ISSN: 0974-5823 Vol. 7 No. 3 March, 2022

International Journal of Mechanical Engineering

TRANSACTION FRAUD DETECTION USING ARTIFICIAL BEE COLONY (ABC) BASED FEATURE SELECTION AND ENHANCED NEURAL NETWORK (ENN) CLASSIFIER

N.Geetha, Research Scholar, Department of Computer Science, P.K.R College of Arts & Science For Women, TamilNadu

Dr.G.Dheepa, Assistant Professor, Department of Computer Science, P.K.R College of Arts & Science For Women, TamilNadu

Abstract: Credit Card Fraud (CCF) is a critical issues faced by cardholders and its issuers from a very long time. CCF occurrences have been reported by applications and customers in their transactions. Detection of application level becomes very easy task. Transaction level fraud detection becomes very difficult when considered to application levels as transaction level fraud detection attributes vary from application levels. In the transaction level fraud detection, selection of features becomes very difficult task and reduces the fraud detection level. This work proposes a new feature selection technique for enhancing classifications of credit card fraud. This research work identifies fraudulent accounts using Enhanced Neural Network (ENN) for enhanced accuracy of results using feature selection techniques based on ABC (Artificial Bee Colony) which select relevant features from transaction level credit card datasets. Various elements of utilized dataset have been investigated in this study resulting in descriptions of logical relationships between transaction record attributes by ENN which identifies CCF in attributes based on LGBP (Logical Graph of Behavior Profile) and user's transaction data.

Keywords: Transaction Fraud Detection, Feature Selection, Artificial Bee Colony (ABC), Logical Graph of Behavior Profile (LGBP), Enhanced Neural Network (ENN) classifier.

1. INTRODUCTION

There have been significant surges in fraudulent transactions impacting economies significantly. Customer's credit card passwords, CVV numbers, and other essential information are under constant danger and susceptible in extensive usage of ecommerce and online shopping [1]. Fraudsters have an easy time accessing important information increasing their intensities. Banking systems also suffer from these online dishonest activities. As the number of fraud cases rises, identity theft on social media is on the rise.

There are two major strategies to avoid CCF namely preventions and detections where the former functions as layers of defense and prevent attacks from fraudsters [2]. When preventions fail detections occur. Hence, detections assist in finding and notifying issues as soon as possible. Typical fraud detection systems include associated academic degree automated instruments as well as human interventions. These automated techniques are based on principles of detecting frauds [3]. They evaluate all new incoming transactions and assign them a bogus score. Manual approaches are created by using fraudulent detectives which pay attention to transactions with high fallacy ratings and their evaluations are binary values implying frauds or legalities. These systems use professionally or knowledge driven rules or a mix of rules for detecting frauds. The developed rules attempt to validate individual scams found by fraud investigators [4]. As a result, conventional methods of dealing with fraudulent activities are gradually being supplemented by online fraud detection software using MLTs (Machine Learning Techniques).

Many models have identified CCF modeling different algorithms [5]. Adapting fraud detection systems to new frauds can be difficult, and retraining the MLTs owing to significant changes in fraud trends can be costly and dangerous. LRs (Logistic Regressions) were used to solve the categorization problems. Using GMM (Gaussian Mixture Model), fraudulent transactions were discretized into strategies [6] where synthetic minority oversampling techniques addressed class imbalances. Sensitivity analyses highlighted relevance of estimations in terms of economic values. A formal model and a robust learning technique for addressing' verification delays as well as a 'alert and feedback' mechanisms are among research accomplishments. The most crucial measure, according to tests, is the accuracy of these notifications.

Markov chains are modeled for tailoring transaction based BP (Behavioral Profile) [7]. The primary principle of these fraud detection models are based on explicit transactional attributes like transaction amounts and commodity categories which are treated nodes while transition probabilities between nodes quantify transaction behavioral aspects. Past experiences indicate the success rates of Markov chains in representing user BP when their transaction patterns were consistent. However, as online purchasing becomes more popular, user's transaction behaviours fluctuate and BPs need to cater to transaction varieties [8]. Hence, Markov chain models become less suitable for such users. Proposing new models to describe user BPs while taking into account behaviour variability and then offering fraud detection algorithms based on this new model were executed in this research paper.

In order to increase classifier's performances this research work introduces feature selections from transaction level credit card datasets based on ABC for the detection of CCF. This work identifies fraudulent accounts using ENN in classifications and then improves accuracy of results using feature selections.

The rest of the research work is organized as follows, Section 2 discusses the CCF detection techniques including supervised and unsupervised learning approaches. Section 3 describes the proposed fraud detection technique. Section 4 illustrates the results and discussion. Section 5 deals with the conclusion and future work.

2. LITERATURE REVIEW

In this section some of the techniques for identifying the CCF using advanced strategies are discussed.

Whitrow et al [9] proposed categorizations as well as cost based performance metrics that were practically applicable. Their techniques were used in two case studies with real data. Transaction aggregations have been proven to be beneficial in a variety of situations, though limited to certain cases. Moreover, duration of aggregations have significant influences on performances. While using RF (Random Forest) in classifications, aggregations appear to be particularly successful. Additionally, RF outperformed other classification approaches including SVM (Support Vector Machine), LR (Logistic Regression) and KNN (K-Nearest Neighbour). Aggregations also offer the benefit of not requiring properly labeled data and are more resistant to population drift effects.

Bahnsen et al [10] presented TAS (Transaction Aggregation Technique) that could provide new sets of characteristics based on evaluations of transaction time's periodic pattern using von Mises distributions. Subsequently, using genuine CCF dataset from a prominent European card processing firm the study compared their scheme with other detection methods for assessment of CCF and their selected characteristic set influences on outcomes. Their results revealed 13% improvement in savings when their recommended periodic characteristics were included in techniques.

Van Vlasselaer et al. [11] detected CCF in online business transactions. Their proposed scheme combined incoming transactional information (internal) with the executing user's prior spending historical information (external). The study used RFM (Recency–Frequency–Monetary) foundations which leveraged on networked merchants and cardholder information for projecting suspicion ratings which were based on the time of executions. Their findings demonstrated that intrinsic network based characteristics were two sides of the same coin and performing models have AUC (Area Under Curve) values greater than 0.98 when the aforesaid two types of features are combined.

To detect CCF, Jha et al [12] devised a technique based on transaction aggregation strategy. Their project aggregated transactions to record customer's buying behaviour prior to transactions which were then aggregated to estimate models for detecting fraudulent transactions. Also, for transaction aggregations and model estimations, the study used real time credit card transaction data acquired from international credit card operations.

Systems designed to detect CCF can be very complex mainly due to data's non-stationary distributions, exceptionally skewed distributions of classes and incessant streams of transactions. Dal Pozzolo et al [13] demonstrated that effective algorithms for detecting CCF using MLTs (Machine Learning Techniques). Credit Card transactional data cannot be found in public domains mainly due to their privacy and confidentiality concerns. This leads to arbitrary questioning and assumptions leaving issues unresolved or creating deficiencies in methods that detect CCF.

Correa Bahnsen et al [14] generated new characteristics by applying von Mises distributions on transaction times to analyze behaviours that were not common. In comparisons of detection algorithms of CCF, the study analyzed how different sets of features influence findings using actual CCF dataset from a European card processing firm. Their results revealed 13% improvement in savings when their recommended periodic characteristics were used. Their approach suggested is now being integrated into fraud detection systems of the aforementioned card processing organization.

LSTM (Long Short-Term Memory) networks were introduced by Jurgovsky et al [15] to evaluate transaction sequences. LSTM enhanced detection accuracy on offline transactions where cardholders were physically present at merchant sites when compared to baseline RFs based classification. Manual feature aggregation procedures are beneficial to sequential and non-sequential learning systems. Following reviews of true positives, it was discovered that both methodologies detected various types of frauds, indicating that the two could be used together.

Carcillo et al. [16] proposed SCARFF (SCAlable Real-time Fraud Finder) which was a combination of Kafka, Spark, and Cassandra (Big Data technologies) and MLTs to addresses imbalances, non-stationary attributes and feedback latencies. Their experimental results of voluminous credit card transaction databases indicated that their approach was precise, efficient and scalable on most of the transactions.

Wiese et al [17] simulated inherent time series sequences for transaction similarity where MLTs assessed transactional sequences. The strategy simulating time in this context could be more resilient to modest alterations in genuine buying behaviours. LSTM combined with SVM were experimented on real time credit card transactions. The study focused on proper selection of features, data pre-processing and evaluation metrics to provide clear bases for comparisons where the latter would facilitate comparison of results obtained from MLTs to datasets biased towards CCF.

Dal Pozzolo et al [18] detected frauds based on ensembles and sliding-windows. The study demonstrated that separate classifier training on feedbacks, delayed labels and aggregations can be very successful strategies. Their experiments on

voluminous real world transactional datasets revealed that their suggested technique significantly enhanced alert precisions, a key concern for investigators of frauds.

To capture fraud behavioral patterns while learning from labeled data, Fu et al [19] proposed CNN (Convolution Neural Network) for fraud detections where feature matrices represented large amounts of transactional data. CNN is used to assess dormant samples or patterns in data and when tested on real transactions of a bank, it was demonstrated that their technique outperformed many other existing approaches.

Increased activities in terms of online purchases makes it clear that user's transaction behaviours vary often and their BP should be able to define variability of transactions. Conventional models are inadequate for such consumers and hence this work focuses on detecting fraudulent transaction based on the user's BP.

3. PROPOSED METHODOLOGY

This work proposes a new model to describe a user's BP while taking into account behaviour variety for fraud detections. Main contributions of this work are detailed below:

1. Sorting characteristics of transactional records in alphabetical order and subsequently classifying attribute values for the construction of LGBP or abstracts that cover user's transactions.
2. Relevant features from credit card datasets at the transaction level were selected using ABC.
3. ENN represents logical relationships between transaction record's attributes. LGBP and user's transaction records are used for computing fraud detection values in attributes.

Figure .1. Proposed Transaction Fraud Detection Technique

3.1. BP (Behavior Profile)

BP constructed from transactions and transaction logs.

Definition 1 (Transaction Record):

Transaction records *r encompass m* attributes and mathematically stated as $r = \{a_1, a_2, \ldots\ldots, a_m | a_1 \in A_1, a_2 \in A_2, \ldots\ldots, a_m \in A_m\}$ where $A_i = \{a_1^i, a_2^i, \ldots \ldots a_n^i\}$ implies values of the i th attribute and $n_i = |A_i|$.

For users (*u)*, their transaction logs are sets of information about executed transactions within a period of time denoted as $L_u = \{r_1^u, r_2^u, \ldots \ldots r_{n^u}^u\}$ where $n^u = |L_u|$.

In original records, certain information needs to be pre processed where identical records are retained in *Lu* to describe the user's behaviors.In simple terms, if *Ru stands for* distinct records in *Lu, then Ru stands for* a set while*Lu implies a* multi-set.

The characteristics present in transaction records are listed in Table 1 lists where seven most significantcharacteristics are displayed in the following order: Merchant_id, Average Amount/ transaction/ day, Daily_chargeback_avg_amt, 6_month_avg_chbk_amt, 6-month_chbk_freq,Transaction_amount, Total Number of declines/day.

The construction of LGBP of user based on merchant _id and transaction log which reflects dependent relationships of all attribute values covering all transactions. Initially all attribute values found in user u's transaction records are abstracted using:

$$A_1^u = \{a \in A_1 | \exists r \in R_u : a \in r\} \quad (1)$$

$$A_2^u = \{a \in A_2 | \exists r \in R_u : a \in r\} \quad (2)$$

$$A_3^u = \{a \in A_2 | \exists r \in R_u : a \in r\} \quad (3)$$

……

$$A_m^u = \{a \in A_m | \exists r \in R_u : a \in r\} \quad (4)$$

It can be seen clearly that $A_1^u \subseteq A_1, A_2^u \subseteq A_2, \ldots\ldots and A_m^u \subseteq A_m$ and maintaining generality can be denoted as $A_i^u = \{a_1^i, a_2^i, \ldots\ldots a_{n_i^u}^i\}$ in which $n_i^u = |A_i^u|$for each $i\epsilon\{1,2,\ldots,m\}$.

Definition 2 (LGBP):

Assuming$L_u = \{r_1^u, r_2^u, \ldots\ldots r_{n^u}^u\}$ represents user's (u) transaction logs, *then* LGBPof u is a directed acyclic graph$G_u = (V_u, E_u)$, where:

1) $V_u = \{a_s, a_e\} \cup A_1^u \cup A_2^u \cup \ldots\ldots \cup A_m^u$ and a_s implies first transaction node while a_e represents the last transaction node;
2) $\forall a \in A_1^u, (v_s, a) \in E_u$;
3) $\forall a \in A_1^u, (a, v_e) \in E_u$;
4) $\forall i \in \{1,2,\ldots,m-1\}, \forall a \in A_i^u, \forall a' \in A_{i+1}^u : (a, a') \in E_u$if and only if $\exists r \in R_u : a \in r \Lambda a' \in r$.

Table 1. Example of Transaction Log

Transaction Records	Transaction Attributes						
	Merchant_id	**Average Amount/transaction/day**	**Daily_chargeback_avg_amt**	**6_month_avg_chbk_amt**	**6-month_chbk_freq**	**Transacction_Amount**	**Total Number of declines/day**
r_1^u	M_{id}^1	SM	LO	YE	YE	(0-200)	(0-5)
r_2^u	M_{id}^2	AV	HI	NO	NO	(0-200)	(6-10)
r_3^u	M_{id}^3	SM	LO	YE	NO	(1000-200)	(11-15)
r_4^u	M_{id}^4	AV	HI	NO	YE	(1000-2000)	(16-20)

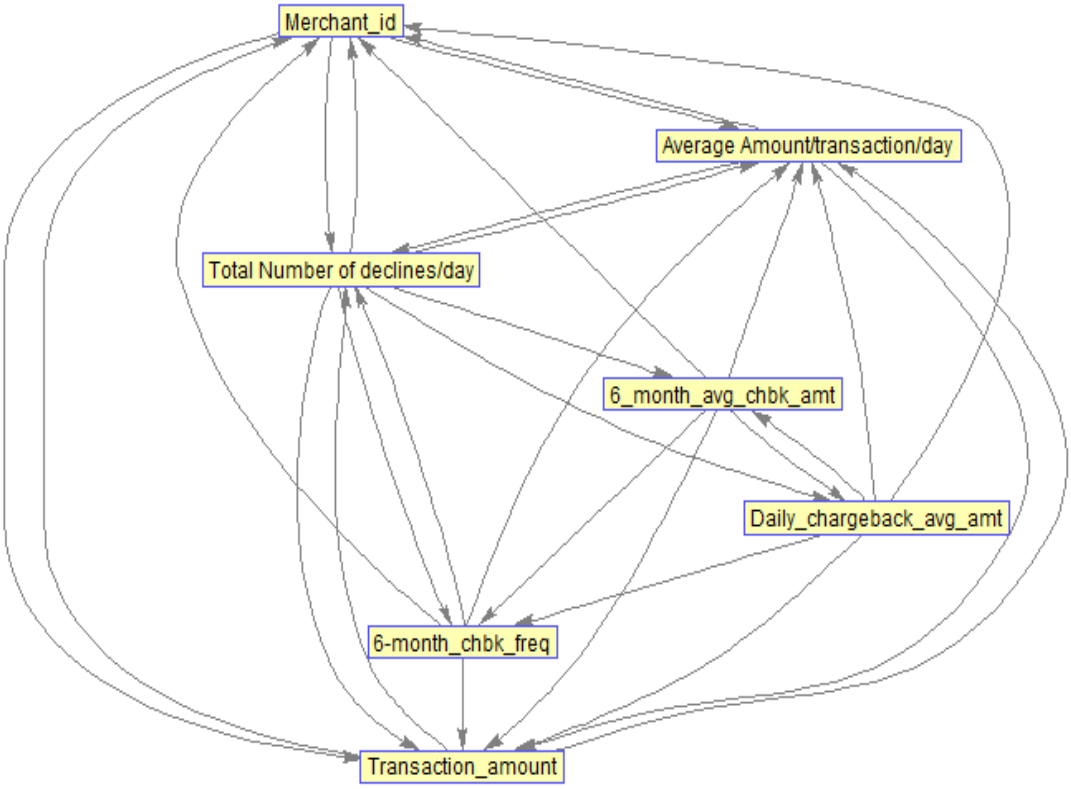

Fig. 2 – User's LGBP built from transaction logs of Table 1.

Definition 3 (Prepaths):

Let *Gu = (Vu, Eu)* be the LGBPs of user u. $\forall v \in Vu$, prepaths(v) is the set of all directed paths from node *as* to node *v* in *Gu*.

Definition 4 (Post nodes):

Let *Gu = (Vu, Eu)* be the LGBPs of u.∀*v* ∈ *Vu*, postnodes(v) is the set of nodes that are directly reached from *v* in *Gu*. For example, there are four directed paths which is illustrated in Figure. 2 and it is derived as,

$$\sigma_1 = a_s.M_{id}^1.SM.LO.YE.YE.(0{,}200).(0{,}5) \quad (5)$$

$$\sigma_2 = a_s.M_{id}^2.AV.HI.NO.NO.(0{,}200).(6{,}10) \quad (6)$$

$$\sigma_3 = a_s.M_{id}^3.SM.LO.YE.NO.(1000{,}200).(11{,}15) \quad (7)$$

$$\sigma_4 = a_s.M_{id}^4.AV.HI.NO.YE.(1000{,}2000).(16{,}20) \quad (8)$$

From the above equation the prepath and the post nodes are defined and its state transition, diversity co-efficient are calculated using the HMM model.

Definition 5 (Behavior Profile):

If *Lu* = $\{r_1^u, r_2^u, \ldots \ldots r_{n^u}^u\}$ represents user *u's* log oftransactions then BP*u = (Vu, Eu, , Mu, ω_u) stands for u's BP* where:

1) *Gu = (Vu, Eu)* representsu's LGBP;

2) *Mu* = {*Mv* |*v*∈ *Vu*} represents transition probabilities based on node paths [20] of *Gu*;

3) $\omega_u$is the diversity coefficient of *u*.

BPs are created for users based on their individual transaction logs. Proposed technique for determining acceptable transaction records to BPs in detailed in the next section.

3.2. Feature selection using ABC

Feature selection aims at finding groups of qualities that can be correlated or linked to classes and not other attributes in the collection This research work uses ABC to minimize class imbalances of inputs.

Algorithmic ABC [21] are recent optimization approaches that mimic complex honey bee's foraging behaviours. Swarms are groups of bees that can successfully complete tasks by working together. Bee swarms in ABC consists of three types of bees namely Employed, onlookers, and scouts. The employed bees forage for food within their immediate environments and bring it to the notice of onlooker bees which select suitable food sources from employed bee discoveries. The food choices are made in terms of quality (higher fitness). A certain number of employed bees turn into scout bees for tracing new food sources. Raw transactional data (diversity coefficients and state transition probabilities) from LGBPs were used in this study's feature selection process for improving the performance level of CCF detections.

In executions of algorithmic ABC, the swarm's initial half is full of employed bees while the other half has onlooker bees and the counts of either category of the bees form the count of swarm's solutions.

The ABC randomly generate starting populations of food sources (solutions) with SN (swarm size) solutions. Assuming Xi = $\{xi,1, xi,2,\ldots,xi,D\}$ represent swarm's *i*th solution with dimension D, then Xi (Employed Bee) develops a new candidate solution Vi in its immediate vicinity, as follows:

$$v_{i,j} = x_{i,j} + \Phi_{i,j}.\left(x_{i,j} - x_{k,j}\right) \quad (9)$$

where Xkrepresents candidate solution selectedrandomly ($i = k$/) while*j*stands for randomized index that is chosen from the set {1, 2, . . . , D}, and $\phi i,j$ is a random number within [−1, 1]. On generation of new candidate solution Vi, a greedy selection is used. Only when fitness of newly generated solutions (Vi) is better than parents (Xi), then parents are updated with new solutions. When all employed bees complete their search, waggle dances are done by them to communicate information on food sources to onlooker bees which then evaluate the amount of available nectar and select sources whose probabilities are proportional to nectars. These probabilistic selections are roulette wheel selections as stated below:

$$P_i = \frac{fit_i}{\sum_{j=1}^{SN} fit_j} \quad (10)$$

where fit*i* stands for ith swarm solution's fitness. It can be observed that better values of i imply that it will be chosen as a food source. When there is no improvement in positions of food sources after specific cycle counts (limits), the food sources are abandoned. Assuming X*i* becomes the abandoned food source, then scout bees search a new food source for replacements of Xi using:

$$x_{i,j} = lb_j + rand(0{,}1).\left(ub_j - lb_j\right) \quad (11)$$

Where lb - lower limits of the ith and *j*th dimensions, ub - upper limit of the ith and *j*th dimensions, rand(0, 1) are numbers generated in the interval [0, 1] based on normal distributions.

Algorithm 1. Feature selection using ABC Algorithm

Input : Raw transaction data (Behavior Profile)
Output: Optimal transaction data

1. Initialize the set of card holders data $x_i, i = 1,2, \ldots..SN$.
2. Evaluate each $x_i, i = 1,2, \ldots..SN$.
3. **While**"good" solutions do not reach predetermined max iterations count **do**
4. **For** i=1 to SN do /* Employed bees phase */
5. Generate $u_i \; with \; x_i$
6. Evaluate u_i
7. **if**fit$(u_i) \geq fit(x_i)$**then**
8. $x_i = u_i$
9. **for** i=1 to SN **do** **/*** Onlooker bees phase ***/**
10. Choose Employed bees
11. Attempt to enhance quality of food sources bassed on the step
12. Create new random food source when they do not improve in successive iterationss /* Scout bees phase */
13. Remember best food sources achieved till now;

3.3. Fraud Detection using Enhanced Neural Network (ENN)

This research work uses enhanced neural network model to detect fraudulent transactions. The Artificial Neural Network (ANN) frequently struggle to manage limited amounts of high-dimensional inputs, resulting in a computational strain. To solve this problem, this research work proposed a Modified Discrete Wavelet Transform (MDWT) for smaller dimensional representation of larger raw vectors for effectively reducing the dimensionality of the problem. ANN can function similarly to human brains, when trained this network properly and learn by example, much like people and are regarded as excellent classifier. Large disparities between legitimate and fraudulent transactions are prominent features of credit card traffics.Simultaneously, public data is also hardly accessibledue to privacy concerns.

NN (Neural Network) can detect CCF by mimicking human brains [22]. They can handle classification issues, it is feasible to identify CCF using them. Customers who pay with credit cards establish predictable trends which is utilized to classify the data. Using previously acquired data, NN are trained to recognize CCF. This data includes information such as the cardholder's occupation, income, credit card number, large purchases and transaction frequency. NNs use this information to determine transactions carried out by cardholders. When a credit card is used, the system compares the transaction information to information kept from prior transactions. If the data follows the pattern, the card is probably certainly used by the owner. If there isn't a match, it doesn't rule out the possibility of fraud, but it certainly raise the stakes.

3.3.1. Working principle (classification)

CCF detections in this work employs the same mechanism as human brain functioning which learn from past experiences and use them to take decisions in day to day lives. When consumers use credit cards, the patterns on credit card usage are set or predictable. NNs are trained about specific patterns of using credit cards by consumers usage data of previous one or two years. NNs are trained using information about several categories concerning card holders including their Average Amount/transaction/day,Daily_chargeback_avg_amt and transaction amounts as illustrated in Figure 3. Regardless of credit card usage patterns, NNs are also trained on numerous CCF that banks have previously faced. Based on the pattern of credit card usage, NNs assess if the transactions are genuine. When unauthorized users use credit cards, NNs based fraud detection systems compare patterns used by fraudsters against patterns of original cardholders on which they are trained, if the patterns match NNs accept transactions as legitimate.

Figure. 3. Layer of Neural Network in Credit Card

Transactions submitted for permissions are accompanied by authentication information including card account numbers and transaction attributes (amounts, merchant ids). Additional data fields from authorization system can be included in a feed (day). Banks, for the most part, do not keep logs of their authorization files and their credit card processing systems archive only transactions that are transmitted by merchants for settlements. Hence, data set of transactions was created using Bank's settlement files as only authorization information that had been archived in settlement files was accessible for model development.

3.3.2. Fraud Detection

Matching patterns do not mean that the transactions must exactly match patterns; rather, NN determine how close transactions are to the patterns; if there is a small difference, the transaction is fine; large differences imply an increase in the likelihood of illegal transaction and NN flag it as a faulty transaction. NN is programmed to provide outputs in the range of 0 to 1. If the NN produce output that is less than or equal to.6 or.7, then the transactions are lawful, but if the output is more than or equal to.7, the likelihood of the transaction being illegal increases. There are some times when legitimate user transactions are substantially different and there's also the possibility that criminals use a card that fits into the pattern for which NN is trained. Despite the fact that it is still uncommon, if the legitimate users are unable to execute transactions owing to these restrictions, it is not a matter of concern. When illegal persons obtain credit cards, they do not use them repeatedly on series of small transactions, but instead make singular or large purchases as quickly as possible. This contradicts the patterns that NN is trained on. History descriptors include information on how the card was used for transactions and how much money was paid into the account in the recent past. Other descriptions might include things like the card's issuance date (or most recent reissue). This might be useful in detecting NRI (non-receipt of issue) fraud.

3.3.3. Problem with the training of the neural network

Existing NNs demand high-dimensional data as inputs, making it difficult to attain high dimensionalities and high transaction availability. A common option is to create derivative features for consumer behaviour patterns based on industry experience that match the consumers' behaviour habits. It is necessary to detect fraud in order to prevent it and investigate several legitimate consumer behaviour patterns as well as fraudster conduct patterns.

Average of transaction amounts, total amounts, differences between current transactions and average transaction amounts and other characteristics can be determined from raw data in selected time frames of inputs. These information can be used to depict relationships between user's transaction amounts and total transaction counts over a period of time. In specific cases, such derived attributes help in accurately describing user's transactional behaviours. In this research work, information entropy based diversity coefficients and wavelet signals were used to train NNs for minimizing error rates and efficiently detecting frauds.

3.3.4.Modified Discrete Wavelet Transform based Neural Network (MDWT-NN)

Analyzing wavelets can provide significant insights into the data’s physical shape as it displays information in time and frequency domains. Studies have found that accurate data pre-processing using wavelet analysis allows models to reflect system’s underlying properties. This work combines ANN with DWT (Discrete Wavelet Transform) where the combined i.e. (DWT-NN) method acquired capabilities of both wavelet analysis and ANNs and efficiently models non-linear and non-stationary time series. ANN integrations with DWT increases precisions of fraud detections. Figure 4 depicts a schematic of the established modeling technique.

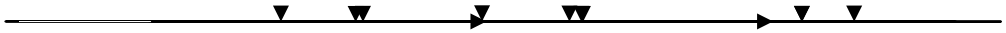

Figure. 4. Schematic of proposed MDWT-ANN models

DWT transforms signals by dividing signals into a number of sets, each of which are time series of coefficients characterising signal's temporal evolutions in the appropriate frequency bands. WTs (Wavelet transforms) as previously stated, are a mathematical paradigm for translating original data into the time-scale domain. Because most financial data are non-stationary, wavelet based models are appropriate models for analysing financial data.

Wavelet theories are based on Fourier analysis, where functions can be represented as sums of sine/cosine functions. Wavelets are functions of time t that obey wavelet's admissibility conditions or simple rules.

$$C_{\varphi} = \int_0^{\infty} \frac{|\varphi(f)|}{f} df < \infty \qquad (12)$$

where $\varphi(f)$ is the Fourier transform and a function of frequency f, of φ(t) . WTs are mathematical tools that can be used for complex tasks including image analysis and signal processing. When encountering non-stationary signals or signals localised by time/space/frequency, they can be handled using Fourier transforms. Within a function/family, there are two types of wavelets. Smoother and low frequency aspects of signals are described by father wavelet while high frequency and detailed components are described by mother wavelet. In a j-level wavelet decomposition, j=1,2,3,..., J symbolises the father wavelet and mother waveletrespectively:

$$\emptyset_{j,k} = 2^{\left(\frac{-j}{2}\right)} \emptyset \left(t - \frac{2^j k}{2^j}\right) \qquad (13)$$

$$\varphi_{j,k} = 2^{\left(\frac{-j}{2}\right)} \varphi \left(t - \frac{2^j k}{2^j}\right) \qquad (14)$$

Where J stands for largest scale supported by data points count and the two types of wavelets mentioned previously, namely father and mother wavelets and meets the following criteria:

$$\int \emptyset(t) dt = 1 and \int \varphi(t) dt = 0 \qquad (15)$$

A wavelet analysis input, time series data or function f(t) can be constructed as a series of projections onto father and mother wavelets indexed by both{k}, k = {0, 1, 2,...} and by{S}=2j, {j=1,2,3,. . .J}. While examining actual discretely sampled data, a lattice for computations must be designed. In mathematics, using a dyadic expansion, as shown in Equation 1, is more convenient (15). The expansion coefficients are provided by the projections:

$$S_{j,k} = \int \emptyset_{j,k} f(t) dt \, , \int \varphi_{j,k} f(t) dt, \quad (16)$$

$$S_j(t) = \sum S_{j,k} \emptyset_{j,k}(t) and \; D_j(t) = \sum S_{j,k} \emptyset_{j,k}(t) \quad (17)$$

WTs are to compute coefficients of wavelet series approximations in Equation (17) for discrete signals, where smooth $S_j(t)$ and details $D_j(t)$ coefficients are the two types of coefficients. Detailed coefficients discover major characteristics in datasets, while smooth coefficients delve deep into data set's most important elements..

The fact that DWT is shift variant in its transformations is the most serious possible flaw. The use of crucial sub-sampling (down-sampling) in DWT result in shift variances. In this method, second wavelet coefficients at decompositions are removed mainly to limit the amount of data that needs to be evaluated while imposing inherent temporal frequency uncertainties of the study (Analyses become more certain about signal's frequency components but less certain about occurrence time). On the contrary wavelet coefficients are significantly dependant on their placement in sub-sampling lattices due to these crucial sub-sampling. As a result, the information entropy-based diversity co-efficient is introduced in this study to prevent shift variance problems.

- **Information entropy-based diversity coefficient**

Entropy is a measure of information content that may be described as an event's unpredictability. As a result, the higher the probability, the lower the unpredictability, implying that the information content is likewise minimal. When an event occurs with a chance of 100%, the unpredictability and information content are both zero. As a result, the suggested information entropy-based diversity coefficient takes use of a property of the entropy equation: the cross-entropy loss function, which is represented in equation (18), may quantify the quality of a classification model.

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{k} 1\{y_i = j\} log \frac{e^{e_j^T x_i}}{\sum_{i=1}^{k} e^{e_j^T x_i}} \qquad (18)$$

It is not impossible to reduce storage and processing costs of linear systems for matrices that correspond to discretizations of higher-dimensional issues by using information entropy based diversity coefficients. When the information entropy-based diversity coefficients are smooth (as they often are), applying an MDWT on the factors and setting negligibly tiny entries to zero can save even more money. When compared to the quadratic loss function, the information entropy based diversity coefficient improves neural network training performance. The proposed MDWT-ANN models for detecting fraud in the provided dataset are represented by algorithm 2.

Algorithm 2. The process of proposed MDWT-ANN models

Input: Optimized transaction data
Output: Prediction of fraud transaction

Begin
Initialize weights;
While not stop criterion do
Calculates $e^p(w)$ for each pattern;
$e1 = \sum_{p=1}^{P} e^p(w)^T e^p(w)$;
Calculates $F^p(w)$ for each pattern;
repeat
Calculates Δw;
$e2 = \sum_{p=1}^{P} e^p(w + \Delta w)^T e^p(w + \Delta w)$;
if $(e1 \leq e2)\ then$
$\mu := \mu * \beta$;
end if;
until $(e2 < e1)$; (Training phase initialization)
/Discrete wavelet transform/
Load transaction data (error value)
Initialize input matrix();
Convolution with scaling vector
Down sample by _2();
/*Thresholding*/
Initialize threshold value ();
Calculated dead value();
/*information entropy-diversity coefficient*/
Update the data using Eq.18;
Perform thresholding on the transform matrix ();
Up sample by_2();
Convolution with wavelet vectors();
/*update values*/
$\mu := \mu/\beta$;
$w = w + \Delta w$;
end while;

4. Performance Evaluation

This section exhibits the performance of the method proposed in this work. First,the data set and set parameters are introduced . Then, the comparison results are illustrated . The MATLAB is used for evaluated.

4.1. Dataset and Parameters

Credit card data is not available for the general public with a few exceptions like https://www.kaggle.com/shubhamjoshi2130of/abstract-data-set-for-credit-card-fraud-detection#creditcardcsvpresent.csv which can be used for detection of CCF.

Transactional record's seven attributes were chosen for study namely Merchant_id, AverageAmount/transaction/day, Daily_chargeback_avg_amt, 6_month_avg_chbk_amt, 6-month_chbk_freq,Transaction_amount, Total Number of declines/day.

The selected attributes were found to be effective for detecting transactional frauds. Transaction amounts/day was divided into SM (small) and AV (Average). In experiments, four merchant id were mentioned and amount frequencies denoted were LO (Low) and HI (High), amounts were divided into four segments: (0-200],(0-200], (1000-200] and (1000- 2000).The data used in experiments were pre-processed based on prior transformations.This proposed work is evaluated in terms of quality metrics of Accuracy, Precision ,Recall and F-measure.

Accuracy is calculated in terms of positives and negatives as follows:

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \quad (19)$$

Precision is defined as the ratio of correctly found positive observations to all of the expected positive observations [23].

$$\text{Precision} = TP/(TP+FP) \quad (20)$$

Recall is defined the ratio of correctly identified positive observations to the over-all observations in [23].

$$\text{Recall} = TP/(TP+FN) \quad (21)$$

F-measureis defined as the weighted average of Precision as well as Recall [23]. As a result, it takes false positives and false negatives.

$$\text{F-measure} = 2*(\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision}) \quad (22)$$

Table 2. Performance results of the proposed and existing methods

Metrics	TAS	LGBPs	LGBP-ENN
Accuracy	85.1500	90.60	93.100
Precision	82.0190	90.4336	93.9557
Recall	84.230	90.7285	91.1612
F-measure	83.0060	90.5808	92.537

Table 2 tabulates performances of the suggested technique with existing methods and it can be clearly identified from table values that the suggested technique outperforms other methods.

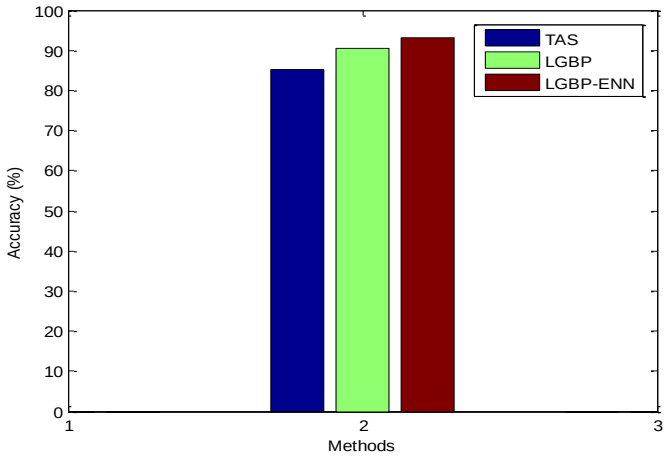

Figure 5. Accuracy comparison between the proposed and existing fraud detection technique

The figure 5 illustrates the Accuracy performances of the proposed LGBP-ENN and existing LGBP and TAS fraud detection techniques where LGBP-ENN's performance is better than the other two techniques.

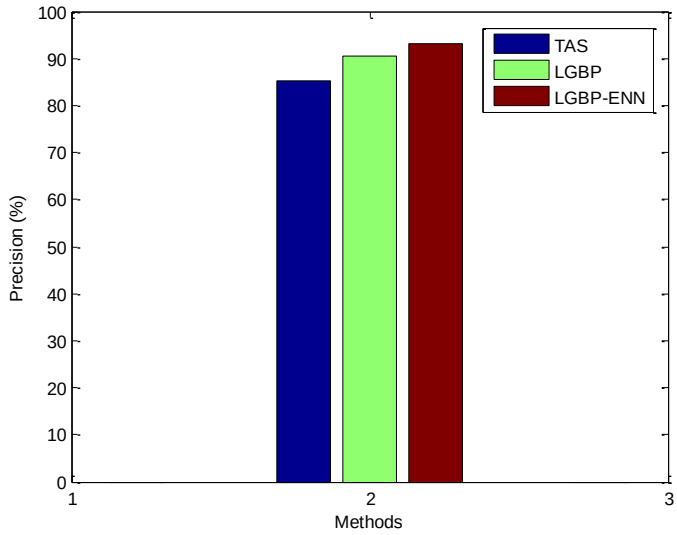

Figure 6. Precision comparison between the proposed and existing fraud detection technique

The figure 6 examines the precision comparisons between the suggested and existing fraud detection techniques. It can be observed from the graph that the suggested method performs better in terms of detecting fraudulent transactions.

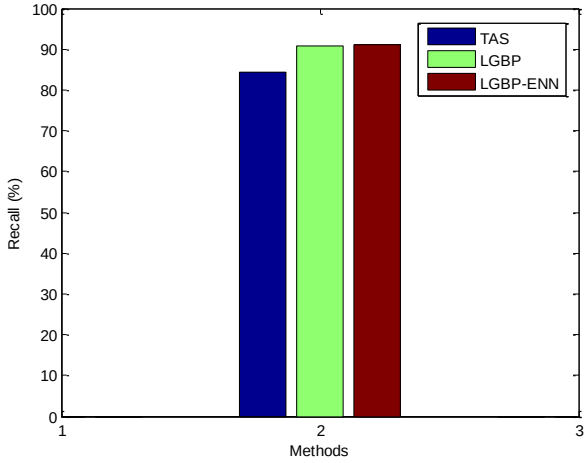

Figure 7. Recall comparison between the proposed and existing fraud detection technique

The figure 7 illustratesrecall performance values of the proposed LGBP-ENN and existing LGBP and TAS fraud detection techniques where LGBP-ENN's performance is better than the other two techniques as optimizations based feature selections are proposed in this work for reducing complexity for classifiers. The proposed ABC have higher selection rates in their attribute selections for fraud detection. Finally when compared to existing LGBP and TAS methods, the proposed LGBP-ENN based model considers diversity of user's behaviors and provides best detection rates.

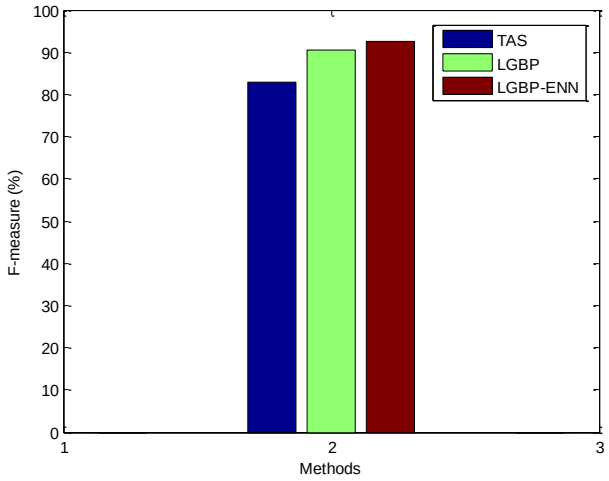

Figure 8. F-measure comparison between the proposed and existing fraud detection technique

The suggested and existing fraud detection techniques are compared using the F-measure in Figure 8. The harmonic mean of precision is given by the F-measure and the fraud and non-fraud accuracies are given by recall. In comparison to the existing LGBP-ENN method, it concludes that the proposed LGBP-ENN method provides best detection of CCF than existing LGBP and TAS techniques.

5. Conclusion

A transaction-based fraud detection approach based on ABC and ENN is proposed in this paper. A recent behavioural profile of a cardholder is built using the behavioural patterns of comparable cardholders. User's BPs are derived from their transaction data in this approach and utilized for detecting transaction frauds in online transactions. The transaction level credit card dataset was utilised to pick relevant features using an ABC based feature selection method. ENN based classifications are utilised to detect fraudulent accounts and a feature selection approach is employed to increase the accuracy of the findings. The variety of a user's transaction behaviour is measured using an information entropy-based diversity coefficient. From the experimental results, the proposed method shows better performance and effectiveness in detecting fraudulent transactions. Further this work is focused on improving the advanced optimization algorithm to improving the performance of the fraud detection.

REFERENCES

1. Tripathi, K. K., & Pavaskar, M. A. (2012). Survey on credit card fraud detection methods. *International Journal of Emerging Technology and Advanced Engineering*, *2*(11), 721-726.
2. Chen, R. C., Chiu, M. L., Huang, Y. L., & Chen, L. T. (2004, August). Detecting credit card fraud by using questionnaire-responded transaction model based on support vector machines. In *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 800-806). Springer, Berlin, Heidelberg.
3. Sahin, Y., & Duman, E. (2011, June). Detecting credit card fraud by ANN and logistic regression. In 2011 International Symposium on Innovations in Intelligent Systems and Applications (pp. 315-319). IEEE.
4. Chaudhary, K., Yadav, J., & Mallick, B. (2012). A review of fraud detection techniques: Credit card. *International Journal of Computer Applications*, *45*(1), 39-44.
5. Şahin, Y. G., & Duman, E. (2011). Detecting credit card fraud by decision trees and support vector machines.
6. Duman, E., & Sahin, Y. (2016). A comparison of classification models on credit card fraud detection with respect to cost-based performance metrics. *Use of Risk Analysis in Computer-Aided Persuasion. NATO Science for Peace and Security Series E: Human and Societal Dynamics*, *88*, 88-99.
7. Sahin, Y., Bulkan, S., & Duman, E. (2013). A cost-sensitive decision tree approach for fraud detection. *Expert Systems with Applications*, *40*(15), 5916-5923.
8. Özçelik, M. H., Duman, E., Işik, M., & Çevik, T. (2010, June). Improving a credit card fraud detection system using genetic algorithm. In *2010 International Conference on Networking and Information Technology* (pp. 436-440). IEEE.
9. Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M. (2009). Transaction aggregation as a strategy for credit card fraud detection. *Data mining and knowledge discovery*, *18*(1), 30-55.
10. Bahnsen, A. C., Aouada, D., Stojanovic, A., & Ottersten, B. (2016). Feature engineering strategies for credit card fraud detection. *Expert Systems with Applications*, *51*, 134-142.
11. Van Vlasselaer, V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. (2015). APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions. *Decision Support Systems*, *75*, 38-48.
12. Jha, S., Guillen, M., & Westland, J. C. (2012). Employing transaction aggregation strategy to detect credit card fraud. *Expert systems with applications*, *39*(16), 12650-12657.

13. Dal Pozzolo, A., Caelen, O., Le Borgne, Y. A., Waterschoot, S., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. *Expert systems with applications*, *41*(10), 4915-4928.
14. Correa Bahnsen, A., Aouada, D., Stojanovic, A., & Ottersten, B. (2015). Detecting credit card fraud using periodic features. In *IEEE International Conference on Machine Learning and Applications*.
15. Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P. E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. *Expert Systems with Applications*, *100*, 234-245.
16. Carcillo, F., Dal Pozzolo, A., Le Borgne, Y. A., Caelen, O., Mazzer, Y., & Bontempi, G. (2018). Scarff: a scalable framework for streaming credit card fraud detection with spark. *Information fusion*, *41*, 182-194.
17. Wiese, B., & Omlin, C. (2009). Credit card transactions, fraud detection, and machine learning: Modelling time with LSTM recurrent neural networks. In *Innovations in neural information paradigms and applications* (pp. 231-268). Springer, Berlin, Heidelberg.
18. Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2015, July). Credit card fraud detection and concept-drift adaptation with delayed supervised information. In *2015 international joint conference on Neural networks (IJCNN)* (pp. 1-8). IEEE.
19. Fu, K., Cheng, D., Tu, Y., & Zhang, L. (2016, October). Credit card fraud detection using convolutional neural networks. In *International Conference on Neural Information Processing* (pp. 483-490). Springer, Cham.
20. Zheng, L., Liu, G., Yan, C., & Jiang, C. (2018). Transaction fraud detection based on total order relation and behavior diversity. *IEEE Transactions on Computational Social Systems*, *5*(3), 796-806.
21. Karaboga, D. (2010). Artificial bee colony algorithm. *scholarpedia*, *5*(3), 6915.
22. Rowley, H. A., Baluja, S., & Kanade, T. (1998). Neural network-based face detection. *IEEE Transactions on pattern analysis and machine intelligence*, *20*(1), 23-38.
23. Powers, D.M., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technologies Vol. 2, no. 1, 2011, pp-37-63.