 | –1.171 | –0.89 |
| 2 | $\propto_{2,} \propto_4$ | 2.356 | 34.38 | 68.58 | 1.098 | 0.281 |
| 3 | $\propto_{3,} \propto_1$ | 1.258 | 49.15 | 48.38 | –1.867 | –0.817 |
| 4 | $\propto_4$ | 3.125 | 69.35 | 63.71 | 2.138 | 1.05 |
| 5 | $\propto_5$ | 0.987 | 56.12 | 53.22 | –0.502 | –1.088 |
| 6 | $\propto_{6,} \propto_4$ | 1.489 | 32.68 | 66.45 | 0.431 | –0.586 |
| 7 | $\propto_7$ | 1.058 | 39.5 | 46.25 | –0.06 | –1.017 |
| 8 | $\propto_8$ | 1.118 | 54.27 | 56.63 | 0.122 | –0.957 |
| 9 | $\propto_{9,} \propto_4$ | 0.996 | 74.47 | 69.86 | –0.201 | –1.079 |
| 10 | $\propto_{5,} \propto_1$ | 1.197 | 61.24 | 49.66 | 1.197 | –0.878 |

**Figure 4. Eigenvalues based on enterprise information security**

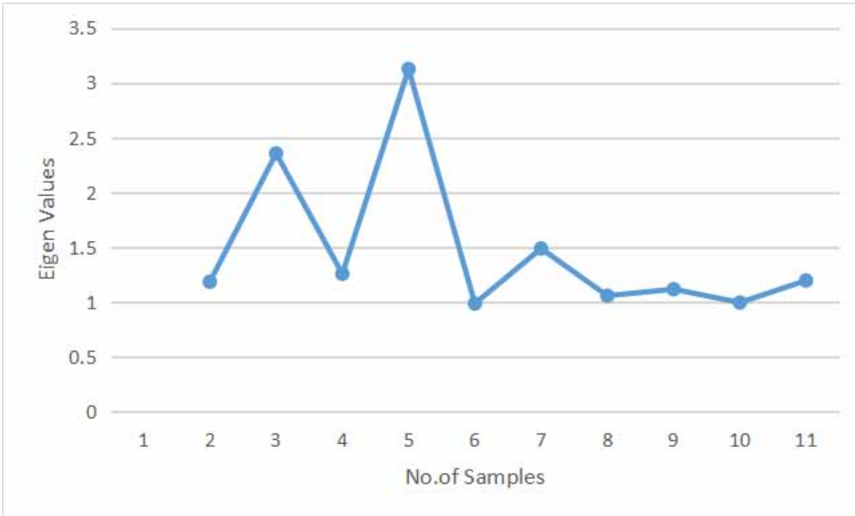
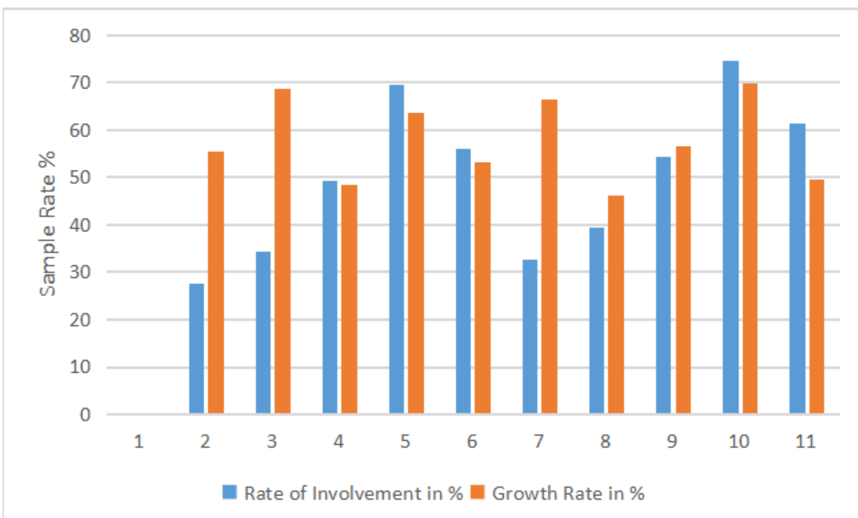



**Figure 5. Involvement vs. growth rate**




The EFCNN was trained with the primitive data sample taken, and by using the MSVM, the risks are classified, and hence the maximum performance is achieved. The banking financial sector data is taken here, and the corresponding risks are considered. In particular, the loan and credit card sections are taken into consideration. Most occurrences of fraud happen in these sections, so they are observed closely. Most enterprise security risks are considered before processing the information, which is at the feature selection stage. Here also, the intermediate processing checks for security-related issues with ultimate care. A total of 1,200 cases are taken for analysis and divided into 10 clusters; each cluster has 120 cases. From the 10 clusters, the valid cases are identified based on this classification. Of these, 752 are fraudulent cases; the remaining are not fraud cases. Table 4 shows the total fraud cases and non-fraud cases observed.

**Table 4. Training dataset performance**

| Number of clusters | Fraudulent % | Intermediate | Non-fraudulent % | Average % |
|---|---|---|---|---|
| 1 | 67.89 | 66.54 | 45.24 | 59.89 |
| 2 | 53.42 | 67.12 | 56.34 | 58.96 |
| 3 | 92.10 | 94.30 | 96.1 | 94.17 |
| 4 | 93.00 | 87.12 | 83.23 | 87.78 |
| 5 | 77.76 | 76.41 | 55.11 | 69.76 |
| 6 | 63.29 | 76.99 | 66.21 | 68.83 |
| 7 | 86.12 | 90.11 | 56.34 | 77.52 |
| 8 | 66.54 | 93.00 | 67.89 | 75.81 |
| 9 | 67.12 | 86.28 | 53.42 | 68.94 |
| 10 | 94.30 | 86.86 | 92.10 | 91.09 |
| 11 | 89.93 | 79.12 | 88.63 | 85.89 |
| 12 | 90.12 | 91.28 | 93.22 | 91.54 |

The enterprise risk model-based performance from the training dataset is used to reduce the biasing phenomenon. Usually, the samples are used for training by learning the data and are not involved in overfitting fraud-related information. Hence, to minimize bias, the estimation of the models is considered for the set of unseen details. Model validation is done using many methods, including a specific method that can divide the training sample by holding the separation of the sample by introducing the 12-fold cross-validation and one-fold hold-out validation. Table 5 shows the performance after doing 12-fold cross-validation.

Here, the algorithms provide specific validation capabilities to create the validation data sets. The sample is divided into 12 folds, each filled with an equal number of fraudulent and non-fraudulent cases. Each fold is trained by using the other folds and is tested by using one hold-out fold, and finally, the average value is computed. Table 5 indicates the benefits of introducing the 12 cross-fold validation. The accuracy of the expected rates is taken to validate the training set, and the models for risk assessment show significantly different results. Figure 6 illustrates the difference between the training and testing sample data after 12-fold cross-validation.

The model accurately classifies the total samples taken. Testing and training are taken as different approaches; thus, about 87% of fraudulent cases and 70% of non-fraudulent cases are identified using this model. The research consumes nearly 100% of the data sets considered, and the results attributed to the whole data sets are only for the entire data sets. The models used here are the MSVM and EFCNN models, and the training accuracy predicted in percentage is shown in Figure 7.

We compared the two proposed methods by considering the number of clusters taken on the x-axis, and the obtained training accuracy proves the favorability of using this method over all other baseline methods. Figure 8 shows the results of comparing the testing phase between the MSVM and EFCNN.

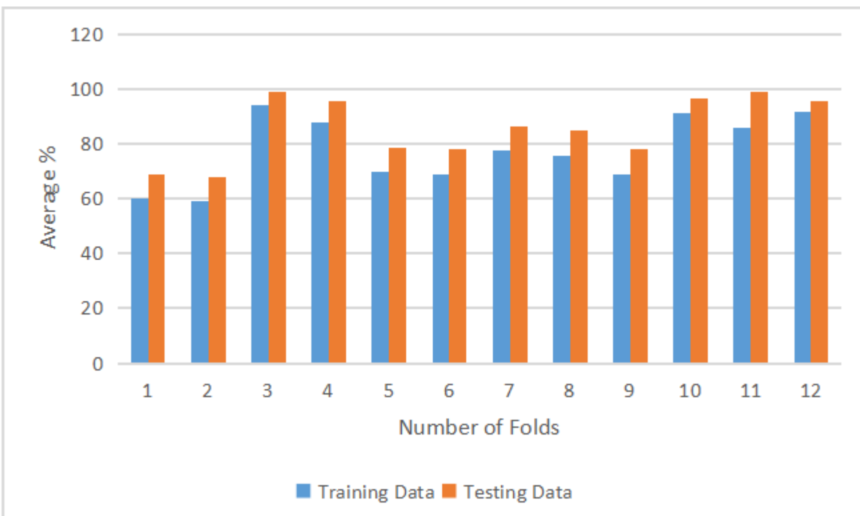
The accuracy of MSVM is somewhat similar to the performance of EFCNN. Hence, combining these two models provides better accuracy and performance in enterprise risk management and assessment. This risk assessment method offers better results than all the other algorithms and techniques used earlier, as depicted in Table 6.

We considered the two evaluation methods used to assess the accuracy of enterprise risk management. The proposed EFCNN and MSVM methods show better performance in terms of precision, f score, and recall value. The ROC and AUC values are also much higher than those of the

Table 5. Testing dataset performance using 12-fold cross-validation

| Number of folds | Fraudulent % | Intermediate | Non-fraudulent % | Average % |
|---|---|---|---|---|
| 1 | 77.01 | 72.21 | 57.47 | 68.90 |
| 2 | 62.54 | 72.79 | 68.57 | 67.97 |
| 3 | 98.22 | 99.97 | 98.33 | 98.84 |
| 4 | 99.12 | 92.79 | 95.46 | 95.79 |
| 5 | 86.88 | 82.08 | 67.34 | 78.77 |
| 6 | 72.41 | 82.66 | 78.44 | 77.84 |
| 7 | 95.24 | 95.78 | 68.57 | 86.53 |
| 8 | 75.66 | 98.67 | 80.12 | 84.82 |
| 9 | 76.24 | 91.95 | 65.65 | 77.95 |
| 10 | 99.42 | 92.53 | 97.33 | 96.43 |
| 11 | 98.22 | 99.97 | 98.33 | 98.84 |
| 12 | 99.12 | 92.79 | 95.46 | 95.79 |

Figure 6. Performance of training sample and testing samples



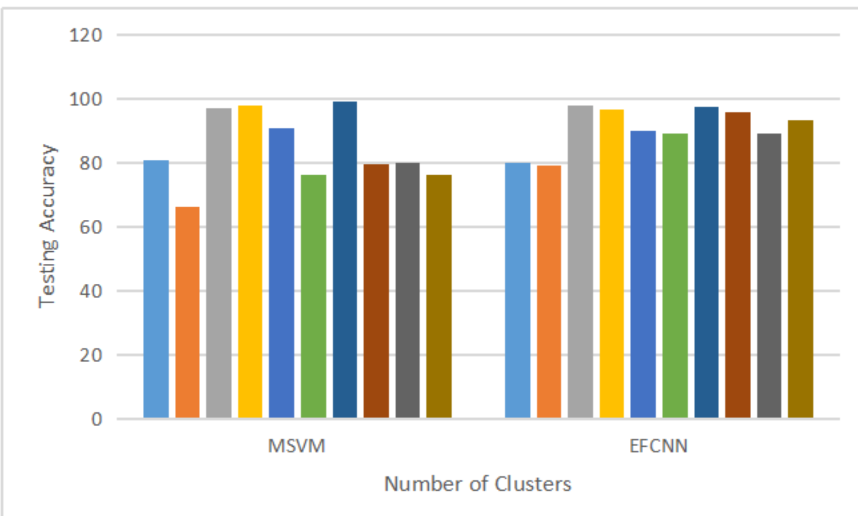

existing methods. The risk assessment also considers security-related aspects. The parameters taken for this research are from the real-time monitored database. Hence, this method is a better enterprise risk assessment and management solution.

## Development Analysis of Data Mining Algorithms in Enterprise Management Risk Assessment

Risk assessment is an essential task in enterprise management. Traditional risk assessment methods often rely on experience and intuition, making it difficult to identify potential risk factors or accurately predict risks. In recent years, with the continuous development of artificial intelligence technology, data mining algorithms have been widely applied in enterprise management, becoming an emerging

Figure 7. Training phase accuracy for MSVM vs EFCNN


Figure 8. Testing phase accuracy for MSVM vs EFCNN




risk assessment method. However, the question of whether any limitations to this algorithm exist in practical applications requires in-depth exploration. The limitations of artificial intelligence data mining algorithms in enterprise management risk assessment mainly include the following aspects:

(1) Data reliability issue: Data is the foundation for data mining, but in practical applications, enterprise management risk assessment involves a large amount of data, which may have problems such as incompleteness, inaccuracy, or inconsistency, which can affect the accuracy and reliability of algorithms.

**Table 6. Comparison of the performance against the baseline methods**

| Algorithms | AUC | ROC | F score | Accuracy | Precision | Recall |
|---|---|---|---|---|---|---|
| RF | 0.88 | 0.67 | 0.78 | 0.80 | 0.89 | 0.80 |
| SVM | 0.57 | 0.87 | 0.72 | 0.91 | 0.83 | 0.75 |
| AdaBoost | 0.98 | 0.64 | 0.81 | 0.98 | 0.92 | 0.84 |
| LSA | 0.87 | 0.85 | 0.86 | 0.87 | 0.97 | 0.89 |
| DTA | 0.68 | 0.99 | 0.83 | 0.76 | 0.94 | 0.86 |
| RFA | 0.78 | 0.76 | 0.77 | 0.87 | 0.88 | 0.80 |
| Naïve Bayes | 0.89 | 0.87 | 0.88 | 0.97 | 0.99 | 0.91 |
| K means clustering | 0.44 | 0.88 | 0.66 | 0.78 | 0.77 | 0.69 |
| Decision tree | 0.85 | 0.76 | 0.80 | 0.86 | 0.92 | 0.83 |
| RNN | 0.89 | 0.78 | 0.84 | 0.85 | 0.95 | 0.86 |
| CNN | 0.55 | 0.56 | 0.55 | 0.84 | 0.67 | 0.58 |
| EFCNN | 0.94 | 0.92 | 0.93 | 0.79 | 0.95 | 0.92 |
| MSVM | 0/98 | 0.98 | 0.98 | 0.87 | 0.97 | 0.98 |

(2) Relying on historical data: Data mining algorithms typically rely on historical data for modeling and prediction. However, enterprise management risk assessment needs to consider future risk situations, and historical data cannot fully represent future situations. Therefore, algorithms may face difficulties dealing with emerging risks or unexpected events.
(3) Model selection and parameter adjustment: When applying data mining algorithms, selecting the appropriate model and adjusting the corresponding parameters is necessary. However, different models and parameter settings may lead to different results, and choosing the most suitable model and parameter settings is a challenge. In addition, the complexity of the model and the demand for computing resources also must be considered in terms of the limitations in practical applications.
(4) Explanatory and credibility issues: Some data mining algorithms are difficult to explain, which may pose challenges for enterprise managers. Managers often need to understand how algorithms make decisions or predictions to understand and accept the results of the algorithms. Therefore, the interpretability and credibility of algorithms are crucial for practical applications.
(5) The influence of human factors: Data mining algorithms can provide auxiliary decision-making information in enterprise management risk assessment, but the final decision still requires the judgment and decision-making of managers. Managers' subjective cognition, experience, and values may impact the final decision, limiting algorithms' effectiveness in practical applications.

In summary, although artificial intelligence data mining algorithms have potential in enterprise management risk assessment, they still face some limitations. To improve the accuracy and reliability of the evaluation, it is necessary to comprehensively consider the actual situation and needs and combine them with other decision-making methods.

Given that the application of artificial intelligence data mining algorithms in enterprise management risk assessment is still in a constantly developing stage, several directions and countermeasures can be explored in the future:

(1) Data quality improvement: To improve the accuracy and reliability of algorithms, data quality management must be strengthened. This includes data cleaning, integration, and annotation to ensure the input data is accurate, complete, and consistent.
(2) Combining multi-source data: In addition to traditional structured data, risk assessment can be conducted by combining unstructured data and external data sources. For example, social media data, news reports, and market trends can provide a more comprehensive perspective on risk analysis, helping businesses better respond to risks.
(3) Introducing deep learning technology: Deep learning technology has achieved significant results in areas such as image recognition and natural language processing and can be further applied in risk assessment. By training and optimizing deep learning models, algorithms' predictive ability and accuracy can be improved.
(4) Explanatory and interpretable algorithms: To enhance the credibility and user acceptance of algorithms, more interpretable and interpretable algorithms can be studied and developed. In this way, managers can better understand the decision-making process of algorithms and adjust and optimize them as needed.
(5) Model fusion and integration: For different types of risks, multiple algorithms can be fused and integrated to consider each algorithm's advantages and characteristics comprehensively. This can improve the comprehensiveness and accuracy of risk assessment.

In summary, the future development direction of artificial intelligence data mining algorithms in enterprise management risk assessment includes improving data quality, introducing multi-source data, applying deep learning technology, developing interpretable algorithms, and model fusion and integration. These development directions can help improve the accuracy, comprehensiveness, and credibility of risk assessment, providing more effective decision support for enterprise managers.

## CONCLUSION

This study tackled enterprise management risk assessment with the help of statistical learning methods using artificial intelligence and data mining algorithms. The inputs were taken from the finance banking sector, with 1,200 information samples from the loan and credit card sectors. This method clearly predicted and indicated the risk status in all aspects. The model comprises an MSVM for real-time prediction of data and accurate classification by estimating the feature values exactly using the J48 decision tree algorithm and the EFCNN for predicting risks accurately. Individually, both models have valid advantages, and their combination provides improved performance compared to all other baseline methods and thus provides the best solution for enterprise management risk assessment. For future research, we highly recommend using intelligent methods to evaluate the security measures to avoid enterprise management risk. These security measures will help reduce the high risk of threats in enterprise management. Future research can also consider expanding the sample size and obtaining more samples from different industries and fields to improve the generalization ability and applicability of the model. At the same time, quality management and data validation should be strengthened to ensure the accuracy and credibility of input data. Although this study has achieved positive results, it still has limitations and directions that require further research. Future research can focus on expanding sample size and sources, processing data quality and integrity, improving model interpretability and interpretability, and validating cross-domain applications to provide more comprehensive and reliable solutions for enterprise management risk assessment.

## DATA AVAILABILITY STATEMENT

The labeled data sets used to support the findings of this study are available from the corresponding author upon request.

## AUTHORS CONTRIBUTION

Lan Huang contributed to the study's conception and design and wrote the first draft of the manuscript. The author contributed to the manuscript revision and read and approved the submitted version.

## COMPETING INTERESTS

The authors declare that there is no conflict of interest.

## FUNDING STATEMENT

This work was not supported by any funds.

## ACKNOWLEDGMENT

The authors would like to sincerely thank those whose techniques have contributed to this research.

## REFERENCES

Alotaibi, E. M. (2023). Risk assessment using predictive analytics. *International Journal of Professional Business Review*, *8*(5), e01723–e01723. doi:10.26668/businessreview/2023.v8i5.1723

Chen, F. H., Hsu, M. F., & Hu, K. H. (2022). Enterprise's internal control for knowledge discovery in a big data environment by an integrated hybrid model. *Information Technology and Management*, *23*(3), 213–231. doi:10.1007/s10799-021-00342-8

Chen, Y., & Zhang, R. (2021). Research on credit card default prediction based on k-means SMOTE and BP neural network. *Complexity*, *2021*, 1–13. doi:10.1155/2021/2574025

Dahooie, J. H., Hajiagha, S. H. R., Farazmehr, S., Zavadskas, E. K., & Antucheviciene, J. (2021). A novel dynamic credit risk evaluation method using data envelopment analysis with common weights and combination of multi-attribute decision-making methods. *Computers & Operations Research*, *129*, 105223. doi:10.1016/j.cor.2021.105223

Fares, O. H., Butt, I., & Lee, S. H. M. (2023). Utilization of artificial intelligence in the banking sector: A systematic literature review. *Journal of Financial Services Marketing*, *28*(4), 835–852. doi:10.1057/s41264-022-00176-7

Guo, Y., Zhang, W., Qin, Q., Chen, K., & Wei, Y. (2023). Intelligent manufacturing management system based on data mining in artificial intelligence energy-saving resources. *Soft Computing*, *27*(7), 4061–4076. doi:10.1007/s00500-021-06593-5

Haq, M. A. (2022a). CDLSTM: A novel model for climate change forecasting. *Computers, Materials & Continua*, *71*(2).

Haq, M. A. (2022b). CNN based automated weed detection system using UAV imagery. *Computer Systems Science and Engineering*, *42*(2).

Haq, M. A. (2022c). Planetscope nanosatellites image classification using machine learning. *Computer Systems Science and Engineering*, *42*(3).

Haq, M. A. (2022d). SMOTEDNN: A novel model for air pollution forecasting and AQI classification. *Computers, Materials & Continua*, *71*(1).

Haq, M. A., Azam, M. F., & Vincent, C. (2021). Efficiency of artificial neural networks for glacier ice-thickness estimation: A case study in western Himalaya, India. *Journal of Glaciology*, *67*(264), 671–684. doi:10.1017/jog.2021.19

Haq, M. A., & Baral, P. (2019). Study of permafrost distribution in Sikkim Himalayas using Sentinel-2 satellite images and logistic regression modelling. *Geomorphology*, *333*, 123–136. doi:10.1016/j.geomorph.2019.02.024

Haq, M. A., Baral, P., Yaragal, S., & Pradhan, B. (2021). Bulk processing of multi-temporal modis data, statistical analyses and machine learning algorithms to understand climate variables in the Indian Himalayan region. *Sensors (Basel)*, *21*(21), 7416. doi:10.3390/s21217416 PMID:34770722

Haq, M. A., Jilani, A. K., & Prabu, P. (2021). Deep learning based modeling of groundwater storage change. *CMC-Computers. Materials & Continua*, *70*, 4599–4617.

Haq, M. A., & Rahim Khan, M. A. (2022). DNNBoT: Deep neural network-based botnet detection and classification. *Computers, Materials & Continua*, *71*(1).

Hu, X. (2023). The evaluation method of college teachers' morality considering intelligent emotion recognition and data mining algorithm. *Applied Artificial Intelligence*, *37*(1), 2186413. doi:10.1080/08839514.2023.2186413

Huang, J. Y., Gupta, A., & Youn, M. (2021). Survey of EU ethical guidelines for commercial AI: Case studies in financial services. *AI and Ethics*, *1*(4), 569–577. doi:10.1007/s43681-021-00048-1

Kunduru, A. R. (2023). Effective usage of artificial intelligence in enterprise resource planning applications. *International Journal of Computer Trends and Technology*, *71*(4), 73–80. doi:10.14445/22312803/IJCTT-V71I4P109

Lin, S. S., Shen, S. L., Zhou, A., & Xu, Y. S. (2021). Risk assessment and management of excavation system based on fuzzy set theory and machine learning methods. *Automation in Construction*, *122*, 103490. doi:10.1016/j.autcon.2020.103490

Luo, N., Yu, H., You, Z., Li, Y., Zhou, T., Jiao, Y., Han, N., Liu, C., Jiang, Z., & Qiao, S. (2023). Fuzzy logic and neural network-based risk assessment model for import and export enterprises: A review. *Journal of Data Science and Intelligent Systems*, *1*(1), 2–11. doi:10.47852/bonviewJDSIS32021078

Milana, C., & Ashta, A. (2021). Artificial intelligence techniques in finance and financial markets: A survey of the literature. *Strategic Change*, *30*(3), 189–209. doi:10.1002/jsc.2403

Moscato, V., Picariello, A., & Sperlí, G. (2021). A benchmark of machine learning approaches for credit score prediction. *Expert Systems with Applications*, *165*, 113986. doi:10.1016/j.eswa.2020.113986

Omidipoor, M., Toomanian, A., Neysani Samany, N., & Mansourian, A. (2020). Knowledge discovery web service for spatial data infrastructures. *ISPRS International Journal of Geo-Information*, *10*(1), 12. doi:10.3390/ijgi10010012

Owonifari, V. O., Igbekoyi, O. E., Awotomilusi, N. S., & Dagunduro, M. E. (2023). Evaluation of artificial intelligence and efficacy of audit practice in Nigeria. *Asian Journal of Economics*. *Business and Accounting*, *23*(16), 1–14.

Pan, Y., & Zhang, L. (2021). A BIM-data mining integrated digital twin framework for advanced project management. *Automation in Construction*, *124*, 103564. doi:10.1016/j.autcon.2021.103564

Prokofieva, M. (2023). Integrating data analytics in teaching audit with machine learning and artificial intelligence. *Education and Information Technologies*, *28*(6), 7317–7353. doi:10.1007/s10639-022-11474-x

Regin, R., Rajest, S. S., & Shynu, T. (2023). A review of secure neural networks and big data mining applications in financial risk assessment. *Central Asian Journal of Innovations on Tourism Management and Finance*, *4*(2), 73–90.

Shah, H. M., Gardas, B. B., Narwane, V. S., & Mehta, H. S. (2023). The contemporary state of big data analytics and artificial intelligence towards intelligent supply chain risk management: A comprehensive review. *Kybernetes*, *52*(5), 1643–1697. doi:10.1108/K-05-2021-0423

Sharma, K., & Mukhopadhyay, A. (2023). Cyber-risk management framework for online gaming firms: An artificial neural network approach. *Information Systems Frontiers*, *25*(5), 1757–1778. doi:10.1007/s10796-021-10232-7

Shen, F., Zhao, X., Kou, G., & Alsaadi, F. E. (2021). A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique. *Applied Soft Computing*, *98*, 106852. doi:10.1016/j.asoc.2020.106852

Thayyib, P. V., Mamilla, R., Khan, M., Fatima, H., Asim, M., Anwar, I., Shamsudheen, M. K., & Khan, M. A. (2023). State-of-the-art of artificial intelligence and big data analytics reviews in five different domains: A bibliometric summary. *Sustainability (Basel)*, *15*(5), 4026. doi:10.3390/su15054026

Wang, D., & Yu, A. (2023). Supply chain resources and economic security based on artificial intelligence and blockchain multi-channel technology. [IJITSA]. *International Journal of Information Technologies and Systems Approach*, *16*(3), 1–15. doi:10.4018/IJITSA.326552

Wang, Y., Xiuping, S., & Zhang, Q. (2021). Can fintech improve the efficiency of commercial banks?—An analysis based on big data. *Research in International Business and Finance*, *55*, 101338. doi:10.1016/j.ribaf.2020.101338

Yi, X. (2023). Application of data mining in enterprise financial risk prediction based on genetic algorithm and linear adaptive optimization. *Soft Computing*, *27*(14), 10305–10315. doi:10.1007/s00500-023-08308-4

Zeng, Z., Luo, W. G., Wang, Z., & Yi, F. C. (2021). Water pollution and its causes in the Tuojiang River Basin, China: An artificial neural network analysis. *Sustainability (Basel)*, *13*(2), 792. doi:10.3390/su13020792

Zhao, J., & Sun, L. (2021). Research on financial control of enterprise group based on artificial intelligence and big data. *Annals of Operations Research*, 1–16.

Zheng, M., Tang, W., & Zhao, X. (2019). Hyperparameter optimization of neural network-driven spatial models accelerated using cyber-enabled high-performance computing. *International Journal of Geographical Information Science*, *33*(2), 314–345. doi:10.1080/13658816.2018.1530355

*Juntao Zhu was born in Ji Yuan, Henan, PR Chian, in 1981. She received the master's degree from Henan Normal University, College of Computer and Information Engineering, PR China. Now, he works in School of Zhengzhou Railway Vocational and Technical College. His main research area is Computer artificial intelligence.*