International Journal of
Advances in Scientific Research and Engineering (ijasre)

E-ISSN : 2454-8006

DOI: 10.31695/IJASRE.2019.33512

Volume 5, Issue 9
September - 2019

Hybrid Design using Counter Propagation Neural Network-Genetic Algorithm Model for the Anomaly Detection in Online Transaction

Amusan, D.G, Olabode, A.O, Ojo, O.S, Folowosele, A.O and Oyediran, M.O

Department of Computer Science and Engineering,

Ladoke Akintola University of Technology, Ogbomoso

Nigeria

ABSTRACT

In e-commerce, credit card fraud is an evolving challenge. The increase in the number of credit card transactions provides more opportunity for fraudsters to steal credit card numbers and execute fraud. Fraud detection is a continuously evolving discipline to tackle ever changing tactics to commit fraud. Existing fraud detection systems have not been so much efficient to reduce fraud transaction rate. Improvement in fraud detection practices has become essential to maintain existence of payment system. This research designed hybrid of Counter Propagation Neural Network and genetic algorithm (CPNN-GA) for the detection of anomaly in any online transactions.

Keywords: Anomaly Detection, Counter propagation neural network, Credit card fraud, Genetic algorithm, Model, Online transactions.

1. INTRODUCTION

The increase in the popularity of e-commerce in our daily lives, credit card usages have dramatically increased over the years. Credit card frauds have also been observed to surge as the number of online transactions have increased [1]. Anomaly detection refers to the problem of finding patterns in data that do not conform to expected behavior. Anomalies in credit card transaction data could indicate credit card fraud or identity theft [2; 3]. As the scale of electronic commerce transactions has grown, it has become very attractive to criminals, and the volume of fraudulent e-commerce transactions is growing rapidly. Therefore, there has been an increase in the amount of attention given to the security of the payment systems used to process online transactions [4].

Counter Propagation Neural Network (CPNN) is a multilayer feed-forward Artificial Neural Network (ANN) based on the combination of the input, competitive, and output layers. Model of CPNN is instar-outstar. It is three-layer neural network that performs input-output data mapping, that is, producing output in the response to an input vector on the basis of Competitive Learning [5]. Genetic Algorithms (GA) are computer-based search techniques patterned after the genetic mechanisms of biological organisms that have adapted and flourished in changing highly competitive environment. GA is the solution for optimization of hard problems quickly, reliably and accurately [6].

Fraud detection is a continuously evolving discipline to tackle ever changing tactics to commit fraud and there is need for special methods of intelligent data analysis to detect and prevent it [7]. Existing fraud detection systems have not been so much efficient to reduce fraud transaction rate. Improvement in fraud detection practices has become essential to maintain existence of payment system [8]. To overcome the challenges posed by existing techniques, this paper aim to propose hybrid CPNN for anomaly detection system in online transactions, where GA is integrated for optimal parameter selection so as to provide better solution quality in terms of classification with high fraud catching rate and low false alarm rate. Subsequently, the rest of this paper is organized in the following sections: some reviews on related anomaly detection, methodology of a proposed CPNN-GA system, followed by results and discussion. The final section concludes the paper along with some recommendations for future research.

II. RELATED REVIEW ON CREDIT CARD FRAUD DETECTION

Credit card fraud detection has drawn a lot of research interest and a number of techniques, with special emphasis on neural networks; data mining and distributed data mining have been suggested [9]. The detection of fraud is a complex computational task and still there is no system that surely predicts any transaction as fraudulent. They just predict the likelihood of the transaction to be a fraudulent [10]. In 2000 [11] designed a system based on genetic programming. A Genetic algorithm is used to establish logic rules capable of classifying credit card transactions into suspicious and non-suspicious classes. The result has scalability issue. [12] designed the hidden Markov model (HMM) to detect the credit card fraud. A HMM is initially trained with the normal behaviour of the cardholder. If the current transaction is not accepted by the trained HMM with high probability, it is considered to be fraudulent. [13] designed a methodology and resulting system prototype for fraud detection on credit card transaction data. The detection engine was based on Artificial Neural Networks (ANNs). The ANNs were tuned in three aspects by Genetic Algorithms (GAs), namely in the determination of the optimum set of input factors to the ANN, the determination of the optimum topology of the ANN, and the determination of the optimum weights connecting the ANN neurons. The results of the study investigations were encouraging in that GAs applied to ANNs for credit card fraud detection improved detection engine performance. [14] suggested a behaviour based credit card fraud detection system. Here the historical behaviour pattern of a customer is used to detect fraud. The transaction record of a single credit card is used to build the system. In this system, unsupervised Self organizing map method is used to detect the outliers from the normal ones. [15] designed an outlier mining method to detect the credit card frauds. This system detected outlier sets by computing distance and setting threshold of outliers. It efficiently detected the overdrafts and is also used to predict the fraudulent transactions but it took longer prediction time. [16] investigated the effectiveness of Artificial Immune Systems (AIS) for credit card fraud detection using a large dataset obtained from an on-line retailer. Three AIS algorithms were implemented and their performance was benchmarked against a logistic regression model. The results suggested that AIS algorithms have potential for inclusion in fraud detection systems but that further work is required to realize their full potential in this domain. [17] designed a probabilistic credit card fraud detection system in online transactions. The designed probabilistic based system serves as a basis for mathematical derivation for adaptive threshold algorithm for detecting anomaly transactions. The system was optimized with Baum-Welsh and hybrid posterior-Viterbi algorithms. The results obtained from the evaluation showed the overall average of accuracy and precision are about 84% and about 86% respectively. [18] designed a support vector machine (SVM) learning for credit card fraud detection. The Support Vector Machine based method with multiple kernel involvement, which also includes several fields of user profile instead of only spending profile. The simulation result showed improvement in TP (true positive), TN (true negative) rate, and also decreases the FP (false positive) and FN (false negative) rate.

From this literature, it was observed that previous fraud detection systems have not been so much capable to reduce fraud transaction rates. This paper proposed CPNN being a variant of ANN is tuned by GA to develop an anomaly detection system for online transactions to improve FAR, FRR, miss rate, prediction time, NPV, prediction accuracy and hit rate.

III. METHODOLOGY

The architectures of the designed system are; data preparation and implementation phase. A dataset of one thousand and three hundred (1300) transactions were acquired from thirteen (13) cardholders. Seven hundred and eighty (780) transactions were used for training while five hundred and twenty (520) transactions were used for testing. The accumulated data were prepared and presented in the form acceptable to the designed system (CPNN-GA) with respect to its parameters.

A. Design of CPNN-GA Model for Anomaly Detection

CPNN, a variant of ANN was used for classification due to its capacity for generalization because of its refined network and experimentally proven better learning rate. GA's optimization was integrated into this system in order to optimize the CPNN training parameters so that the best chromosome having optimal parameter setting can be obtained, and used by CPNN for classification purposes. The flowchart depicted the CPNN-GA for anomaly detection system is shown in Fig 1. The system operated in two stages; in the first phase, GA formed clusters. Clustering was done by dot product, while in second phase, the weights between the cluster units and the output units were adjusted. Minimizing error function; error function being the average error incurred when CPNN classifies large input data was considered. Initial weights were randomly selected between 0 and 1, with an assumed initial population size.

DOI: 10.31695/IJASRE.2019.33512

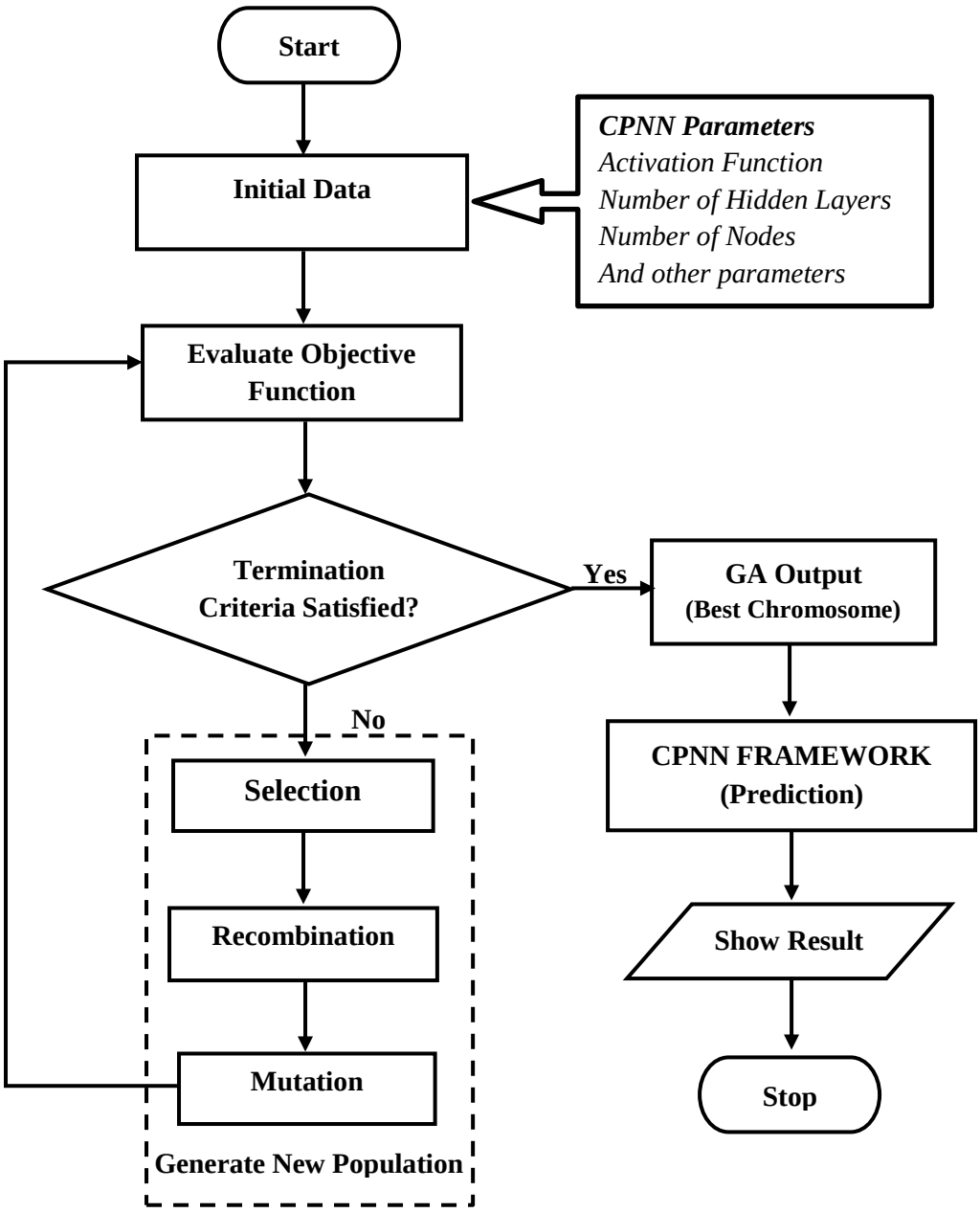

Fig 1: Flowchart showing CPNN-GA Model for Anomaly Detection System

b. Optimization of Features using Genetic Algorithm

Genetic algorithm performed optimization with respect to determination of the network topology, determination of the set of input attributes and determination of the neuron weights. GA tried to optimize the network topology as it evaluates the genomes in its population for candidate network topologies, and tries to optimize that specific topology for set of input features. For each of these input feature combinations, a CPNN test was constructed and trained. The construction took place for each candidate solution, given the fixed topology as determined by GA. In addition, GA optimized the weights for the constructed CPNN.

The input factors, topology, and weights were encoded into a single genome for optimization. One-dimensional array of real numbers was used for encoding. The number of input factors, the number of layers, and the number of nodes in each layer determine the length of the genome. The total length of the genome L was calculated as;

$$L = n_{input} * n_1 + \sum_{i=1}^{k-1} (n_j * n_{i+1}) + n_k \qquad eqn\ (1)$$

DOI: 10.31695/IJASRE.2019.33512

Where n_{input} is the number of input attributes for the CPNN, k is the number of internal layers, and n_i is the number of nodes in layer i. The last term in the equation (1) is for the weights between the last internal layer and the output layer, which consists of a single node. The assumption is that each node in one layer is connected with every node in the subsequent layer. The encoding of the genome representing input features is done via simple binary encoding. . A zero in a specific bit in the genome means an input attribute is not chosen for the CPNN design, whereas a one in that bit means that the input attribute is chosen for the CPNN design.

c. Classification using CPNN algorithm

The input vector was fed into the network with adjusted weights to obtain desired output vector as training mode. The cluster unit does not assume any topology, but the winning unit was allowed to learn. The first phase and the second phase learning algorithm were summarized as follows:

i. Normalize the input vector.
ii. The highest Kohonen layer neuron is declared the winner and its weight is adjusted to yield unity output.
iii. Then the weight vector of the winning Kohonen neuron is equsl to the input vector with the best approximation value. Kohonen neuron is unsupervised.

The flow diagram of the proposed design model of CPNN-GA anomaly detection system is shown in Fig 2.

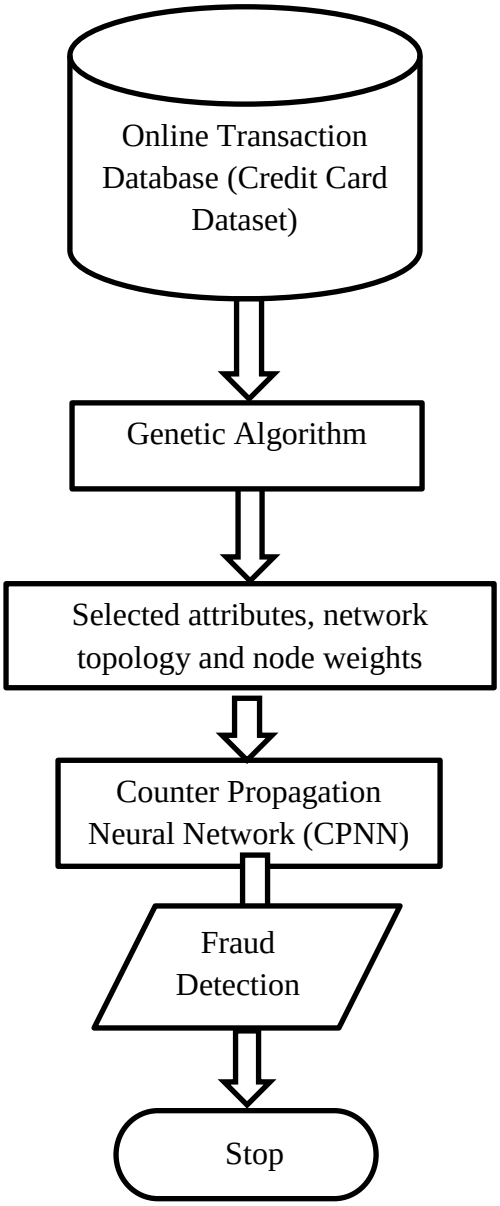

Fig 2: Flow diagram of the designed Model of CPNN-GA Anomaly Detection system

DOI: 10.31695/IJASRE.2019.33512

iv. The output of the Grossberg layer is calculated using dot product method.

$$g_i = \sum_j v_{ij} k_j = v_{ih} k_h = v_{ih} \qquad eqn\ (2)$$

v.

vi. Weights from non-zero kohonen neurons (non-zero Grossberg layer inputs adjusted. Weight adjustment follows the relation in equation 3.3

$$v_{ij}(n+1) = v_{ij}(n) + \beta\left[T_i - v_{ij}(n) k_j\right] \qquad eqn\ (3)$$

vii. The weights converged to the average value of the desired outputs, that is, best match an input-output (x-T) pair.

d. EVALUATION METRICS

CPNN initialization was achieved using; seven hundred (700) epochs; two hidden layers; eight neurons, while GA was initialized using a population size of sixty (60), crossover probability of 0.01 and generation of one hundred. In a fraud detection domain, the metrics deemed best for evaluation of the designed system include False Acceptance Rate (FAR), False Rejection Rate (FRR), Prediction Accuracy, Hit rate, Miss rate, Negative Predictive Value (NPV) and prediction time.

True Positive (TP) shows the number of genuine transactions correctly identified as non-fraudulent. False Positive (FP) is the number of fraudulent transactions incorrectly identified as genuine. False Negative (FN) denotes a mistaken instance of considering genuine transaction as fraudulent. True Negative (TN) shows the number of fraudulent transactions correctly identified as fraudulent.

False Acceptance Rate (FAR) denotes the rate at which the designed system incorrectly accepts a fraudulent transaction as genuine.

$$False\ Acceptance\ Rate\ (FAR) = \frac{\mathrm{FP}}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})}\ x\ 100 \qquad eqn(4)$$

False Rejection Rate (FRR) denotes the rate at which the designed system erroneously flags a genuine transaction as fraud.

$$False\ Rejection\ Rate\ (FRR) = \frac{\mathrm{FN}}{\mathrm{TN} + \mathrm{FP} + \mathrm{TP} + \mathrm{FN}}\ x\ 100 \qquad eqn(5)$$

Prediction accuracy (ACC) represents the percentage ratio of the total number of transactions that were correctly identified.

$$Prediction\ Accuracy\ (ACC) = \frac{TP + TN}{TP + TN + FP + FN} \qquad eqn(6)$$

Hit rate denotes the exactness of the designed system at spotting genuine transactions in a pool of genuine transactions.

$$Hit = \frac{TP}{TP + FN} \qquad eqn(7)$$

Miss rate denotes the ratio at which the system erroneously rejects genuine transaction in a midst of genuine transactions.

$$Miss = \frac{FN}{TP + FN} \qquad eqn(8)$$

Negative predictive value (NPV) is the rate at which a fraudulent transaction is correctly identified in ratio to all negatively assigned instances.

$$Negative\ predictive\ value\ (NPV) = \frac{TN}{TN + FN} \qquad eqn(9)$$

DOI: 10.31695/IJASRE.2019.33512

IV. RESULTS AND DISCUSSION

The implementation tool used was MATLAB R2012a version on Windows 7 Ultimate 32-bit operating system, Intel®Pentium® B960@2.20GHZ, 4GB Random Access Memory and 500GB hard disk drive. The overall simulation results of the CPNN-GA based system was compared with existing research results. In terms of false rejection rate, the designed CPNN-GA system had the least FRR of 2.69%. It implies that the design system is tolerant in falsely accepting impostor that could have access to cardholder's account as illustrated in Fig 3. In terms of false acceptance rate, the design system of CPNN-GA has least FAR of 2.69%. The design system demonstrated that is highly tolerant of the behavioral pattern upon which it was trained. The design system of CPNN-GA has ability to exhibit highest resistant to impostors (least FAR) and least resistant (least FRR) against genuine transactions. In terms of prediction accuracy, the CPNN-GA has highest predictive ability to correctly identify transaction types because it combines optimization ability of GA and learning strength of CPNN. CPNN-GA has prediction accuracy of 95.58%.

The behavioral pattern exhibited by the design system was further examined using hit rate, miss rate and negative predictive value (NPV). CPNN-GA has the best ability to spot genuine transaction in pool of genuine transactions. Therefore, CPNN-GA has highest hit rate of 97.19%

Table 1: Table showing performance evaluation comparison over existing research results

	FAR (%)	FRR (%)	ACC (%)	HIT RATE (%)	MISS RATE (%)	NPV (%)	PREDICTION TIME (s)	TRAINGING TIME (s)
GA	6.73	8.85	84.42	90.08	10.07	34.87	8.94	9.09
CPNN	3.85	6.35	89.42	93.20	6.88	31.54	3.42	5.67
CPNN-GA	1.73	2.69	95.58	97.19	2.81	42.18	13.59	15.42

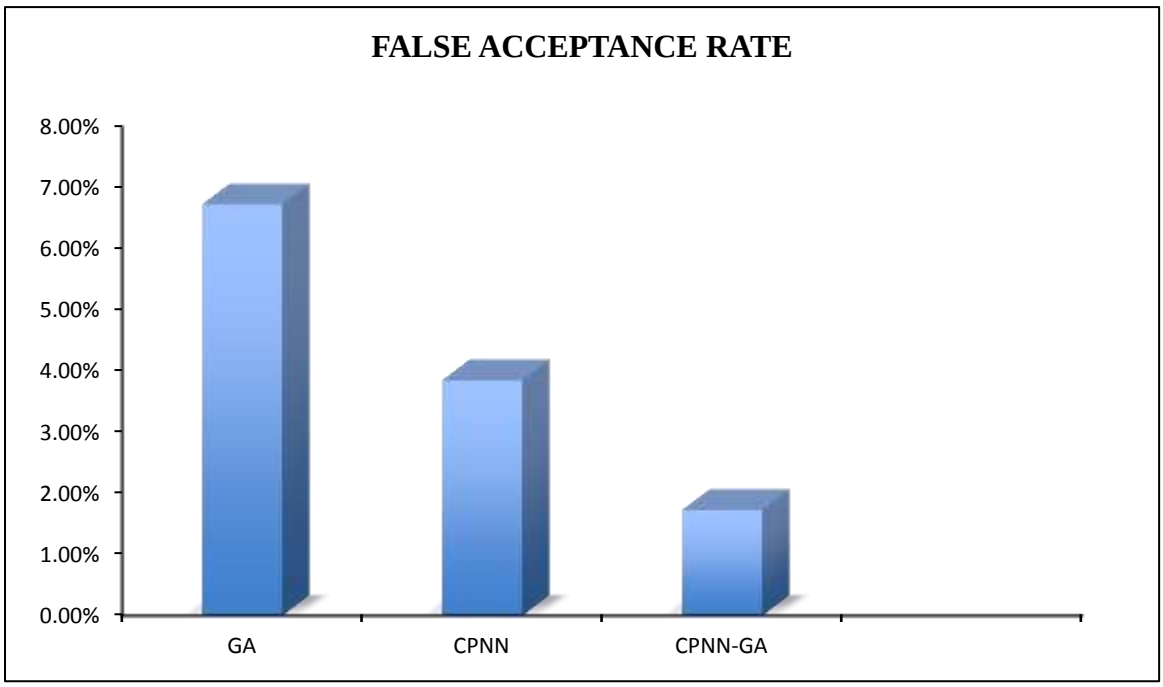

Fig 3: Bar chart showing false acceptance rate.

DOI: 10.31695/IJASRE.2019.33512

as compared with existing research results. Because of the greater ability of CPNN-GA to correctly identify fraudulent transactions in ratio to all negative assigned instances. CPNN has highest negative predictive value.

V. CONCLUSION AND FUTURE WORK

The simulation result of CPNN-GA based system was compared with the GA based system and CPNN based system to ascertain the effectiveness of the designed CPNN-GA system. The performance evaluation result from the considered systems deduced that the CPNN-GA based system outperformed the GA based system and the CPNN based system, as it had the least false acceptance rate, least false alarm rate, highest prediction accuracy, highest hit rate, lowest miss rate and highest negative predictive value. Although the designed CPNN-GA system took the most time at training and testing but it still performed within experimentally compliable time. Future work can be carried out by comparing the effect of other artificial neural network algorithms with another optimization algorithm.

REFERENCES

[1] Lee M., Ham S. and Jiang Q. (2014). E-commerce Transaction Anomaly Classification. *Statistics Department Stanford University*, pp 1-5.

[2] Varun C., Arindam B. and Vipin K. (2009). Anomaly Detection: A Survey. *ACM Computing Surveys*, **41** (3):1-58.

[3] Neda N., Leila B. and Ebrahim N. (2012). Surveying Different Aspects of Anomaly Detection and Its Applications. *The Journal of Mathematics and Computer Science*, **4 (**2):129-138

[4] Malini N. and Pushpa M. (2017). Analysis on Credit Card Fraud Detection Techiniques By Data Mining and Big Data Approach. *International Journal of Research in Computer Applications And Robotics*, **5 (**5):38-45.

[5] Vandana S., Sanjeev J., Vilas S., and Dev A. (2015). Fuzzy Counter Propagation Neural Network Control for a Class of Nonlinear Dynamical Systems. *Computational Intelligence and Neuroscience*, pp 1-12

[6] Malhotra R., Singh N. and Yaduvir S. (2011). Genetic Algorithms: Concepts, Design for Optimization of Process Controllers. *Computer and Information Science*, **4** (2):39-54.

[7] Razak T. A. and Ahmed G. N. (2014). A Comparative Analysis on Credit Card Fraud Techniques Using Data Mining. *International Journal of Data Mining Techniques and Applications, Integrated Intelligent Research (IIR)*, **3** (2): 398-400.

[8] Khan M. Z., Pathan J. D., Ahmed A. H. E. (2014). Credit Card Fraud Detection System Using Hidden Markov Model and K-Clustering. *International Journal of Advanced Research in Computer and Communication Engineering*, **3** (2):5458-5461.

[9] Priya B., Malvika D., Shweta P., Nivedita S. and Dhake B. G. (2014). Survey on Credit Card Fraud Detection Using Hidden Markov Model. *International Journal of Advanced Research in Computer and Communication Engineering*, **3** (5):6445-6448.

[10]Zareapoor M., Seeja. K. R. and Alam M. A. (2012). Analysis of Credit Card Fraud Detection Techniques: based on Certain Design Criteria. *International Journal of Computer Applications*, **52** (3):35-42.

[11]Bentley P. J., Kim J. J., Gil-Ho and Choi, J., (2000). Fuzzy Darwinian Detection of Credit Card Fraud. *Proceedings of 14th Annual Fall Symposium of the Korean Information Processing Society*, 1-4

[12]Srivastava A., Kundu A., Sural S. and Majumadar A. K. (2008). Credit Card Fraud Detection Using Hidden Markov Model. *IEEE Transactions on Dependable and Secure Computing*, **5** (1):37-48.

[13]Carsten, A.W. (2008). Credit Card Fraud Detection Using Artificial Neural Networks Tuned by Genetic Algorithms. *Unpublished PhD Thesis, Hong Kong University of Science and technology*, 1-226.

DOI: 10.31695/IJASRE.2019.33512

[14]Zhang Y, Fucheng Y. and Liu H.,(2009). Behavior–Based Credit Card Fraud Detection Model. *Fifth International Joint Conference on INC, IMS and* IDC, pp. 855-858.

[15]Wen-Fang Y. and Wang N. (2009). Research on Credit Card Fraud Detection Model Based on Distance Sum. *Proceedings of the International Joint Conference on Artificial Intelligence*, pp 353-356.

[16]Brabazon A., Cahill J., Keenan1 P., and Walsh D. (2010). Identifying Online Credit Card Fraud using Artificial Immune Systems. *IEEE Congress on Evolutionary Computation (CEC),* 1-7.

[`17] Falaki S. O., Alese B. K., Adewale O. S., Ayeni J. O., Aderounmu G. A. and Ismaila W. O. (2012). Probabilistic Credit Card Fraud Detection System in Online Transactions. *International Journal of Software Engineering and Its Applications* **6**, pp 4.

[18] Patel S. and Gond S. (2014). Supervised Machine (SVM) Learning for Credit Card Fraud Detection. *International Journal of Engineering Trends and Technology (IJETT),* **8** (3):137-139.

DOI: 10.31695/IJASRE.2019.33512