



Boosting Financial Fraud Detection Using Parameter Tuned Ensemble Machine Learning Model

Reem Atassi^{*1}, Aziz Zikriyov², Nurbek Turayev², Sagdullayeva Gulnora Botirovna²

¹ Higher Colleges of Technology, United Arab Emirates

² Tashkent state University of economics, Uzbekistan

Emails: ratassi@hct.ac.ae; a.zikriyov@tsue.uz; n.turayev@tsue.uz; gulbotir82@gmail.com

Abstract

Fraud detection in the financial industry is a challenging area as financial transactions gradually shift to digital platforms. More and more businesses such as the financial industry are operationalizing their services online as the usage of the internet is growing exponentially. Accordingly, financial fraud can increase in number and forms worldwide leading to remarkable financial losses that make financial fraud a main challenge. Threats such as irregular attacks and unauthorized access must be identified through a financial fraud detection system. Over the past few years, data mining and machine learning (ML) approaches have been widely used to address these issues. However, this technique has yet to be enhanced in terms of speed computation, identifying unknown attack patterns, and dealing with big data. This study presents Financial Fraud Detection using the Parameter Tuned Ensemble Machine Learning (FFD-PTEML) method. The FFD-PTEML incorporates multiple advanced components, such as z-score normalization for feature scaling and ensemble classification employing Artificial Neural Networks (ANN), Multilayer Perceptron (MLP), and Radial Basis Function (RBF) networks. The use of z-score normalization ensures uniformity in feature distribution, improving the effectiveness and interpretability of the fraud detection technique. Furthermore, the ensemble classification model combines the strength of different neural network architectures to enhance the detection performance and resilience to complicated fraud patterns. FFD-PTEML demonstrates better performance than the classical technique through extensive experimentation on real-time financial datasets, exhibiting high sensitivity and specificity in fraudulent activity detection.

Keywords: Financial Fraud Detection; Machine Learning; Radial Basis Function; Z-Score Normalization; Fintech Industry

1. Introduction

The fintech industry has perceived amazing growth and distraction currently, redesigning the scenery of economic services globally. With advanced platforms presenting whole digital experiences, fintech has developed traditional banking, payments, investment, and providing progressions [1]. The author well-known that this quick progress has also transported forth novel tasks, particularly in the realm of security and fraud. Financial fraud is a major fear for both fintech companies and financial organizations [2]. The growth of digital trades and the rising trust in skill have delivered fraudsters with novel ways to utilize exposures in the method. The fintech platforms manage huge sizes of sensitive financial information and enable many transactions every day; they become major objectives for fake actions [3]. Furthermore, the fast-paced nature of fintech processes demands present fraud recognition and prevention abilities, demanding innovative tools and models.

Despite numerous struggles to decrease financial fraudulent actions, its durability disturbs the economy and society adversely, as huge amounts of money go to fraud daily [4]. Numerous fraud recognition techniques were presented several years ago. Most classical models are physical, and this is time-consuming, imprecise expensive, and impractical. More research has been conducted to decrease losses resulting from fraudulent actions, but they are not effective [5]. With the development of the artificial intelligence (AI) technique, data mining (DM) and machine

learning (ML) must be utilized for identifying fake actions in the economic part. Both unsupervised and supervised approaches were used for forecasting fraud actions. Classification techniques have been the most popular model for identifying economic fraudulent dealings [6]. In this situation, the 1st phase of model training utilizes a database with feature vectors and class labels. Then, the trained technique is employed in order to categorize test samples in the subsequent step [7].

In the latter period, technical developments joined with access to an extensive range of data which have enhanced business fraud recognition [8]. Previous research on business fraud recognition concentrated mainly on utilizing logistic regression. However, with the improvement in ML technique and the accessibility of high computing influence in current years, more researchers have developed organized ML approaches [9]. The latest studies presented that ML models are efficiently used to identify economic fraud and overtake the traditionally employed fraud recognition techniques utilized in preceding studies. However, it is noticeable that most of the preceding ML works on corporate fraud recognition mainly trusted utilizing individual ML classifiers to build their fraud detection technique and did not take advantage of ensemble learning (EL) [10].

This study presents Financial Fraud Detection using the Parameter Tuned Ensemble Machine Learning (FFD-PTEML) method. The FFD-PTEML incorporates multiple advanced components, such as z-score normalization for feature scaling and ensemble classification employing Artificial Neural Networks (ANN), Radial Basis Function (RBF), and Multilayer Perceptron (MLP) networks. The use of z-score normalization ensures uniformity in feature distribution, improving the effectiveness and interpretability of the fraud detection technique. Furthermore, the ensemble classification model combines the strength of different neural network architectures to enhance the detection performance and resilience to complicated fraud patterns. FFD-PTEML demonstrates better performance than the classical technique through extensive experimentation on real-time financial datasets, exhibiting high sensitivity and specificity in fraudulent activity detection.

2. Literature Survey

Khalid et al. [11] developed an innovative ensemble model that combines a Random Forest (RF), K-Nearest Neighbor (KNN), Boosting classifiers, Support Vector Machine (SVM), and Bagging. The ensemble method deals with the dataset imbalance complexity related to maximum credit card databases by applying the Synthetic undersampling and Oversampling technique (SMOTE) model in a few ML algorithms. Zhao and Bai [12] introduced a novel technique dependent upon the ML method. This technique demonstrates 5 distinct classification methods and 3 ensemble algorithms for predicting the financial fraud data of registered companies, comprising DT, RF, LR, XGBOOST, SVM, and ensemble methods with the voting classifier. In conclusion, the optimum distinct architecture is elected from five ML techniques and the preeminent ensemble model between all hybrid techniques. In [13], a fraud feature-improving method was developed with a spiral oversampling balancing technique (SOBT). Particularly, a compound grouping elimination approach has been offered for feature extraction. Moreover, a multifactor synchronous embedding technique was developed that integrates the effectiveness assessment metrics of the embedding method for every feature. Besides, an SOBT was developed to balance the ratio of illegal to fraudulent transactions.

Duan et al. [14] presented an innovative technique for credit card fraud detection, the Casusal Temporal Graph Neural Network (CaT-GNN) method which comprises two main mechanisms: Causal-Inspector and Causal-Intervener. The causal intervener executes a causal mixed-up improvement under setting nodes reliant on the set nodes. Causal-Inspector applies attention weights at the temporal attention mechanism for recognizing the causal and surrounding nodes. Next, the. In [15], an enhanced Financial Statement Fraud (FSF) identification technique was proposed. An FSF method was presented by implementing a great ensemble method, the XGBoost method. The problem of class imbalance in the database is resolved by employing the SMOTE method. The technique also employs various ML approaches in Python for predicting the FSF. The method enhances the XGBoost technique. Zhao et al. [16] presented self-attention generative adversarial networks (SAGANs) technique. Leveraging self-attention mechanisms, SAGANs will differentiate noticeable features and patterns with wide-ranging transaction databases, thus raising further insightful consideration and improved recognition. By integrating the GANs, this technique can be proficient in producing data that reflects real fraudulent behaviour that subsequently enriches and improves fraud detection methods.

3. The Proposed Model

In this study, we have presented a novel FFD-PTEML method. The FFD-PTEML incorporates multiple advanced components, such as z-score normalization for feature scaling and ensemble classification employing ANN, MLP, and RBF networks. Fig. 1 defines the entire procedure of the FFD-PTEML method.

For the header and the footer, just change the journal name and the abbreviation, then leave all other information for our production team at the ASPG editorial office to be updated after your paper acceptance.

This article gives linear model, which is the direct simplex method using neutrosophic logic, the logic that is the new vision of modelling and is designed to effectively address the uncertainties inherent in the real world founded by the Romanian mathematician Florentine Smarandache [1, 2]. In addition to that, Ahmed A. Salama presented the theory of neutrosophic classical categories as a generalization of the theory of classical categories [12,20], also, he developed, introduced, and formulated new concepts in the various disciplinary of mathematics, statistics, computer science by neutrosophic theory [17,18,19,22,28].

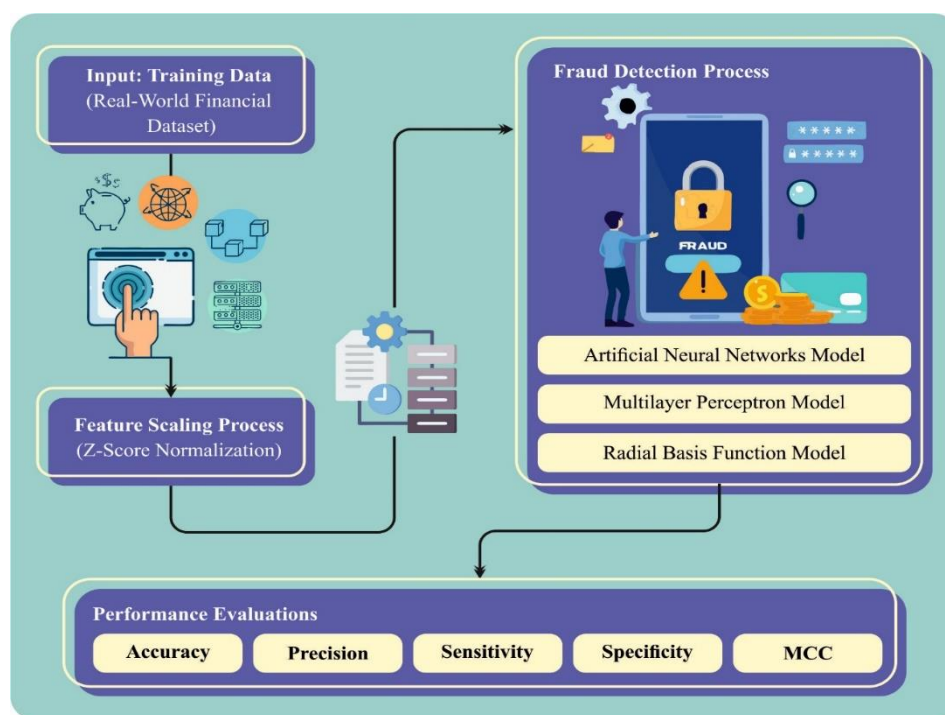


Figure 1: Overall process of FFD-PTEML method

A. Z-score Normalization

The Z-score normalization method was used to scale the feature values to have a mean zero and a variance of 1 [17]. This can be performed by subtracting the mean from all the values and later dividing by the variance. To perform a z-score normalization on every value in a dataset, we use the subsequent equation:

$$\text{New value} = \frac{(x - \mu)}{\sigma} \quad (1)$$

Where x : Original value, μ : Mean, σ : variance.

B. Ensemble Learning Process

The ensemble classification model combines the strength of different neural network architectures to enhance detection performance and resilience to complicated fraud patterns.

1. Artificial Neural Networks

Neural networks (NNs) are named as electronic methods of the neural architecture of the human begin's brain [18]. The method of training and learning has been essentially dependent upon involvement. Electronic methods of natural NNs have been also dependent upon similar techniques, and the mode of systems dealing with issues varies from the computational techniques that are generally employed by computer systems. Virtual or computer-NN have been only capable of simulating a smaller portion of these characteristics and features of biological neural networks (BNNs). Additionally, the main aim of make an ANN instead of imitating the human brain, which can be a technique to resolve engineering complexity motivated by the behavioral pattern of BNNs. In BNNs, neurons

have been linked to one another in a 3D architecture. The interconnections among neurons in BNNs are several and complicated it cannot be possible to develop an identical ANN. Nowadays, incorporated circuit technology permits us to make NNs in 2D architecture. The human being's brain comprises a huge amount of neurons for processing different data and identifying the surrounding world. Alternatively, neurons of the human beings brain obtain data exists new neurons over dendrites. Such neurons collect the input data organized, and once it surpasses a threshold, it will be activated (Fire). Besides, the activated signal can be linked to alternative neurons via axons.

Mostly, ANNs employ complex instances as inputs and outputs for the arithmetical model and offer an accurate outcome in the forthcoming unidentified inputs. Back-propagation (BP) method has been a major prevalent method for resolving the error values of the ANNs that can be an adapted form of the Least Mean Squares (LMS) method. With the consideration of output, χ for an ANN, y for the inputs, the computed equation will be expressed as Eq. (2):

$$y(t) = f_a \left(\sum_i \omega_i x_i + b_i \right) \quad (2)$$

Here, ω_i refers to the i^{th} member of weight, f_a defines the activation function, b_i denotes the i^{th} member of the bias, and χ and y indicate, the input and the output as earlier stated. According to earlier indications, the major popular technique for feed-forward neural networks (FFNNs) represents BP. The method utilizes the gradient descent method and needs numerous rounds, measures the weights for producing the desirable outcome, and evaluates the network errors. The local optima is completely dependent upon the initial weight values, the major problem with utilizing the gradient descent method. A global technique must be employed to balance such drawbacks. In this study, gradient descent could be changed with a recently proposed optimization technique to overcome the assessment.

2. Multilayer Perceptron

In ML as an ensemble method, meta-learners are models established to study from the assumptions or forecasts of several base models to improve entire forecast efficiency [19]. During this case, MLP was utilized as a meta-learner. MLP is a controlling and fully connected (FC) dense layer data-driven modeling tool in ANNs that receives inputs and creates outcomes to the chosen size. An MLP is an NN composed of several layers; it takes an input layer, many hidden layers (HL), and an output layer, all of which have many nodes. Employing an MLP as a meta-learner offers the benefit of modeling difficult connections among the base model forecasts and targeted variables. During the ensemble, it captures intricate forms that can be unexploited by linear models; making them appropriate to handle difficult connections from the integrated forecasts of the base learners. MLPs naturally acquire appropriate features in the input data that are highly effective once the base learners create a huge feature count. MLPs are concentrated on the most helpful aspects, increasing the effectiveness and efficiency of the ensemble technique.

MLPs are many parameters that are adjusted, permitting a maximum degree of customization. Appropriate tuning of the MLP structure can lead to enhanced solutions and optimum alteration to the certain problem at hand. However, it is essential to research distinct structures and parameters for the MLP to determine the formation that works better for your certain time sequence predicting problem. To declare the ensemble's solution on unobserved data, suitable testing, and validation are also important.

3. Radial Basis Function Networks

The RBF-NN is represented by a direct three-layer model that has the ability to estimate some constant operation with random accuracy [20]. In this architecture, the intermediary layer performs as the HL, utilizing the RBF as the activation method. Determined as a monotonously improving performance of the Euclidean distance in space among the specific center c and specified point x , the RBF function represents diverse categories, with the popular Gaussian function. It can be represented as given below:

$$p_j = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{(x - c_j)^2}{2\sigma_j^2} \right) j = 1, 2, \dots, n \quad (3)$$

Here c_j means the center of the j th node of HL, x characterizes the input; σ_j represents the width of the j th node at the HL. Fig. 2 portrays the infrastructure of the RBF network.

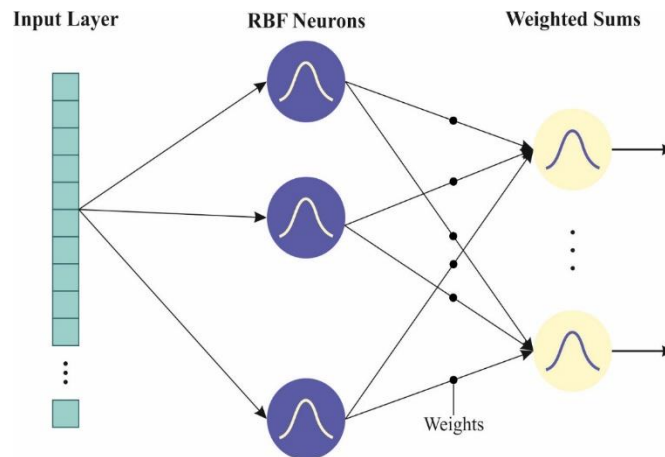


Figure 2: Structure of RBF network

Although the conventional RBF-NN method exhibits a simple architecture, rapid computation, and strong nonlinear processing abilities, the efficiency is dependent upon determining crucial parameters in the learning stage. Accordingly, recognizing the optimum parameters for the network develops the leading one.

4. Result Analysis and Discussion

The performance outcome of the FFD-PTEML method is examined using the Kaggle dataset [21], comprising 900 instances with 2 classes as illustrated in Table 1.

Table 1: Details on database

Class	No. of Instances
Non-Fraud	450
Fraud	450
Total Samples	900

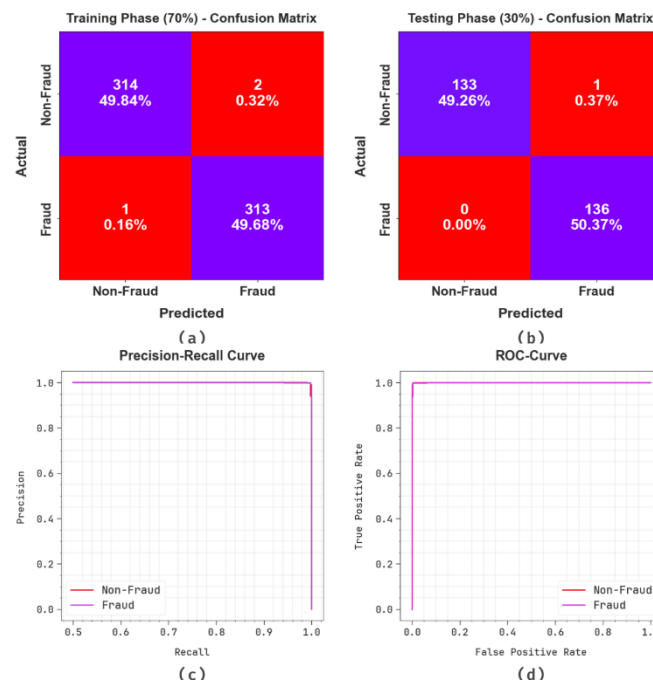


Figure 3: Classifier result of (a-b) Confusion matrices and (c-d) PR and ROC curves

Fig. 3 displays the classifier outcomes of the FFD-PTEML system at the test dataset. Figs. 3a-3b showcases the confusion matrices offered by the FFD-PTEML model on 70:30TRAS/TESS. This figure pointed out that the FFD-PTEML technique can be identified and categorized into two classes correctly. Meanwhile, Fig. 3c illustrates the PR result of the FFD-PTEML system. The figure noted that the FFD-PTEML method provides maximal PR effectiveness in each class. In conclusion, Fig. 3d represents the ROC result of the FFD-PTEML method. This figure represents that the FFD-PTEML technique offers effective outcomes with improved ROC values with different class labels.

The fraud detection results of the FFD-PTEML algorithm are provided in terms of distinct measures in Table 2 and Fig. 4. These experimentation outcome values pointed out that the FFD-PTEML technique pointed out the fraud and non-fraud transactions proficiently. Based on 70%TRAS, the FFD-PTEML technique offers average $accu_y$, $prec_n$, $sens_y$, $spec_y$, and MCC of 99.52%, 99.52%, 99.52%, 99.52%, and 99.05%, correspondingly. Also, with 30%TESS, the FFD-PTEML method gains average $accu_y$, $prec_n$, $sens_y$, $spec_y$, and MCC of 99.63%, 99.64%, 99.63%, and 99.26%, respectively.

Table 2: Fraud detection outcome of FFD-PTEML technique at 70:30 of TRAS/TESS

Classes	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$	MCC
TRAS (70%)					
Non-Fraud	99.37	99.68	99.37	99.68	99.05
Fraud	99.68	99.37	99.68	99.37	99.05
Average	99.52	99.52	99.52	99.52	99.05
TESS (30%)					
Non-Fraud	99.25	100.00	99.25	100.00	99.26
Fraud	100.00	99.27	100.00	99.25	99.26
Average	99.63	99.64	99.63	99.63	99.26

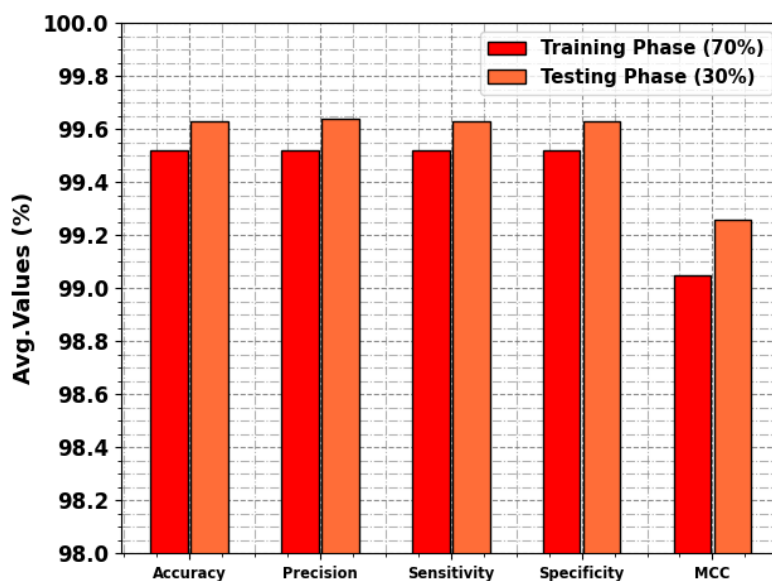


Figure 4: Average of FFD-PTEML method at 70:30 of TRAS/TESS

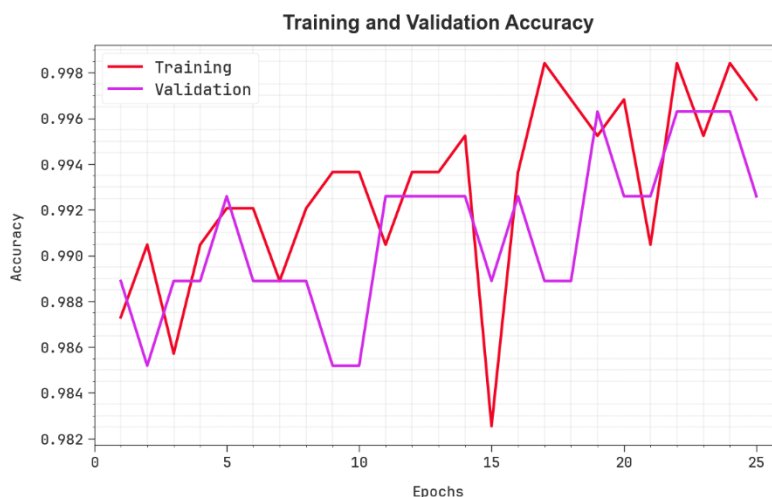


Figure 5: $Accu_y$ curve of the FFD-PTEML method

The efficiency of the FFD-PTEML method is graphically demonstrated in Fig. 5 in the form of training accuracy (TRAA) and validation accuracy (VALA) curves. The figure displays a valuable analysis of the behaviour of the FFD-PTEML technique over varying epoch counts, representing its learning development and generalization capabilities. Mostly, the figure denotes a constant improvement in the TRAA and VALA with development in epochs. It ensures the adaptive aspects of the FFD-PTEML algorithm in the pattern recognition process under TRA and TES data. The increased trends in VALA outline the proficiency of the FFD-PTEML algorithm in adjusting to the TRA data and exceeding to provide precise classification in undetected data, showing robust generalization proficiencies.

Fig. 6 illustrates an extensive view of the training loss (TRLA) and validation loss (VALL) outcomes of the FFD-PTEML method over diverse epochs. The progressive minimization in TRLA emphasizes the FFD-PTEML technique enhancing the weights and diminishing the classification error under TRA and TES data. The figure denotes a better understanding of the FFD-PTEML algorithm related to the TRA data, emphasizing its ability to capture patterns. Significantly, the FFD-PTEML method incessantly increases its parameters by decreasing the variances among the prediction and real TRA class labels.



Figure 6: Loss curve of the FFD-PTEML system

To underline the enhanced results of the FFD-PTEML algorithm, a wide-ranging comparative study is made with other models in Table 3 and Fig. 7 [22]. These obtained outcomes implied that the DSGBT, DTDS, RFGBT, and DTNB models have obtained poor performance with minimal MCC and $accu_y$ values. Meanwhile, the DTBBT, CCFDC-GRFOEL, and OCSODL-CCFD techniques have reached moderately closer values of $accu_y$ and MCC. Nevertheless, the FFD-PTEML technique has highlighted improved performance with maximum $accu_y$ and MCC

values of 99.63% and 99.26%, correspondingly. Thus, the FFD-PTEML technique can be used to detect financial fraud proficiently.

Table 3: Comparative outcomes of FFD-PTEML model other recent techniques

Methods	Accuracy	MCC
FFD-PTEML	99.63	99.26
CCFDC-GRFOEL	99.58	99.12
DSGBT Model	98.27	98.94
DTGBT Model	99.00	98.52
DTDS Model	98.91	98.17
RFGBT Model	98.95	98.22
DTNB Model	98.31	98.34
OCSODL-CCFD	99.24	98.80

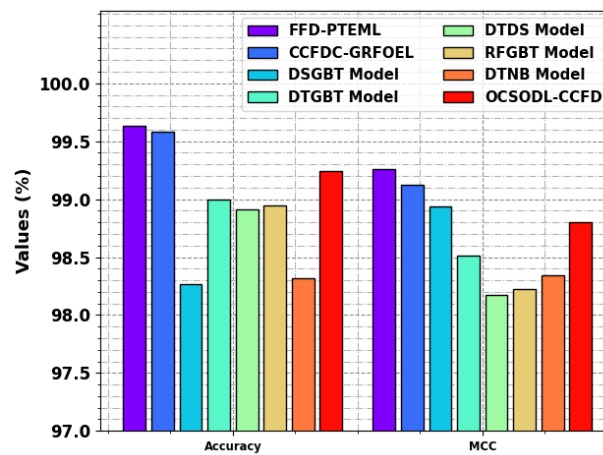


Figure 7: Comparative result of FFD-PTEML approach other recent techniques

6. Conclusion

In this study, we have presented an innovative FFD-PTEML approach. The FFD-PTEML incorporates multiple advanced components, such as z-score normalization for feature scaling and ensemble classification employing ANN, MLP, and RBF networks. The use of z-score normalization ensures uniformity in feature distribution, improving the effectiveness and interpretability of the fraud detection technique. Furthermore, the ensemble classification model combines the strength of different neural network architectures to enhance the detection performance and resilience to complicated fraud patterns. FFD-PTEML demonstrates better performance than the classical technique through extensive experimentation on real-time financial datasets, exhibiting high sensitivity and specificity in fraudulent activity detection.

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

- [1] Islam, S., Haque, M.M. and Karim, A.N.M.R., 2024. A rule-based machine learning model for financial fraud detection. *International Journal of Electrical & Computer Engineering* (2088-8708), 14(1).
- [2] Noviandy, T.R., Idroes, G.M., Maulana, A., Hardi, I., Ringga, E.S. and Idroes, R., 2023. Credit Card Fraud Detection for Contemporary Financial Management Using XGBoost-Driven Machine Learning and Data Augmentation Techniques. *Indatu Journal of Management and Accounting*, 1(1), pp.29-35.

- [3] Balbaa, M. E. (2024). Socio-Economic Indicators and their Impact on Sustainable Economic Development: An In-depth Analysis of Egypt. *International Journal of Economics and Financial Issues*, 14(2), 136–145. <https://doi.org/10.32479/ijefi.16016>
- [4] Vashistha, A., Tiwari, A.K., Singh, P., Yadav, P.K. and Pandey, S., 2024. A Robust Framework for Fraud Detection in Banking using ML and NN. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, pp.1-12.
- [5] Valavan, M. and Rita, S., 2023. Predictive-Analysis-based Machine Learning Model for Fraud Detection with Boosting Classifiers. *Computer Systems Science & Engineering*, 45(1).
- [6] Nahri Aghdam Ghalejoogh, J., Rezaei, N., Aghdam Mazarae, Y. and Abdi, R., 2024. Detecting financial fraud using machine learning techniques. *International Journal of Nonlinear Analysis and Applications*, 15(1), pp.199-214.
- [7] Hajek, P., Abedin, M.Z. and Sivarajah, U., 2023. Fraud detection in mobile payment systems using an XGBoost-based framework. *Information Systems Frontiers*, 25(5), pp.1985-2003.
- [8] Saheed, Y.K., Baba, U.A. and Raji, M.A., 2022. Big data analytics for credit card fraud detection using supervised machine learning models. In *Big data analytics in the insurance market* (pp. 31-56). Emerald Publishing Limited.
- [9] Olaleye, V.O., Odeniyi, O.A. and Alese, B.K., 2024. Ensemble-based Predictive Model for Financial Fraud Detection. *update*, 12(42).
- [10] D. H. M. de Souza and C. J. Bordin Jr, "Ensemble and Mixed Learning Techniques for Credit Card Fraud Detection," *arXiv preprint arXiv:2112.02627*, 2021
- [11] Khalid, A.R., Owoh, N., Uthmani, O., Ashawa, M., Osamor, J. and Adejoh, J., 2024. Enhancing credit card fraud detection: an ensemble machine learning approach. *Big Data and Cognitive Computing*, 8(1), p.6.
- [12] Zhao, Z. and Bai, T., 2022. Financial fraud detection and prediction in listed companies using SMOTE and machine learning algorithms. *Entropy*, 24(8), p.1157.
- [13] Ni, L., Li, J., Xu, H., Wang, X. and Zhang, J., 2023. Fraud feature boosting mechanism and spiral oversampling balancing technique for credit card fraud detection. *IEEE Transactions on Computational Social Systems*.
- [14] Duan, Y., Zhang, G., Wang, S., Peng, X., Ziqi, W., Mao, J., Wu, H., Jiang, X. and Wang, K., 2024. CaT-GNN: Enhancing Credit Card Fraud Detection via Causal Temporal Graph Neural Networks. *arXiv preprint arXiv:2402.14708*.
- [15] Ali, A.A., Khedr, A.M., El-Bannany, M. and Kanakkayil, S., 2023. A Powerful Predicting Model for Financial Statement Fraud Based on Optimized XGBoost Ensemble Learning Technique. *Applied Sciences*, 13(4), p.2272.
- [16] Zhao, C., Sun, X., Wu, M. and Kang, L., 2024. Advancing financial fraud detection: Self-attention generative adversarial networks for precise and effective identification. *Finance Research Letters*, 60, p.104843.
- [17] Fei, N., Gao, Y., Lu, Z. and Xiang, T., 2021. Z-score normalization, hubness, and few-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 142-151).
- [18] Li, B. and Khayatnezhad, M., 2024. Modified artificial neural network based on developed snake optimization algorithm for short-term price prediction. *Heliyon*, p.e26335.
- [19] Awan, A.A., Majid, A., Riaz, R., Rizvi, S.S. and Kwon, S.J., 2024. A Novel Deep Stacking-based Ensemble Approach for Short-term Traffic Speed Prediction. *IEEE Access*.
- [20] Xia, K., Lou, Y., Yuan, Q., Zhu, B., Li, R. and Du, Y., 2024. Optoelectronic Torque Measurement System Based on SAPSO-RBF Algorithm. *Sensors*, 24(5), p.1576.
- [21] <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>
- [22] Maashi, M., Alabdullah, B. and Kouki, F., 2023. Sustainable financial fraud detection using garra rufa fish optimization algorithm with ensemble deep learning. *Sustainability*, 15(18), p.13301.