SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V10I11P105

Volume 10 Issue 11, 45-54, November 2023

Original Article

Enhancing Credit Score Analysis: A Novel Approach with Random Forest and Kernel SVM

Annie Chacko[1], D. John Aravindhar[2], A. Antonidoss[3]

[1,2]*Department and Computer Science and Engineering, Hindustan Institute of Technology and Science, Tamilnadu, India.*
[3]*Department and Computer Science and Business Systems, S.A Engineering College (Autonomous), Tamilnadu, India.*

[1]*Corresponding Author : annievargh@gmail.com*

Received: 16 September 2023 Revised: 27 October 2023 Accepted: 14 November 2023 Published: 30 November 2023

Abstract - *Credit score analysis systematically evaluates an individual or entity's financial history and behaviour to determine their creditworthiness. Traditional methods for credit score analysis have several challenges, such as privacy concerns, lack of flexibility, vulnerability to identity theft, limited data, and real-time analysis. To overcome these complexities, this paper proposes a novel method combining the advantages of Random Forest and kernel Support Vector Machine (SVM). The proposed method has three phases: data preprocessing, feature extraction, and classification. In the preprocessing phase, the proposed method eliminates the noise and errors from the raw data based on obtaining quality input for the analysis. In this study, Random Forest is utilized to extract the most significant features based on the domain and credit data analysis also, kernel SVM is employed for classification by analyzing the components and their impact on credit scoring. Also, the study conducted experiments on the German Credit Risk dataset. The performance evaluation of the proposed method involves analyzing its effectiveness based on evaluation metrics and comparing its performance with existing methods. The experimental results depict that the proposed method obtained better outcomes and achieved high efficiency for credit score analysis.*

Keywords - *Credit score analysis, Random Forest, Kernel Support Vector Machine, German Credit Risk dataset, Feature extraction, Classification.*

1. Introduction

Due to the emergence of rural markets, India, as a developing economy, recognized it as a growth engine, drawing attention to credit penetration among the farming community. Bank access is still a concern and requires manipulation, especially for small farmers, poor households, and individual female farmers [1]. Rural credit is one of the significant inputs to worldwide farm production [2].

The dependent variable is contradictory and associated with credit scoring; the non-failed loans are denoted as '1', and the failed loans are represented as '0'. Therefore, the proper evaluation of credit applicants is essential from the point of view of the banking department [3]. Over the past years, the topic of credit scoring in finance and economics has been at the forefront of employing machine learning methods, namely SVM and decision trees. To develop the credit-scoring models, financial institutions and banks provide more consideration [4].

Credit scoring is a statistical analysis executed by financial organizations or credit bureaus to assess the borrower's creditworthiness, depending on the demographic data, credit history, and behaviour [5]. Typically, borrowers with increased credit scores may get a loan and charge a low-interest rate, and due to the lack of credit data, customers are excluded from credit.

To the credit bureaus, non-bank lenders and financial organizations routinely which can provide credit information reports from the bureau to assess the creditworthiness of new loan applications and investigate accounts for each membership and application fee [6].

Determining the features and choosing the task are challenging in constructing credit scoring models. The data applied for models can gathered from several resources, and sometimes, the data size will be reduced [7]. A popular problem is credit rating prediction research about the financial positions of organizations from particular economic rates [8]. Various studies found that real-world data are roughly dispensed among multiple categories, it is difficult to manage the continuous aspects of credit rating problems, and it cannot be fulfilled with prospects.

Recent financial scenarios deal with challenges such as computational complexity and cost expenses. Current techniques such as DGHNL, GSCI-based Ensemble, Fuzzy

BWM-TOPSIS, and IBA-DE methods have poor performance rates and lack of efficiency. The extraction-based Random Forest is proposed to overcome these issues, which can achieve better performance rate, accuracy, and superior classification efficiency. The proposed extraction-based Random Forest is employed for credit score analysis by motivating this. A novel extraction-based Random Forest method is proposed to enhance the credit score analysis performance, and the primary contribution of this work is explained as follows.

Preprocessing steps are added in this work to get a proper input that will fed into the technique. Data are gathered from the German Credit Risk dataset; normalization, data transformation, and handling missing values steps are applied to these collected data. A feature extraction process is employed to convert the proposed model to a more accurate level. Here, Random Forest and kernel SVM techniques are implemented for feature extraction and classification, a significant process for building a model to remove and classify irrelevant features and obtain better results.

The remaining sections of the paper are deliberated as follows. In section 2, the past literature works based on credit score is depicted. The proposed methodology describing various phases is described in section 3. In section 4, the results and comparative analysis are presented. Finally, the conclusion is provided in section 5 with future works.

2. Related Works

Plawiak et al. [9] introduced Credit Scoring (CS) prediction using a Deep Genetic Hierarchical Network of Learners (DGHNL). CS plays an essential role in banks and many commercial organizations, leading to the disbursement of loans in the financial sector. It decreases the risk, so the banks use many methods to secure the customers and face the challenges in CS. This method included tasks with many learners, such as SVM, KNN, etc. Further, the ML repository has thousands of records, and the German Credit dataset was analyzed using the DGHNL.

On the other hand, it achieves 94.60% accuracy, and the merit of the method was ease of use and more time needed to train learners. Chen et al. [10] established the credit scoring model using Generalized Shapley Choquet Integral (GSCI) with an ensemble method. Credit scoring was more essential in banks as well as commercial organizations. This method developed a new ensemble method by the integration of Shapley as well as the Choquet Integral.

Further, based on the accuracy and the objectivity of diversity activity, ambiguous size linearity of the determination programming method was constructed. Normal vague numbers were used during training to preserve the original data. On the other hand, the GSCI method was introduced to compute the forecasting value of ensemble methods. As a result, the drawback was that the dataset in the technique was minimal.

Roy et al. [11] implemented the credit scoring method for Small and Middle-sized Enterprises (SMEs), which was crucial in every country's economic growth. Still, access was obstructed for proper funding, and many risks were faced for the disbursement of funds. To avoid this problem, this method was developed. In this method, BWM was used to control the quality of weights, and TOPSIS marks were utilized for SMEs. Further real-time analyses were done to prove the efficiency of the developed method.

On the other hand, this method achieved more accuracy, and the demerit cost of computation was high. Jelinek et al. [12] illustrated the credit ratings using the Interpolative Boolean Algebra (IBA) based Differential Evolution (DE) method. Credit scoring was not only the country's economic growth but also an indicator of growth and the reason for trust in the government. In this method, utilizing the DE process commonly available, receive a history of indicators from the data of the IBA framework based on the pseudo-logic activity, explain these functions, and then implement subtle quality between the countries. On the other hand, this method enhances prediction performance but lacks security in credit ratings.

Tezerjan et al. [13] developed a credit score using hybrid methods. Many size ideas were created for credit rating for bank customers. The world was changing economy, so the size models did not contain the exact accuracy the detailed model was necessary to calculate the accuracy. This method provides the client results utilizing the five C criteria, and then the various economic shocks damage the consumer criteria. In addition, using the historical information based on the ANFIS and RNN, the hybrid method was utilized to forecast the various stock market division shocks. As a result, this method takes more time for scoring the consumer.

Tripathi et al. [14] elaborated on the Collaborative Feature Ranking (CFR) to enhance performance. CS plays an essential role in banks and many commercial organizations, leading to the disbursement of loans in the financial sector. The dataset of credit score included year income, working status, etc., connected to the user credentials, so the dataset was very high-dimensional.

Further, in the high-dimensional dataset, many aspects were unnecessary, which caused some fundamental issues. So, a helpful aspect exam model overcame this issue, and CRF was utilized to enhance the performance of the categorization function. On the other hand, this method uses five datasets and contains a large dataset; the drawback is that it is expensive for data categorization.

2.1. Research Gap

In machine and deep learning, numerous methods are generated for credit score analysis, in which the methods extract and classify the aspects. That has the advantages of extracting elements and classification; meanwhile, they have disadvantages, which means they require a lot of time to extract the elements, have low efficiency, have higher implementation time, and more. Hence, this work proposes an extraction-based Random Forest. This method resolves issues and provides better results. The research gap is presented as follows:

2.1.1. Efficient Extraction and Classification

By overcoming the issues of existing methods, the proposed model extracts and classifies the analysis effectively and requires less time for these extractions and classifications.

2.1.2. Improved Model Performance

To enhance the model performance in terms of both feature extraction and classification. This significantly analyzes the credit score outcomes by enabling early analysis. Thereby, the performance of the proposed performance is enhanced.

3. Proposed Methodology

Figure 1 represents the overall workflow of the proposed model. For credit score analysis, the German Credit Risk dataset is employed. These datasets are to be preprocessed using the data preprocessing steps: normalization, data transformation, and handling missing values.

After that, the feature extraction phase can take place with the Random Forest method to extract the irrelevant features. The Kernel SVM is employed for classification. Finally, the performance evaluation involves analyzing its effectiveness based on evaluation metrics. A detailed explanation of the proposed methodology is given in the section below.

3.1. Preprocessing

Preprocessing is the necessary step that efficiently removes the noise and errors from the raw data based on obtaining quality input for the analysis. In the data preprocessing phase, the credit score analysis data is chosen from the German Credit Risk dataset, and three types of techniques are data normalization, handling missing data, and data transformation to enhance the data quality.

3.1.1. Data Normalization

Normalization is erasing the distortion and possible dependencies caused by various quantities. Further normalization is one of the steps of preprocessing and strategies for data augmentation to enhance the diversity of the dataset [15].

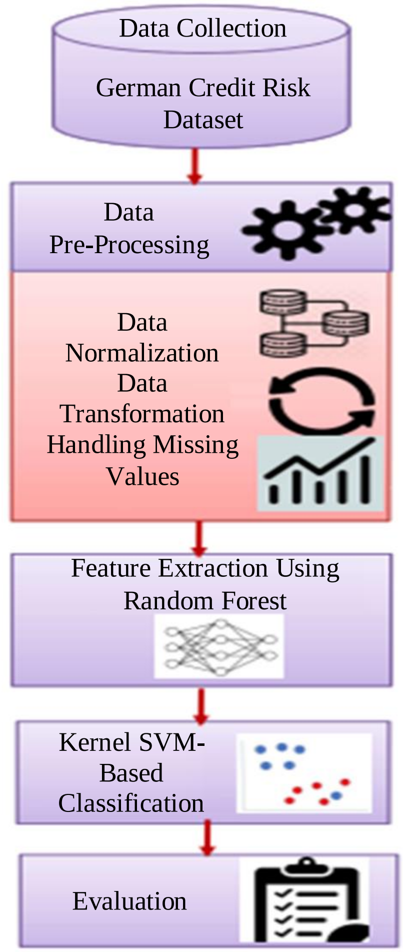

Fig. 1 The overall workflow of the proposed model

This process between 0 and 1 or between -1 and +1 to a certain extent sending data generally, the normalization technique is used in the preprocessing stage and is expressed as,

$$y' = \frac{y - MIN}{MAX - MIN} \tag{1}$$

3.1.2. Handling Missing Values

One of the significant difficulties throughout the preprocessing is handling missing data, which should be dealt with carefully. A dataset holding missing values is created for many reasons, such as equipment failure, human mistakes, absence of data, data not being up-to-date, or conflicting with other data. Handling the missing data fills the linear interpolation, improving the credit score analysis accuracy.

3.1.3. Data Transformation

These processes support the decision-making processes of an established one, usable from design to changing the data in structure, construction, and cleansing.

The transformation function is utilized to improve interoperability and change data.

$$x_j = e(w_j) \tag{2}$$

Here e, it denotes the transformation function.

3.2. Feature Extraction

Features are significant based on the domain and credit data analysis. They also allow one to select exact data from the high-dimensional data and transfer it into low-dimensional data in the crucial data without any data loss. Suppose there is a tremendous amount of noise available in the data. This dimensionality is minimized without losing necessary data for coherent categorization by feature extraction. The Random Forest method is utilized for feature extraction, as discussed below.

3.2.1. Random Forest

In the subsection of similar ensemble learning, Random Forest [16] is regarded as one of the ensemble learning techniques. To access a random subset of feature vectors of each decision tree and base in a Random Forest. Therefore, the feature vector can be expressed as:

$$W = (w_1, w_2, \ldots . W_o) \tag{3}$$

Where the dimension property of the base learner denotes O. To discover the prediction function *f(w)* describes the primary task that forecasts the parameter Z. The mathematical expression of the prediction function is,

$$K(Z, f(w)) \tag{4}$$

Where the loss functions denote K reducing the desired value of the loss function is the primary goal. Zero-one loss and squared error loss are general categorizations and regression application possibilities. These functions are represented below.

$$K(Z, f(w)) = (Z - f(w)^2) \tag{5}$$

$$K(Z, f(w)) = J(Z \neq f(w)) = \begin{cases} 0, \ if\ Z = f(w) \\ 1, \ otherwise \end{cases} \tag{6}$$

A collection of base learners come together to make an ensemble. The base learners are expressed as,

$$g_1(w), g_2(w), \ldots ., g_i(w), \tag{7}$$

The averaging and classification applications are denoted by Equation (8) and Equation (9) for regression applications.

$$f(w) = \frac{1}{I}\sum_{i=1}^{I} g_i(w) \tag{8}$$

$$f(w) = \arg\max i \sum_{i=1}^{I} J(z = g_i(w)) \tag{9}$$

To detect a subset by applying multiple substitutions of features and training data that increase the performance and precision of the outcome in the Random Forest. Depending on the classifier’s performance, all the resources are combined into the RF algorithm to calculate the performance.

3.3. Classification

Classification is one of the tasks in credit score analysis; the central problem of credit scoring is binary classification. [17] Many researchers utilize this classification process to predict credit risk. In this method, analysis of the aspects of Kernel SVM used for classification is discussed in the sections below.

3.3.1. Kernel SVM

A kernel SVM [18] can be defined similarly. $\phi_k : \Re^m \rightarrow \mathrm{H}_s$ is an imagined non-linear function employed on the input to convert it to a greater dimensional region where inner products can be assumed as H_s, which denotes a Hilbert space. The induction of the mapping $\emptyset$ is to focus on the necessary correspondence between linear SVMs and kernel SVMs. Consider an input $w \in \Re^m$ may accomplish two-label categorization and the mathematical formulation is,

$$z = \langle\ x, \phi_k(w)\ \rangle_{H_s} + a \tag{10}$$

Where $x \in H_s$ and $a \in \Re$. Kernel SVM requires a suitable kernel K_s, in preference to directly executing the mapping and appealing inside products of Hilbert space. This is referred to as the kernel trick. The kernel function is associated with the non-linear mapping via the expression (11).

$$K_s(w_1, w_2) = \langle \phi_k(w_1), \phi_k(w_2) \rangle_{H_s} \tag{11}$$

The Gaussian Radial basis kernel is generally utilized kernel function, expressed by,

$$K_s(w_1, w_2) = \exp(-\eta \parallel w_1 - w_2 \parallel^2) \tag{12}$$

Consider that *X* place in the span of $\phi_k(\nu_j)$ specific values $\nu_j \in \Re^m$.

$$x = \sum_{j=1}^{l} b_j \phi_k(\nu_j) \tag{13}$$

Where $b_j \in \Re$ denotes the scalars. The support vectors refer V_j, and they can be considered as,

$$z = \langle x, \phi_k(w) \rangle_{H_s} + a \tag{14}$$

$$= \left\langle \sum_{j=1}^{l} b_j \phi_k(\nu_j), \phi_k(w) \right\rangle_{H_s} + a \tag{15}$$

$$= \sum_{j=1}^{l} b_j \langle \phi_k(\nu_j), \phi_k(w) \rangle_{H_s} + a \tag{16}$$

$$= \sum_{j=1}^{l} b_j K_s(\nu_j, w) + a \tag{17}$$

In the RBF kernel, this changes

$$z(w) = \sum_{j=1}^{l} b_j \exp(-\eta \| w - \nu_j \|^2) + a \tag{18}$$

The Gaussian centres are selected from the training samples {Wi}, but the trained parameter is V_j in normal SVMs. The kernel SVM is a better classifier and the most straightforward technique than linear SVM. Kernel SVM categorizes data depending on their complex surfaces.

4. Results and Discussions

The effectiveness of the proposed method for credit score analysis and the results achieved from the study are illustrated in this section.

The proposed method is evaluated with different evaluation measures namely recall, AUC-ROC, F1-score, specificity, accuracy and precision, and the results are compared with existing methods such as Deep Genetic Hierarchical Network of Learners (DGHNL) [9], Generalized Shapley Choquet Integral (GSCI) based Ensemble [10], Fuzzy Best Worst Method with the Technique for Order Preferences by Similarity to an Ideal Solution (Fuzzy BWM-TOPSIS) [11] and Interpolative Boolean Algebra with Differential Evolution (IBA-DE) [12].

4.1. Experimental Setup

In this study, Matlab is implemented, and the system uses Intel Core i5-3470 with 2 GB RAM operating at a clock speed of 3.20 GHz. Furthermore, the computing platform meets minimum software and hardware requirements, including sufficient storage capacity, computational power, and compatibility with the proposed method. Considering these system requirements and platform specifications, this study ensures efficient and reliable implementation of the proposed method for credit score analysis.

4.2. Parameter Setting

The parameter setting process enhances the effectiveness of the proposed method, and Table 1 depicts the parameter setting of the study. In this process, the optimal parameter values are generated to enhance the performance of the proposed method.

Table 1. Parameter setting

Parameter	Value
Kernel	Radial Basis Function
Gamma	0
Step size	0.005
Batch size	64
Drop out	0.5
Momentum	0.9
Degree	3
Trees	50
Regularization Parameter	1

In this work, the parameter setup contains a batch size of 64, an out value of 0.5, a momentum of 0.9, a degree of 3, a step size of 0.005, several trees is 50, a regularization parameter of 1, a gamma is 0, and the kernel is Radial basis function. This work ensures reliable and efficient implementation of the proposed method for credit score analysis.

4.3. Dataset Description

In this study, the German Credit Risk dataset [19] is utilized to implement the proposed method for credit score analysis. The dataset is used for assessing the credit risk of loan applicants. It was initially collected from a German bank. It consists of 20 attributes, which include both numerical and categorical variables. These attributes provide information about the applicant's personal and financial details, such as age, gender, job, housing, and existing credits. The primary target variable in this dataset is binary, denoting whether a credit applicant is a good or bad credit risk. This binary classification is based on the credit applicant's loan repayment ability. In this work, 5000 instances are collected from the dataset, and these instances are split into training and testing in the ratio of 80:20 to evaluate the proposed method for credit score analysis.

4.4. Evaluation Measures

The performance of the proposed method in credit score analysis is evaluated through different evaluation measures such as specificity, accuracy, precision, recall, AUC-ROC, and F1-score. The performance evaluation of these metrics is conducted based on the mathematical formulations mentioned below.

4.4.1. Accuracy

The proportion of correctly predicted instances to the total cases is defined as accuracy (a_y). It is a widely used performance metric for credit score analysis, and it is formulated as,

$$a_y = \frac{t_{pos} + t_{neg}}{t_{pos} + t_{neg} + f_{pos} + f_{neg}} \tag{19}$$

4.4.2. Precision

Precision is the ratio of correctly predicted positive instances to all positive ones. Precision can be represented as,

$$p_n = \frac{t_{pos}}{t_{pos} + f_{pos}} \tag{20}$$

4.4.3. Recall

Recall (r_e) is the proportion of correctly predicted positive instances to all positive ones. The recall can be calculated as,

$$r_e = \frac{t_{pos}}{t_{pos} + f_{neg}} \tag{21}$$

4.4.4. F1-Score

F1-Score (*f1*) is the harmonic mean of precision and recall, which is useful when there is an imbalance in the dataset. The F1-score can be formulated as,

$$f_1 = 2 \times \frac{(p_n \times r_e)}{(p_n + r_e)} \tag{22}$$

4.4.5. Specificity

The proportion of negative instances predicted correctly to all actual negative models is called specificity (S_{py}). The mathematical representation of specificity is,

$$s_{py} = \frac{t_{neg}}{t_{neg} + f_{pos}} \tag{23}$$

4.4.6. AUC-ROC

AUC-ROC measures the model's ability to discriminate between different instances. It plots the true positive rate (t_{pos} Rate) against the false positive rate (f_{pos} Rate) at various classification thresholds and calculates the area under the curve.

$$t_{pos}\ Rate = \frac{t_{pos}}{t_{pos} + f_{neg}} \tag{24}$$

$$f_{pos}\ Rate = \frac{f_{pos}}{f_{pos} + t_{neg}} \tag{25}$$

Equations (19) to (25) t_{pos} represents true positive, t_{neg} indicates true negative, f_{pos} denotes false positive, and f_{neg} represents false negative respectively.

4.5. Performance Analysis

The performance analysis of the proposed method for credit score analysis using the specified performance metrics, namely recall, AUC-ROC, specificity, F1-score, precision, and accuracy, provides a comprehensive evaluation of its effectiveness.

The performance is evaluated by comparing the proposed method with the existing techniques, namely DGHNL, GSCI-based Ensemble, Fuzzy BWM-TOPSIS, and IBA-DE. Figures 2 to 7 depict the comparative graphical representation of the proposed method and the existing techniques for different evaluation metrics based on credit score analysis.

The accuracy of the proposed method and the existing techniques is illustrated by the graphical analysis shown in Figure 2. The proposed method achieved a high accuracy of 98.76% while the existing methods such as DGHNL, GSCI-based Ensemble, Fuzzy BWM-TOPSIS, and IBA-DE obtained low accuracy of 97.64%, 96.57%, 95.43% and 94.32% respectively.

Figure 3 illustrates the graphical analysis to depict the precision of the proposed method and the existing methods. The proposed method achieved a high precision of 97.86% while the existing methods such as DGHNL, GSCI-based Ensemble, Fuzzy BWM-TOPSIS, and IBA-DE obtained low accuracy of 96.74%, 95.62%, 94.58% and 93.45% respectively.

In Figure 4, the graphical analysis represents the recall of the proposed and existing methods. The proposed method achieved a high recall of 97.84% while the existing methods such as DGHNL, GSCI-based Ensemble, Fuzzy BWM-TOPSIS, and IBA-DE obtained low recall of 96.72%, 95.61%, 94.56% and 93.47% respectively.

Figure 5 presents the graphical analysis to illustrate the F1-score of the proposed method and the existing methods. The proposed method achieved a high F1-score of 97.82% while the existing methods such as DGHNL, GSCI-based Ensemble, Fuzzy BWM-TOPSIS, and IBA-DE obtained low F1-score of 96.75%, 95.63%, 94.51% and 93.46% respectively.

The specificity of the proposed method and the existing techniques is depicted by the graphical analysis illustrated in Figure 6. The proposed method achieved a high specificity of 97.83% while the existing methods such as DGHNL, GSCI-based Ensemble, Fuzzy BWM-TOPSIS, and IBA-DE obtained low specificity of 96.79%, 95.68%, 94.56% and 93.47% respectively.

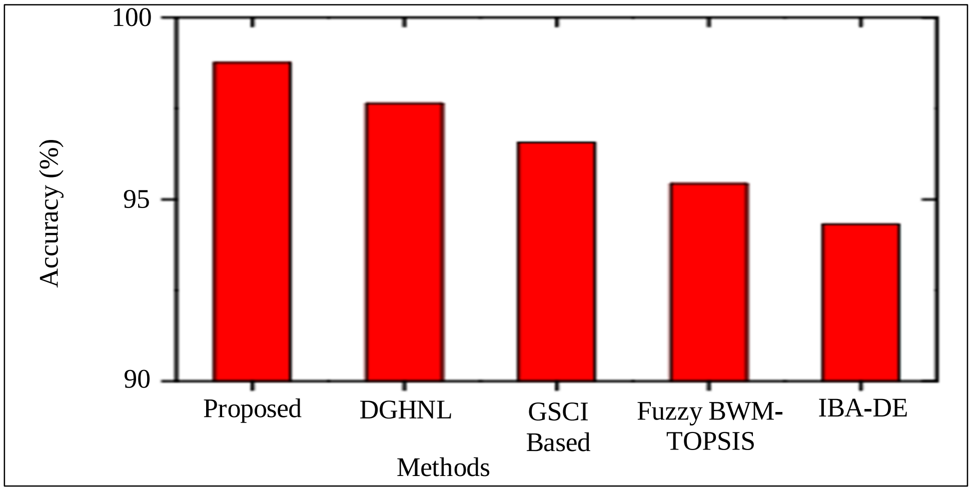

Fig. 2 Graphical representation based on accuracy

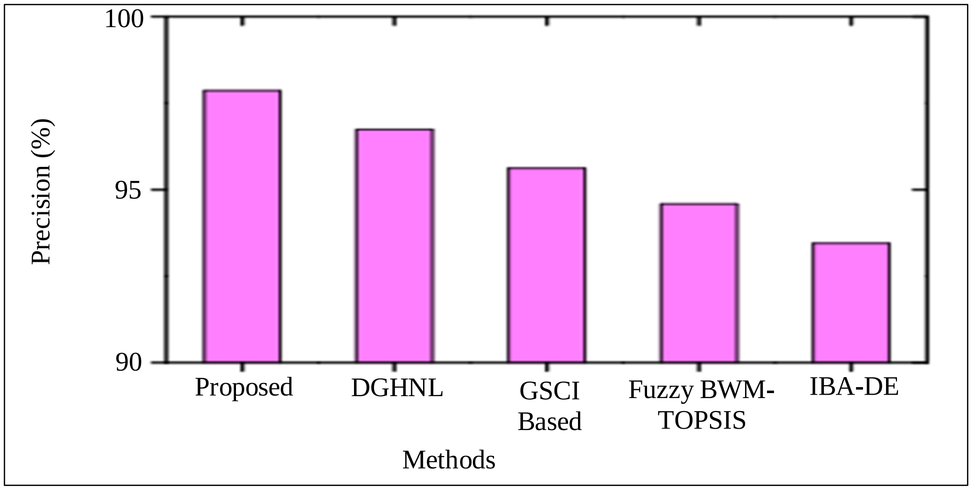

Fig. 3 Precision analysis for performance validation

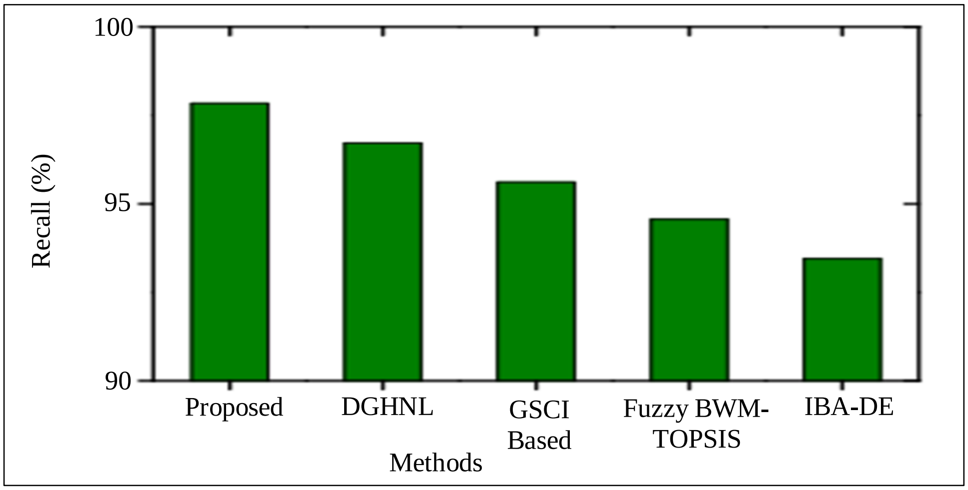

Fig. 4 Performance evaluation based on recall

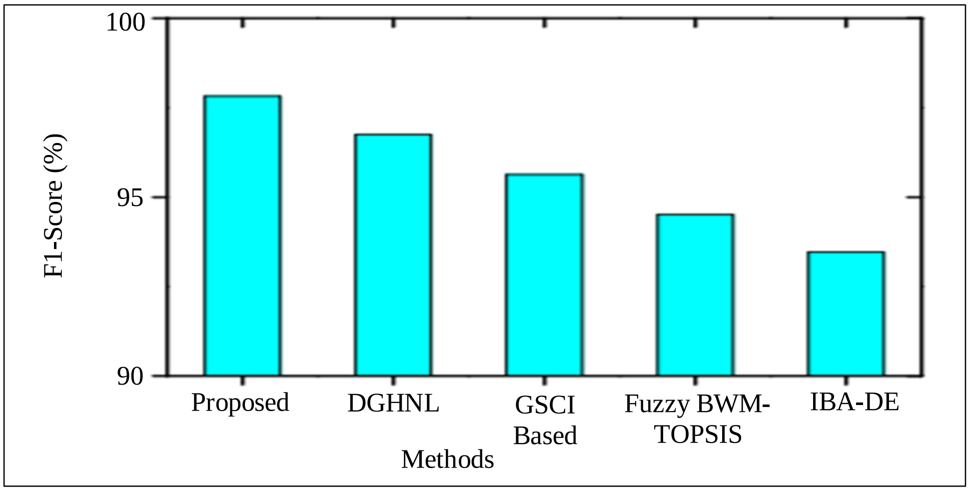

Fig. 5 Graphical representation based on F1-score

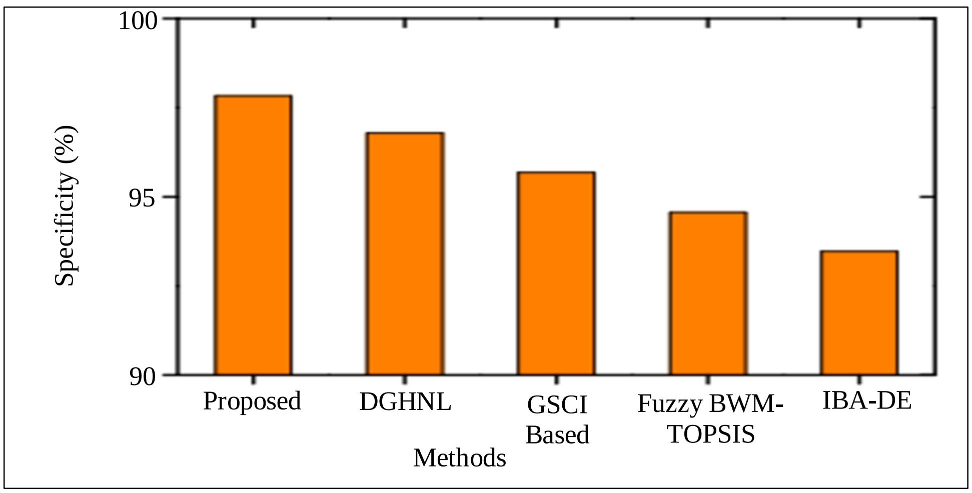

Fig. 6 Specificity analysis for performance validation

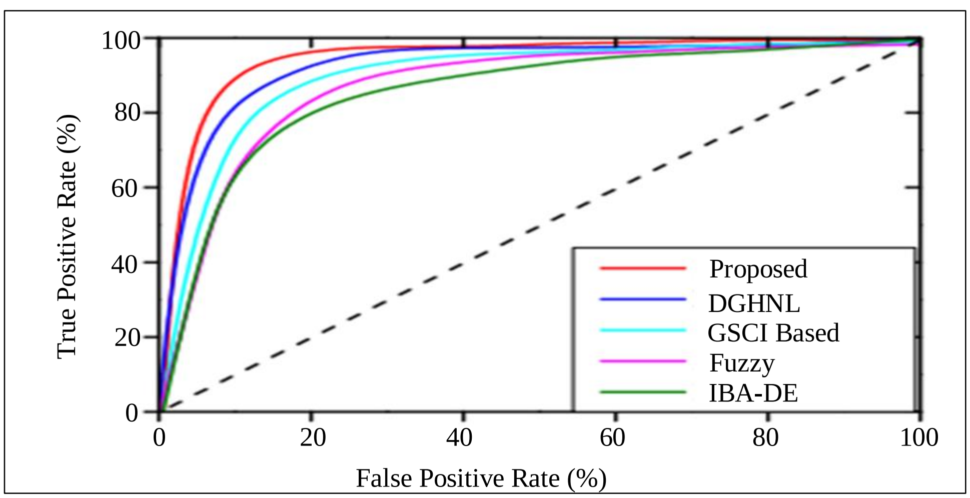

Fig. 7 Performance evaluation based on AUC-ROC

Figure 7 depicts the graphical analysis to illustrate the AUC-ROC of the proposed method and the existing methods. The proposed method obtained a high AUC-ROC of 97.81% while the existing methods such as DGHNL, GSCI-based Ensemble, Fuzzy BWM-TOPSIS, and IBA-DE obtained low AUC-ROC of 96.73%, 95.64%, 94.57% and 93.45%

respectively. The performance analyses evaluate the effectiveness of the proposed method for credit score analysis. The results illustrate that the proposed method achieved a higher recall, AUC-ROC, F1-score, specificity, accuracy, and precision than the existing methods.

5. Conclusion

In this paper, the proposed method holds significant advantages for credit score analysis by assessing various factors such as types of credit accounts, credit utilization, and payment history.

In this study, Random Forest is utilized to extract features from the dataset by understanding which parts have the most significant importance, and kernel SVM is employed for classification by analyzing the components and their impact on credit scoring.

The study is validated on the German Credit Risk dataset, and the effectiveness of the proposed method is evaluated using different evaluation measures such as specificity, AUC-ROC, accuracy, precision, F1-score, and recall. The results are compared with techniques such as DGHNL, GSCI-based Ensemble, Fuzzy BWM-TOPSIS, and IBA-DE. The proposed method achieved a high accuracy of 98.76%. The results depict that the proposed method achieved better results for credit score analysis.

In future work, Artificial Intelligence (AI) and machine learning models will continue to evolve in credit score analysis, providing more accurate and dynamic creditworthiness assessments. Also, the study will focus on making credit scoring models more transparent and fairer, addressing possible biases, and giving clear explanations for decisions.

References

[1] Anil Kumar, Suneel Sharma, and Mehregan Mahdavi, “Machine Learning (ML) Technologies for Digital Credit Scoring in Rural Finance: A Literature Review,” *Risks*, vol. 9, no. 11, pp. 1-15, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[2] Yoichi Hayashi, “Emerging Trends in Deep Learning for Credit Scoring: A Review,” *Electronics*, vol. 11, no. 19, pp. 1-30, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Yiheng Li, and Weidong Chen, “A Comparative Performance Assessment of Ensemble Learning for Credit Scoring,” *Mathematics*, vol. 8, no. 10, pp. 1-19, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Beibei Niu, Jinzheng Ren, and Xiaotao Li, “Credit Scoring Using Machine Learning by Combing Social Network Information: Evidence from Peer-to-Peer Lending,” *Information*, vol. 10, no. 12, pp. 1-14, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[5] Apostolos Ampountolas et al., “A Machine Learning Approach for Micro-Credit Scoring,” *Risks*, vol. 9, no. 3, pp. 1-20, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Juan Laborda, and Seyong Ryoo, “Feature Selection in a Credit Scoring Model,” *Mathematics*, vol. 9, no. 7, pp. 1-22, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Sunghyon Kyeong, Daehee Kim, and Jinho Shin, “Can System Log Data Enhance the Performance of Credit Scoring?-Evidence from an Internet Bank in Korea,” *Sustainability*, vol. 14, no. 1, pp. 1-12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Lars Ole Hjelkrem, Petter Eilif de Lange, and Erik Nesset, “The Value of Open Banking Data for Application Credit Scoring: Case Study of a Norwegian Bank,” *Journal of Risk and Financial Management*, vol. 15, no. 12, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Paweł Pławiak et al., “DGHNL: A New Deep Genetic Hierarchical Network of Learners for Prediction of Credit Scoring,” *Information Sciences*, vol. 516, pp. 401-418, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Xiaohong Chen et al., “A Novel GSCI-Based Ensemble Approach for Credit Scoring,” *IEEE Access*, vol. 8, pp. 222449-222465, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[11] Pranith Kumar Roy, and Krishnendu Shaw, “A Multicriteria Credit Scoring Model for SMEs Using Hybrid BWM and Topsis,” *Financial Innovation*, vol. 7, pp. 1-27, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Srđan Jelinek et al., “A Novel IBA-DE Hybrid Approach for Modeling Sovereign Credit Ratings,” *Mathematics*, vol. 10, no. 15, pp. 1-21, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Mostafa Yousofi Tezerjan, Azamdokht Safi Samghabadi, and Azizollah Memariani, “ARF: A Hybrid Model for Credit Scoring in Complex Systems,” *Expert Systems with Applications*, vol. 185, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[14] Diwakar Tripathi, B. Ramachandra Reddy, and Alok Kumar Shukla, “CFR: Collaborative Feature Ranking for Improving the Performance of Credit Scoring Data Classification,” *Computing*, vol. 104, no. 4, pp. 893-923, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Peshawa J. Muhammad Ali, and Rezhna H. Faraj, “*Data Normalization and Standardization: A Technical Report*,” Koya University, Ebril, Iraq, Machine Learning Technical Reports, vol. 1, no. 1, pp. 1-6, 2014. [Google Scholar]

[16] Mohammad Savargiv, Behrooz Masoumi, and Mohammad Reza Keyvanpour, “A New Random Forest Algorithm Based on Learning Automata,” *Computational Intelligence and Neuroscience*, vol. 2021, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] Diwakar Tripathi et al., Experimental Analysis of Machine Learning Methods for Credit Score Classification," *Progress in Artificial Intelligence*, vol. 10, pp. 217-243, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[18] Shihao Jiang, Richard Hartley, and Basura Fernando, "Kernel Support Vector Machines and Convolutional Neural Networks," *2018 Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1-7, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[19] German Credit Risk, Credit Classification. [Online]. Available: https://www.kaggle.com/datasets/uciml/german-credit