Journal of Positive School Psychology
2022, Vol. 6, No. 3, 6822-6831
Gopikrishna Panda et al. http://journalppw.com

An Intensified Social Spider Optimization (ISSO) based Progressive Kernel Ridge Regression (PKRR) Classification Model for Automobile Insurance Fraud Detection

Gopikrishna Panda[1], Sunil Kumar Dhal[2] andSudipta Dash[3]

[1]Sri Sri University, Cuttack
[2]Business Studies, Sri Sri University, Cuttack
[3]Computer Application, Regional College of Management, Bhubaneswar

Email: ergopipanda@gmail.com

ABSTRACT

Automobile insurance fraud detection is one of the most essential and highly demanding process need to be accomplished in the insurance industries. The existing works are highly concentrated on developing the fraud detection system by using various detection with the major issues of high complexity in computational operations, requires more time consumption, reduced speed of processing, and high error outputs. The proposed work objects to implement an efficient optimization based classification model for designing the automobile insurance fraud detection system. For this purpose, an Intensified Social Spider Optimization (ISSO) based Progressive Kernel Ridge Regression (PKKR) classification mechanism is developed in this work, which helps to accurately detect the insurance frauds from the given datasets. Initially, the data preprocessing and clustering operations have been performed by using the Distance based Fuzzy Clustering (DFC) approach, which produces the normalized clustered dataset for further processing. Then, the ISSO technique is deployed to select the best features for training the model of classifier based on the optimal fitness value. By using these features, the PKKR classifier accurately predicts the insurance frauds with reduced computational complexity and time consumption. During performance analysis, various evaluation indicators are used to validate the results of proposed detection mechanism.

Keywords: Automobile Insurance Fraud Detection (AIFD), Fuzzy Distance based Clustering (FDC), Intensified Social Spider Optimization (ISSO), ProgressiveKernel Ridge Regression (PKRR)

1. INTRODUCTION

Due to the rapid growth of insurance industries, the fraudulent activities are highly increasing in the recent days [1, 2]. According to the recent reports, it is studied that around 20% to 35% of automobile insurance privileges are identified as the fraudulent. Normally, the insurance frauds target [3, 4]to increase the economic and reputation loss of insurance industries, which is considered as the major crisis need to be resolved by using an effective prediction approach. The insurance fraud detection is considered as one of the most challenging task in both technical and operational sides. Conventionally, an inspection audit[5]has been conducted for the identification and detection of insurance frauds. Also, it is considered as a kind of anomaly detection problem, which is solved by the use of detection and classification techniques[6-8]. In thispresentstudy, various machine learning and deep learning based classification techniques are used for identifying and accurately detecting the fraud in automobile insurance sector. The typical detection framework comprises the working modules of dataset collection, normalization, clustering, feature selection, and classification[9, 10]. In which, the preprocessing is one of the most essential process in the detection system, because the noisy data can degrade the performance of entire system with increased misclassification results. Hence, it must be preprocessed [11, 12]by missing value prediction, noise elimination, and normalization. After that, the clustering is performed to group the similar attributes into the form of clusters, which helps to improve the prediction accuracy of classifier.

Moreover, the optimization and classification[13, 14] are the most essential stages of insurance fraud detection system, in which the optimization techniques are mainly used to select the particular number of features by computing the optimal global solution. For this purpose, there are different types of techniques[15] have been utilized in the conventional works, which includes Whale Optimization (WO), Artificial Bee Colony (ABC) optimization, Cat Swarm Optimization (CSO), Genetic Algorithm (GA), and other meta-heuristic techniques. Consequently, the classification approaches[16] are increasingly used for predicting the classified label as whether normal or anomaly. For automobile insurance fraud detection system, the different types of classifiers such as Support Vector Machine (SVM), Decision Tree (DT), Naïve Bayes (NB), K-Nearest Neighbor (KNN), Gradient Boost (GB), and other machine learning/deep learning techniques are used in the present days. Still, the conventional techniques facing the problems[17, 18] of increased error rate in detection, high time consumption for training and testing the data, reduced convergence speed, and reduced detection performance. In order to solve these problems, the proposed work intends to develop an intelligent optimization based classification technique for designing an automobile insurance fraud detection system. The major contributions of the proposed work are as follows:

- To develop an intelligent automobile insurance fraud detection system, an advanced optimization based classification methodologies are implemented.
- To cluster the normalized dataset for simplifying the process of classification, a Fuzzy Distance based Clustering (FDC) technique is employed.
- To optimally select the suitable number of features based on the best fitness function, an Intensified Social Spider Optimization (ISSO) technique is deployed, which helps to improve the accuracy of classifier.
- To accurately predict the classified label based on the optimal set of features, the Progressive Kernel Ridge Regression (PKRR) based classification technique is utilized.
- To validate the results of the proposed fraud detection system, various performance measures are validated and compared with the recent state-of-the-art models.

The remaining portions of this paper are segregated into the following units: Section II reviews some of the existing works related to the data clustering, optimization, and machine learning classification techniques used for an automobile insurance fraud identification and detection. Also, the advantages and disadvantages of each technique according to its working operations and characteristics has been discussed. Then, the clear description about the research methodology with its working flow and algorithmic models are presented in Section III. The performance analysis of both conventional and proposed fraud detection techniques are validated and compared by using various measures in Section IV. Finally, the overall paper is concluded with its future work in Section V.

2. Literature Review

This section reviews some of the existing works related to the concept of automobile fraud detection system using optimization and machine learning based classification techniques. Also, it discusses about the benefits and demerits of each technique according to its key features and characteristics.

Itri, et al [19]utilized an Extreme Gradient Boost (XGB) classification technique for developing an automated fraud detection system. The main contribution of this work was to reduce the human intervention and financial losses in the automobile insurance system by identifying the fraudulent claims with the help of machine learning classification approach. The major stages involved in this work were data preprocessing, exploration, privacy preservation, and classification. The key benefits of this work were better detection rate and increased accuracy. However, it facing the difficulties related to the factors of high computational complexity and time consumption for classification.

Majhi, et al [20] suggested a fuzzy clustering incorporated with modified whale optimization algorithm for developing an automobile insurance fraud detection framework. The main focus of this work was to utilize the clustering based optimization technique for obtaining the best optimal solution to improve the performance of classifier. Also, various machine learning based classification approaches were compared in this work for validating the increased performance of optimization based classification technique. However, this work has the major limitations of increased time consumption for training and testing the data. *Subudhi, et al* [21] utilized a hybrid optimization based clustering model, named as, Genetic Algorithm (GA) with Fuzzy C-Means (FCM) clustering for the detection of frauds in an automobile industries. Here, the 10-cross fold validation method was used to validate the performance of the detection system. *Subudhi, et al* [22]implemented an optimized FCM technique for developing a two stage insurance fraud detection system. This work mainly intends to increase the accuracy of classification by implementing an optimization based clustering approach. Also, some of the extensively used machine learning approaches such as Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), and Decision Tree (DT) have been validated and compared in this work for analyzing the better performance of detection mechanism.

Liu, et al [23] introduced a Maximum Likelihood Evidential Reasoning (MAKER) framework for an accurate automobile insurance fraud detection. This work mainly objects to utilize the Evidential Reasoning (ER) rule for increasing the overall detection efficiency. Here, the data driven and experience based likelihood functions were estimated for improvising the process of classification. Yet, this framework has the major limitations of increased computational cost and complexity, which affects the effectiveness and performance of the entire system. *Pathi, et al* [24] validated the efficiency of three different classification techniques such as SVM, MLP and K-Nearest Neighbor (KNN) for au automobile insurance fraud detection system. Here, the Synthetic Minority Oversampling (SMOTE) analysis was performed for solving the data imbalance problem. During this process, the minority and majority classes have been segregated from the unbalanced dataset, and the synthetic samples were randomly generated for the removal of imbalance class. The key benefit of this work was, the SMOTE analysis could improve the detection performance with reduced false positives.

Pranavi, et al [25] implemented a Random Forest (RF) integrated with KNN classification algorithm for an accurate detection of insurance fraud. The main intention of this work was to utilize a hybrid machine learning technique for detecting the insurance frauds with increased accuracy and reduced complexity. The stages involved in this framework were data analysis, transformation, model training, testing and classification. However, this work does not has an increased ability to handle the large dimensional datasets with reduced time consumption. *Yan, et al* [26] employed an Extended Learning Machine (ELM) incorporated with the Artificial Fish Swarm Optimization (AFSO) technique for an automobile insurance fraud detection system. This paper objects to use an optimization based classification technique for accurately detecting the insurance frauds from the given datasets based on the optimal number of features. Also, an empirical analysis has been conducted in this work for evaluating the performance of this work by using various measures.

Dhieb, et al [27] suggested an AI driven XGB classification technique for identifying and detecting the automobile insurance frauds from the datasets. Here, the blockchain methodology was implemented in this detection framework for increasing the speed of processing with reduced cost consumption. In addition to that, the risk management and handling events were also performed in this work by using the XGB classification approach. The major benefits of this approach were ensured accuracy, security, and optimal performance outcomes. *Chen, et al* [28] employed a Deep Convolutional Neural Network (DCNN) classification approach for developing the credit card fraud detection system. The main intention of this work was to utilize an enhanced classification model for improving the accuracy of detection. But, it facing the problems of increased computational complexity, requires high processing time, and cost consumption.

In this present study, it is examined that the conventional fraud detection frameworks are highly using the different types of optimization based classification approaches for processing. Yet, it has the major issues related to the terms of complexity of computational operations, increased training and testing time, reduced efficiency, and inability of handling of large dimensional datasets. Hence, the proposed work intends to develop a new fraud detection framework by using an advanced optimization and classification techniques.

3. RESEARCH METHODS AND MODELS

The main contribution of this paper is to develop an enhanced automobile insurance fraud detection system by using an advanced optimization based classification methodologies. For this purpose, a novel Intensified Social Spider Optimization (ISSO) and Progressive Kernel Ridge Regression (PKRR) Classification methodologies are developed, which helps to detect the automobile insurance fraud with increased accuracy and reduced computational complexity. The major benefits of the proposed ISSO-PKRR based automobile insurance fraud detection system are as follows: reduced computational complexity, requires minimal time consumption, increased detection efficiency, high accuracy, and optimal performance rate. The overall flow of the proposed system is shown in Figure 1.

Here, the input datasets are preprocessed at first for normalizing the dataset by identifying the missing values and eliminating the irrelevant attributes. Normally, the data preprocessing is one of the most essential stage for the detection and classification systems, because the overall performance of the classifier is highly depends on the quality of input data. Hence, it must be preprocessed for eliminating the noisy contents and, the normalized dataset can be used for further processing. After that, the data clustering can be performed for grouping the attributes of information based on the distance value, which is done by using the Fuzzy Distance based Clustering (FDC) approach.

Moreover, the data clustering helps to simplify the process of classification by grouping the data into the form of clusters. In this technique, the data matrix is initialized at first for computing the centroids, and the degree of membership can be estimated according to the distance value. Then, the clustered data is fed to the input of feature selection, which is also one of the most essential operation in the detection application systems. Because, the optimization techniques are mainly used for selecting the best number of attributes based on the optimal fitness function, which helps to improve the accuracy and reduce the time consumption of overall detection system. In addition to that, it is used to train the data model using the optimal number of features for improving the efficiency of classifier. Finally, the Progressive Kernel Ridge Regression (PKRR) based machine learning technique is utilized to predict the classified label based on the optimal feature set. For evaluating the performance of this system, various measures such as sensitivity, specificity, accuracy, precision, recall, similarity coefficients, and error rate have been computed during analysis. Moreover, the obtained results are compared with some recent state-of-the-art models for proving the betterment of the proposed approach over the other techniques.

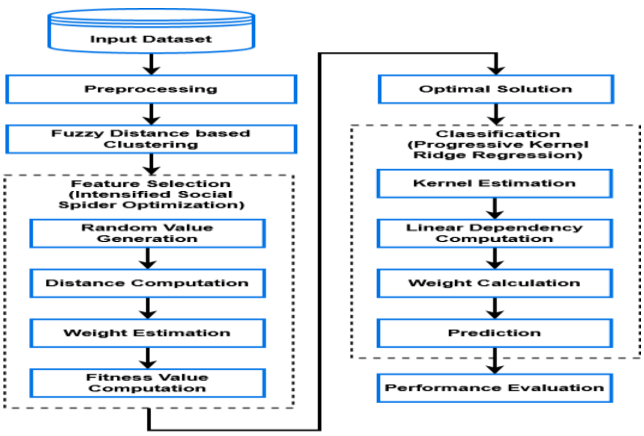

Fig 1. Flow of the proposed system

A. Distance based Fuzzy Clustering (DFC)

Initially, the automobile insurance dataset has been obtained as the input for processing, which is preprocessed at first by normalizing the attribute information. In general, the raw or original dataset comprises noisy contents, missing information, and irrelevant attribute information. Also, the noisy data can affect the detection accuracy of classifier with increased false positives, because the performance of detection system is highly depends on the input data. Hence, the data preprocessing is performed in this work for filling missing attributes and filtering the noisy contents for further processing. Moreover, the original imbalanced data can be balanced by using the Distance based Fuzzy Clustering (DFC) approach. The main purpose of using this technique is to construct the group of clusters based on the normalized attributes of information. Here, the membership functions are calculated for identifying the meaningful clusters in the given dataset, which also supports to reduce the objective function as shown in below:

$$Ob_F = \sum_{i=1}^{N} \sum_{j=1}^{cl} d_{ij}^{r} \left\| a_i - cl_j \right\|^2 \tag{1}$$

Where, Ob_F indicates the objective function, cl is the centroid of cluster, r denotes the real number i.e. >1, and d_{ij} defines the degree of membership in the set a_i of cluster. Then, the steps involved in the DFC mechanism is represented in below:

Algorithm I – Distance based Fuzzy Clustering

Step 1: Initially, form the data matrix of cluster by using,

$$D = [d_{ij}];$$

Step 2: Then, the centroid of vector is estimated as shown in below:

$$Cl^{(v)} = [cl_j];$$

$$Cl_j = \frac{\sum_{i=1}^{N} d_{ij}^{r} a_i}{\sum_{i=1}^{N} d_{ij}^{r}}$$

Step 3: After that, the data matrices such as $D^{(v)}$ and $D^{(v+1)}$ are estimated and updated by using the following model:

$$Cl_j = \frac{1}{\sum_{v=1}^{cl} \left(\frac{\|a_i - cl_j\|}{\|a_i - cl_v\|} \right)^{\frac{2}{r} - 1}}$$

Step 4: If the condition is satisfied, $\|D^{(v+1)} - D^{(v)}\| < \varepsilon$

Stop;

Else

Go to Step 2;

B. Intensified Social Spider Optimization (ISSO)

In this stage, the set of features are extracted from the clustered data by using an Intensified Social Spider Optimization (ISSO) technique. The main contribution of using this technique is to optimally select the best features based on the estimated global optimum value. Then, the selected optimal number of features are fed to the classifier for training the data model, which helps to increase the accuracy of overall detection system. Because, the high number of features require more time for training and detection, which also reduces the efficiency of the system. Hence, it is more essential to optimally select the best features according to the best fitness value computed by using the ISSO technique. Moreover, it is a kind of meta-heuristic optimization model and, works based on the cooperative nature of social spiders. In this model, the searching space can be identified according to the best location of spiders. The key benefits of using this technique are as follows:

- Increased convergence speed
- Bets optimal solution with reduced number of iterations
- High accuracy
- Optimal performance efficiency

In this model, the total number of female spiders are maximum in the communication, when compared to the make spiders. Then, its count is randomly selected based on the least proportion value as shown in below:

$$FS_C = [(0.9 - rand * 0.25) * L_P] \tag{2}$$

Where, FS_C is the female spiders, L_P indicates the least proportion value, and *rand* denotes the random value lies in the range of [0, 1]. Consequently, the male spiders in the community are computed by using the following model:

$$MS_C = L_P - FS_C \tag{3}$$

After that, the weight value is estimated based on the fitness value of all spiders as shown in below:

$$\omega_e = \frac{Fi_s - W_F}{B_F - W_F} \tag{4}$$

Where, ω_e indicates the weight value, Fi_s denotes the fitness value of spider, W_F is the worst fitness value and B_F is the best fitness value of whole population. Furthermore, the tiny vibrations are estimated based on the weight and distance values of spiders as represented in below:

$$V_{s,t} = \omega_e e^{-k_{s,t}^2} \tag{5}$$

Where, $V_{s,t}$ denotes the vibrations between the spider s and t, ω_e is the estimated weight value, $k_{s,t}$ indicates the estimated distance value of spiders *s* and *t*. Moreover, the fitness is computed by using the cross-fold validation model represented in below:

$$Fi_s = 1 - CF_{Va} \tag{6}$$

$$CF_{Va} = 1 - \frac{1}{10}\sum_{i=1}^{10} |acc| \times 100 \tag{7}$$

Based on the fitness value, the optimal solution is identified for selecting the best features used for training the classifier to detect the insurance frauds.

C. Progressive Kernel Ridge Regression (PKRR) Classification

During the classification process, the optimal number of selected features obtained from the previous stage are considered as the inputs for processing. Then, these features are used to train the classifier model for accurately detecting the insurance frauds from the given dataset, which is accomplished by using the Progressive Kernel Ridge Regression (PKRR) model. It is a kind of machine learning classification approach and, mainly used for detecting the automobile insurance frauds from the desired datasets. The main operations involved in this mechanism are kernel estimation, linear dependency computation, weight value estimation, and output label prediction. In this technique, the dependencies between the covariance and response variables have been estimated for computing the linear function. For the given set of training samples, $\{(a_i, b_i) | a_i \in X^N, b_i \in Y^M, i = 1,2, \dots S\}$, the ordinary least square model is applied to reduce the loss of factor as represented in below:

$$\sum_i (b_i - \rho_i a_i)^2 \tag{8}$$

Where, b_i indicates the row vector, ρ_i is the least square value, and a_i is the training sample. Then, the ridge regression is performed to reduce the cost consumption and squared error values as represented by using the following model:

$$R(\rho) = \sum_i (b_i - \rho^T a_i) + \frac{||\rho||^2}{\sigma} \tag{9}$$

Where, σ indicates the regularization parameter that helps to minimize the estimate variance by maintaining the tradeoff between the bias and variance of the estimated value. In this model, the cross validation is also performed to identify the optimal parameters used for reducing the error values. By using this technique, the multivariate label has been predicted according to the kernel function. After estimating the linear dependency, the variance of estimated can be minimized with the total cost function by using equation (9). Consequently, the total derivatives are estimated by using the following model:

$$\rho' = \left(AB^T + \frac{1}{\sigma}\right)^{-1} AB^T \tag{10}$$

After that, the kernel matrix is constructed based on the mercer condition as shown in below:

$$\begin{cases} \tau_{NR} = RM^T \\ \tau_{NR_{c,d}} = a_c \cdot a_d = K(a_c, a_d) \end{cases} \tag{11}$$

Here, the kernel matrix τ can be replaced with value of random matrix RM^T, then the input samples are mapped with the kernel function according to the input of feature space. Subsequently, the kernel function of the ridge regression model is estimated by using the following model:

$$K(\mu, vec) = \exp(-\delta ||\mu - vec||^2) \tag{12}$$

Where, δ defines the kernel parameter, and its output function of ridge regression is estimated as follows:

$$o(x) = \begin{bmatrix} K(a, a1) \\ \vdots \\ K(a, an) \end{bmatrix} \left(\frac{1}{\sigma} + \tau NR\right)^{-1} Z^T \tag{13}$$

$$\rho^* = \left(\frac{1}{\sigma} + \tau NR\right)^{-1} Z^T \tag{14}$$

Based on this output kernel function and parameter, the output predicted label has been produced as whether normal or fraud. The key benefits of the proposed ISSO-PKRR based automobile insurance fraud detection system are listed as follows:

- Increased classification accuracy
- Reduced time consumption for data training and testing due to optimization
- Minimized error value and false positives
- Minimal computational complexity
- High efficiency in operation

4. Result Discussion

This section evaluates the performance analysis of the proposed ISSO-PKRR based automobile insurance fraud detection system by using various measures. For validating this system, the automobile insurance dataset has been utilized in this work, and some of the existing detection systems are compared with the proposed model during evaluation by using the measures of sensitivity, specificity, accuracy, precision, recall, error rate, and similarity coefficients. In which, the sensitivity, specificity, and accuracy are the most widely used performance measures for validating the efficiency of classification. The increased values of these measures assure the better classification performance of the detection system. Consequently, the other measures precision, recall, f1-score, and similarity coefficients are also used for testing the overall effectiveness and accurateness of the detection techniques, which are calculated by using the following models:

$$Sensitivity = \frac{TP}{TP + FN} \tag{15}$$

$$Specificity = \frac{TN}{TN + FP} \tag{16}$$

$$\Pr ecision = \frac{TP}{TP + FP} \tag{17}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

$$F1\_Score = \frac{2 \times \Pr ecision \times \operatorname{Re} call}{\Pr ecision + \operatorname{Re} call} \tag{19}$$

$$\Pr ecision = \frac{TP}{TP + FP} \tag{20}$$

$$Recall = \frac{TP}{TP + FN} \tag{21}$$

$$F1\_Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{22}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{23}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{24}$$

$$Error\_Rate = 1 - Accuracy \tag{25}$$

$$Kappa\_Coeff = \frac{P_o - P_e}{1 - P_e} \tag{26}$$

Table 1: Sensitivity, specificity, and accuracy analysis

Techniques	***Sensitivity***	***Specificity***	***Accuracy***
MLP	49.98	28.75	74.98
KNN	71.5	41.26	69.02
SVM	90.53	38.92	58.41
DT	92.27	58.35	79.24
ISSO-PKKR	95.68	97.8	98.5

Table 2: Training time of existing and proposed classification techniques

Techniques	***Training time (ms)***
DT	471
NB	155
KNN	1254
XGB	995
ISSO-PKKR	985

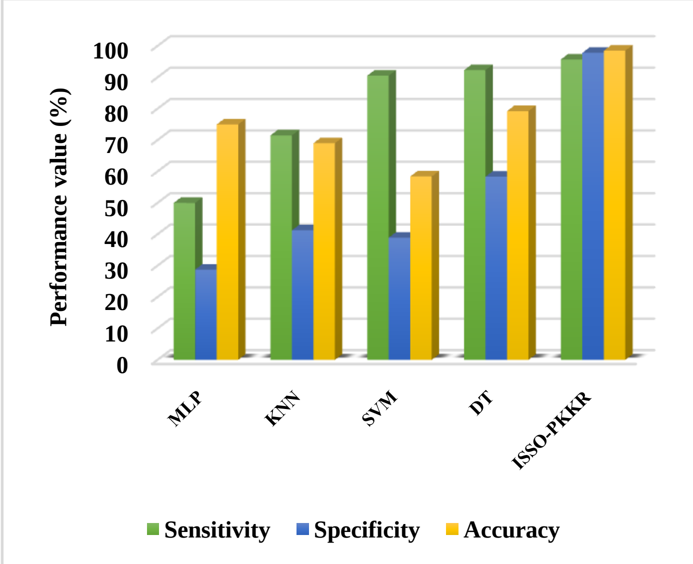

Fig 2. Comparative analysis based sensitivity, specificity, and accuracy

Fig 2 and Table 1 compares the existing[24] and proposed classification approaches used for an automobile insurance fraud detection based on the measures of sensitivity, specificity, and accuracy. The existing techniques considered for this analysis are Multilayer Perceptron (MLP), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Decision Tree (DT). Based on the obtained results, it is analyzed that the proposed the ISSO-PKKR technique outperforms the other approaches with increased sensitivity, specificity, and accuracy values by exactly detecting the insurance frauds from the given datasets.

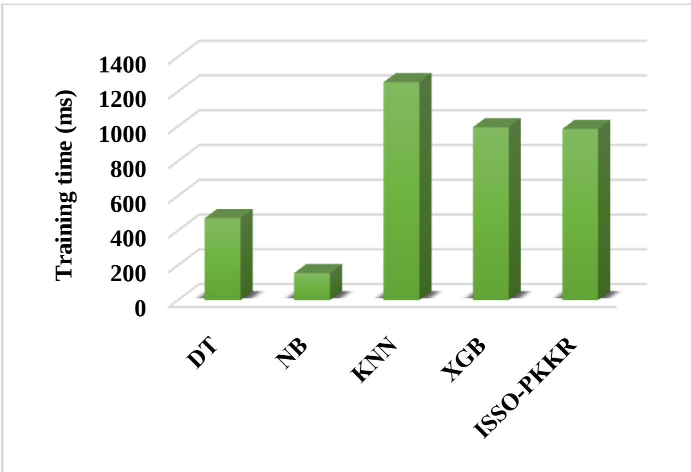

Fig 3. Comparative analysis based on training time

Fig 3 and Table 2 compares the training time of both existing [27] and proposed insurance fraud detection techniques, which includes Decision Tree (DT), Naïve Bayes (NB), K-Nearest Neighbor (KNN), and Extreme Gradient Boost (XGB). Typically, the training time is defined as the amount of time required for training the model of classifier. When compared to the other approaches, the proposed ISSO-PKKR technique requires the reduced amount of time consumption. Because, the proposed mechanism selects the optimal number of features based on the best fitness value, which are further used for training the classifier.Similar to that, the precision, recall, accuracy and f1-score of both existing and proposed classification models are validated as shown in Fig 4 and Table 3. Normally, the overall improved performance of the detection and classification technique is determined based on the increased values of these measures. According to the obtained results, it is analyzed that the proposed ISSO-PKKR technique outperforms the other techniques with the increased values of these measures. Because, the distance based clustering model helps to normalize the unbalanced data by performing the missing values prediction, and irrelevant attributes removal. Then, the clustering has been applied on the normalized dataset for grouping the information, and this type of clustering based optimization and classification approach could efficiently increase the performance of detection system.

Table 3:Precision, recall, accuracy and f1-score of existing and proposed techniques

Techniques	*Accuracy (%)*	*Precision (%)*	*Recall (%)*	*F1-score (%)*
DT	86.99	81	86	83
NB	52.06	37.3	52	42.5
KNN	42.70	22.3	42.7	25.5
XGB	99.25	99.28	99.2	99.26
ISSO-PKKR	99.5	99.3	99.3	99.35

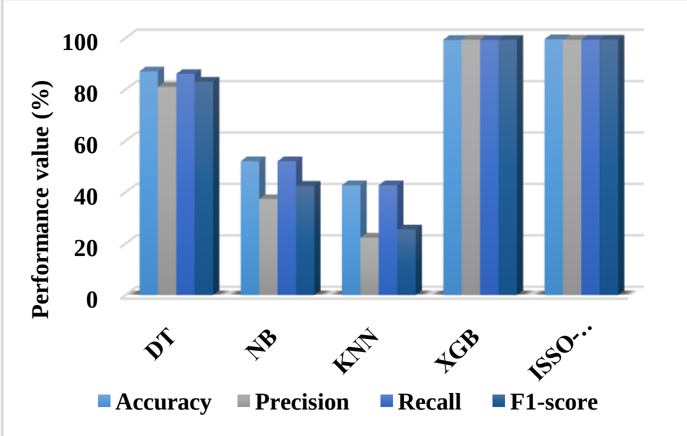

Fig 4. Comparative analysis based on accuracy, precision, recall and f1-score

Table 4: Analysis on Prediction accuracy and Error rate

Techniques	*Prediction accuracy*	*Error rate*
BP	77.78	0.130
GA-BP	88.89	0.040
IAGA-BP	100	0.020
NAGA-BP	100	0.010
ISSO-PKKR	100	0.009

Table 4 compares the prediction accuracy and error rate of existing [29] and proposed classification techniques, and its corresponding graphical representation are shown in Fig 5 and 6 respectively. The prediction accuracy of classifier is determined based on the amount of samples that are exactly predicted as the fraudulent. Similar to that, the error rate is estimated according to the number of samples that are incorrectly predicted by the detection technique. From this analysis, it is evident that the proposed ISSO-PKKR provides the reduced error rate and increased prediction accuracy values, when compared to the other techniques. By using an efficient optimization based classification process, the proposed detection could efficiently reduce the error rate.

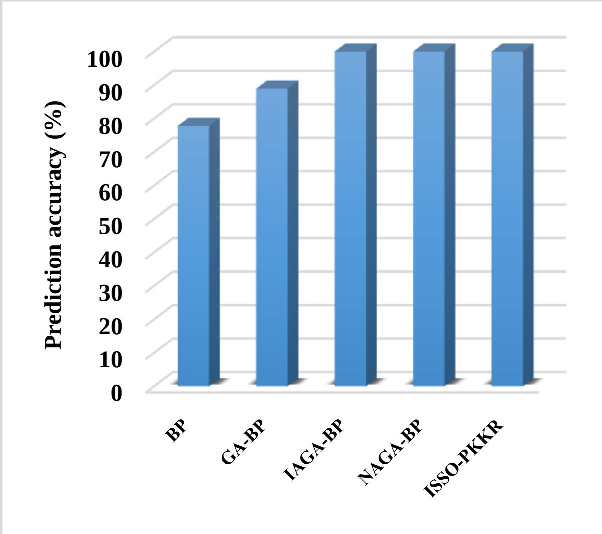

Fig 5. Comparative analysis based on prediction accuracy

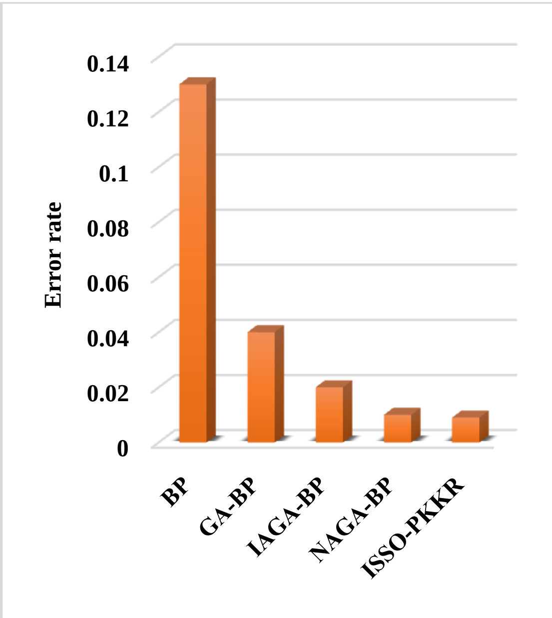

Fig 6. Comparative analysis based on error rate

5. Conclusion

This paper presented an intelligent optimization based classification framework for accurately detecting the automobile insurance frauds.The major stages involved in this framework are data preprocessing, clustering, optimization, and classification. After normalization, the data clustering can be performed for grouping the attributes of information based on the distance value by using the FDC approach, which also helps to simplify the process of classification by grouping the data into the form of clusters. Typically, the high number of features require more time for training and detection, which also reduces the efficiency of the system. Therefore, the ISSO technique is used to optimally select the best features based on the estimated global optimum value. Then, the selected optimal number of features are fed to the classifier for training the data model. Finally, the PKKR classification approach is used to predict the classified label based on the optimal feature set. The major benefits of the proposed ISSO-PKRR based automobile insurance fraud detection system are as follows: reduced computational complexity, requires minimal time consumption, increased detection efficiency, high accuracy, and optimal performance rate. For validating the results of this technique, various evaluation measures have been used during performance analysis. Based on the assessment, it is observed that the proposed ISSO-PKKR technique outperforms the other existing techniques with improved performance results.

References

[1] Z. Shaeiri, and S. Kazemitabar, “Fast unsupervised automobile insurance fraud detection based on spectral ranking of anomalies,” *International Journal of Engineering*, vol. 33, no. 7, pp. 1240-1248, 2020.

[2] Y. Wang, and W. Xu, “Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud,” *Decision Support Systems*, vol. 105, pp. 87-95, 2018.

[3] A. Bodaghi, and B. Teimourpour, "Automobile insurance fraud detection using social network analysis," *Applications of Data Management and Analysis*, pp. 11-16: Springer, 2018.

[4] G. Kowshalya, and M. Nandhini, "Predicting fraudulent claims in automobile insurance." pp. 1338-1343.

[5] Y. Li, C. Yan, W. Liu, and M. Li, "Research and application of random forest model in mining automobile insurance fraud." pp. 1756-1761.

[6] M. Salmi, and D. Atif, "Using a Data Mining Approach to Detect Automobile Insurance Fraud." pp. 55-66.

[7] Y. Li, C. Yan, W. Liu, and M. Li, “A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification,” *Applied Soft Computing*, vol. 70, pp. 1000-1009, 2018.

[8] K. G. Al-Hashedi, and P. Magalingam, “Financial fraud detection applying data mining techniques: A

comprehensive review from 2009 to 2019," *Computer Science Review*, vol. 40, pp. 100402, 2021.

[9] X. Zhou, S. Cheng, M. Zhu, C. Guo, S. Zhou, P. Xu, Z. Xue, and W. Zhang, "A state of the art survey of data mining-based fraud detection and credit scoring." p. 03002.

[10] B. Zewdu, and G. Belay, "Demystifying Predictive Analytics with Data Mining to Optimize Fraud Detection in the Insurance Industry." pp. 432-442.

[11] C. Yan, J. Liu, W. Liu, and X. Liu, "Research on automobile insurance fraud identification based on fuzzy association rules," *Journal of Intelligent & Fuzzy Systems*, no. Preprint, pp. 1-14, 2021.

[12] K. Parnian, F. Sorouri, A. N. Souha, A. Molazadeh, and S. Mahdavi, "Fraud Detection in Health Insurance Using a Combination of Feature Subset Selection Based on Squirrel Optimization Algorithm and Nearest Neighbors Algorithm Methods," 2019.

[13] I. M. N. Prasasti, A. Dhini, and E. Laoh, "Automobile Insurance Fraud Detection using Supervised Classifiers." pp. 47-52.

[14] S. Dasgupta, "Detecting Anomalous Insurance Claims with Hybrid Feature Optimisation and Classification Techniques," Dublin, National College of Ireland, 2019.

[15] S. Misra, S. Thakur, M. Ghosh, and S. K. Saha, "An autoencoder based model for detecting fraudulent credit card transaction," *Procedia Computer Science*, vol. 167, pp. 254-262, 2020.

[16] X. Liu, J.-B. Yang, and D.-L. Xu, "Fraud detection in automobile insurance claims: A statistical review." pp. 1003-1012.

[17] G. Jaya Krishna, and V. Ravi, "Anomaly detection using modified differential evolution: An application to banking and insurance." pp. 102-111.

[18] H. Xia, Y. Zhou, and Z. Zhang, "Auto insurance fraud identification based on a CNN-LSTM fusion deep learning model," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 39, no. 1-2, pp. 37-45, 2022.

[19] B. Itri, Y. Mohamed, Q. Mohammed, and B. Omar, "Performance comparative study of machine learning algorithms for automobile insurance fraud detection." pp. 1-4.

[20] S. K. Majhi, "Fuzzy clustering algorithm based on modified whale optimization algorithm for automobile insurance fraud detection," *Evolutionary intelligence*, vol. 14, no. 1, pp. 35-46, 2021.

[21] S. Subudhi, and S. Panigrahi, "Use of optimized Fuzzy C-Means clustering and supervised classifiers for automobile insurance fraud detection," *Journal of King Saud University-Computer and Information Sciences*, vol. 32, no. 5, pp. 568-575, 2020.

[22] S. Subudhi, and S. Panigrahi, "Two-Stage Automobile Insurance Fraud Detection by Using Optimized Fuzzy C-Means Clustering and Supervised Learning," *International Journal of Information Security and Privacy (IJISP)*, vol. 14, no. 3, pp. 18-37, 2020.

[23] X. Liu, J.-B. Yang, D.-L. Xu, K. Derrick, C. Stubbs, and M. Stockdale, "Automobile insurance fraud detection using the evidential reasoning approach and data-driven inferential modelling." pp. 1-7.

[24] S. Padhi, and S. Panigrahi, "Decision Templates based Ensemble Classifiers for Automobile Insurance Fraud Detection." pp. 1-5.

[25] P. S. Pranavi, H. Sheethal, S. S. Kumar, S. Kariappa, and B. Swathi, "Analysis of Vehicle Insurance Data to Detect Fraud using Machine Learning," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 8, no. 7, pp. 2033-2038, 2020.

[26] C. Yan, M. Li, and W. Liu, "Improved ELM optimization model for automobile insurance fraud identification based on AFSA," *Intelligent Data Analysis*, vol. 23, no. S1, pp. 67-85, 2019.

[27] N. Dhieb, H. Ghazzai, H. Besbes, and Y. Massoud, "Extreme Gradient Boosting Machine Learning Algorithm For Safe Auto Insurance Operations," *IEEE Access*, vol. 8, pp. 58546-58558, 2020.

[28] J. I.-Z. Chen, and K.-L. Lai, "Deep convolution neural network model for credit-card fraud detection and alert," *Journal of Artificial Intelligence*, vol. 3, no. 02, pp. 101-112, 2021.

[29] C. Yan, M. Li, W. Liu, and M. Qi, "Improved adaptive genetic algorithm for the vehicle Insurance Fraud Identification Model based on a BP Neural Network," *Theoretical Computer Science*, vol. 817, pp. 12-23, 2020.