

Zhiyue Gao^{1*}

Accounting Analysis by Incorporating Apriori Association Rule Algorithm



Abstract: - Electronic payment systems are increasingly being used globally for online business transactions. Accounting analysis provides domain expertise to assist feature selection, preprocessing, and model interpretation, hence improving anomaly identification in credit card data using deep learning. Deep learning models can effectively identify anomalies, detect fraudulent activities, and continuously improve fraud detection capabilities in credit card transactions by integrating accounting principles and financial knowledge. This guarantees strong fraud prevention measures for financial institutions. In this manuscript, Accounting Analysis by Incorporating Apriori Association Rule Algorithm (AA-IAARA-PMNN-PCBESA) is proposed. Initially, the input datas collected from credit card dataset are given as input. The input datas are fed to pre-processing using Confidence Partitioning Sampling Filtering (CPSF) for identifying the missing values from the input data. Afterward the pre-processed datas were given to feature selection using Black Winged Kite Algorithm (BWKA) for selecting the transaction features. Then selected features were given to Port-Metriplectic Neural Network (PMNN) optimized with Polar Coordinate Bald Eagle Search Algorithm (PCBESA) for accurate detection of anomaly in the credit card data and classify the detected anomaly as fraud and non-fraud. The proposed AA-IAARA-PMNN-PCBESA approach is implemented in Python. The performance of the proposed AA-IAARA-PMNN-PCBESA technique attains 19.5%, 24% and 23% higher accuracy, 24.6%, 23.15% and 24.8% higher Precision and 21.4%, 27.36% and 21.08% higher recall compared with existing methods such as a Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost (PE-CCFD-SVM), machine learning based credit card fraud detection using the GA algorithm for feature selection (CC-FD-RF) ensemble learning approach for anomaly detection in credit card data with imbalanced and overlapped classes (AD-CCD-XGB) models respectively.

Keywords: Anomaly, Confidence Partitioning Sampling Filtering, Black Winged Kite Algorithm, Credit Card Data, Detection, Polar Coordinate Bald Eagle Search Algorithm and Port-Metriplectic Neural Network.

I. INTRODUCTION

Although accounting analysis makes it possible to spot odd trends, abnormalities, and discrepancies in financial transactions, it is necessary for the detection of credit card fraud. Accounting procedures can assist in identifying patterns of both valid and fraudulent transactions by reviewing historical transaction data. Any departure from regular spending habits or strange transaction sequences may signal the need for additional research. Anomaly detection methods, such as neural networks and clustering algorithms, can be used to find transactions that drastically differ from typical behavior. These irregularities might point to possible fraud cases that need more research [1, 2]. Credit cards are commonly utilized for everyday financial transactions as well as instantaneous services. Electronic devices with Internet access, such as smartphones, tablets, and PCs, are used by consumers to make purchases [3]. The increased use of electronic banking and online payment systems has led to a rise in fraudulent transactions, which has cost the economy billions of dollars annually in losses. A survey estimates that credit card theft will affect cards, customers, and businesses worldwide in 2020, costing the global economy 28.58 USD billion [4-6]. Furthermore, it was found that the US economy lost 11 billion dollars in 2020 as a result of merely fraudulent credit card transactions. The situation is alarming because, in the last ten years, credit card theft has tripled globally, according to recent estimates [7-9]. It is estimated that losses would rise from \$9.84 billion USD in 2011 to 32.39 billion USD in 2021. The high false positive rates and incapacity of traditional rule-based anomaly detection techniques to adjust to evolving fraud trends often render them ineffectual. Therefore, in order to reliably identify abnormalities in credit card data, contemporary anomaly detection technologies rely on sophisticated machine learning and data mining algorithms. According to the prediction, businesses globally may experience cumulative losses greater than 408.50 billion throughout the following ten years. The Malaysian banking industry suffered a significant setback in 2016 when fraudulent credit card transactions resulted in a loss of RM 51.3 million [10-12]. Furthermore, the Malaysian government expressed alarm upon learning that 12.8% of credit cards in the country found it difficult to make the minimum payment. Financial institutions have to cope with a variety of credit card-related issues, including (i) late

¹ Mrs. Zhiyue Gao^{**} ¹Shanxi Polytechnic College, Taiyuan, Shanxi, 030006, China
Email: 15235365596@163.com

payments resulting from cardholders' forgetting about refunds, and (ii) the possibility of fraud by unidentified third parties [13–15].

An automatic anomaly detection system can be a valuable tool in resolving this issue. Data mining is one approach that looks promising for solving these issues [16–18]. However, because of the following characteristics that are frequently found in the data, financial data poses particular difficulties for professionals in data mining: (i) Class samples that overlap; (ii) an unequal distribution of classes. Credit card statistics are often under representative of important domains such as fraud and late payments. If learning algorithms have enough data, they can effectively recognise and learn patterns from these classes. But when these classes have few instances, the process becomes more challenging, making it more challenging to define their decision boundaries accurately and generally using conventional learning approaches [19]. One of the most crucial tasks in the field of fraud detection and prevention is the identification of anomalies in credit card data. Financial institutions and credit card companies need to identify unusual activity in credit card transactions to safeguard their customers and reduce losses due to the growth of online transactions and the increasing expertise of fraudsters. The issue gets more complex when there is a significant amount of attribute value overlap in multiple regular transactions. If the minority classes are linearly separable, the learning algorithm's performance may not be significantly affected by a significant imbalance in the dataset [20].

Several works were have presented previously in literatures were depending on anomaly detection in credit card data using deep learning. Few of them were mentioned here.

Islam, et.al, [21] has introduced an ensemble learning method for detecting anomalies in credit card data that contains overlapping and unbalanced classes. This paper proposes a model called CCAD, which leverages meta-learning ensemble techniques and the base learner's paradigm to increase the rate at which credit card irregularities are detected. Use the XGBoost algorithm as the Meta learner and four outlier detection techniques as base learners in the suggested stacked ensemble strategy to identify abnormalities in credit card transactions. To solve the issues of overfitting and data imbalance, apply the stratified sample technique and k-fold cross-validation procedure. Additionally, the discordance rate is computed to improve the precision of group learning tasks.

Ileberi, et al. [22] have recommended using AdaBoost and SMOTE for Machine Learning Methods' Performance Evaluation in Credit Card Fraud Detection. Here, skewed real-world datasets from European credit cardholders were used to create a CC-FD-RF. The class imbalance problem was resolved by resampling the dataset using the Synthetic Minority over-sampling technique. The machine learning methods of support vector machines, decision trees, logistic regression, random forests, and additional trees were used to assess this system. The Adaptive Boosting approach was used to these ML algorithms in order to enhance their classification quality.

Ileberi, et al. [23] have investigated a ML -based credit card fraud detection system that selects characteristics using the GA method. Credit card fraud is among the financial fraud incidents that have increased as a result of recent advancements in electronic payment and e-commerce. Thus, it was imperative to implement systems that are capable of detecting credit card fraud. When utilising ML to detect credit card fraud, special attention must be paid to the attributes of the scam. This study presents a genetic algorithm based feature-selective ML credit card fraud detection engine. Once the best characteristics were chosen, the recommended detection engine uses machine learning classifiers like Random Forest , ANN, Naive Bayes , Decision Tree , Logistic Regression .

Bakhtiari, et al. [24] have presented “Credit card fraud detection using ensemble data mining methods”. Credit card use was becoming more and more common, which has led to an increase in security issues and an increase in fraud aimed at obtaining unapproved financial benefits. Different approaches have been successfully developed by researchers to identify and forecast credit card fraud. ML and data mining are two of these techniques. In this sense, the problem prediction accuracy was crucial. This study looks at ensemble learning techniques, such as gradient boosting (LiteMORT and LightGBM), combining them with simple and weighted averaging techniques, and evaluating them.

Hemdan and D.H. Manjaiah, [25] have suggested “Anomaly Credit Card Fraud Detection Using Deep Learning”. The yearly costs associated with fraudulent credit card transactions were approaching the billion-dollar mark, making them a serious worldwide problem. Customers, businesses, and the financial industry all start to have serious security concerns about it. Financial institutions may be able to identify fraudulent transactions more accurately by utilising state-of-the-art prediction systems. We evaluated our deep learning

model with other already employed machine learning approaches in order to identify anomalies related to credit card theft in financial transactions. The proposed theory was tested using a dataset that was gathered over the course of two days in September 2013 throughout Europe.

Esenogho, et al. [26] have demonstrated a feature-engineered neural network ensemble for better credit card fraud detection. According to this study, a neural network ensemble classifier in conjunction with a hybrid data resampling approach may effectively detect credit card fraud. The basis learner in adaptive boosting, which generates the ensemble classifier, is a LSTM neural network. SMOTE-ENN, or the edited nearest neighbour, was employed in conjunction with the synthetic minority oversampling technique to create the hybrid resampling. Using publicly accessible real-world credit card transaction data, the efficacy of the suggested approach was illustrated.

Alarfaj, et al. [27] have suggested using “state-of-the-art deep learning and machine learning techniques for detecting credit card fraud”. The use of the most recent advancements in deep learning algorithms has received the majority of attention. To find the best outcomes, a comparison between ML and deep learning algorithms was carried out. Fraud was found by a comprehensive empirical investigation using the European card benchmark dataset. A machine learning approach was initially used to the dataset, leading to a slight improvement in the accuracy of fraud detection. Afterwards, the performance of fraud detection was enhanced by the use of three convolutional neural network-based designs. Further layer additions increased detection precision even further. A thorough empirical investigation has been conducted with the most recent models, epochs, and hidden layer counts.

The literature presents a variety of methods for credit card fraud detection, leveraging both traditional machine learning techniques and more advanced deep learning algorithms. Each approach comes with its own set of strengths and limitations, along with potential research gaps that could be addressed in future studies. One prevalent method involves ensemble learning techniques, such as the use of meta-learning ensemble approaches like XGBoost combined with outlier detection methods. While these methods often demonstrate improved detection rates for creditcard anomalies, they may still face challenges with Overfitting and class imbalance. Additionally, further exploration could be done on the effectiveness of different meta-learners and base learners within the ensemble framework. Another approach involves resampling techniques like SMOTE combined with boosting algorithms like AdaBoost. These methods aim to mitigate the issue of class imbalance in credit card fraud datasets; however, they may not fully capture the complexity of real-world fraud patterns. Future research could explore more sophisticated resampling techniques or ensemble combinations to better address this issue. Feature selection methods, such as genetic algorithms (GA), are also commonly employed to improve the efficiency of fraud detection models. While these techniques help in identifying relevant features, they may overlook important interactions between variables or fail to adapt to evolving fraud patterns. Exploring dynamic feature selection approaches or hybrid methods combining feature engineering with deep learning could be fruitful avenues for future research. Deep learning models, particularly neural network ensembles, have shown promise in capturing intricate patterns in creditcard transaction data. However, these models often require large amounts of data and computational resources, which may limit their scalability and practical implementation. Moreover, interpretability remains a significant challenge with deep learning approaches, hindering trust and understanding of model decisions. These above mentioned drawbacks are motivated to do this research work.

Below is a summary of this research work's principal contributions.

- In this research, Accounting Analysis by Incorporating Apriori Association Rule Algorithm (AA-IAARA-PMNN-PCBESA) is proposed.
- Develop a Confidence Partitioning Sampling Filtering (CPSF) based preprocessing method for identifying the missing values from the input data.
- Transaction Features are selected using Black Winged Kite Algorithm (BWKA).
- Port-Metriplectic Neural Network (PMNN) is constructed for the anomaly detection in credit card data.
- Propose a PCBESA to optimize Port-Metriplectic Neural Network (PMNN).
- The proposed model's effectiveness is evaluated using current techniques, including as AD-CCD-XGB, PE-CCFD-SVM and CC-FD-RF models respectively.

The remaining manuscripts are arranged as follows: Part 2 describes the proposed methodology, Part 3 proves the result with discussion and Part 4 concludes this manuscript.

II. PROPOSED METHODOLOGY

In this proposed methodology, Accounting Analysis by Incorporating Apriori Association Rule Algorithm (AA-IAARA-PMNN-PCBESA) is proposed. This process consists of five steps: Dataset, Pre-processing using Confidence Partitioning Sampling Filtering (CPSF), Feature Selection using Black Winged Kite Algorithm (BWKA), Anomaly Detection using Port-Metriplectic Neural Network (PMNN) and Optimization using PCBESA. The Block diagram of AA-IAARA-PMNN-PCBESA is confirmed by Figure 1.

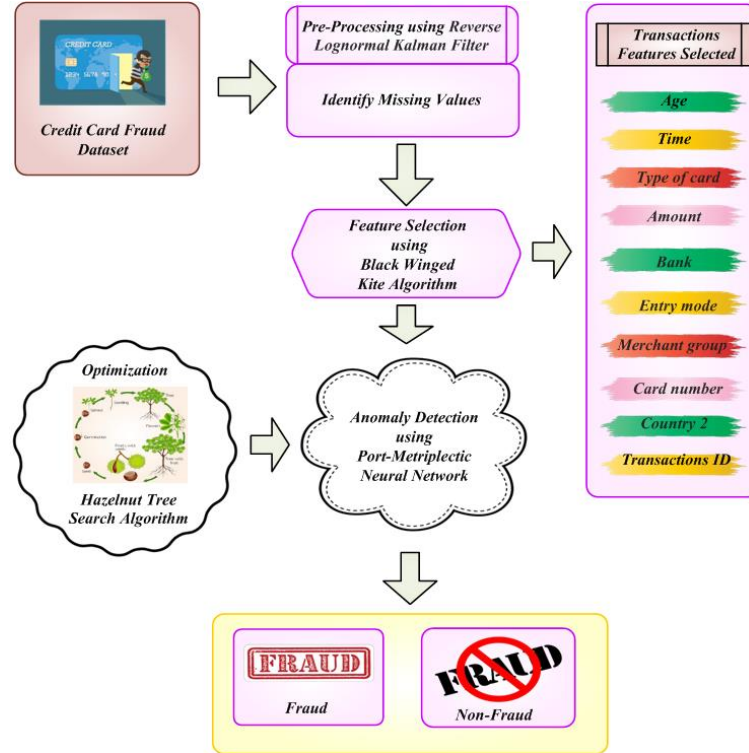


Figure 1: Block Diagram of proposed AA-IAARA-PMNN-PCBESA method

A. Data Acquisition

The Credit Card Fraud Dataset [28] is where the input data are first gathered. The dataset is made up of credit card transactions that cardholders across Europe conducted in September 2013. The transactions that took place over the period of 2 days are displayed in this dataset. 284,807 transactions total; 492 of them were fake. As a result of the dataset's severe imbalance, 0.172% of all transactions are classified as positive (fraud) transactions. Numbers from a PCA transformation are the only variables it accepts. Unfortunately, due to privacy considerations, we are unable to provide the original features and further context for the data. The primary components that PCA was able to extract are features V1 through V28. The two parameters that remain unaltered are "Time" and "Amount." The amount of seconds that pass between each transaction and the initial transaction is stored in the dataset's "Time" attribute. For cost-sensitive learning based on examples, the feature "Amount," which reflects the transaction amount, may be utilised. When fraud is present, the response variable "Class" has a value of 1, and when it is not, it has a value of 0. The feature of credit card fraud dataset is listed in the table 1.

Table 1: Features of Credit Card Fraud Dataset

Sl.No	Features Name	Description
1	Gender	Gender of the card holder
2	Type of card	MasterCard, American express, Visa debit
3	Time	The transaction's date and time
4	Account number	Customer's identification number
5	Country	Country of tax
6	Bank	Issue bank of the card
7	Transaction type	ATM , POS ,Internet
8	Entry mode	magnetic stripe , Pin and chip
9	Transaction ID	Transaction identification number
10	Merchant group	Merchant group identification
11	Card number	The credit card's identification

12	Country 2	Country of residence
13	Merchant code	Identification of the merchant type
14	Age	Card holder age
15	Amount	Amount of the transaction in euros

B. Pre-Processing using Confidence Partitioning Sampling Filtering (CPSF)

In this segment, Pre-Processing using CPSF [29] is discussed. The proposed CPSF is used to identify the missing values from collected data. CPSF concentrates on areas where the model has the highest level of confidence in an effort to increase anomaly detection accuracy. It can improve detection accuracy by giving priority to samples that are more likely to be anomalies by dividing the data according to confidence levels. In order to lower false positive rates, CPSF can help exclude samples with poor model confidence. This is important for credit card fraud detection since it reduces false alarms while correctly identifying fraudulent transactions. Increasing the sensitivity of anomaly detection systems is the main objective of CPSF. It can identify subtle or hitherto undetected abnormalities that conventional approaches might miss by concentrating on samples with higher likelihoods of anomalies. Initially, the Confidence Partitioning Sampling Filtering is given in equation (1),

$$\omega_d = \frac{q(\hat{j}_d)}{\sum_{d=1}^D q(\hat{j}_d)} \tag{1}$$

Where q represents the sampling interval, \hat{j} is used as a dimension for sampling interval, $q(\hat{j}_d)$ is the partitioning sampling input and the probability space may be considerably compressed into a limited probability space ω_d with just a little loss of probability information. D is the maximum space under the probability information. The identification of missing values from the input data is given in equation (2),

$$i(j) = \sum_{d=1}^D \omega_d \delta(j - \hat{j}_d) \tag{2}$$

Where $\sum_{d=1}^D$ represents the weighted impulse function, ω_d represents the bounded space, $(j - \hat{j}_d)$ represents the variety of noise and distortion, δ is the probability data that their weights describe, d is the input of the noisy data and $i(j)$ is the minimized data value from the given (j) input data. The partial prior impulse function is given in equation (3),

$$i_f(j_g | n_{1:g-1}) = \int i(j_g | j_{g-1}) \omega_{g-1,f} \delta(j_{g-1} - \hat{j}_{g-1,f}) dj_{g-1} \tag{3}$$

Where j_g represents the filtering of process noise, j_{g-1} denotes the additive property of the impulse function, ω_{g-1} is the sequences of probability density function, δ is the increase in the accuracy and reliability of the detecting system and i_f is the partial prior impulse function. Then the identified missing values are given in equation (4),

$$\bar{O}_{i(j_g|n_{1:g-1})}^\alpha = \bigcup_{r=1}^{R_{g-1}} (\bar{j}_{g,r} + O_{iw}^\alpha) \tag{4}$$

Where $\bar{O}_{i(j_g|n_{1:g-1})}$ represents the weighted parameters which merged the flat files into data frame of the given data, $\bar{j}_{g,r}$ is the sampled input data, $\bigcup_{r=1}^{R_{g-1}}$ represents the variable for identifying the missing data, O_{iw}^α is the data size reduces undesired changes and \bar{O}^α denotes the number of data values from the given input data. Finally the input data is preprocessed successfully by identifying the missing values using CPSF. Next, the feature selection step receives the pre-processed data..

C. Feature Selection using Black Winged Kite Algorithm (BWKA)

The Black Winged Kite Algorithm (BWKA) is utilized for the feature selection[30]. The BWKA is employed to select the transaction features. The most relevant features are efficiently chosen from a huge pool of transaction data by BWKA. It is possible to increase the anomaly detection system's overall efficiency by concentrating on

its essential aspects. The architecture of BWKA allows it to adjust to changes in the properties of credit card transaction data over time. Finding the transaction features that are most important for spotting irregularities in credit card data is the major purpose of BWKA. The algorithm's goal is to improve the system's overall detection skills by concentrating on these features. Here, step by step procedure for obtaining the transaction features using BWKA is described here. To creates a uniformly distributed population for selecting transaction features. The entire step method is then presented in below,

Step 1: Initialization

Initial population of BWKA is, initially generated by randomness. Then the initialization is derived in equation (5).

$$WA = \begin{bmatrix} WA_{1,1} & WA_{1,2} & \dots & \dots & WA_{1,dim} \\ WA_{2,1} & WA_{2,2} & \dots & \dots & WA_{2,dim} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ WA_{pop,1} & WA_{pop,2} & \dots & \dots & WA_{pop,dim} \end{bmatrix} \tag{5}$$

Here, *WA* is denotes the Black Winged Kite; *pop* is denotes the potential solutions and *dim* is denotes the size of the provided problem dimension.

Step 2: Random Generation

Following setup, random parameters were generated for the input. The fitness value is dependent upon a feature requirement.

Step 3: Fitness Function

From initialised values, a random solution is produced. It is calculated by selecting the transaction features. Then the formula is derived in equation (6).

$$Fitness\ Function = Selecting\ Transaction\ Features \tag{6}$$

Step 4: Attacking Behaviour

Black-winged kites are predators on tiny grassland animals and insects. During a fight, they alter the angles of their wings and tails to suit the wind speed. They hover silently to study their prey, then swiftly plunge and strike. Various assault behaviours are included in this technique for worldwide search and investigation. BWKA quickly explores the search space by combining many search operators, akin to a bird's agile movements as it pursues its prey. Thus it is given in equation (7)

$$z_{t+1}^{i,j} = \begin{cases} z_t^{i,j} + n(1 + \sin(s)) \times z_t^{i,j} & q < s \\ z_t^{i,j} + n \times (2s - 1) \times z_t^{i,j} & else \end{cases} \tag{7}$$

Here, $z_{t+1}^{i,j}$ and $z_t^{i,j}$ is denotes the i^{th} Black-winged kites' t position in the j^{th} dimension in the position and $(t + 1)^{th}$ iteration steps correspondingly; s denotes the random number; q is denotes the count of iteration and n is denotes the count of iteration. This encompasses haphazard investigation, regional focus on viable remedies, and worldwide investigation to include the whole landscape. The mathematical model of black-winged kites' attack behaviour is given in equation (8)

$$n = 0.05 \times d^{-2 \times \left(\frac{t}{T}\right)^2} \tag{8}$$

Here, n is denotes the number of iteration; $\hat{\partial}$ is denotes the constant of detection; d is denotes the dimension of the attacking behaviour; T is denotes the entire count if iteration and t is denotes the constant value.

Step 5: Migration Behaviour

Bird migration is a complicated habit that is influenced by the temperature and availability of food. The purpose of bird migration is to adjust to seasonal variations. During the winter, numerous birds move from the north south in search of better supplies and living circumstances. Leaders are often in charge of migration, and their ability to navigate is essential to the group's success. A continuous probability distribution with two parameters is called a one-dimensional Cauchy distribution. The one-dimensional Cauchy distribution's probability density function is given in equation (9)

$$g(y, \phi, \eta) = \frac{1}{\tau} \frac{\phi}{\sigma^2 + (y - \eta)^2} \tag{9}$$

Here, ϕ and η is denotes the probability density functions; τ is denotes the distributed element; σ is denotes the random position and y is denotes the runtime complexity. In contrast, the population will be guided till it reaches its target if the present population's fitness value is higher than that of the random population. With this approach, great leaders may be dynamically chosen to guarantee a smooth changeover. The mathematical model of black-winged kites' migrate behaviour is given in equation (10)

$$g(y, \phi, \eta) = \frac{1}{\tau} \frac{\phi}{y^2} \tag{10}$$

Here, τ is denotes the distributed element; ϕ and η is denotes the probability density functions and y is denotes the runtime complexity.

Step 6: Termination Criteria

Here, the BWKA is used to select the Transaction feature, and step 3 is repeated iteratively till halting criterion is satisfied. The chosen features are provided as input for anomaly detection in Credit card fraud data. The chosen features are given in table 2.

Table 2: List of selected features by Black Winged Kite Algorithm(BWKA)

Sl.No	Features Name	Description
1	Age	Card holder age
2	Time	The transaction's date and time
3	Type of card	MasterCard, American express , Visa debit
4	Amount	Amount of the transaction in euros
5	Country 2	Country of residence
6	Bank	Issue bank of the card
7	Entry mode	Magnetic stripe , Chip and pin
8	Merchant group	Merchant group identification
9	Card number	The credit card's identification
10	Transaction ID	Transaction identification number

D. Anomaly Detection using Port-Metriplectic Neural Network (PMNN)

In this section, Anomaly Detection in credit card data using Port-Metriplectic Neural Network (PMNN) [31] is discussed. The detected anomaly is classified as Fraud and Non-fraud. From credit card data, PMNN can automatically extract complicated traits that may be challenging to find using conventional techniques. This enables them to identify subtler irregularities. High-dimensional data, which is frequently included in credit card transactions and includes numerous aspects like location, time, amount, etc., can be handled by PMNN. The main objective is to highlight transactions that appear suspicious or that differ from the cardholder's typical spending patterns. This lessens the chance of monetary losses. PMNN is a neural network architecture that combines principles from port-Hamiltonian systems theory with Metriplectic mathematical terms, allowing it to accurately mimic the dynamics of complicated systems such as transaction features is given in equation (11)

$$\dot{y} = M(y) \frac{\partial D}{\partial y} + P(y) \frac{\partial R}{\partial x} \tag{11}$$

Where, $M(y)$ denotes the symmetric, positive semi-definite dissipation matrix.; ∂ is represent the governing variable of time, R is represent the precisely this 2nd potential entropy; $P(y)$ is represent the poison matrix, continues to be skew-symmetric; This formulation to be PMNN can consist of numerous levels, including input, hidden, and output layers, with every layer encompassing nodes that are interconnected or neurons is given in equation (12)

$$B_h = \left\{ (y_s, y_{u+2}) \right\}_{u=1}^U \tag{12}$$

Where, B_h is represent a simple-step state vector that contains labelled pairs; y_u is represent the single-step state vector; $y_{u=1}$ is represent the evolution in time; to train PMNN model is assessed on validation and testing

datasets to determine its ability to detect the anomaly in credit card data. The PMNN model is trained using input data and selected transaction features. Each data is paired with a label it is given in equation (13)

$$\dot{y} = \{z, C\} + [y, P] \tag{13}$$

Where, C is represent the entropy of the pendulum mass; P is represent the second pendulum. PMNN architecture is specifically developed to detect anomaly information acquired from data to classify the anomaly is given in equation (14)

$$L_{bluk}(x) \frac{\partial D_{bluk}}{\partial y} = 0 \tag{14}$$

Here, D_{bluk} represents an irreversible flow of energy at the barrier, while L_{bluk} represents a reversible flux of entropy at the boundary. The PMNN method in anomaly detection and it classifying such as fraud and non-fraud. As a result, equation (15) provides it.

$$\dot{y} = M \frac{\partial D}{\partial y} + P \frac{\partial R}{\partial y} \tag{15}$$

Where, M is represent the skew-symmetry; P is represent the positive semi-definiteness. PMNNs in detected anomaly classification, when anomaly of datas are detected and classified, it helps in improved secured. The classified output are calculated in the equation (16)

$$\dot{y} = -M_{boun} \frac{\partial D_{boun}}{\partial y} - P_{boun} \frac{\partial R_{boun}}{\partial y} \tag{16}$$

Here $-M_{boun}$ and $-P_{boun}$ indicates the calculation of classified detected anomaly. Finally, PMNN classifies detected anomaly as fraud and non-fraud. In this work, PCBESA for accurate classification of anomaly detection in credit card data, this method optimizes the PMNN optimum parameter \dot{y} and $\hat{\partial}y$. Here SETOA is employed for tuning the weight and bias parameter of PCBESA.

E. Optimization using Polar Coordinate Bald Eagle Search Algorithm (PCBESA)

In this segment, PCBESA [32] is described. The PCBESA optimize the PMNN weight parameters \dot{y} and $\hat{\partial}y$ in order to improve the ability to detect and classify anomaly in credit card data. Deep learning models with many parameters require a method for quickly searching through a wide parameter space to discover optimal solutions, and PCBESA is built to do global optimization. The main objective is to use deep learning models to improve credit card data anomaly detection capabilities. PCBESA aims to reduce false positives and increase the accuracy of detecting fraudulent transactions through efficient parameter optimization.

Step 1: Initialization

When the PCBESA is initializing, the polar angle and polar diameter of the person are instantly initialized and saved as an array in the polar coordinate system. Moreover, a non-uniform distribution is the outcome of mapping the starting points Cartesian space is converted from polar coordinate space. It introduces Archimedes theory and cumulative density function in order to prevent the distortion that occurs during the transformation. To get the initialization formula, the CDF is inversed, as shown in equation (17).

$$\rho = ran * (urb - lrb) + lrb, \quad \theta = \beta * \cos^{-1}(2 * ran - 1) \tag{17}$$

Here, ρ represent the polar diameter, urb denotes the upper bound in PBES, lrb denotes the lower bound in PBES, θ represent the polar angle, β represent the disturbance coefficient and ran represent the random number.

Step 2: Random Generation

Weight parameters are established at random. The values generated randomly between 0 and 1.

Step 3: Fitness Function

Fitness function creates random solution form initialized values. It calculated using optimizing parameter. Consequently, equation (18) illustrates it,

$$Fitness\ Function = optimizing[\dot{y}\ and\ \hat{\partial}y] \tag{18}$$

Step 4: Exploration phase for optimizing \dot{y}

During the PCBESA exploring phase, every person's location is updated. The individual location in PCBESA is updated through the renewal of the polar diameter. The formula for updating the polar diameter is given by equations (19) and (20).

$$\rho_{h, new} = \rho_h + l_1 * (\rho_h - \rho_{mean}) + m_1 * \dot{y}(\rho_h - \rho_{h+1}) \quad (19)$$

$$l_1 = \frac{\omega_1}{\max(|\omega_1|)}, m_1 = \frac{\eta_1}{\max(\eta_1)}, \omega_1 = q_1 * \sin(\theta), \eta_2 = q_1 * \cos(\theta), q_1 = z * \pi * ran \quad (20)$$

Here, ρ represent the polar diameter, ran represent the random number, ρ_{mean} show the mean polar diameter following the prior search, $\rho_{h, new}$ indicate the most recent polar dimension, ρ_h denotes the h-th polar diameter, q_1 denotes the control parameter, z represent the the spiral trajectory's controlling parameter, l_1, m_1, ω_1 and η_1 denotes the normalization parameters and θ represent the polar angle.

Step 5: Exploitation phase for optimizing $\hat{\partial}y$

In the PCBESA exploitation phase, the updated polar diameter is swooped in order to enhance the FLNN performance as stated in equations (21) and (22).

$$\rho_{hnew} = ran * \rho_{best} + l_2 * (\rho_h - b_1 * \rho_{mean}) + m_2 * \partial y(\rho_h - b_2 * \rho_{best}) \quad (21)$$

$$l_2 = \frac{\omega_2}{\max(|\omega_2|)}, m_2 = \frac{\eta_2}{\max(\eta_2)}, \omega_2 = q_2 * \sinh(\theta), \eta_2 = q_2 * \cosh(\theta) \quad (22)$$

Here, ρ represent the polar diameter, ran represent the random number, ρ_{mean} show the mean polar diameter following the prior search, $\rho_{h, new}$ represent the updated polar dimension, ρ_h denotes the h-th polar dimension, ρ_{best} serve as the ideal polar diameter at the moment. q_2 Denotes the control parameter, b_1 & b_2 represent the enhancement coefficient, l_2, m_2, ω_2 and η_2 denotes the normalization parameters and θ represent the polar angle. Equation (23), which represents the polar angle mathematically, is the formula.

$$\theta_{h+1} = \beta * \theta_h \pm 2 * \cos^{-1}(2 * ran - 1) \quad (23)$$

Here, θ represent the polar angle, ran represent the random number and β represent the disturbance coefficient. The convergence efficiency and individual updating speed will both be significantly increased by the PCBESA approach. By modifying the weight parameter of PMNN, the PCBESA approach successfully improves its performance for the detection and classification of anomaly in credit card data.

Step 6: Termination Criteria

In the PMNN, the weight parameters for generators are optimized using the PCBESA, dynamically adjusting weights inspired by celestial mechanics. The iterative refinement guided by halting criteria $\rho = \rho + 1$, ensures optimal weight convergence, maximizing PMNN generator performance. Then flowchart of PCBESA for optimizing the weight parameters of PMNN for enhanced anomaly detection and classification of credit card data is shown in figure 2.

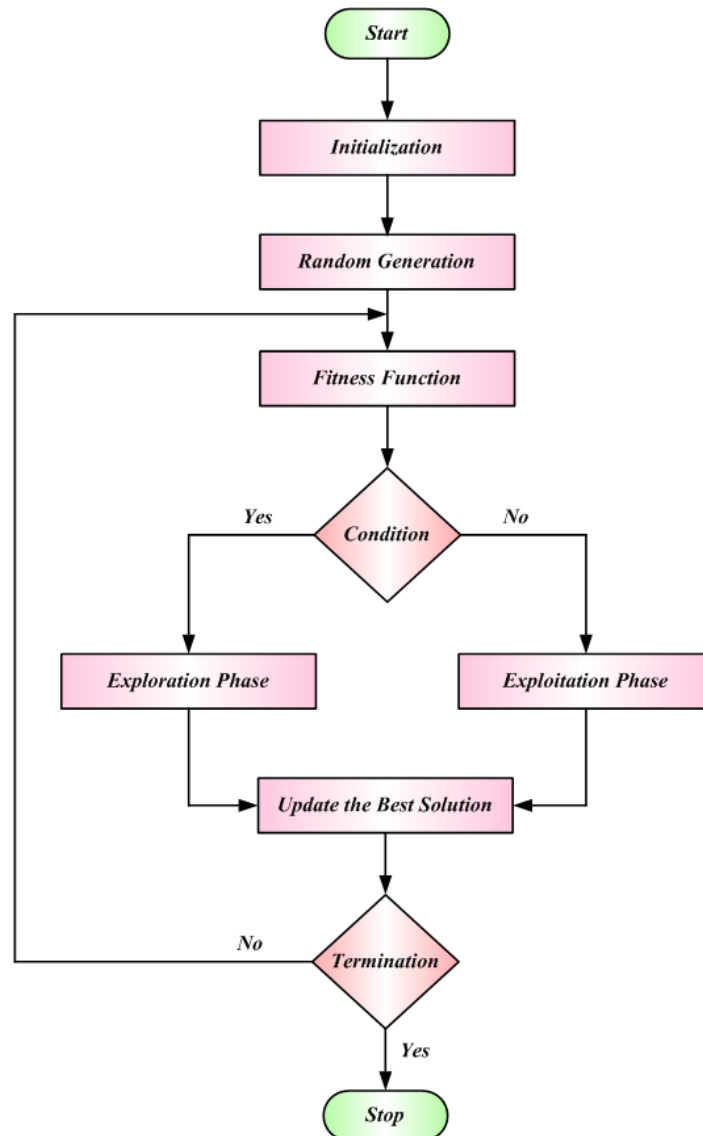


Figure 2: Flowchart of Polar Coordinate Bald Eagle Search Algorithm (PCBESA)

III. RESULT AND DISCUSSION

The outcome of the proposed AA-IAARA-PMNN-PCBESA technique have detected the anomaly in the credit card data. This proposed method is implemented using Python and evaluated by using several performance analysing metrics like, Recall, F1-score, , Precision, Accuracy Mathews Correlation Coefficient, Detection Rate and ROC are analysed. The results of the proposed AA-IAARA-PMNN-PCBESA method are contrasted to those existing methods such as AD-CCD-XGB [21], PE-CCFD-SVM [22] and CC-FD-RF [23].

A. Performance Metrics

Performance measures include Mathews Correlation Coefficient, F1-score , Accuracy, Recall, Precision, Detection Rate and ROC. The confusion matrix will be used to scale the performance parameters, it is decided.

- True Negative (TN): Accurately categorized Anomaly detection in Credit Card Data as Fraud.
- True Positive (TP): Accurately categorized Anomaly detection in Credit Card DataNon-fraud.
- False Positive (FP): Inaccurately categorized Anomaly detection in Credit Card Data as Fraud.
- False Negative (FN): Inaccurately categorized Anomaly detection in Credit Card Dataas Non-fraud.

1) Accuracy

The definition of accuracy is "the ratio of accurately identified fraudulent transactions to the total number of fraudulent transactions in the test dataset." To compute the metric, use equation (24)

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (24)$$

2) Precision

"The number of fraudulent transactions that are actually fraud divided by the total number of fraud estimated by the model as fraud" is the definition of precision: "from all the fraudulent transactions which model estimated as fraud." Equation (25) is used to construct this measure, which is used to detect true fraud from fraudulent transactions.

$$Precision = \frac{TP}{(TP + FP)} \quad (25)$$

3) Sensitivity

"The number of fraudulent transactions actually are fraud divided by the total number of transactions labelled as fraudulent" is the definition of recall: "from all the fraudulent transactions which model estimated as fraud." This statistic aids in providing a response to the query of what percentage of fraudulent transactions are identified and quantified by equation (26)

$$Sensitivity = \frac{TP}{TP + FN} \quad (26)$$

4) F1-score

"Harmonic average of recall and precision from the fraudulent transactions dataset, where an F1-score reaches its worst value at 0 and best at 1 (perfect recall and precision) and worst at 0" is the definition of the F1-score. Equation (27) is used to compute this metric, which is used to assess the model's entire performance.

$$F1 - Score = \frac{Precision * Recall * 2}{(Precision + Recall)} \quad (27)$$

5) AUC

AUC is the computation of the complete two-dimensional region beneath the ROC curve, which can be obtained by evaluating the joint performance of all potential classification thresholds. Thus, it is given in equation (28)

$$AUC = \frac{1}{|l_+| * |l_-|} \sum_1^{|l_+|} \sum_1^{|l_-|} \alpha(l_+, l_-) \quad (28)$$

6) Matthews Correlation Coefficient

The degree of correlation between the estimation and actual class labels is computed using the Matthews Correlation Coefficient .It takes one if and only if the actual and estimated class labels coincide. The formula for calculating it is (29)

$$MCC = \frac{(TN * TP) - (FP * FN)}{\sqrt{(TP + FN)(TN + TP)(TP + FN)(TP + FP)}} \quad (29)$$

7) Detection Rate

Equation (30) may be used to quantify detection rate, which is defined as "how accurately the model is estimating the true positive cases or how precisely the model is tackling the fraudulent cases in the fraudulent transactions."

$$Detection\ rate = \frac{TP}{(TP + FN)} \quad (30)$$

B. Performance analysis

Figure 3 to 9 shows simulation result of AA-IAARA-PMNN-PCBESAmethod. The performance metrics are analyzed with existing likes AD-CCD-XGB, PE-CCFD-SVM and CC-FD-RF methods.

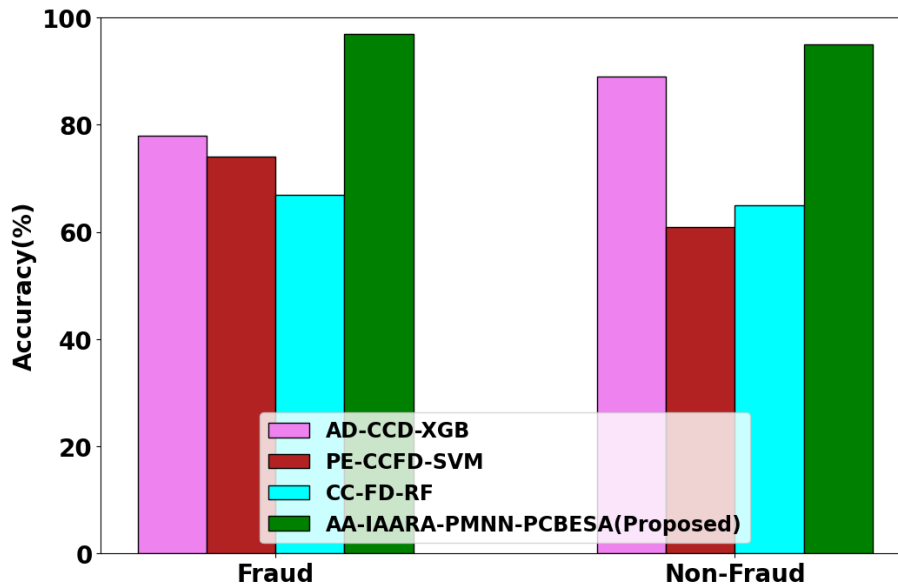


Figure 3: Performance Analyses of Accuracy

The performance analyses of accuracy is depicted in figure 3. An accuracy graph could demonstrate how a proposed AA-IAARA-PMNN-PCBESA model accuracy fluctuates with various models, for identifying transactions as fraudulent or non-fraudulent, in the context of credit card fraud detection or anomaly detection. The trade-offs between false positives and true positives can be better understood with the aid of this representation, which also assist can determine which model parameters are best for getting the required degree of accuracy. The proposed AA-IAARA-PMNN-PCBESA method attains 19.5%, 24% and 23% higher accuracy for fraud and 17.5%, 22.33% and 27.21% higher accuracy for non-fraud when compared to existing methods such as AD-CCD-XGB, PE-CCFD-SVM and CC-FD-RF respectively.

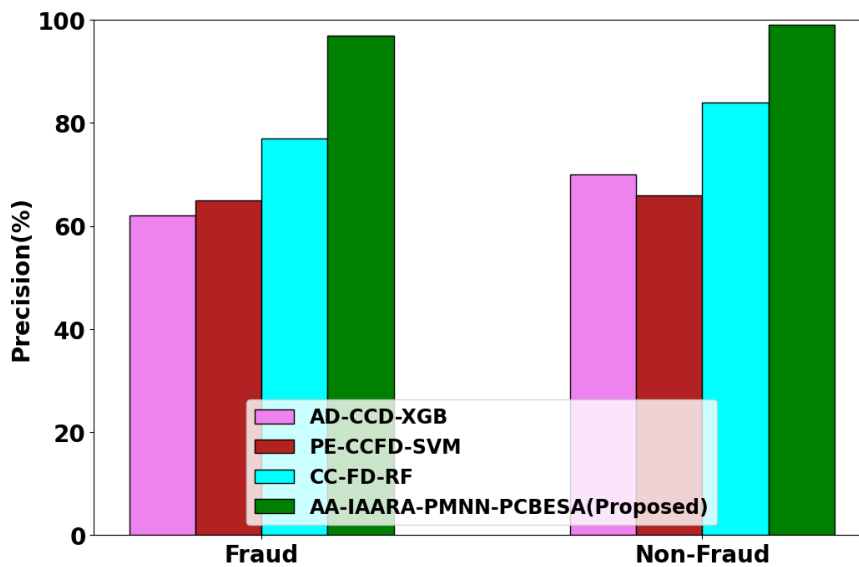


Figure 4: Performance Analyses of Precision

The performance Analyses of precision is depicted in figure 4. In all circumstances, the proposed AA-IAARA-PMNN-PCBESA model obtains the best precision of 99.86%. Additionally, it improves its precision by 99.97% in non-fraud situations and 97.82% in fraud cases. This suggested model can improve the performance to detect from credit card datasets when compared to other previous efforts. The proposed AA-IAARA-PMNN-PCBESA method attains 24.6%, 23.15% and 24.8% higher Precision for fraud and 27.5%, 28.33% and 25.21% higher precision for non-fraud when compared to existing methods such as AD-CCD-XGB, PE-CCFD-SVM and CC-FD-RF respectively.

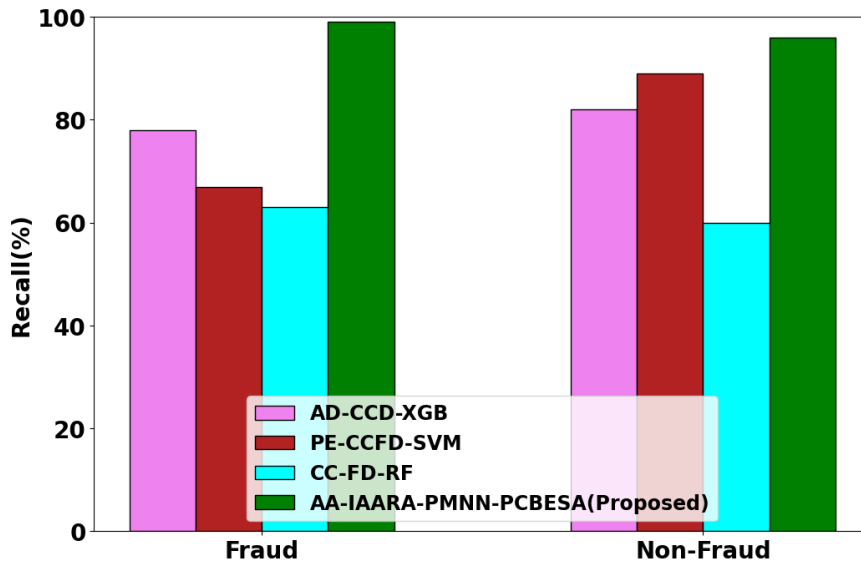


Figure 5: Performance Analyses of Recall

The performance analysis of recall is illustrated in figure 5. When a recall graph is used in conjunction with a credit card fraud dataset, it can offer important information on how well the proposed AA-IAARA-PMNN-PCBESA model detects fraudulent transactions and help with decision-making when it comes to refining the model for fraud detection. The proposed AA-IAARA-PMNN-PCBESA method attains 21.15%, 26.33% and 27.25% higher recall for fraud and 21.4%, 27.36% and 21.08% higher recall for non-fraud when compared to existing methods such as AD-CCD-XGB, PE-CCFD-SVM and CC-FD-RF respectively.

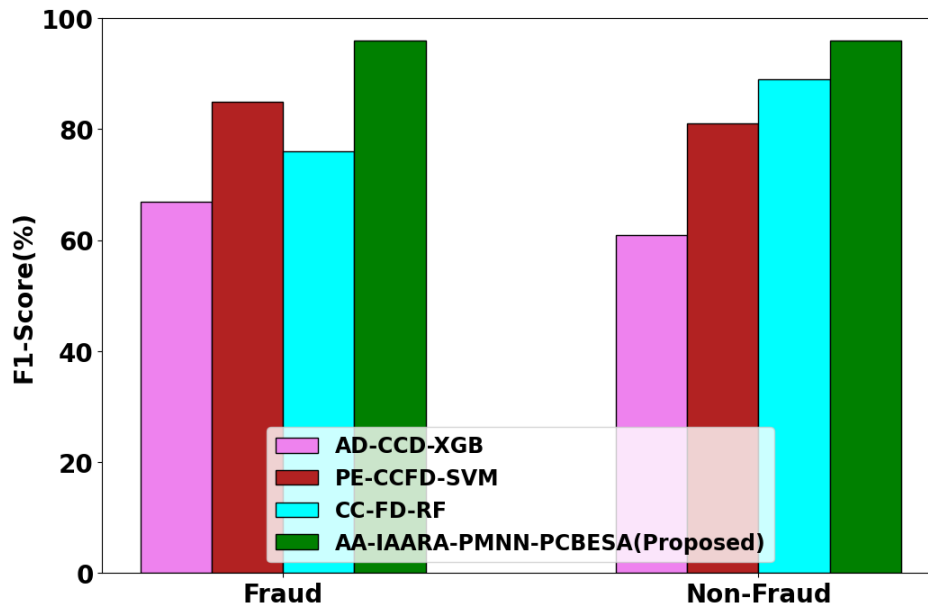


Figure 6: Performance Analyses of F1-Score

The performance analyses of F1-Score is depicted in figure 6. A high F1-score in credit card fraud detection indicates that the proposed AA-IAARA-PMNN-PCBESA model can detect fraudulent transactions with a minimal amount of misclassifications. It is a thorough metric that takes into account both the precision of fraud forecasts and the capacity to identify fraud. The proposed AA-IAARA-PMNN-PCBESA method attains 28.3%, 27.4% and 31.0% higher F1-score for fraud and 22.25%, 23.13% and 24.33% higher F1-Score for non-fraud when contrasted to existing methods such as AD-CCD-XGB, PE-CCFD-SVM and CC-FD-RF respectively.

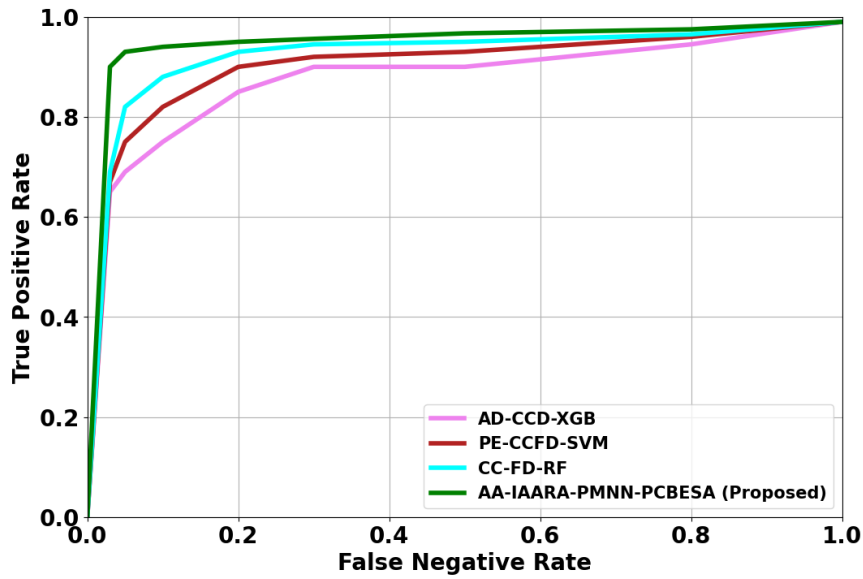


Figure 7: Performance Analyses of ROC

The performance analyses of ROC is depicted in figure 7. When evaluating the effectiveness of credit card fraud detection models, the ROC curve is a useful tool since it offers information on how well the proposed AA-IAARA-PMNN-PCBESA models can differentiate between fraudulent and authentic transactions when compared to other approaches. The proposed AA-IAARA-PMNN-PCBESA method attains 6.433%, 2.332%, and 3.223% greater ROC when compared to existing methods such as AD-CCD-XGB, PE-CCFD-SVM and CC-FD-RF respectively.

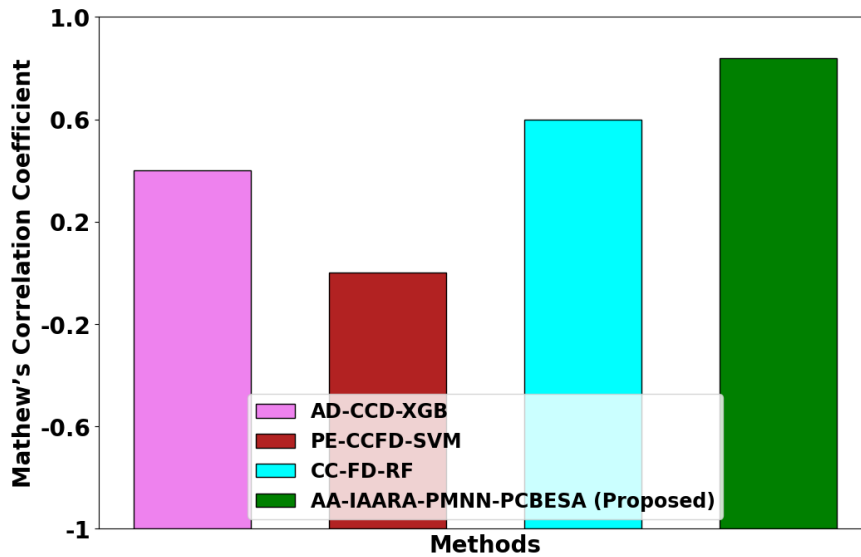


Figure 8: Analyses of Mathew's Correlation Coefficient Performance

Analyses of Mathew's Correlation Coefficient Performance is depicted in figure 8. The Matthews Correlation Coefficient, which accounts for true negatives, false negatives, true positives, and false positives, in situations involving the detection of credit card fraud, is a trustworthy indication of the efficacy of the suggested AA-IAARA-PMNN-PCBESA classification model. The proposed AA-IAARA-PMNN-PCBESA method attains 7.43%, 5.33%, and 4.27% greater Mathew's Correlation Coefficient when compared to existing methods such as AD-CCD-XGB, PE-CCFD-SVM and CC-FD-RF respectively.

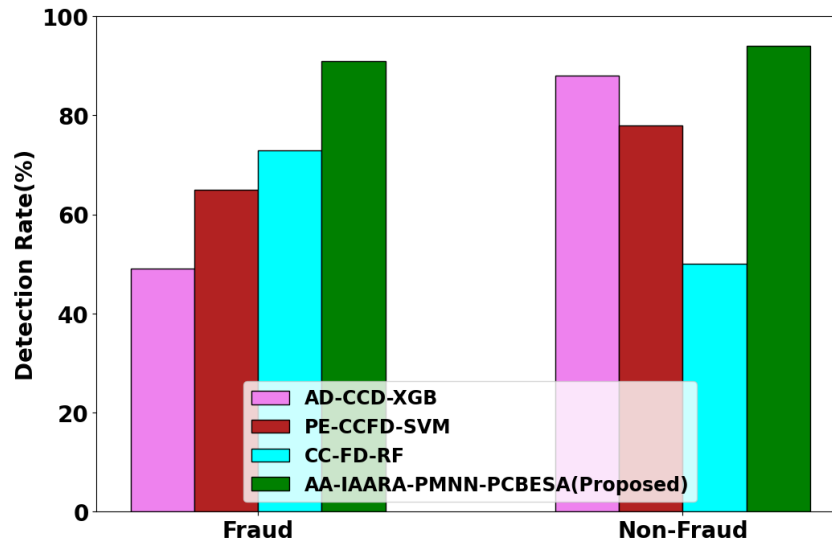


Figure 9: Performance Analyses of Detection Rate

The performance analyses of Detection Rate is depicted in figure 9. When assessing the effectiveness of proposed AA-IAARA-PMNN-PCBESA classification model, detection rate is a crucial parameter. This is especially true for credit card fraud detection, where it is crucial to correctly identify fraudulent transactions. The proposed AA-IAARA-PMNN-PCBESA method attains 27.43%, 25.33%, and 24.27% higher detection rate for fraud and 25.32%, 26.35% and 24.54% higher detection rate for non-fraud when compared to existing methods such as AD-CCD-XGB, PE-CCFD-SVM and CC-FD-RF respectively.

C. Discussion

A novel Accounting Analysis by Incorporating Apriori Association Rule Algorithm (AA-IAARA-PMNN-PCBESA) is developed in this paper. To protect financial institutions, cardholders, and the larger financial ecosystem from fraud and illegal activity, anomaly detection in credit card data is crucial. By facilitating proactive risk management, operational efficiency, regulatory compliance, and client protection, it eventually contributes to a safe and reliable payment environment. In the realm of automated research, anomaly detection of credit card data is a crucial subject that needs the greatest focus. In this work, the input data was collected from credit card fraud dataset. Our primary focus in this study has been to detect anomaly in the credit card data and classify such as fraud and non-fraud. The pre-processing technique Confidence Partitioning Sampling Filtering (CPSF) is used to identify the missing values. The transaction features are selected by using Black Winged Kite Algorithm (BWKA). The highly effective deep learning approach called Port-Metriplectic Neural Network (PMNN) optimized with PCBESA is proposed to guarantee excellent accuracy. For all of the performance metrics looked at, this classifier performs better than the others at detecting the anomaly from the credit card data. AA-IAARA-PMNN-PCBESA performs better in terms of F1-score, Recall, Precision, Accuracy, Detection Rate, Mathew's Correlation Coefficient and ROC are analysed when compared to AD-CCD-XGB, PE-CCFD-SVM and CC-FD-RF methods respectively.

V. CONCLUSION

In this manuscript, Accounting Analysis by Incorporating Apriori Association Rule Algorithm (AA-IAARA-PMNN-PCBESA) was successfully implemented. Accounting analysis offers a distinct benefit when it comes to identifying irregularities in credit card data since it makes use of domain-specific knowledge, guarantees adherence to legal requirements, and presents a comprehensive picture of financial activities. This methodology improves interpretability and transparency, fosters contextual understanding, and allows for ongoing development, all of which contribute to increased efficacy in credit card transaction fraud detection and prevention. This study has described a classification of detected anomaly in credit card data based on PMNN optimized with Polar Coordinate Bald Eagle Search Algorithm (PCBESA) and the different feature descriptors. PMNN-PCBESA is used to classify the detected anomalies in the credit card data as fraud and non-fraud. The proposed AA-IAARA-PMNN-PCBESA approach is implemented in Python. The performance of the proposed AA-IAARA-PMNN-PCBESA approach contains 28.3%, 27.4% and 31.0% higher F1-score, 6.433%, 2.332%,

and 3.223% greater ROC and 25.32%, 26.35% and 24.54% higher detection rate when analyzed to the existing methods like AD-CCD-XGB, PE-CCFD-SVM and CC-FD-RF methods respectively. Future research utilizing accounting analysis to identify abnormalities in credit card data may use unstructured data sources, including transaction descriptions, and cross-disciplinary cooperation to create interdisciplinary techniques. Anomaly detection accuracy and effectiveness could be increased by enhancing the interpretability of anomaly detection models, putting dynamic risk assessment frameworks into place, utilizing behavioral analytics, investigating block chain technology, and taking temporal patterns into account. These actions would ultimately support fraud prevention measures in financial transactions.

Acknowledgement

Project of the 14th Five Year Plan for Education Science in Shanxi Province in 2021: Research on the Talent Training Model of Higher Vocational Education under the Background of Supply Side Reform (Project Number: GH-21255)

REFERENCES

- [1] Carcillo, F., Le Borgne, Y.A., Caelen, O., Kessaci, Y., Oblé, F. & Bontempi, G. (2021). Combining unsupervised and supervised learning in credit card fraud detection. *Information sciences*, 557, 317-331.
- [2] Karthik, V.S.S., Mishra, A. & Reddy, U.S. (2022). Credit card fraud detection by modelling behaviour pattern using hybrid ensemble model. *Arabian Journal for Science and Engineering*, 47(2), 1987-1997.
- [3] Osegi, E.N. & Jumbo, E.F. (2021). Comparative analysis of credit card fraud detection in simulated annealing trained artificial neural network and hierarchical temporal memory. *Machine Learning with Applications*, 6, 100080.
- [4] Vanini, P., Rossi, S., Zvizdic, E. & Domenig, T. (2023). Online payment fraud: from anomaly detection to risk management. *Financial Innovation*, 9(1), 66.
- [5] Zioviris, G., Kolomvatsos, K. & Stamoulis, G. (2022). Credit card fraud detection using a deep learning multistage model. *The Journal of Supercomputing*, 78(12), 14571-14596.
- [6] Shi, P., Zhao, Z., Zhong, H., Shen, H. & Ding, L. (2021). An improved agglomerative hierarchical clustering anomaly detection method for scientific data. *Concurrency and Computation: Practice and Experience*, 33(6), e6077.
- [7] Leevy, J.L., Hancock, J. & Khoshgoftaar, T.M. (2023). Comparative analysis of binary and one-class classification techniques for credit card fraud data. *Journal of Big Data*, 10(1), 118.
- [8] Forough, J. & Momtazi, S. (2021). Ensemble of deep sequential models for credit card fraud detection. *Applied Soft Computing*, 99, 106883.
- [9] Mienye, I.D. & Sun, Y. (2023). A deep learning ensemble with data resampling for credit card fraud detection. *IEEE Access*, 11, 30628-30638.
- [10] Benchaji, I., Douzi, S., El Ouahidi, B. & Jaafari, J. (2021). Enhanced credit card fraud detection based on attention mechanism and LSTM deep model. *Journal of Big Data*, 8, 1-21.
- [11] Asha, R.B. & KR, S.K. (2021). Credit card fraud detection using artificial neural network. *Global Transitions Proceedings*, 2(1), 35-41.
- [12] Madhurya, M.J., Gururaj, H.L., Soundarya, B.C., Vidyashree, K.P. & Rajendra, A.B. (2022). Exploratory analysis of credit card fraud detection using machine learning techniques. *Global Transitions Proceedings*, 3(1), 31-37.
- [13] Langevin, A., Cody, T., Adams, S. & Belling, P. (2022). Generative adversarial networks for data augmentation and transfer in credit card fraud detection. *Journal of the Operational Research Society*, 73(1), 153-180.
- [14] Gao, F., Li, J., Cheng, R., Zhou, Y. & Ye, Y. (2021). Connet: Deep semi-supervised anomaly detection based on sparse positive samples. *IEEE Access*, 9, 67249-67258.
- [15] Paldino, G.M., Lebichot, B., Le Borgne, Y.A., Sibli, W., Oblé, F., Boracchi, G. & Bontempi, G. (2024). The role of diversity and ensemble learning in credit card fraud detection. *Advances in Data Analysis and Classification*, 18(1), 193-217.
- [16] Zhang, X., Han, Y., Xu, W. & Wang, Q. (2021). HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture. *Information Sciences*, 557, 302-316.
- [17] Dumitrescu, B., Băltoiu, A. & Budulan, Ş. (2022). Anomaly detection in graphs of bank transactions for anti money laundering applications. *IEEE Access*, 10, 47699-47714.
- [18] Itoo, F., Meenakshi & Singh, S. (2021). Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection. *International Journal of Information Technology*, 13(4), 1503-1511.
- [19] Habibpour, M., Gharoun, H., Mehdipour, M., Tajally, A., Asgharnezhad, H., Shamsi, A., Khosravi, A. & Nahavandi, S. (2023). Uncertainty-aware credit card fraud detection using deep learning. *Engineering Applications of Artificial Intelligence*, 123, 106248.

- [20] Xie, Y., Liu, G., Yan, C., Jiang, C., Zhou, M. & Li, M. (2022). Learning transactional behavioral representations for credit card fraud detection. *IEEE Transactions on Neural Networks and Learning Systems*.
- [21] Islam, M.A., Uddin, M.A., Aryal, S. & Stea, G. (2023). An ensemble learning approach for anomaly detection in credit card data with imbalanced and overlapped classes. *Journal of Information Security and Applications*, 78, 103618.
- [22] Ileberi, E., Sun, Y. & Wang, Z., (2021). Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost. *IEEE Access*, 9, 165286-165294.
- [23] Ileberi, E., Sun, Y. & Wang, Z. (2022). A machine learning based credit card fraud detection using the GA algorithm for feature selection. *Journal of Big Data*, 9(1), 24.
- [24] Bakhtiari, S., Nasiri, Z. & Vahidi, J., (2023). Credit card fraud detection using ensemble data mining methods. *Multimedia Tools and Applications*, 82(19), 29057-29075.
- [25] Hemdan, E.E.D. & Manjaiah, D.H., (2022). Anomaly credit card fraud detection using deep learning. *Deep Learning in Data Analytics: Recent Techniques, Practices and Applications*, 207-217.
- [26] Esenogho, E., Mienye, I.D., Swart, T.G., Aruleba, K. & Obaido, G., (2022). A neural network ensemble with feature engineering for improved credit card fraud detection. *IEEE Access*, 10, 16400-16407.
- [27] Alarfaj, F.K., Malik, I., Khan, H.U., Almusallam, N., Ramzan, M. & Ahmed, M., (2022). Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. *IEEE Access*, 10, 39700-39715.
- [28] <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>
- [29] Qiang, X., Xue, R. & Zhu, Y. (2024). Confidence partitioning sampling filtering. *EURASIP Journal on Advances in Signal Processing*, 2024(1), 24.
- [30] Wang, J., Wang, W.C., Hu, X.X., Qiu, L. & Zang, H.F. (2024). Black-winged kite algorithm: a nature-inspired meta-heuristic for solving benchmark functions and engineering problems. *Artificial Intelligence Review*, 57(4), 1-53.
- [31] Hernández, Q., Badías, A., Chinesta, F. & Cueto, E. (2023). Port-metriplectic neural networks: thermodynamics-informed machine learning of complex physical systems. *Computational Mechanics*, 72(3), 553-561.
- [32] Zhang, Y., Zhou, Y., Zhou, G., Luo, Q. & Zhu, B., (2022). A curve approximation approach using bio-inspired polar coordinate bald eagle search algorithm. *International Journal of Computational Intelligence Systems*, 15(1), 30.