Detection of credit card fraud by applying genetic algorithm and particle swarm optimization

Debachudamani Prusti[1], Jitendra Kumar Rout[2], Santanu Kumar rath[1]

[1] Department of CSE, NIT Rourkela, India
[2] Department of CSE, NIT Raipur, India
debaprusti@gmail.com, jitu2rout@gmail.com, skrath@nitrkl.ac.in

Abstract. Fraudulent activities associated with the credit card is a pertinent problem often occurring in a global level. The customers are losing their trust with the financial institutions and the financial institutions are in a difficult state to win the goodwill of customers. A substantial number of researchers show interest to work on fraud detection in order to develop an optimized method or model to identify the fraudulent activities that are happening in a regular and continuous form with the credit card in our everyday life. Genetic algorithm (GA) and the potential solution-based particle swarm optimization (PSO) are two optimization algorithms, which can be considered along with the neural network to analyze the possible fraudulent transactions. The optimization algorithms help to make the learning process faster and optimized with a superior and better predictive accuracy value. The PSO based neural network has been trained thoroughly and performance values are compared with GA based neural network, by increasing the number of iterations and the population or number of swarms. It has been observed that algorithm based on PSO gives an optimized result for fraudulent transaction detection.

Keywords: Credit card fraud detection, GA, PSO, Neural Network, Predictive Performance.

1 Introduction

Online financial transaction facilities help in eliminating the burden of completing a transaction without visiting the onsite branch of a bank [1] [2]. The usage of online transaction services viz., credit card transactions attract more number of users for a hassle free online experience, but simultaneously it also attracts the fraudulent customers with an intention of hijacking money from other account holders. The malicious users or the fraudsters pretend themselves as authorized users to steal the account as well as transaction information in an intelligent manner.

In this proposed research, attempt has been made to detect frauds in credit card transactions, so that the transactions can occur in a secured manner [3] [4]. Detection of malicious behavior is a subtle problem to find the original identity of card users, attempting to intrude into the online credit card transactions by customers. A good approach to identify the malicious activities, which reside among various credit card

transactions, is to find the divergence that happens with continuous and frequent data transaction. Different approaches are applied in the past research methodologies with various techniques to identify the divergent behavior i.e., supposed to be analyzed for better identification of the fraudulent transaction behaviors [5].

In this proposed study, the credit card frauds are analyzed by considering the financial transaction data of several financial institutions associated with the credit cards. In the dataset, the fraudulent and legitimate classes are highly imbalanced. This higher imbalance between the two classes prompts the fraud detection system a very challenging job. Fraud detection technique can be conceptualized as a data mining technique with an objective to correctly predict the classification of online transactions as genuine or fraudulent. Basically, for classification problems, a good number of performance measures are specified and most are associated with correct number of cases, which can able to classify correctly.

Application of optimization techniques such as GA and PSO to detect the fraud in credit card transactions have been considered in this proposed approach [6] [7] [8]. The optimization techniques help to find various parameters, which is likely to be maximum or minimum (optimal) value of any target function. Optimization techniques are often applied with artificial neural network and SVM to evaluate the performance parameters for classification algorithms, since they provide certain coefficients that are often identified by trial and error method or by using the exhaustive search method.

Motivation of the study. This study is motivated to improve a credit card fraud detection solution that can optimize the chance of identifying the fraudulent transaction with better performance. Although the solution has been regarded as a successful one, still we consider that it can further be improved by taking the weighted values that can be better adjusted by considering the recent transaction behaviors and frauds occurred into it. The optimization methods such as GA and PSO help to improve the optimal values from the input by taking the various weight parameters.

2 Literature survey

It has been observed in the literature that a good number of researchers as well as practitioners have presented their studies on fraud detection using several data mining and machine learning algorithms. The research includes random forest, support vector machine, regression analysis, artificial neural network applications, as these algorithms are very much helpful to classify the legitimate and fraudulent activities in the financial transactions. Quah and Srinagesh (2008) have suggested a framework for outlier analysis separately for each customer that can be applied in real time, later on a predictive algorithm has been applied to classify the suspicious transactions. Panigrahi et al. (2009) have suggested the methodology for fraud detection based on four components and are connected serially. A set of abnormal and suspicious transactions are initially identified and later on Bayesian learning algorithm is considered to predict the fraudulent transaction.

Sanchez et al. (2009) have represented several approaches and have applied association rule mining (ARM) method to characterize new card patterns for regular usage and identifying the ones that are not fitted to the suspicious patterns. The survey of Bolton and Hand (2002) have provided a brief description of literature related to fraudulent transaction detection issues. Mitchell (1998) has commented about genetic algorithms in terms of evolutionary algorithms with an objective to obtain better solutions against the increase in time. Aote et al. (2013) have presented a detailed work on application of PSO along with its limitations. Bratton and Kennedy (2007) defined a standard PSO algorithm with the recent developments that help to improve the performance on standard measures to extend original PSO.

3 Artificial Neural Network application

Artificial neural network (ANN) works in the similar way as a human brain does and it can be very well considered for detection of fraudulent transactions [9] [10] [11].

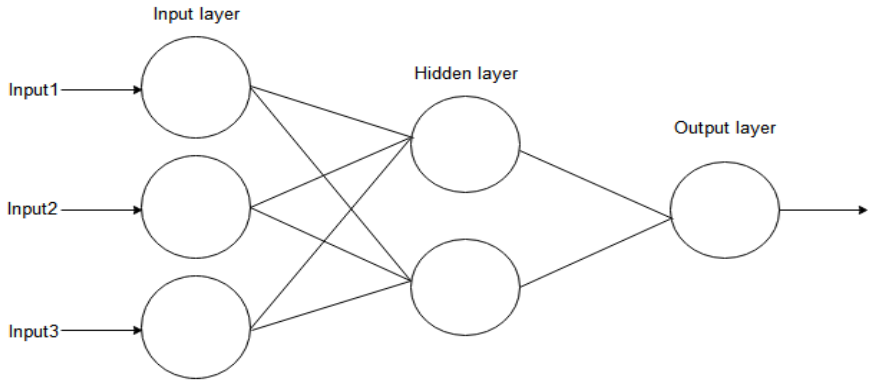

Fig. 1. Layers of neural network with single hidden layer

By using the previous one or two years data of various transactions, neural network is trained to identify a fixed pattern of using credit card by a specific customer. As shown in Fig. 1, the neural networks are trained in a multi-layered architecture style having one input layer, one output layer and at least one hidden layer. Except the input nodes, other nodes or neurons behave as activation function. It helps in mapping the input signals with the response variables.

4 Genetic Algorithm

The Genetic Algorithm as shown in Fig. 2 follows the procedure being inspired from the natural evolution [3] [12]. The whole objective is that with the evolution of generation, the survival chance of a stronger member in the population is greater than that of the weaker members. Starting with a number of initial given solutions, the genetic

algorithm acts as the parent of current generation evolution. Crossover and mutation operators generate new solutions from these solutions. The unfit members are eliminated from the generation and more fit members are selected for the next generation as the parents. This process continues until a number of generations have passed and subsequently the best solution is obtained. But in some cases, genetic algorithm does not give any guaranty for identifying the global maxima and also there is chances to be trapped in local maxima.

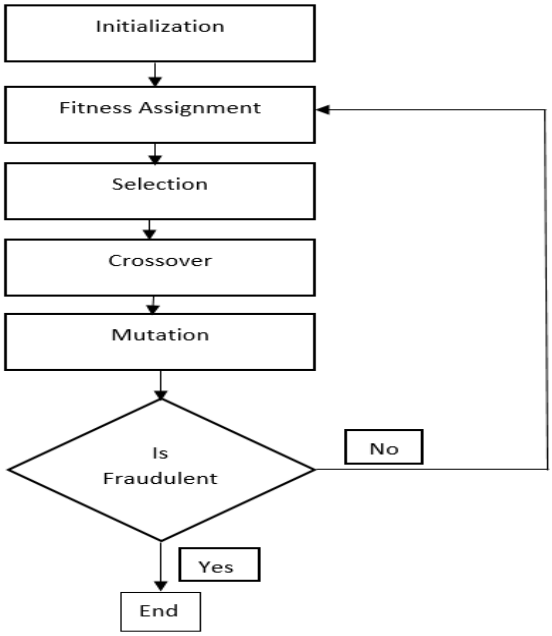

Fig. 2. Steps for applying genetic algorithm

GA with Neural Network. Hybridization of genetic algorithms with artificial neural networks present a better performance where GA is used to find the various performance parameters [12]. The main objective is how accurately GA and ANN can be hybridized, i.e., how the neural network should be represented with the genetic algorithm for a better predictive result.

In the initialization step, a large number of random individuals are generated to begin the algorithm procedure. Then the empirical values of the parameters are evaluated, by applying artificial neural network according to the genome information. After training with back-propagation, its performance is determined. Rather considering the individual's performance, the fitness evaluation considers a greater number of cases. In order to generate small networks, few approaches consider about the network size for the better evaluation of the parameters. Later, crossover and mutation replace the worst members of the population by creating new individuals. Initial population is generated by randomizing weight matrices rather by randomizing the chromosome strings allowing the initial weights to be distributed in a closed range. The message or

information in the neural network is encoded with neural network algorithm within the genome of the genetic algorithm. Initially, the random individuals are generated and their parameters are evaluated based on the genome information. Finally, its performance values are determined empirically post training with the back-propagation neural network algorithm.

5 Particle Swarm Optimization

Particle Swarm optimization (PSO) is one of the stochastic optimization techniques. Rather inspired by the natural evolution similar to other larger class of evolutionary algorithms like evolutionary strategies, genetic algorithms, genetic programming, PSO is mainly prompted from the stimulation of social behavior of the swarm particles in large group [13] [14]. PSO is based on sociological behavior and mainly inspired sociologically is visualized with bird flocking [15]. It is a kind of evolutionary algorithm similar to others, which is initialized with the population through random solutions.

The algorithm holds a swarm of particles, in which every particle tries to solve an optimization problem by providing a potential solution to the problem. Unlike other evolutionary algorithms, in PSO the individual's potential solution are passed through the problem space [16]. Let *S* be the swarm size and the particle *i* has several characteristics for obtaining the particle solution at each iteration. A swarm of particles is first initialized with the random position x_i and velocity v_i and the objective function *f(x)* is calculated by considering the particles coordinates as input measures. The disadvantage of PSO is that, it loses swarm diversity with low convergence rate during the iteration process.

5.1 Personal best (p-best)

The p-best or particle best position is the individual best position *Pi*, of particle *i*. It is the best position of a particle that it visits (Prior value of x_i) and yields the fitness value which is regarded as highest. For any minimization method, if the position yields the smaller function value, then it is considered as having highest fitness. It denotes *f(x)* as the objective function, should be minimized for the particle.

5.2 Global best (g-best)

The robustness of g-best or global best is, it provides a faster rate of convergence at lower expense. This g-best particle has only a single best solution known as global best particle *Pg* among all the particles in the population or swarm. It behaves as an attractor and helps to pull the particles in the group towards it. Slowly all particles are being converged to g-best position. The swarm may converge in a premature manner unless updated regularly.

PSO algorithm mainly consists of three steps such as initialization, velocity updating and position updating.

Initialization. A swarm of particles are initialized randomly with positions and velocities in the problem space. Once the lower and upper limits of the decision variable are specified, the search space is getting confined. The coordinates are initialized with both positions as well velocities within a certain permissible range by fulfilling both equality and inequality constraints.

Velocity updating. The velocity of each particle in PSO is calculated using the distance traveled by the particle. It depends on the particle memory that is the previous best position and the swarm memory, which is the previous best solution. The velocity of each particle is updated by using the velocity updation equation by considering the particle memory and swarm memory.

$$v_i(t+1) = wv_i(t) + c_1r_1[\hat{x}(t) - x_i(t)] + c_2r_2[g(t)- x_i(t)] \quad (1)$$

Where,

i is considered for particle index
w is the inertia weight to balance the local and global coefficient
c_1 and c_2 are considered as acceleration coefficients
r_1 and r_2 are random values generated with every velocity updation
$x_i(t)$ is the particle's position at time t
$\hat{x}(t)$ is the particle's individual best solution as of time t
$v_i(t)$ is the particle's velocity at time t
$g(t)$ is the swarm's best solution as of time t

Position updating. Between the successive iterations, the coordinates of all particles are updated according to the given equation:

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (2)$$

Where, $x_i(t+1)$ is the new position, $x_i(t)$ is the previous position and $v_i(t+1)$ is the new velocity.

PSO with Neural Network. PSO has a vast application area in evolutionary system to evolve artificial neural networks and other classification methods based on PSO algorithm [17] [18]. PSO-NN coordinates the architecture and the weights of neural network as shown in Fig. 3. PSO algorithm is used to predict the positioning errors, caused due to their geometric parameters [19]. A hybrid approach of PSO and NN training has been proposed in this research study for a better predictive result by considering different performance parameters for both training and testing. PSO based NN is applied on the neurons of the neural network to optimize the parameter values and it helps to minimize the mean square error (MSE) iteratively. PSO helps for optimizing the weight matrices of the neural network and it is used to produce an output through an axon to another neuron. A correctly trained neural network is considered as an expert in categorizing the information that to be analyzed.

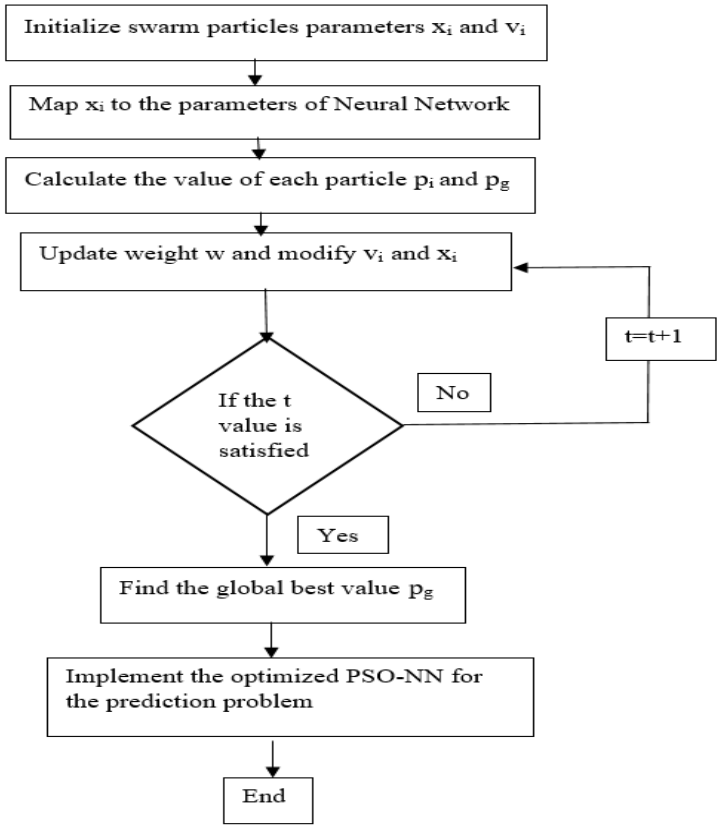

Fig. 3. PSO based neural network learning algorithm

6 Result and Performance analysis

To detect the fraudulent transactions, we have implemented both GA based neural network as well as PSO based neural network to find the prediction accuracy values and other performance parameters.

6.1 Dataset used for experiment

The optimal usage of dataset is a prior requirement to conduct the classification methodology. The size and volume of dataset most probably affect both training as well as testing data. Fraudulent classification data has been applied for the proposed classification model with optimization techniques has been retrieved from Kaggle website (https://www.kaggle.com/mlgulb/creditcardfraud). The dataset has a dimension of 31 columns and 284807 rows. From it, 70% data instances are applied for training and remaining 30% for testing purpose. The predictive accuracy value and other evaluation metrics have been optimized with 70% training data and 30% testing data.

6.2 Experimental setup

GA based neural network and PSO-NN optimization algorithms have been implemented in Matlab platform version R2019a. During the implementation, the system configuration was noted as core i7 processor and 3.4 GHz clock speed. The secondary and main memory space were 1TB and 8GB respectively.

6.3 Confusion matrix

A confusion Matrix helps to represent various evaluation metrics in a classification model. It shows the correct and incorrect classification samples with actual and predictive results in the test data. It is designed to count the number of all four results for the two-class classification and denoted as true positive, false positive, true negative and false negative.

6.4 Performance Parameters

Different performance metrics such as accuracy, sensitivity, precision, F-measure, specificity and mean square error (MSE) are evaluated by using the values in confusion matrix. The parameter values for the PSO based neural network are compared by setting the iteration values and the predictive accuracy value is calculated and the mean square error value is observed with PSO base neural network.

We have implemented Genetic algorithm and PSO with neural network technique and their performance parameters are critically analyzed. In the GA based neural network the predictive accuracy is observed to be 89.91%. The genetic algorithm simulation result as shown in Fig. 4 represents the maximum instances lies between -5 to +5 values and few instances are sparse away.

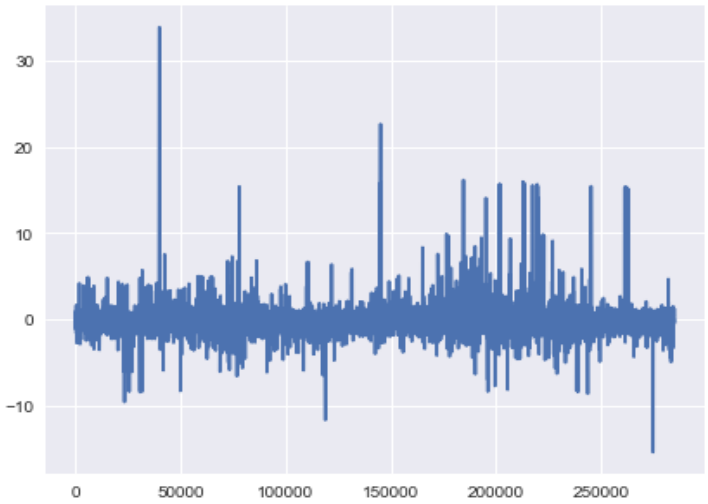

Fig. 4. Simulation result by implementing genetic algorithm

In the PSO based neural network, various performance parameters are critically evaluated by considering five phases with five different iterations and the mean square error is observed for each phase. We have considered the population or swarm size 100.

In Table 1, the experimental values of various performance parameters have been presented. They have been assessed by considering 70% and 30% of training and testing data respectively. The performance measures are critically assessed for accuracy, sensitivity, specificity, precision, F-measure and mean square error. We have considered five different phases to find the result of various parameters. The prediction accuracy for phase 5 is observed to be 91.58%, which is significantly improved comparing to other four phases. The mean square error has been reduced in phase 5 and observed to be 0.67%. In PSO based neural network technique, the optimized value of predictive accuracy with other performance measures have been achieved with reduced false alarm.

Table 1. Performance results for PSO based Neural Network

Performance parameters	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5
Accuracy	90.01	90.21	90.34	91.13	91.58
Sensitivity	91.11	90.06	91.05	90.00	91.69
Specificity	90.07	93.84	90.75	91.78	92.81
Precision	98.92	96.01	96.35	98.09	97.93
F-measure	95.29	94.85	94.92	95.36	95.50
MSE	5.46	3.03	1.45	1.07	0.67

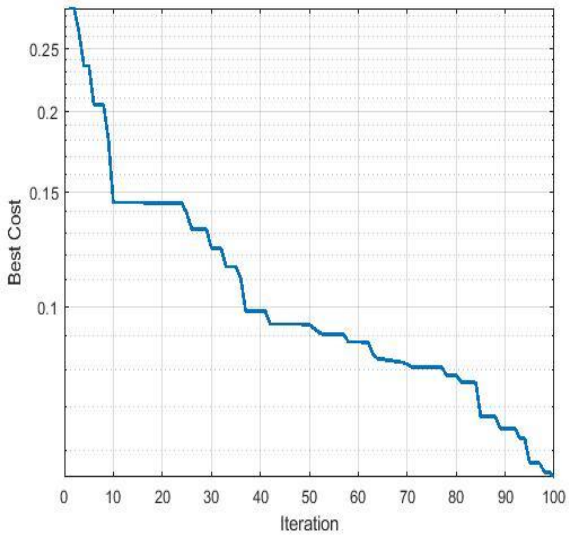

Fig. 5. Best cost is calculated with 100 iterations

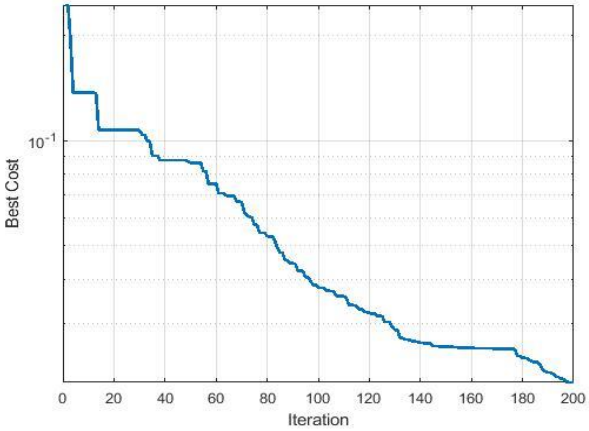

Fig. 6. Best cost is calculated with 200 iterations

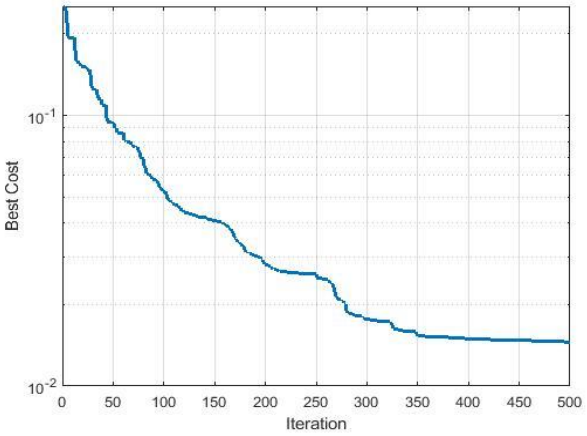

Fig. 7. Best cost is calculated with 500 iterations

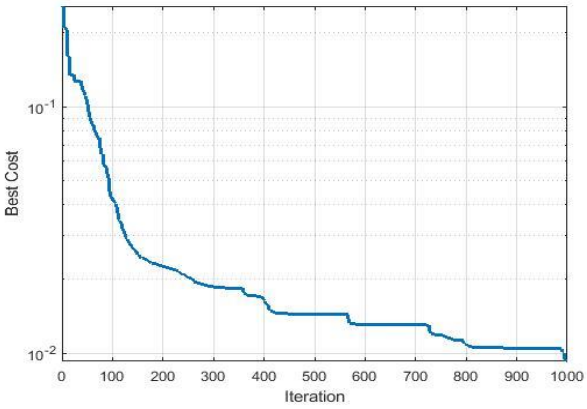

Fig. 8. Best cost is calculated with 1000 iterations

By using the PSO based neural network the best cost is calculated in terms of mean square error value. From Fig. 5 to Fig. 9, the X-label shows the total number of iterations and the Y-label shows the best cost value. In phase 1, phase 2, phase 3, phase 4, phase 5, we have taken the iterations 100, 200, 500, 1000 and 2000 respectively having the population or swarm size 100. We observed that the best cost value in terms of mean square error value is reduced with more number of iterations. In phase 5, the MSE value is observed to be 0.67% when 2000 number of iterations are taken into consideration. Also, the predictive accuracy value is significantly increased and is observed to be highest i.e., 91.58% when 2000 number of iterations are considered.

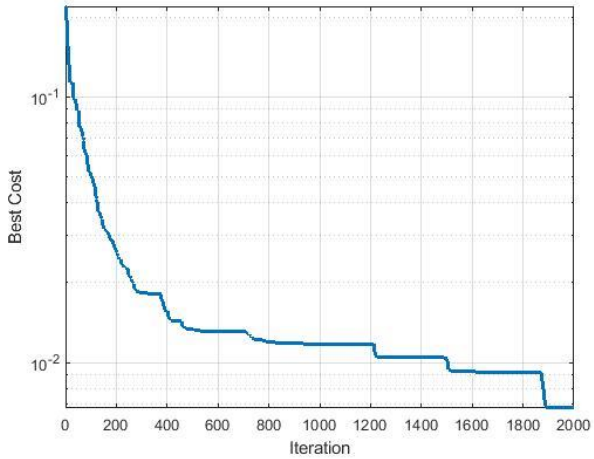

Fig. 9. Best cost is calculated with 2000 iterations

7 Conclusion

In this proposed study, GA and PSO are employed along with neural network to make the learning process faster. The credit card fraud detection technique will be more efficient when the GA and PSO algorithms use the machine learning classification technique such as neural network. Among the GA based neural network and PSO based neural network, the latter has an improved prediction accuracy of 91.58%, when 2000 iterations are considered. The best cost is optimized with increasing the number of iterations and the mean square error value has been reduced to 0.67%, when the total number of iterations are 2000 considered.

References

1. Behdad, Mohammad, Luigi Barone, Mohammed Bennamoun, and Tim French.: "Nature-inspired techniques in the context of fraud detection." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6), 1273-1290 (2012).
2. Gayathri, C., Umarani, R., "Efficient detection of financial fraud detection by selecting optimal ensemble architecture using optimization approaches." *Indian Journal of Innovations and Developments*, 4(8), 1-9 (2015).

3. Gadi, Manoel Fernando Alonso, Xidi Wang, and Alair Pereira do Lago. "Credit card fraud detection with artificial immune system." In *International Conference on Artificial Immune Systems*, pp. 119-131. Springer, Berlin, Heidelberg (2008).
4. Prusti, Debachudamani.: "Efficient intrusion detection model using ensemble methods." PhD diss., (2015).
5. Prusti, Debachudamani, SS Harshini Padmanabhuni, and Santanu Kumar Rath. "Credit card fraud detection by implementing machine learning techniques." In *Safety, Security, and Reliability of Robotic Systems*, pp. 205-216. CRC Press (2020).
6. Duman, Ekrem, and M. Hamdi Ozcelik. "Detecting credit card fraud by genetic algorithm and scatter search." *Expert Systems with Applications* 38(10), 13057-13063 (2011).
7. Alam, Shafiq, Gillian Dobbie, Patricia Riddle, and M. Asif Naeem. "A swarm intelligence based clustering approach for outlier detection." In *IEEE Congress on Evolutionary Computation*, pp. 1-7. IEEE (2010).
8. Shahreza, M. Lotfi, D. Moazzami, B. Moshiri, and M. R. Delavar. "Anomaly detection using a self-organizing map and particle swarm optimization." *Scientia Iranica* 18(6), 1460-1468 (2011).
9. Brause, R., T. Langsdorf, and Michael Hepp. "Neural data mining for credit card fraud detection." In *Proceedings 11th International Conference on Tools with Artificial Intelligence*, pp. 103-106. IEEE (1999).
10. Prusti, D., and Rath, S. K.: "Fraudulent transaction detection in credit card by applying ensemble machine learning techniques." In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1-6. IEEE, (2019).
11. Ghosh, Sushmito, and Douglas L. Reilly. "Credit card fraud detection with a neural-network." In *System Sciences, 1994. Proceedings of the Twenty-Seventh Hawaii International Conference on*, vol. 3, 621-630. IEEE (1994).
12. Patidar, R., and Sharma L.: "Credit card fraud detection using neural network." *International Journal of Soft Computing and Engineering (IJSCE)* 1, 32-38 (2011).
13. Zhang, Yudong, Shuihua Wang, Lenan Wu, and Yuankai Huo. "PSONN used for Remote-Sensing Image Classification." *Journal of Computational Information Systems*, 6(13), 4417-4425 (2010).
14. Kennedy, J., Eberhart, R. C.: "Particle swarm optimization". Proceedings of IEEE International Conference on Neural Networks, Piscataway, NJ, pp.39-43 (1995).
15. https://shodhganga.inflibnet.ac.in/bitstream/10603/181389/12/12_chapter%204.pdf, "Particle Swarm Optimization".
16. Elías, Arturo., Alberto Ochoa-Zezzatti, Alejandro Padilla, and Julio Ponce.: "Outlier analysis for plastic card fraud detection a hybridized and multi-objective approach." In *International Conference on Hybrid Artificial Intelligence Systems*, pp. 1-9. Springer, Berlin, Heidelberg, (2011).
17. Das, M., Taylan, and L. Canan Dulger. "Off-line signature verification with PSO-NN algorithm." In *2007 22nd international symposium on computer and information sciences*, pp. 1-6. IEEE (2007).
18. Zhang, C., Shao, H., Li, Y.: "Particle swarm optimization for evolving artificial neural network", Proceedings of the IEEE International Conference on Systems, Man. and Cybernetics 2000, pp. 2487- 2490, (2000).
19. J.R. Zhang, J. Zhang, T-M. Lok, M.R. Lyu, "A Hybrid particle swarm optimization-back propagation algorithm for feedforward neural network training". Applied Mathematics and Computation, in Press, (2006).