Grenze International Journal of Engineering and Technology, June Issue

DF-NFLF-TO: A Deep Fusion Neuro-Fuzzy–Logistic Model for Fraud Detection under Extreme Class Imbalance

Kirti Akash Nimbhorkar[1] and Vinod M. Patil[2]
[1-2]Department of Computer Science and IT, Shri Shivaji College of Art, Commerce and Science, Akola, Maharashtra, India
Email: kirtiakashnimbhorkar@gmail.com, vinmpatil21@gmail.com

***Abstract*— All over the world people suffer from financial fraud in credit cards so need robust model to detect the fraudulent transaction over the large amount of records. There is a problem of highly imbalanced transaction data because very small number of fraudulent records in datasets. Conventional deep learning models report high accuracy under such conditions but fail to reliably identify minority fraud instances, and lead to poor recall and limited practical utility. To address this challenge, this paper proposes Deep Fusion Neuro-Fuzzy–Logistic model with validation-driven threshold optimization for robust fraud detection under extreme class imbalance. The proposed model combines the probabilistic logistic learning, neuro-fuzzy inference, and deep representation learning into a unified late-fusion model to enable the complementary decision evidence from heterogeneous models. Unlike fixed decision thresholds commonly adopted in prior studies, DF-NFLF-TO used to validation-based threshold optimization strategy to explicitly control false positive rates while preserving meaningful fraud capture. The experimental model conducted on a large-scale financial transaction dataset shows that the proposed model significantly improves imbalance-aware performance with the accuracy score of 99.88%. The final outcome shows that combined deep fusion with threshold optimization gives more trustworthy solution for actual fraud detection.**

***Index Terms*— Fraudulent Transaction, Fusion Model, neuro-fuzzy inference, fraud detection, threshold optimization.**

I. INTRODUCTION

The frequency and number of financial transactions expanded globally due to the rapid development of digital payment systems and online banking facilities. The developments increased the attack opportunities for fraudulent activity and also improved the accessibility and usability [1]. The significant financial losses the customer lost the confidence due to credit card fraud [2]. The problem is more difficult when dataset is extremely unbalanced in nature. The fraudulent events are very small percentage of all transactions but necessitate prompt and precise identification [3]. Traditional fraud detection model based on rule-driven and machine learning works manually to capture the features. The model is computationally efficient and interpretable, but struggle to capture complex features and patterns in fraudulent behaviors [4]. The deep learning (DL) model shows the superior learning ability as compared to the individual models but not sufficient to address all the real time issues [5].

Grenze ID: 01.GIJET.12.2.143_22

The fraud detection model needs the balance data and false alarm rates show the stable probabilistic outputs. Consequently, obtained the high recall without overwhelming analysts with false positives remains a central problem [6]. A critical limitation of many existing studies lies in the use of a fixed decision threshold, typically set at 0.5, to convert predicted probabilities into class labels. In extremely imbalanced settings, this practice often leads to suboptimal performance, either by missing a large proportion of fraud cases and by generating excessive false alerts. The most models conflate the learning of decision boundaries with the selection of an operating point, limiting their adaptability to different deployment scenarios. These issues motivate the need for a model that not only improves discrimination between fraud and non-fraud transactions but also explicitly accounts for threshold selection in a principled and data-driven manner. To address these challenges, this paper proposed a Deep Fusion Neuro-Fuzzy–Logistic model with validation-based threshold optimization (DF-NFLF-TO) to detect the credit card frauds. The proposed model combines the several distinct signals through a late fusion strategy, combined the stability of probabilistic linear modeling, the expressive power of DNN model, and the sensitivity of anomaly detection via autoencoder reconstruction errors. This study is motivated by three key considerations. First, combined the models to improve robustness and generalization across the class imbalance. Second, optimized the decision threshold on a validation set to predict the actual risk. Third, combined the discriminative and anomaly-based views to enhances the detection rate across the unseen fraud data. Together, these design choices aim to bridge the gap across the accuracy and practical deployment requirements. Following points highlights the main contributions of study:

- To investigate the late-fusion model that combine the probabilistic logistic regression, deep neural network predictions, and autoencoder-based anomaly scores to enhance fraud detection under highly imbalanced conditions.
- A systematic threshold selection strategy is proposed to optimize the operating points based on MCC, F1-score, and false positive rate constraints, thereby improving real-world applicability and decision reliability.
- The performance of proposed model is evaluated using threshold-dependent and threshold-independent metrics, along with calibration and decision curve analyses shows that improved discriminative performance and operational utility on an imbalanced test set.

The rest of the structure of paper is Section 2 presents the previous work in credit card fraud detection. Section 3 describes the complete the methodology of models that contain data preprocessing, model construction, fusion strategy, and threshold optimization. Section 4 presents the experimental outcome of proposed models. Finally, Section 5 conclude the study and discussed the potential directions of future study.

II. Related Work

Last few decades raised the usage of deep learning (DL) techniques to detect the fraud in credit card and financial because of its superior ability of model to capture the temporal transactional features and trends. Ayub et al. (2025) suggested the fraud detection model based on LSTM that obtained the accuracy of 99.38. The results show the effectiveness of sequential learning models to captured the temporal dependencies inherent in transaction streams, thereby outperforming conventional ML models [7]. Sulaiman et al. (2024) investigated multiple DL models such as Autoencoders, CNN, and LSTM model that combined with systematic hyperparameter optimization. The model obtained an accuracy of 97.2% across the european credit card [8]. Zouaoui et al. (2025) suggested the ensemble model to detect the fraud that combined the ANN, CNN, RNN, LSTM, and GRU models. Although DL models show strong performance, and the RF model obtained the highest accuracy of 99.99% indicate the ensemble learning highly effective in real-time anomaly detection scenarios [9]. Similarly, Polytarchos (2025) reported that LSTM-based models obtained 92.00% accuracy to predict the fraudulent customer behavior; however, their dependence on extensive historical data limits scalability in dynamic real-time [10]. Alrasheedi (2025) suggested the MLP and ANN model to identify the fraud and detection rate of 96.00 that shows the capability of feedforward neural networks to identify fraudulent patterns within high-dimensional transaction data [11]. In contrast, Reddy et al. (2024) proposed a hybrid JNBO–SpinalNet model that integrates quantile normalization, distance-based feature selection, and the Jellyfish–Namib Beetle Optimization algorithm, achieving an accuracy of 89.10%. [12]. Qayoom et al. (2024) suggested the Deep Q-Network model to detect the fraud to capture the transactional behaviors. The model obtained a validation accuracy of 97.10% that shows the potential of autonomous learning model in dynamic fraud detection [13]. Additionally, Al-Imran et al. (2024) suggested the LSTM model to capture long-term temporal dependencies and obtained an accuracy of 98.5% and outperforming traditional machine learning classifiers [14].

TABLE I: COMPARATIVE ANALYSIS OF EXISTING DL–BASED FRAUD DETECTION METHOD

Authors	Models	Dataset	Accuracy (%)	Key Observations
Ayub et al. (2025)	LSTM	Transaction sequences	99.38	Strong temporal modeling; high precision and recall
Sulaiman et al. (2024)	AutoEncoder, CNN, LSTM	European credit card dataset	99.20	Hyperparameter tuning improves detection rate
Zouaoui et al. (2025)	ANN, CNN, RNN, LSTM, GRU, RF	Ensemble learning	99.20	Random Forest outperformed DL models
Polytarchos (2025)	LSTM	Behavioral transaction data	92.00	Requires extensive historical data
Alrasheedi (2025)	MLP, ANN	Financial transactions	99.00	Effective for non-linear feature relationships
Reddy et al. (2024)	JNBO–SpinalNet	Optimized feature selection	89.10	Lower accuracy despite advanced optimization
Qayoom et al. (2024)	DQN	Reinforcement learning	97.10	Adaptive learning for evolving fraud patterns
Al-Imran et al. (2024)	LSTM	Sequential transaction data	98.50	Key Observations

Table I shows the comparative analysis of existing DL models that shows the advancements in many existing approaches either rely heavily on imbalanced datasets or require large volumes of historical data that lead to biased performance or reduced generalization. These limitations motivate the exploration of generative and data-balancing techniques combined with DL models to enhance robustness and detection reliability.

III. METHODOLOGY

The proposed Deep Fusion Neuro-Fuzzy–Logistic model with Threshold Optimization (DF-NFLF-TO) is designed to address extreme class imbalance and complex decision boundaries in credit card fraud detection. The model combines the heterogeneous predictive signals using a late-fusion strategy (i) a linear probabilistic classifier, (ii) a deep neural network for nonlinear feature learning, and (iii) an autoencoder-based anomaly score.

A. Dataset Description

In this study used publicly available financial transactional dataset contain total of 6,362,620 samples across 10 distinct attributes which was gather from Kaggle. Each record in dataset was labeled as is Fraud and legitimate to classify the legitimate and fraudulent activities. The dataset contains 6,354,407 legitimate transactions and 8,213 fraud transactions.

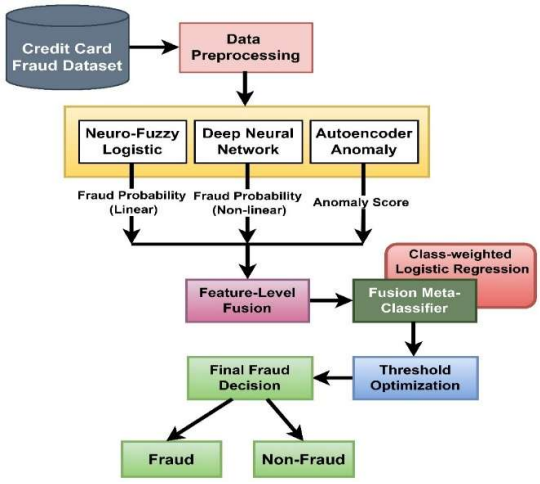

Figure 1. Architectural block diagram of fraud detection Model

Figure 1 show the complete architectural block diagram of the DF-NFLF model to detect the fraud based on the threshold optimization strategy. The model combined the probabilistic, DL, and anomaly using the fusion approach and optimize the threshold for improving the performance across the imbalance data.

Data Preprocessing: The original dataset be denoted as $D = \{(x_i, y_i)\}_{i=1}^{N}$, where, $x_i \in R^d$ shows the feature vector of the i-th transaction and $y_i \in \{0,1\}$ shows the non-fraud and fraud classes, respectively.
Feature Scaling: All numerical features are normalized to the range [0,1] using Min–Max normalization:

$$\tilde{x}_{ij} = \frac{x_{ij} - min(x_j)}{max(x_j) - m \quad (x_j)} \tag{1}$$

where, x_{ij} shows the j^{th} feature of the i^{th} sample. The scaling parameters are estimated on D_{tr} and applied consistently to D_{val} and D_{te}.

B. Model Construction

The DF-NFLF model consists of three branches, each producing a continuous fraud likelihood or anomaly score.
Probabilistic Linear based on Neuro-Fuzzy–Logistic model: A class-weighted logistic regression model is trained on the normalized training data to stable and interpretable baseline probability:

$$p^{(1)}(x) = \sigma(w^{\top}x + b), \tag{2}$$

where $\sigma(\cdot)$ is the sigmoid function. Class weighting compensates for the skewed class distribution and improves sensitivity to minority instances.
Deep Neural Network: A feed-forward DNN capture nonlinear interactions among transaction attributes. The network consists of multiple fully connected layers with $ReLU$ activations and batch normalization, followed by a sigmoid output layer:

$$p^{(2)}(x) = \sigma(f_\theta(x)), \tag{3}$$

where, $f_\theta(\cdot)$ denotes the learned nonlinear transformation parameterized by θ. Early stopping based on validation performance is applied to prevent overfitting.
Autoencoder-Based Anomaly: An autoencoder is trained on non-fraud samples from D_{tr} to model normal transaction behavior. Given an input x, the reconstruction error is computed as:

$$r(x) = \frac{1}{d} \parallel x - \hat{x} \parallel_2^2 , \tag{4}$$

where $\hat{x}$ is the reconstructed output. Fraudulent transactions exhibit higher reconstruction errors due to deviation from normal patterns.
Fusion Mechanism: The outputs of the three branches are concatenated to form a fusion feature vector:

$$z(x) = [p^{(1)}(x), p^{(2)}(x), r(x)]. \tag{5}$$

A class-weighted logistic meta-classifier is trained on $z(x)$ to estimate the final fraud probability:

$$p_f(x) = \sigma(\beta^{\top} z(x) + c) \tag{6}$$

This late-fusion strategy enables adaptive weighting of probabilistic confidence and anomaly evidence.
Threshold Optimization: Rather than using a fixed decision threshold, the DF-NFLF model optimizes the threshold on the validation set to achieve robust real-world performance.
MCC-Optimized Threshold: The threshold t^*_{MCC} is selected to maximize the Matthews Correlation Coefficient:

$$t^*_{MCC} = \arg\max_{t \in (0,1)} MCC(y_{val}, I[p_f \geq t]) \tag{7}$$

F1-Optimized Threshold: The alternative threshold t^*_{F1} maximizes the F1-score on D_{val}
FPR-Constrained Threshold: For deployment-oriented evaluation, a threshold $t^*_{FPR}*$ is selected such that:

$$FPR(t^*_{FPR}) \leq \alpha, \tag{8}$$

where α shows the predefined false positive rate of 0.1% and 0.01%. The optimized threshold is then fixed and applied to the imbalanced test set.

III. Result Analysis

In this section present the performance of proposed model. All experiments were conducted on Google Colab Pro using a high-memory runtime to stable training of proposed model. The dataset was divided into three sets such as 64% training set, 16% validation set, and 20% test set to preserve the class distribution to shows the

realistic fraud prevalence. The probabilistic branch used a class-weighted logistic regression with a maximum of 1,000 iterations to address class imbalance. The deep neural network branch consisted of fully connected layers with ReLU activation and batch normalization. The Adam optimizer with early stopping is used to prevent losses and overfitting. The autoencoder branch was trained solely on non-fraud samples to learn normal transaction patterns and compute reconstruction-based anomaly scores. Measure the performance of model using a set of evaluation parameters as shown in equation.

$$Accuracy = \frac{T\_P + T\_N}{T\_P + T\_N + F\_P + F\_N} \quad (9)$$

$$Precision = \frac{T\_P}{T\_P + F\_P} \quad (10)$$

$$Recall = \frac{T\_P}{T\_P + F\_N} \quad (11)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (12)$$

MCC: Correlation between predicted and actual binary classes

TABLE II. PERFORMANCE OF THE DF-NFLF MODEL ACROSS IMBALANCED TEST SET UNDER VALIDATION-OPTIMIZED OPERATING THRESHOLDS

Parameters	DF-NFLF TEST (MCC-Optimized)	DF-NFLF TEST (F1-Optimized)	DF-NFLF TEST (FPR ≤ 0.1%)	DF-NFLF TEST (FPR ≤ 0.01%)
Threshold	0.999000	0.999000	0.996694	1.000000
Accuracy	0.998822	0.998822	0.998642	0.999342
Precision	0.534123	0.534123	0.482180	0.855379
Recall	0.685940	0.685940	0.699939	0.590383
F1-score	0.600586	0.600586	0.571003	0.698596
MCC	0.604719	0.604719	0.580308	0.710337
Kappa	0.600005	0.600005	0.570346	0.698277
ROC-AUC	0.958957	0.958957	0.958957	0.958957
Avg Precision	0.632457	0.632457	0.632457	0.632457
Log Loss	0.182242	0.182242	0.182242	0.182242
Brier Score	0.035334	0.035334	0.035334	0.035334

Table II shows the performance of DF-NFLF model on the original imbalanced test set under four operating points derived from validation-based threshold selection. The MCC-optimized and F1-optimized settings converge to the same threshold (0.999) and yield identical results and obtained the accuracy of 0.998822, Precision = 0.534123, Recall = 0.685940, and F1-score = 0.600586, with balanced agreement indicated by MCC = 0.604719 and Kappa = 0.600005. Under the FPR ≤ 0.1% constraint (threshold 0.996694), recall increases slightly (0.699939) at the expense of precision (0.482180), reflecting a more sensitive screening configuration. In contrast, the FPR ≤ 0.01% setting (threshold 1.000) substantially increases precision (0.855379) while reducing recall (0.590383), representing a stricter, low-false-alarm deployment mode. Across all operating points, threshold-independent discrimination remains stable (ROC-AUC = 0.958957) and ranking quality is consistent (Avg Precision = 0.632457), while probabilistic reliability is supported by low Log Loss (0.182242) and Brier Score (0.035334).

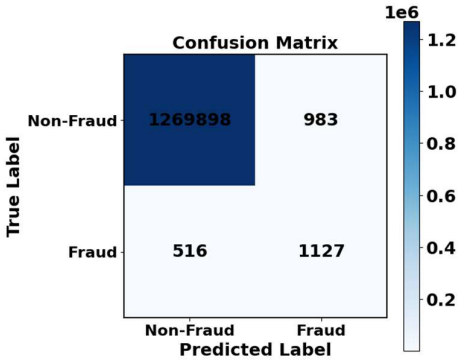

Figure 2. Classification result of the proposed DF-NFLF model across imbalanced test set at the optimized decision threshold

Figure 2 shows the classification result of the proposed DF-NFLF model across imbalanced test set at the selected operating threshold. Out of a large volume of legitimate transactions, 1,269,898 non-fraud cases were correctly identified as non-fraud, while 983 legitimates transactions were incorrectly flagged as fraud that shows very low false positive rate across the dataset size. For fraudulent transactions, the model detected 1,127 fraud cases, whereas 516 fraud instances were misclassified as legitimates.

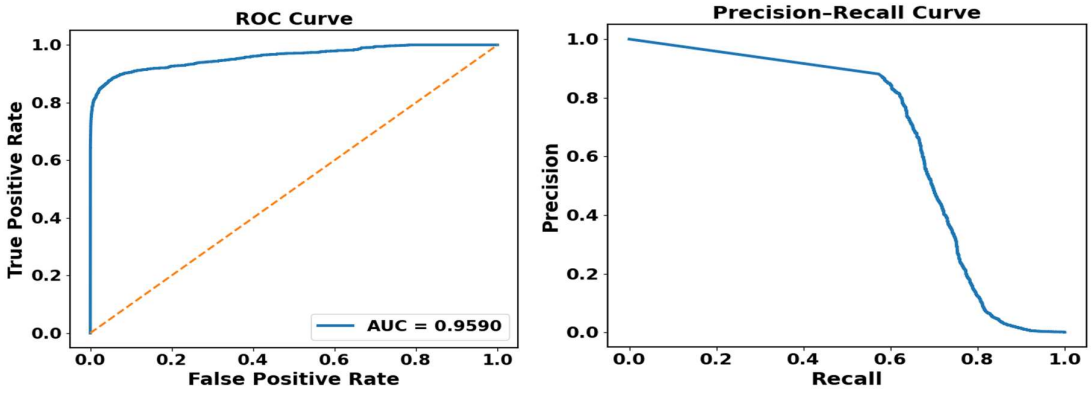

Figure 3. ROC and Precision–Recall curves of the proposed DF-NFLF model

Figure 3 shows the ROC and PR curves of the DF-NFLF model over the imbalanced test set. The ROC curve shows the strong convex shape with an AUC of 0.9590 that indicate the high TPR across a wide range of FPR and shows the robust class separability. The PR curve indicate the model's effectiveness under severe class imbalance to maintain the near-perfect precision at lower recall levels and achieving a favorable trade-off as recall increases.

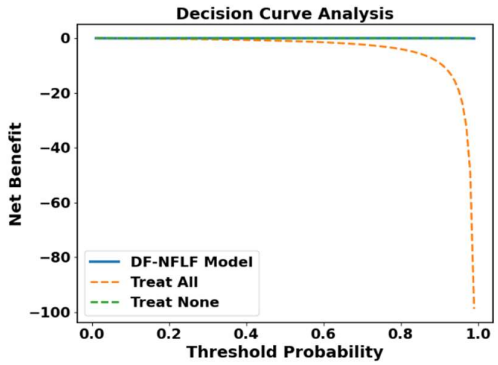

Figure 4. Decision curve analysis of the proposed DF-NFLF model

Figure 4 shows the decision curve analysis of the proposed DF-NFLF model on the imbalanced test set that indicate the superior net benefit across a broad range of threshold probabilities compared to treat-all and treat-none strategies.

TABLE III. COMPARISON OF EXISTING MODELS WITH PROPOSED MODELS

Authors	Models	Accuracy (%)
Ayub et al. (2025)	LSTM	99.38
Sulaiman et al. (2024)	AutoEncoder, CNN, LSTM	99.20
Zouaoui et al. (2025)	ANN, CNN, RNN, LSTM, GRU, RF	99.20
Polytarchos (2025)	LSTM	92.00
Alrasheedi (2025)	MLP, ANN	99.00
Reddy et al. (2024)	JNBO–SpinalNet	89.10
Qayoom et al. (2024)	DQN	97.10
Al-Imran et al. (2024)	LSTM	98.50
Proposed	DF-NFLF-TO	99.88

Table III shows the comparison of the existing models with proposed model. This is clearly indicated that proposed DF-NFLF-TO model performed better as compared to previous models.

IV. Conclusion and Future Scope

This study presented a DF-NFLF-TO model to credit card fraud detection across extreme class imbalance. The proposed model integrates complementary decision signals derived from a probabilistic linear model, a DNN, and an autoencoder-based anomaly detector through a late fusion strategy. By explicitly decoupling model learning from operating-point selection, the model enables flexible deployment across diverse risk tolerance settings. Extensive experiments conducted on a large-scale transactional dataset that show the DF-NFLF-TO model obtained the strong discriminative performance over the evaluation parameters and maintained the robust agreement parameters on the imbalanced test set. Validation-driven threshold optimization further enhanced practical effectiveness by enabling controlled trade-offs between precision and recall. The proposed obtained the accuracy score of 0.99% across all the FPR-constrained threshold across the imbalanced dataset. The decision curve analysis shows the practical significance and outperforms naive treat-all and treat-none methods in terms of net benefit. The future work of the proposed model will combine with fusion strategy to enhanced performance and fraud detection rate using attention-based mechanism that allows the dynamic weighting of branch outputs. Second, combine utility-based and cost-driven loss functions to correlate improvement to financial threat categories. Third, to account for conceptual drift in changing transaction categories, flexible or online threshold modification techniques could be investigated.

References

[1] Moradi, F., Tarif, M., & Homaei, M. (2025). A Systematic Review of Machine Learning in Credit Card Fraud Detection. https://doi.org/10.20944/preprints202507.1085.v1

[2] Krishnaraj, G., Ravi, C. N., & Ahmed, M. (2025). Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms. International Journal of Computer Communication and Informatics, 7(1), 1–10. https://doi.org/10.34256/ijcci2511

[3] Kim, J., Seol, J., Kim, S., & Chen, L. (2025). Deep Learning Approaches for Credit Card Fraud Detection: A Data Balancing Perspective. 393–402. https://doi.org/10.1109/sera65747.2025.11154522

[4] Shah, N. P., Panchal, K. B., Jambhale, V. A., Kharat, G. K., & Galinde, S. N. (2023). Online Fraudulent Transaction Detection Through Machine Learning and Deep Learning Algorithms. International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering. https://doi.org/10.17148/ijireeice.2024.12102

[5] Akouhar, M., Abarda, A., El-Fatini, M., & Ouhssini, M. (2024). A Deep Learning and Resampling Approach to Credit Card Fraud Detection. 10, 1–6. https://doi.org/10.1109/wincom62286.2024.10655834

[6] Bian, C. (2025). Credit Card Fraud Detection: Machine Learning and Deep Learning Advances, Challenges, and Future Directions. ITM Web of Conferences, 78, 02023. https://doi.org/10.1051/itmconf/20257802023

[7] Ayub, M. I., Bhattacharjee, B., Akter, P., Uddin, M. N., Gharami, A. K., Islam, Md. S., Suhan, S. I., Khan, M. S., & Chambugong, L. (2025). Deep Learning for Real-Time Fraud Detection: Enhancing Credit Card Security in Banking Systems. The American Journal of Engineering and Technology, 07(04), 141–150. https://doi.org/10.37547/tajet/volume07issue04-19

[8] Sulaiman, S. S., Nadher, I., & Hameed, S. M. (2024). Credit Card Fraud Detection Using Improved Deep Learning Models. Cmc-Computers Materials & Continua, 78(1), 1049–1069. https://doi.org/10.32604/cmc.2023.046051

[9] Zouaoui, H., & Naas, M.-N. (2025). Credit card fraud detectionand risk management strategies: A deep learning-based approach for EU banks. Research Papers in Economics and Finance, 9(1), 55–80. https://doi.org/10.18559/ref.2025.1.2108.

[10] Polytarchos, E. (2025). Credit Card Fraud Detection Through Deep Learning and Real-Time Data Streams: A Comparison and New Directions. https://doi.org/10.47852/bonviewfsi52026108

[11] Alrasheedi, M. A. (2025). Enhancing Fraud Detection in Credit Card Transactions: A Comparative Study of Machine Learning Models. Computational Economics. https://doi.org/10.1007/s10614-025-11071-3

[12] Reddy, V. V. K., Reddy, R. M., Munaga, M. S. K., Karnam, B., Maddila, S. K., & Kolli, C. S. (2024). Deep learning-based credit card fraud detection in federated learning. Expert Systems with Applications, 124493. https://doi.org/10.1016/j.eswa.2024.124493

[13] Qayoom, A., Khuhro, M. A., Kumar, K., Waqas, M., Saeed, U., Réhman, S. U., Wu, Y., & Wang, S. (2024). A novel approach for credit card fraud transaction detection using deep reinforcement learning scheme. PeerJ, 10, e1998. https://doi.org/10.7717/peerj-cs.1998

[14] Al-Imran, M., Ayon, E. H., Islam, M. R., Mahmud, F., Akter, S., Alam, M. K., Hasan, M. T., Afrin, S., Shorna, J. F., & Aziz, Md. A. (2024). Transforming banking security: the role of deep learning in fraud detection systems. The American Journal of Engineering and Technology, 06(11), 20–32. https://doi.org/10.37547/tajet/volume06issue11-04.