e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

AN EFFICIENT FRAUDULENT ACTIVITY RECOGNITION FRAMEWORK USING DECISION TREE ENABLED DEEP ARTIFICIAL NEURAL NETWORK

Gaurav Anand*1, Bharatwaja Namatherdhal *2

*1TDS Telecommunications LLC, USA

*2Adobe Inc, USA

DOI : https://www.doi.org/10.56726/IRJMETS36978

ABSTRACT

Customers today prefer using credit cards as the most common form of payment because it makes online shopping more comfortable and bill paying simpler. The hazards associated with credit card fraud transactions are a major issue that needs to be avoided. There are numerous data mining approaches that can be used to successfully reduce these risks, but they all had poor accuracy and high computing complexity when it came to detecting credit card fraud. Regarding this, the decision tree enabled deep artificial neural network (DT enabled deep ANN) is used in this research to detect fraudulent credit card activity. The raw data, which includes the cardholder, terminal, and time delta information, is first gathered and preprocessed. Using the decision tree enabled ANN; the users are classified as genuine and fraudulent after extracting the acquired data. The decision tree aids in decision-making based on past data, and the ANN is used in this context to analyze complex patterns. The combined use of these classifiers aids in the efficient analysis of the data and produces a result that is more precise and effective. The proposed DT enabled deep ANN achieved 93.46% accuracy, 93.16% sensitivity, and 92.68% specificity, which is significantly higher efficiency than other existing approaches.

Keywords: Credit card fault detection, Decision Tree, ANN, Hidden Markov Model and DT enabled deep ANN

I. INTRODUCTION

Due to a recent rise in popularity of this method of payment, the majority of people now utilize it instead of cash when making routine payments in their everyday lives. Credit cards are present in the wallets of the majority of Americans, claim [9] [8]. 7 out of 10 Americans currently hold one or more credit cards. Customers who use credit cards can simply keep track of their spending and understand where their money is going [1]. Clients are not constrained in their spending, unlike the cash strategy, which is limited to the money in your wallet [5]. Additionally, the majority of enterprises and organizations are now focusing on online services as a result of the quick expansion of the usage of current technology in all disciplines. As a result, a consumer needs a credit card in order to use services and make transactions online [2]. This makes making purchases with cash challenging and time-consuming. According to the 2015 Nilson Report, damages from debit and credit card theft rose by 19% to $16.31 billion in 2014. Internet fraud is expected to have cost the economy a total of $3.5 billion in 2012 [10] [4], according to Cyber Sources (2013). The two basic types of fraud involving credit cards are behavior fraud and application fraud [11, 12]. When an application for a credit card is fraudulent, that is called application fraud. It occurs when a fraudster applies for a new credit card using fraudulent identification, and the issuer of the card approves the application [6]. Behavior fraud occurs after the credit card is approved and issued. It makes reference to credit debit card transactions that appear to be fraudulent. The identification and avoidance of fraud remain important challenges for credit card companies as well as important academic areas [4] since identifying and halting even a small portion of criminal activity would prevent hundreds of millions of millions in damages. Recently, the ML community has paid a lot of attention to deep learning (DL), which has also demonstrated significant positive results in a number of uses, including recognition of words, language processing, and machine vision [7]. To detect fraud in the use of credit cards, a variety of techniques are applied, such as information mining and artificial intelligence algorithmic approaches. However, the findings are typically unsatisfactory. Consequently, it is necessary to build effective and efficient algorithms that work considerably. By applying artificial neural network method and comparing it to a few other machine learning algorithms, the use of our credit card by fraudsters can be stopped before the purchase is authorized [9]. The primary goal of the research is to identify credit card errors using deep artificial neural networks (DT enabled deep ANNs). The raw data, which includes the cardholder, terminal, and time delta information, is first gathered and preprocessed. After the data has been preprocessed, the Hidden Markov Model (HMM) is used to extract the features, which aids in

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

resolving the temporal probabilistic reasoning that was independent of transition. Using the decision tree enabled ANN; the users are classified between genuine and fraudulent after extracting the acquired data. The main contributions of the research are as follows,

- **Decision tree enabled ANN:** T

he users are classified as genuine or fraudulent by utilizing the decision tree enabled ANN. The decision tree aids in decision-making based on past data, and the ANN is used in this context to understand complex patterns. The combined use of these classifiers aids in the efficient analysis of the data and produces a result that is more precise and effective.

II. METHODOLOGY

The review on credit fault detection is discussed in the below section, Xinwei Zhang et al. [4] presented a novel feature extraction system with an architecture for deep learning for fraud with credit cards detection. Despite having a high calculating cost, this technique was an effective and useful tool for identifying credit card fraud. Emmanuel Ileberi et al. [5] developed a genetic algorithm-based detection of credit card fraud engine that uses machine learning to recognize attributes. The class imbalance dataset issue was resolved by this strategy; however it had a low detection accuracy rating. Using a deep learning algorithm, Joy Iong-Zong Chen and Kong-Long Lai [6] devised a financial fraud detection technique based on the DCNN scheme. Higher performance was gained using this strategy, although training was extremely challenging. In order to identify fraudulent behavior, Javad Forough, Saeedeh Momtazi and colleagues [7] developed an innovative system for voting based on artificial artificial neural networks and an ensemble framework based on sequential data modeling. The accuracy and AUC score of this technique were greater, but only linearly separable patterns were learned. A fraud detection technique utilizing Random Forest was presented by Rashmi S. More et al. [8] and can assist in resolving this issue in the real world. Although this approach improved performance and accuracy, it had an overfitting issue.

2.1 Challenges

The challenges faced in credit card fault detection is discussed in the below section,

- An unbalanced dataset and a dynamic environment present major obstacles for fraud detection systems.
- Misclassified information could be a significant problem as not all fraudulent behaviour is detected or recorded.
- The main issues faced by financial fraud detection schemes include constantly changing fraudulent behavior, a lack of a mechanism for tracking fraud transaction information, the limitations of machine learning algorithms, and the difficulty of training other models and algorithms with highly skewed financial fraud datasets.

III. MODELING AND ANALYSIS

In this research, a deep artificial neural network with decision trees enabled (Decision tree enabled deep ANN) is used to identify credit card fraud. The raw data, which includes the cardholder, terminal, and time delta information, is first gathered and preprocessed. The word card holder refers to a specific person, terminal denotes the interface that facilitates electronic payments, and time delta tells us how different times are expressed in different units. After the data has been preprocessed, the HMM is used to extract the features, which aids in resolving the temporal probabilistic reasoning that was independent of transition. After extracting the data, the users are classified as genuine or fraudulent by utilizing the decision tree enabled ANN. The decision tree aids in decision-making based on past data, and the ANN is used in this context to understand complex patterns. The combined use of these classifiers aids in the efficient analysis of the data and produces a result that is more precise and effective. Figure 1 shows the proposed credit card fault detection model.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

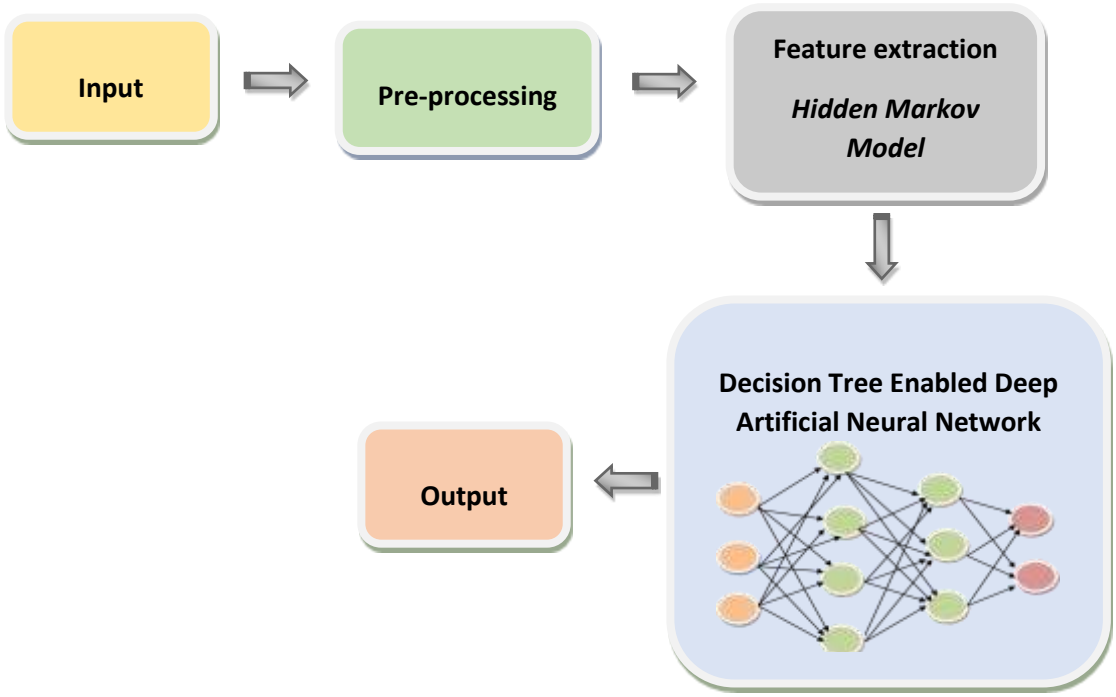


Figure 1: Credit card fault detection model

3.1 Input

Initially, the data is collected from the dataset which consists of information about the card holder, terminal and time delta which is mathematically represented as follows,

$C = \sum_{n=1}^{z} C_n$ (1)

where, the dataset is denoted as C, the dataset information's is denoted as C_n, which ranges from 1 to z.

3.2 Pre-processing

This dataset might contain duplicate, incomplete, or null values; as a result, the preprocessing stage removes these values, formats the data, and enhances it to a standard. This step also improves the dataset's accuracy and quality and raises its dependability. The preprocessed equation is represented mathematically as,

$C = \sum_{n=1}^{z} C_n^*$ (2)

Where, C_n^*represent the preprocessed data.

3.3 Feature extraction

Feature extraction is a process that converts raw data into manageable numerical features while preserving the original data set's information. The accuracy of learned models is improved by the collection of characteristics from the input data. The feature extraction process used in this research is hidden Markov model and is discussed in the section below.

3.3.1 Hidden Markov Model

Following preprocessing, features are extracted using a HMM, which aids in resolving the temporal probabilistic reasoning that was independent of transition. A group of statistical models called HMM are used to describe a signal's statistical characteristics. A set of density functions for probability corresponding to each state, as well as an underlying, invisible Markov chains with a limited number of states, a transition from one state to another probability matrix, and a starting probability distribution, are the two interrelated processes that make up an HMM. Among an HMM's components are, M is the model's total number of states. $D = \{D_1, D_2, \ldots . D_M\}$ if D is the collection of states. $w_s \in D, 1 \leq s \leq L$, where L is the observation length sequence, gives the state of the model at time s. The variety of observation symbols isN. $H = \{H_1, H_2, \ldots . H_N\}$ if H is the collection of all observation potential symbols, often known as the model of codebook. $B = \{b_{ij}\}$, where B is the state transition probability matrix,

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

$b_{ij} = Q[w_s = D_j | w_{s-1} = D_i] 1 \leq i, j, \leq M$ (3)

With the constraint, $0 \leq b_{ij} \leq 1$, and $\sum_{i=1}^{M} b_{ij} = 1, 1 \leq j \leq M$. A is the symbol probability observation matrix, that is $A = \{a_j(f)\}$, where

$a_j(f) = Q[U_s = h_f | w_s = D_i], 1 \leq i \leq M, 1 \leq f \leq N$ (4)

WhereC is the starting state distribution and $U_s$is the observation symbol at times, $C = \{\pi_j\}$, which is stated mathematically as,

$\pi_j = Q[w_1 = D_i], 1 \leq j \leq M$ (5)

A HMM is denoted by the triplet form in a shorthand notation and is shown as,

$\beta = (B, A, C)$ (6)

The description given above describes a discrete HMM, in which the discrete observations symbols selected at random from a finite alphabet, where$H = \{H_1, H_2, \ldots . H_N\}$. Continuous observation density functions serve as the states' identifiers in a continuous density HMM. A finite mixture with the following formula is the most generic expression of the model probability density function,

$a_j(U) = \sum_{f=1}^{N} d_{jf} M(U, \mu_{if}, p_{if}), 1 \leq j \leq M$ (7)

where, $d_{jf}$is the coefficient of mixture of the k^{th} mixture in state j. It is assumed that $M(U, \gamma_{if}, p_{if})$ is a mean vector of Gaussian μ_{if} and a matrix of covariance p_{if} without losing generality.

3.4 Decision tree enabled ANN

After extracting the data, the users are classified either as genuine or fraudulent by utilizing the decision tree enabled ANN. The decision tree aids in decision-making based on previous information, and the ANN is used in this context to analyze complex patterns. The combined use of these classifiers aids in the efficient analysis of the data and produces a result that is more precise and effective. The following section discusses a thorough explanation of the classifiers.

3.4.1 Decision tree

The decision trees approximation method for discrete valued objective functions uses a decision tree for expressing the learned function. The decision tree's nodes each represent a test of an instance property, and every branch descending from those nodes represents one of the attribute's potential values. The categorization of the occurrences is provided by the decision tree. The procedures involved in categorizing an instance involve beginning at the lowest node of the tree, assessing the property described by the node, and then moving along the tree branch that matches the numerical value of the attribute in the example supplied. The same process is then applied to the sub tree that is now rooted on the new node. The usage of decision trees is one of the efficient strategies widely used in a range of fields, including as ML, processing images, and pattern recognition [35]. The DT, a sequential approach in which each test evaluates a numerical attribute to a threshold value, effectively and cogently unites a series of fundamental tests [36]. Conceptual rules are much easier to construct than numerical weights in a neural network of interconnections between nodes [37, 38]. DT is primarily utilized for grouping. Additionally, data mining commonly employs the categorization model known as DT [39]. There are nodes and branches in every tree. Each subset provides a value that a node may accept, and each node represents an attribute in an identified category [40, 41]. Due to their simple analysis and correctness across a wide range of data sources, decision tree models have found use in several application domains [42]. When there are a few nodes present, a DT is demonstrated to be the best alternative for more precise categorization of the dataset. The local wasteful search model, which assumes data gain as the aim function, is widely employed in the DT to divide the classes, is expressed as.

$C_{DT} = 1 - \sum_{\chi} V_{DT_\chi}^2$ (8)

where V_{DT_χ} represents the probability of the χ^{th}class.

3.4.2 Artificial Neural network

A supervised machine learning technique called an ANN takes inspiration from how the human brain functions. Input, hidden, and output layers make up the core elements of the simplest ANN. The size of the input layer depends on the amount of features in a particular dataset. The complexity of a task can influence the hidden layer

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

size, and the sort of problems that need to be solved can influence the output layer size. A node or neuron is the fundamental part of an ANN. The network first gets the various input variables corresponding to known observations through the input layer during the training process. The hidden layer neurons receive these input variables and process and extract important information from them in order to forecast output values. Through an initialization function, weights are set for each connection. Using a learning function, the network iteratively modifies these weights during training so that the projected output value for each training point corresponds as nearly as feasible to the known goal output value. The network takes the unknown point data for the entire area during classification and classifies those using calibrated weights. A graphic illustration of a straightforward ANN is shown in Figure 2.

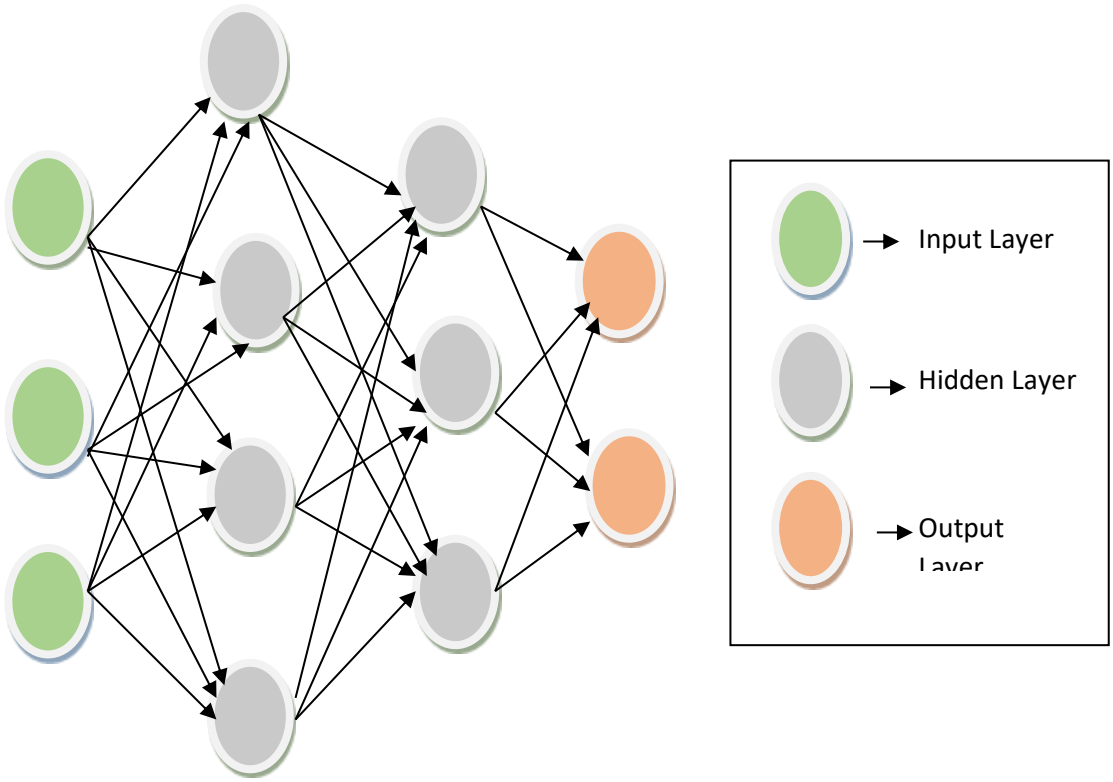


Figure 2: ANN architecture

IV. RESULTS AND DISCUSSION

This section describes the accuracy, sensitivity, and specificity for the training percentage (TP) performance measures for the fraudulent activity recognition framework-based DT enabled ANN classifier.

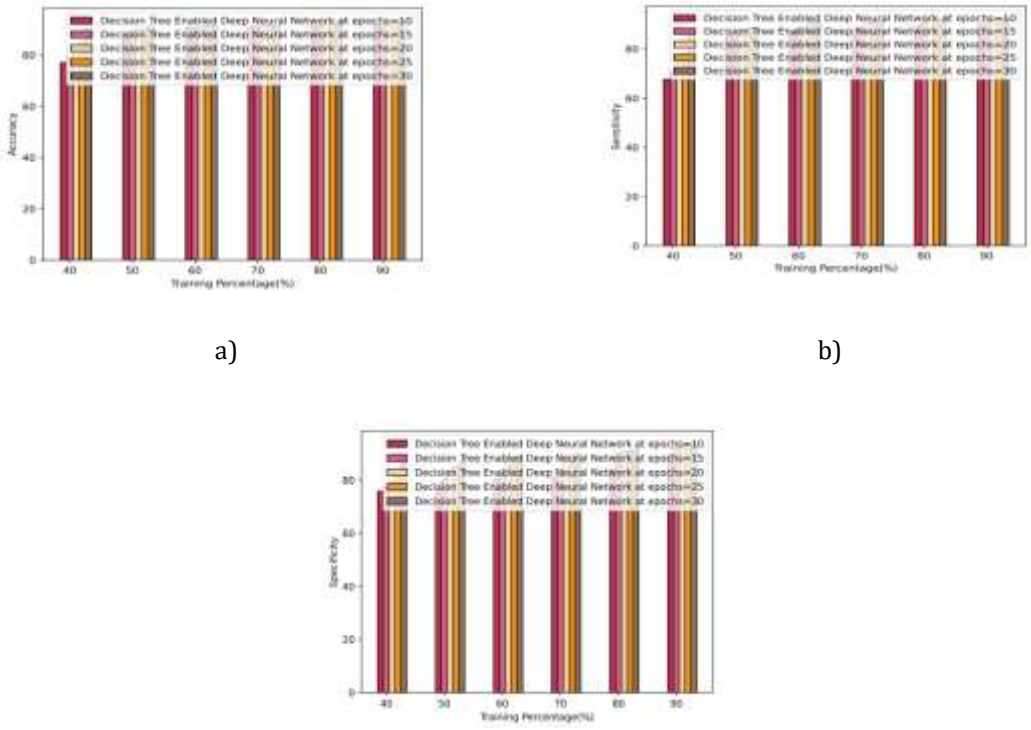
4.1 Experimental setup

The implementation of credit card fault detection determined by DT enabled ANN classifier uses the PYTHON tool in the Windows 10 System with 8GB RAMS.

4.2 Performance evaluation based on TP

Figure 3 displays the results of the performance research for the suggested credit card defect detection based on DT enabled ANN classifier for the varied epochs 10, 15, 20, 25, and 30 for the TP 40, 50, 60, 70, 80, and 90. The DT enabled ANN classifier illustrated in figure 3 a) for the TP 90 obtained values of 91.83%, 91.99%, 92.03%, 92.54%, and 93.60% at the beginning when the accuracy of the techniques was measured. The sensitivity of the DT enabled ANN classifier is also evaluated; it yielded values of 89.89%, 92.01%, 92.04%, 92.58%, and 92.70% for the TP of 80 shown in figure 3 b). Attained values of 89.66%, 92.61%, 93.48%, 93.65%, and 93.83% for the TP 80 are presented in figure 3 c). Lastly, the specificity of the DT enabled ANN classifier is evaluated.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com



a) b)

c)

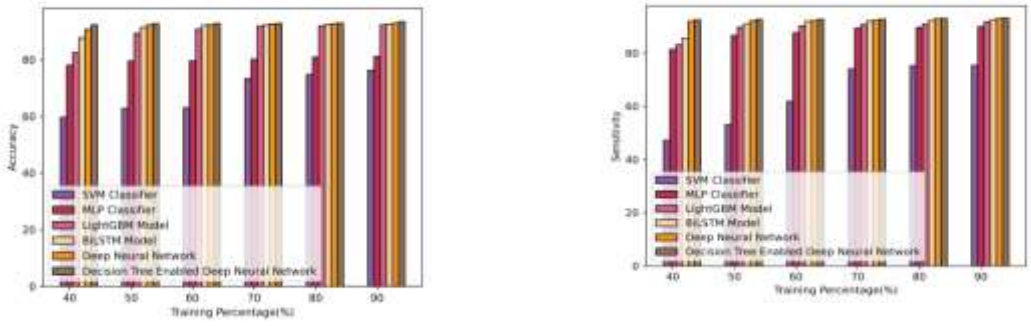
Figure 3. Performance evaluation with TP a) accuracy b) sensitivity c) specificity

4.3 Comparative methods

The proposed method of the DT enabled ANN classifier is compared with Support Vector Machine (SVM) [CD-1] [13], Multi-layer Perceptron (MLP) [CD-2] [14], Light gradient-boosting machine (LightGBM) [CD-3] [15], Bidirectional long short-term memory (BiLSTM) [CD-4] [16], and Deep Neural Network [CD-5] [17].

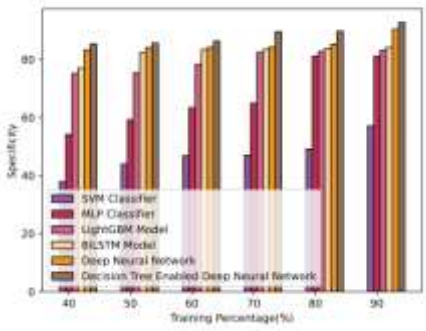
4.3.1 Comparative evaluation based on TP

Figure 4 illustrates schematically how the parameter metrics are assessed and compared using the training percentage. For ease of measurement, the proposed DT enabled ANN classifier's improvement percentage is used. In terms of accuracy, the DT enabled ANN classifier improved by 0.57% over DNN during 90% training. In terms of similarity, the DT enabled ANN classifier improved by 0.21% over DNN during 90% training. Finally, in terms of specificity, the DT enabled ANN classifier improved by 2.54% over DNN during 90% training. The analysis demonstrates that the proposed DT enabled ANN classifier method outperformed the previously used strategies.



a) b)

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com



c)

Figure 4. Comparative evaluation with TP a) accuracy b) sensitivity c) sensitivity

4.3.2 Comparative discussion

The best results are shown in table 1 together with a comparison of the credit fault identification model based on deep neural networks with DT capabilities. In comparison to all other ways, the suggested method yields the best outcomes.

Table 1. Comparative discussion of the proposed credit fault detection model based on DT enabled deep NN

Methods/ Metrics	TP 90		
	Accuracy (%)	Sensitivity (%)	Specificity (%)
CD-1	76.31	75.36	57.06
CD-2	81.10	89.86	81.12
CD-3	92.27	91.65	83.15
CD-4	92.60	92.46	84.06
CD-5	92.93	92.97	90.32
Proposed	93.46	93.16	92.68

V. CONCLUSION

In this research, a deep artificial neural network with decision trees enabled (DT enabled deep ANN) is used to detect credit card fraud. The raw data, which includes the cardholder, terminal, and time delta information, is first gathered and preprocessed. After the data has been preprocessed, the HMM is used to extract the features, which aids in resolving the temporal probabilistic reasoning that was independent of transition. Using the decision tree enabled ANN; the users are classified between genuine and fraudulent after extracting the acquired data. The decision tree aids in decision-making based on past data, and the ANN is used in this context to analyze complex patterns. The combined use of these classifiers helps in the efficient analysis of the data and produces a result that is more precise and effective. The accuracy, sensitivity, and specificity of the suggested DT enabled deep ANN were 93.46%, 93.16%, and 92.68%, respectively. This is a significant improvement over other approaches already in use. Future research can use the proposed method to develop and evaluate massive real-time data sets with other machine learning techniques.

VI. REFERENCES

[1] Roy, Abhimanyu, Jingyi Sun, Robert Mahoney, Loreto Alonzi, Stephen Adams, and Peter Beling. "Deep learning detecting fraud in credit card transactions." In 2018 Systems and Information Engineering Design Symposium (SIEDS), pp. 129-134. IEEE, 2018.

[2] Najadat, Hassan, Ola Altiti, Ayah Abu Aqouleh, and Mutaz Younes. "Credit card fraud detection based on machine and deep learning." In 2020 11th International Conference on Information and Communication Systems (ICICS), pp. 204-208. IEEE, 2020.

[3] Trivedi, Naresh Kumar, Sarita Simaiya, Umesh Kumar Lilhore, and Sanjeev Kumar Sharma. "An efficient credit card fraud detection model based on machine learning methods." International Journal of Advanced Science and Technology 29, no. 5 (2020): 3414-3424.

[4] Zhang, Xinwei, Yaoci Han, Wei Xu, and Qili Wang. "HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture." Information Sciences 557 (2021): 302-316.

[5] Ileberi, Emmanuel, Yanxia Sun, and Zenghui Wang. "A machine learning based credit card fraud detection using the GA algorithm for feature selection." Journal of Big Data 9, no. 1 (2022): 1-17.

[6] Chen, Joy Iong-Zong, and Kong-Long Lai. "Deep convolution neural network model for credit-card fraud detection and alert." Journal of Artificial Intelligence 3, no. 02 (2021): 101-112.

[7] Forough, Javad, and Saeedeh Momtazi. "Ensemble of deep sequential models for credit card fraud detection." Applied Soft Computing 99 (2021): 106883.

[8] More, Rashmi, Chetan Awati, Suresh Shirgave, Rashmi Deshmukh, and Sonam Patil. "Credit card fraud detection using supervised learning approach." Int. J. Sci. Technol. Res 9 (2021): 216-219.

[9] M. Zareapoor, P. Shamsolmoali et al., "Application of credit card fraud detection: Based on bagging ensemble classifier," Procedia computer science, vol. 48, no. 2015, pp. 679–685, 2015.

[10] N. Mahmoudi, E. Duman, Detecting credit card fraud by modified fisher discriminant analysis, Expert Syst. Appl. 42 (5) (2015) 2510-2516.

[11] C. Phua, R. Gayler, V. Lee, K. Smith-Miles, On the communal analysis suspicion scoring for identity crime in streaming credit applications, Eur. J. Oper. Res. 195 (2) (2009) 595-612

[12] U. Fiore, A.D. Santis, F. Perla, P. Zanetti, F. Palmieri, Using generative adversarial networks for improving classification effectiveness in credit card fraud detection, Inform. Sciences 479 (2019) 448-455.

[13] Youssef, Ahmed M., Biswajeet Pradhan, Abhirup Dikshit, and Ali M. Mahdi. "Comparative study of convolutional neural network (CNN) and support vector machine (SVM) for flood susceptibility mapping: a case study at Ras Gharib, Red Sea, Egypt." Geocarto International (2022): 1-28.

[14] Al Bataineh, Ali, Devinder Kaur, and Seyed Mohammad J. Jalali. "Multi-layer perceptron training optimization using nature inspired computing." IEEE Access 10 (2022): 36963-36977.

[15] Liang, Junchao, Yude Bu, Kefeng Tan, Jingchang Pan, Zhenping Yi, Xiaoming Kong, and Zhou Fan. "Estimation of stellar atmospheric parameters with light gradient boosting machine algorithm and principal component analysis." The Astronomical Journal 163, no. 4 (2022): 153.

[16] Yang, Biao, Yinshuang Wang, and Yuedong Zhan. "Lithium Battery State-of-Charge Estimation Based on a Bayesian Optimization Bidirectional Long Short-Term Memory Neural Network." Energies 15, no. 13 (2022): 4670.

[17] Borisov, Vadim, Tobias Leemann, Kathrin Seßler, Johannes Haug, Martin Pawelczyk, and Gjergji Kasneci. "Deep neural networks and tabular data: A survey." IEEE Transactions on Neural Networks and Learning Systems (2022).