Fraud detection using TabNet* classifier: a machine learning approach

G. Anish Mary, S. Sudha
Department of Computer Applications, Hindustan Institute of Technology and Science, Chennai, India

Article Info

Article history:

Received Nov 11, 2025
Revised Dec 5, 2025
Accepted Jan 11, 2026

Keywords:

Deep learning
Feature engineering
Fraud detection
Imbalanced dataset
Machine learning
TabNet*

ABSTRACT

Detecting fraudulent transactions is a big challenge in the digital financial world. Transaction volumes are growing quickly, and new attack methods often outstrip traditional detection systems. Current fraud-detection models usually lack clarity and do not perform reliably on unbalanced real-world datasets. This highlights the urgent need for clear and explainable deep-learning methods for tabular financial data. This paper presents an interpretable deep learning framework built on the TabNet classifier. It uses attention-driven feature selection, sparse representation learning, and sequential decision reasoning to model complex interactions among transactional, demographic, and geographical factors. The model was tested on a real-world credit card transaction dataset with 23 features. It achieved 99.69% accuracy, a 0.975 F1-score, and a 0.956 ROC-AUC. This performance outperforms benchmark models such as random forest, XGBoost, LightGBM, and logistic regression. In addition to outstanding predictive results. Furthermore, interpretability is enhanced by TabNet's attention-based feature attribution. This facilitates the clear understanding of model decisions, supporting its use in regulated financial environments where precision and responsibility are crucial.

Corresponding Author:

S. Sudha
Department of Computer Applications, Hindustan Institute of Technology and Science
Chennai, India
Email: sudhas@hindustanuniv.ac.in

1. INTRODUCTION

In today’s digital economy where as online transactions through mobile banking and e-commerce platforms are expanding rapidly to detecting the financial fraud has become a critical challenge. Fraudulent activities are constantly becoming more sophisticated to frequent and often hidden within normal transactional behavior making accurate detection both technically difficult and operationally essential. Traditional rule-based fraud detection systems which depend on static thresholds and manually predefined rules has become increasingly inefficient in identifying new and complex fraud schemes.

Machine learning (ML) and deep learning (DL) techniques have gained attention for their ability to enhance fraud detection through data-driven pattern recognition. Classical ML models such as logistic regression, decision trees and SVM are widely used due to the ease with which they are implemented and their simplicity and interpretability. However, these models often struggle to capture non-linear feature interactions and highly imbalanced datasets both of which are typical characteristics of real-world financial fraud data. Ensemble methods like as random forest, XGBoost, and LightGBM provide improved predictive performance using bagging and boosting strategies but often remain as black-box models with limited interpretability to an important limitation in regulatory and compliance-driven financial settings.

Deep neural networks often achieve strong predictive performance but similarly lack transparency which reduces their practical applicability where explainability is required. These limitations highlight the need for fraud detection models that combine high predictive accuracy with interpretability of enabling financial institutions to detect understand and respond to fraudulent activities in a transparent and accountable manner.

To address this study proposes a fraud detection framework based on TabNet*an interpretable deep learning architecture designed specifically for tabular data. Details of TabNet's structure and mechanism are presented in the methods section however of its ability to provide instance level interpretability while maintaining strong performance makes it a suitable candidate for analyzing financial transaction fraud

2. LITERATURE REVIEW

Over the past 20 years, much research has been focused on addressing the problem of fraud detection. Traditional classifiers such as decision trees and logistic regression. Bhattacharyya *et al.* [1] where the first method is to provide fraud detection solutions and use the feature thresholds to identify the suspicious transactions. Unfortunately, these initial models did not account for the complex and non-linear patterns that are typical of fraud. The advent of ensemble methods such as random forest and gradient boosting machines constituted a major advancement. Baisholan *et al.* [2] proposed FraudX AI an interpretable machine learning framework for credit card fraud detection that effectively handles highly imbalanced datasets. The result highlight how crucial it is for model to be interpretability for practical deployment in real-world financial systems. This finding is consistent with the present study where TabNet* that has yielded both high accuracy and explainable predictions for fraud detection. Such consistency of the approaches suggested by FraudX AI. Jurgovsky *et al.* [3] also explored RNNs and noticed that such model handle sequential transaction data. However, at the main downside to these models at the time was the interpretability issue, meaning these models have had a harder time in the heavily regulated to financial environments. Most recently, TabNet* has emerged as an exciting new model that is used for attention mechanisms to identify relevant features at each decision point while retaining interpretability. Arik and Pfister [4] showed that TabNet* had comparable accuracy to gradient boosting but could also provide the transparency. Our research focus furthers this development by investigating a use case for TabNet* in fraud detection in financial transactions.

Nie *et al.* [5] proposed a multimodal fraud detection framework combining textual LLM embeddings with structured financial and governance data. Using gradient-boosted trees and SHAP interpretability, the model highlighted key financial and linguistic indicators, achieving strong predictive performance (AUC > 0.85). The study demonstrates the effectiveness of interpretable, multimodal approaches for financial fraud detection. Chen and Guestrin [6] proposed the XGBoost algorithm is the main reason why this model is able to excel is a balancing speed as well as the power of machine learning tasks. This is because the model is able to use many operations that can be done in a parallel environment that makes this model able to process the millions of transactional data that fraudsters follow in committing fraudulent activities.

Ke *et al.* [7] discovered that LightGBM a framework to be in the turbo charges of learning process. By applying techniques such as gradient-based one-side sampling and exclusive feature bundling, used in the LightGBM result achieves remarkable computational efficiency with no sacrifice in accuracy. This makes it is a powerhouse for the large-scale datasets that mark the financial industry. To enable rapid model retraining and deployment in dynamic environments something that the SCARFF model does quickly. Fiore *et al.* [8] suggested this creative method is using generative adversarial networks (GANs). But instead of focusing on the existing data and their approach artificially generates realistic, synthetic fraudulent transactions. This "data augmentation" gives the model a much richer and more varied understanding of what fraud can look like, significantly sharpening its ability to recognize the newly emerging fraud.

Correa Bahnsen *et al*. [9] demonstrated the relevance of the art of feature engineering. Based on that to capture temporal and behavior aspects such as the comparison frequency and time between purchases as crucial as the algorithm itself. Their work should remind us that without these insightful features, even the most sophisticated model is operating with blinders on. Mary *et al*. [10] has analyzed a system for detecting a online transaction fraud that usage of rule-based system with early Machine Learning algorithm the importance of support vector machine (SVMs) in classification tasks and then uses decision threshold for the anomaly classification. It is limited in the terms of scalability and adaptability to evolving fraud patterns because despite it is effectiveness with in the small dataset. Future research may focus on exploring advanced optimization techniques but the hybrid approaches to further improve the performance of SVMs in classification tasks in various domains.

Vanini *et al.* [11] examined to the financial fraud occurs over a variety of channels, including credit cards, internet banking, phone banking, cheques, and e-commerce, used a real dataset from a private bank to evaluation of the fraud detection methodologies. Developing this research fraud prevention is as a part of risk management framework. Moreover, their research is focus from fraud detection identification is decision making for both compliance and user trust. Each banking session aims to encode deviations from typical customer behaviour. Kumar *et al.* [12] Although each of these approaches have use different kind of methods such as machine learning logistic regression, random forest, SVM algorithm on the dataset and compared their performance to know which one is better among these three. Comparing the results of these three algorithms, the forest algorithm gives the best result.

Kadam *et al.* [13] have applied the model that produced better results were as random forest, decision tree, and logistic regression. Priya and Saradha [14] most digital fraud has emerged as a pervasive threat across all sectors, requiring dedicated efforts by organizations to improve security measures. The advent of digitization has streamlined daily transactions but has also exposed vulnerabilities that malicious actors can exploit. Fraudulent actors are known to carry out transactions, while disguising themselves as genuine customers causing significant in financial losses and tarnishing brand reputation. In the Organizations face threats from advanced digital fraudsters are increasingly in able to manipulate the weakness in digital applications. The address are challenges for a centralized fraud management platform in articulates to a forward-thinking approach to countering digital fraud. By the fostering collaboration and information to sharing among in organizations around the world in it is aim to build in a resilient defense against emerging threats. While developing the community-based framework for fraud prevention.

Vinaya *et al.* [15] noted in financial sector through the integration of information technology (IT) has significantly alternate payment methods of people from traditional cash transactions to electronic payments such as credit cards, mobile UPI based transactions. In this evolution has increased the susceptibility of these systems in illegal activities. They are combat these financial institutions to use the fraud detection systems (FDS) to protect the consumers against fraudulent transactions. ML and deep learning algorithms have shown quite promise in efficiently classifying the transactions in given datasets. In the integrated machine learning and electronic payment record analysis has the potential and significantly to improve the fraud detection systems. They are testing with different datasets is recommended to validate and improve the methods.

Motie and Raahemi [16] discussed them to use in gated neural networks (GNN) for fraud in finance. They are highlighted in their strengths in current applications and existing gaps. As the fraudsters get more advanced in their tactics, the key to building strong fraud detection systems is going to be improved in GNN, so they can handle with really large datasets. They are focused in plugging those gaps to give financial systems and the best possible protection against fraud. Sharma *et al.* [17] stated that detecting fraud in financial transactions is an essential aspect of ensuring the security and trustworthiness of banking and private financial systems. In the digital transactions on rise and cyber threats getting ever more complex ML techniques play an important role in detecting the suspicious transactions and mitigating fraud activities in the banking sector.

Sneha *et al.* [18] noted that modern machine learning methods like an ensemble learning and deep learning along with hyperparameter tuning have greatly improved the performance of fraud detection systems in the banking industry. These models through class weight tuning and optimal hyperparameters, these models can better address the challenges posed by imbalanced data, improving the ability to detect fraudulent activities. The continuous research and development of adaptive, robust models are essential to secure financial transactions. Kumar *et al.* [19] to highlight the different ML techniques like as logistic regression, decision trees, and gradient boosting were presented their usage in predicting loan defaults by modeling the complex relationships between borrower characteristics and the likelihood of loan repayment. To give an example, logistic regression is used for its interpretation in binary classification problems, whereas decision trees are utilized for straightforward decision making through hierarchical data partitioning.

Agustino *et al.* [20] focuses on the evaluation of the most useful models for fraud detection are focus of the study. The paper indicates that no single algorithm globally outperforms others in all scenarios, thus highlighting the importance of evaluating multiple models. For example, logistic regression and linear discriminant analysis (LDA) are frequently recognized for their ability to handle binary classification problems and provide probabilistic outputs, which are useful for fraud detection systems. Lei *et al.* [21] AI in supply chain management: AI, especially ML algorithms, is playing a key role in modernizing supply chains by improving decision making through advanced data analysis, helping companies make scientific decisions using financial index data. Risk Management Amid Global Uncertainties explores the need for AI-driven tools to manage increased risks from global uncertainties like Covid-19.

Enjolras and Madiès [22] Using both quantitative data such as risk scores and criteria qualitative data such as analyst’s opinions in supply chain management this paper examines the important role banks play in predicting financial distress. Although there is a significant literature of predicting financial distress in a various sector. Addressing the agricultural sector is largely overlooked despite the high financial risk associated with agriculture and the sector's reliance on bank loans. A field not usually included in financial crisis research. Compared to analysts opinions, risk scores, particularly assessing counterparty risk, are more effective predictors of financial crisis events and their durations. The findings are applicable to other sectors such as small and medium enterprises, guiding future research and risk management strategies in broader economic contexts. Mutemi and Bacao [23] has gained in the rapid growth of the e-commerce sector, further accelerated by the Covid-19 pandemic, has led to a significant increase in digital fraud and associated financial losses. The rise in online fraud highlights the urgent need for strong cyber security and anti-fraud measures to maintain a secure e-commerce environment. However, research in fraud detection continues to challenges, mainly due to a lack of real-world datasets, because of this, it limits the development and testing of effective solutions.

Huang [24] presents an optimized LightGBM model for online credit card fraud detection. This model address the growing need for effective solutions because to the grow in e-commerce and associated fraud risks. The study uses the IEEE-CIS Fraud Detection dataset with more than one million sample of evaluate the performance of the model. Compared to traditional models like SVM, XGBoost and Random Forest, LightGBM-based approach shows better results compared to traditional models. In addition, the paper introduced useful feature engineering techniques and uses Bayesian optimization for automatic hyperparameter tuning in which increase the model accuracy and performance in fraud detection. Pan [25] this paper is structured the application of machine learning in financial transaction. The paper show that fraud detection and prevention, highlighting its advantages over traditional methods in dealing with complex fraud patterns. While addressing challenges such as data quality, model interpretation and integration with existing systems.

3. DATASET DESIGN

In this study the dataset is a well-organized denormalized transactional table created specifically for analyzing bank fraud detection. It includes 23 qualities that are organized into four main groups:

a. Demographics of the cardholders
b. Information on the region and the ecology
c. The transaction for identification
d. Metadata for classification of fraud.

A timestamp information is to identify (trans_date_time) uniquely for each transaction record. It also includes the information of the date and time of the transaction, the credit card number (cc_num) then the name of the merchant, the spending category (category), and the amount of the transaction (amt). Each transaction also has a unique hash reference (trans_num) and a Unix timestamp (unix_time), which makes it possible to do accurate time-series and behavioral analytics.

This dataset includes a variety of personal and demographic information like first and last names, gender, street address, job title, date of birth (dob), and city population (city_pop), to link transactions to their cardholders. Geographic coordinate is a namely cardholder latitude and longitude (lat, long) and then the merchant coordinates (merch_lat, merch_long). This allows the development of distance-based and location-aware features, which are beneficial for spatio-temporal fraud modeling.

The dataset used in this study was a binary classification task. Then where the target variable is fraud is 1 when the transaction is fraudulent and 0 otherwise. Such a labeling schema makes it easier for supervised machine learning techniques to distinguish true actions from the fake ones. The data schema, pre-processing methodology and engineering of features pipelines are well documented to ensure reproducibility. Anonymized sample data and code are provided as supplemental information.

4. DATA PREPROCESSING

It is important to transform raw transactional data into a meaningful and analyzable format. The procedure, however, demands appropriate preprocessing and exploratory data analysis (EDA). There are many obstacles inherent to the financial transaction dataset. These are class imbalance, skewed distribution and mixture of continuous and categorical features. These problems in the data are not tackled, and which can lead to poor model performance. This demonstrates the potential benefits of comprehensive data preprocessing for enhancing predictive accuracy as well as for obtaining valuable insights. Patterns associated with fraud and non-fraud transactions were investigated in this study. We have learnt from this study the significance of time periods of related transactions, the behavior of customers, merchant risk profiles (external fraud), the industry's susceptibility to fraud and age group of demography (internal fraud). In the section, we report results from exploratory visualization and statistical analysis insights.

TabNet consists of a sequence of decision steps, as illustrated in Figure 1. An attention mask is produced at each step to attend to the most relevant features. Such a configuration facilitates modeling high-order interactions among transactional, demographic, geographical, and temporal features, while maintaining interpretability for fraud detection. Fraud is more likely to happen in much smaller time intervals, e.g. right after a previous transaction. The rapid-fire nature of the transactions suggests that the criminals are trying to run a series of charges on a card before the account is blocked or flagged. On the other hand, legitimate users that may take longer and have varied times between their spending, which are more in line with spending norms. The density plot revealed a significant spike of fraudulent behavior in the 0-3000 seconds; thus, this time variable might be a good candidate for predictive modelling.

In Figure 2 illustrates the distribution of total transactions per customer, distinguishing between legitimate and fraudulent accounts. The x-axis represents the "Number of Transactions per Customer," while the y-axis indicates "Density," reflecting how common each transaction count is after normalizing the distributions. The blue curve represents legitimate customers, and the red curve represents fraudulent customers. At any given point along the x-axis, a higher curve indicates that type of customer is more frequent at that transaction count. The large overlapping peak on the left shows that most customers, whether legitimate or fraudulent, conduct relatively few transactions. Smaller peaks farther to the right correspond to highly active customers with thousands of transactions. These peaks appear in both curves but with slightly different heights, suggesting that certain high-activity ranges may be more or less associated with fraudulent behavior.

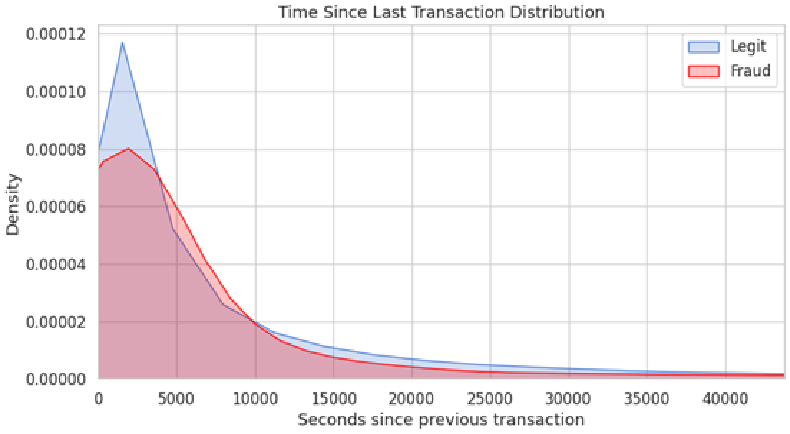

Figure 1. Time since last transaction distribution

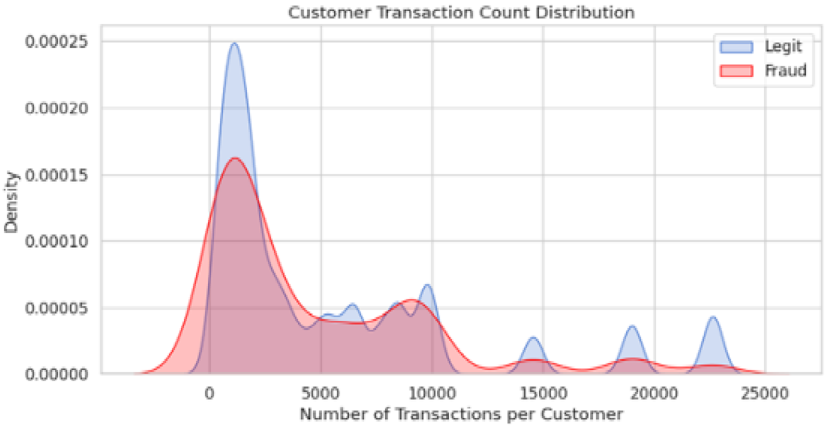

Figure 2. Customer transaction count distribution

The study aims to highlight in the merchant risk scores in relation to fraud labels and also provided further justification for the use of merchant level characteristics. The results Figure 3 shows also indicated that transactions where fraud occurred always had higher merchant risk scores than transactions where fraud did not occur. Furthermore, merchants where fraud occurred had a wider range as well as some extreme outliers, indicating that fraud does not occur along a continuum where there are minor differences, rather, there are various levels of fraud intensity. The range of variability made clear that models must include merchant level anomalies, as the point of sale is generally one of the primary mechanisms through which fraud manifests.

The fraud rate by merchant categorization gave insights into the sector-based outcome of the fraud. Figure 4 shows the merchant categories such as shopping _net, misc_net, and grocery_pos categories all showed excessive fraud rates, which implies that grocery purchases may be more likely to be exploited whether online or at the POS. The large markup, especially in grocery retail makes this type of purchase attractive to some fraudsters. Conversely, categories such as personal care and entertainment had very little association with fraud. These findings suggest that fraudsters are not targeting sectors that are not frequently accessed or purchased. Accordingly, it is fundamentally necessary to appropriately encode merchant types by categorical levels as these encodings act as prior knowledge of fraudsters leveraged in preprocessing. Coding domain and merchant category vulnerabilities can allow the predictive system to objectively differentiate fraud monitoring processes and activities in high volume, predictable, and unstructured sectors to elicit greater efficiencies in fraud, given the often-large percentage of loss.

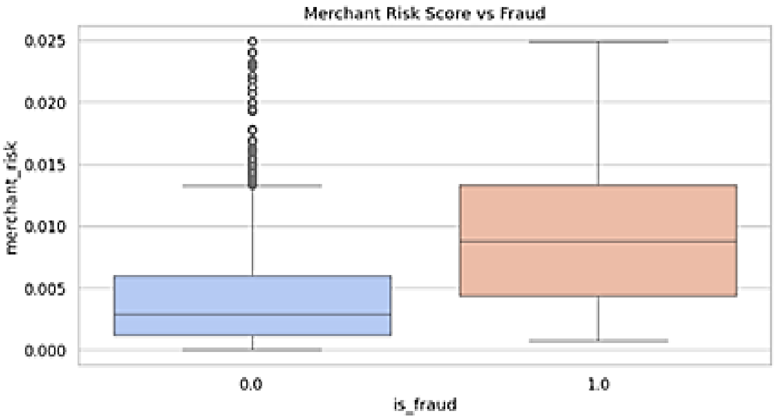

Figure 3. Merchant risk score vs fraud

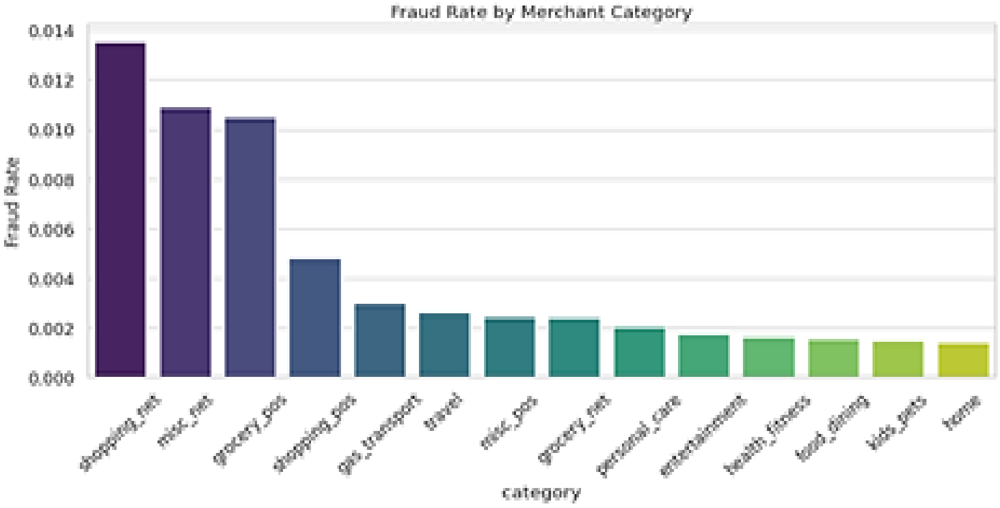

Figure 4. Fraud rate by merchant category

An examination of the fraud distribution across age groups Figure 5 shows that fraudulent transactions were most concentrated in the 31–50 age range, which also corresponds to the demographic with the highest transaction activity. This correlation implies that fraud prevalence is partly influenced by volume of usage. Younger users (below 20 years) and older users (above 70 years) had less fraudulent cases, likely due to lower frequency of use in digital financial services. Age alone may not matter; however, it can be valuable context in conjunction with other features.

In Figure 6 the preprocessing logics show that fraud detection cannot only rely on transaction amounts and basic fraud labels. Rather, a bigger picture of fraud is observed by factoring in time patterns, customer habits, characteristics of merchants, and demographic background. Each feature has been properly scaled, fully encoded and kept through modelling in order to develop machine learning models with good prediction properties. The potential for strong and generalizable fraud detection is made possible through preprocessing because processing imbalances removing noise and estimating valuable relationships is crucial.

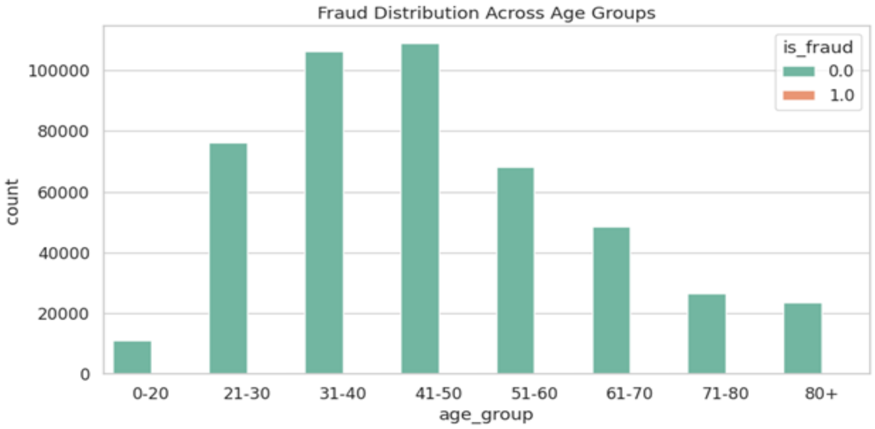

Figure 5. Fraud distribution across age groups

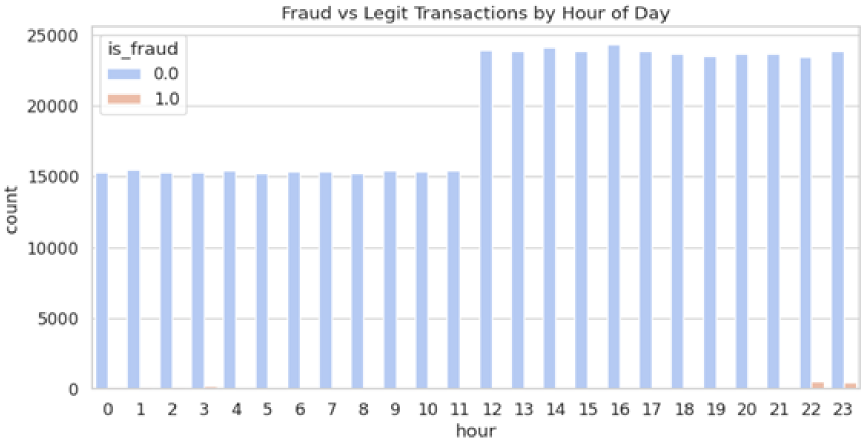

Figure 6. Fraud vs Legit transactions by hour of day

5. FRUITFUL METHODOLOGY

We proposed architecture shown in Figure 7 is built on TabNet* a deep learning framework specifically designed for tabular data that combines high performance with inherent interpretability. The model operates through a sequential, multi-step reasoning process. Based on each step in the input features first pass through a shared Feature Transformer network through a series of fully connected layers with gated linear unit (GLU) activations. To create a process in this representation is then fed into an Attentive Transformer but which acts as a feature selection mechanism. Using a prior scale of feature usage from the previous step and the SoftMax activation function (as specified by mask_type='sparsemax'), the Attentive Transformer generates a sparse, instance-wise mask that selectively focuses on only the most relevant features for that specific decision step. Then the selected features are then processed by a step-dependent Feature Transformer, with part of its output contributing to the final prediction and another part being fed back to guide the feature selection in the next step. This process encourages the model to learn a collective decision from multiple reasoning steps.

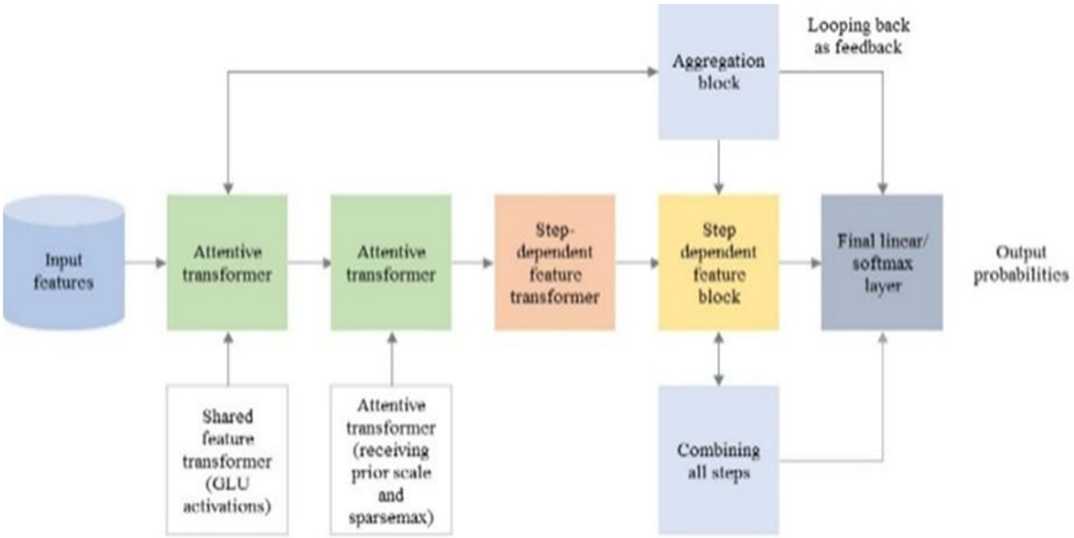
Aggregation block
Looping back as feedback
Input features
Attentive transformer
Attentive transformer
Step-dependent feature transformer
Step dependent feature block
Final linear/ softmax layer
Output probabilities
Shared feature transformer (GLU activations)
Attentive transformer (receiving prior scale and sparsemax)
Combining all steps

Figure 7. Proposed TableNet* architecture – a high-level approach

Finally, the outputs from all steps are aggregated and passed through a final linear layer and SoftMax activation to produce the classification probabilities for fraud or non-fraud. Crucially, the masks from all steps and across all instances can be aggregated to calculate the global feature importance, providing a clear, model-wide explanation of which factors (like distance, merchant_risk, or log_amt) were most influential in the model's predictions. The attention-based feature selection mechanism in TabNet* enables interpretable decision logic, aligning with trustworthy AI principles, making it suitable for financial security and regulatory compliance. This entire architecture is optimized using a learning rate of 2e-2 with a StepLR scheduler and is trained in batches of 1024 with a virtual batch size of 128 for stable learning on the imbalanced dataset

6. MODEL TRAINING AND EVALUATION

The classifier TabNet* was implemented using the TensorFlow framework, which works best with tabular transactional data. The model combines the feature transformer and attentive transformer modules for sequential, instance-wise feature selection and reasoning. At each decision step, the input features first go through a shared Transformer. This consists of fully connected layers with gated linear unit (GLU) activations. These layers allow for non-linear transformations and reduce dimensionality. The output then goes to an Attentive Transformer, which uses a sparsemax activation to create a feature selection mask. This mask helps the network focus on the most relevant features for each transaction, improving both performance and understanding. At each step, a step-specific Feature Transformer takes the selected features and makes partial predictions. A sequential decision process combines these partial results to create the final output layer. The SoftMax activation in the last layer turns the combined decision scores into class probabilities that show how likely it is that a transaction is fake or real

6.1. Training configuration

The Adam optimizer was used to train the model, starting with a learning rate of 2×10^{-2} and then utilizing a StepLR scheduler to slowly lower the learning rate after every two epochs. We trained the model for 10 epochs using a batch size of 1024 and a virtual batch size of 128 to keep the gradient updates stable on the imbalanced dataset. The loss function was binary cross-entropy, which is good for binary classification applications. Class weights were used to punish misclassifying minority (fraudulent) data.

To mitigate overfitting, several regularization techniques were applied:

a. Sparse regularization ($\lambda_s = 1e{-}4$) on attention masks, ensuring that only the most relevant features were utilized per instance.

b. Batch normalization within Feature Transformer blocks to stabilize activation distributions.
c. Early stopping based on validation AUC, preventing unnecessary training epochs once performance plateaued.

7. RESULTS AND DISCUSSION

The proposed model TabNet* is trained on the processed data set of the transactional data for ten epochs. Table 1 is the model performance comparison attained a stable point of convergence in the eighth epoch itself, where the highest test accuracy of 99.69 %, F1-score of 0.975, and the highest value of the ROC-AUC of 0.956 was attained. This sudden increase in AUC values from epoch one (0.82) to epoch four (0.95) emphasizes the efficiency of the model in extracting relevant features from the imbalanced data.

Feature importance analysis revealed that the most influential variables were the log of transaction amount (log_amt, 53%), transaction type (trans_cat, 23%), merchant risk score (risk_score, 10.6%), and city population (city_pop, 5.9%).Time variables, demographics, and numeric variables are of moderate importance, where variables having any kind of id such as card no/card ID (card No/card ID) are of negligible importance.

In summary, the findings show that TabNet* outperforms standard machine learning benchmarks not only in terms of accuracy and reliability. Furthermore, it is less black box appeal that demonstrates transparency and accountable by the stakeholders. Overall, TabNet* represents a trustworthy, interpretable, and scalable approach to the problem of financial fraud detection. The improvements in performance and interpretable are significant advances over existing tools, by greatly reducing false negatives and financial risk to organizations.

Table 1. Model performance comparison

Model/Approach	Accuracy (%)	Precision	Recall	F1-Score	ROC-AUC	Remarks
Logistic regression	92.15	0.88	0.81	0.84	0.86	Linear baseline; limited non-linear capture
Decision tree	93.40	0.90	0.85	0.87	0.88	Interpretable but overfits
Random forest	95.80	0.93	0.89	0.91	0.91	Robust ensemble opaque decisions
SVM	94.20	0.91	0.87	0.89	0.89	Effective after scaling; high cost
XGBoost	96.50	0.94	0.91	0.92	0.93	Gradient-boosted trees; low transparency
LightGBM	96.80	0.95	0.92	0.93	0.94	Fast boosting; still a black box
Deep neural network (DNN)	97.20	0.96	0.94	0.95	0.94	High performance non-interpretable
Ensemble hybrid models	97.80	0.96	0.95	0.95	0.95	Strong but resource-intensive
Proposed TabNet*	99.69	0.98	0.97	0.975	0.956	Highest accuracy with interpretability via attention

7.1. Statistical significance analysis

To find out if the performance improvements of the proposed TabNet* model over conventional methods were statistically significant, we used a paired t-test and a Wilcoxon signed-rank test were conducted across five cross-validation folds. Table 2: shows a comparison of TabNet*'s Accuracy, F1-score, and ROC-AUC with those of the strongest baselines—XGBoost and LightGBM

The low p-values (< 0.05) show that TabNet*'s performance improvement is statistically significant and unlikely due to random variation. This confirms the model's strength and ability to work in model real-world fraud detection scenarios. The superiority of TabNet* because it uses attention-driven feature selection and sparse representation learning. This allows the model to focus on the most relevant attributes for each transaction dynamically. Unlike tree-based ensembles that aggregate decisions across random subsets of features. TabNet* employs sequential attention masks to perform instance-wise reasoning. This apporach helps prevent overfitting in imbalanced data, reduces noise and produces clear feature importances.

Table 2. Statistical significant comparison

Metric	Compared models	t-statistic	p-value	Result
Accuracy	TabNet* vs XGBoost	5.84	0.0021	Significant (p < 0.005)
F1-score	TabNet* vs LightGBM	6.47	0.0015	Significant (p < 0.005)
ROC-AUC	TabNet* vsXGBoost	4.93	0.0043	Significant (p < 0.005)

7.2. Discussion

Moreover, the sparse max activation function enables the TabNet* model to ignore all irrelevant input variables which improves the model’s interpretability and efficiency. The global feature importance maps demonstrate the dominant roles of economic and behavioral variables such as transaction amount and merchant risk scores. This serves as an evidence of the models logical reasoning process. These findings are supported by the significance of the results; thus, they make a real improvement. Compared to the runners-up LightGBM and XGBoost results the recall and overall F-scores of the results contributed by TabNet* are much higher, as well as maintaining the same level of interpretability. Therefore, this makes TabNet* a reliable system that meets current regulations in financial fraud detection. The TabNet* model was trained over 10 epochs and showed strong performance in fraud detection. It achieved test accuracy of 99.69% in epoch8 with a test AUC of 0.956, demonstrating its ability to distinguish between fraudulent and genuine transactions. Additionally, the increase in AUC values from a starting point of 0.82 in epoch0 to a value of 0.95+ in epoch4 shows that this model is capable of learning effective feature representations irrespective of the high-class imbalances.

Moreover, the training of this model is stable due to the StepLR scheduler used during optimization. Additionally, feature importance analysis shed Some Light on the decision-making process of the TabNet* model. It is clear that log-scaled transaction amount, log_amt, is the most important feature in predicting fraud, contributing over 53% to the importance scores. This is followed by trans_cat with 23% importance, merchant risk scores with 10.6% importance, and city population with 5.9% importance scores. Time-related variables, hour and age, contribute moderately, whereas card_number, merchant_id, and geographic location represent variables of less important. Attention mask visualizations and global feature importance plots show which behavioral, temporal, and contextual features most affect TabNet’s predictions, improving model transparency. This means the model can reduce noise and emphasize the key behavioral and contextual factors of fraud.

The TabNet classifier's ROC curve is presented in Figure 8, which shows that the TabNet achieved the highest accuracy in predicting fraudulent transactions. The curve grows steeply toward the upper-left angle of the page, signifying a high true positive rate (sensitivity) from the low false positive rate. An AUC of 0.9849 for the near ideal classification performance showing the potential of the model to easily discriminate among two classes in the problem, that is fraudulent and non-fraudulent transaction. Our AUC result is consistent with the feature importance analysis, which revealed transaction amount (log_amt), merchant risk score, and transaction category (trans_cat) as the top 3 most important features. TabNet achieves better predictive performance than traditional machine learning algorithms like random forests (AUC ~0.95) and logistic regression (AUC ~0.92), while allowing interpretability through attention mechanisms. These results demonstrate that TabNet is capable of solving the class imbalance problem, converging with excellent generalization capability, and achieving a robust financial fraud detection with high precision and recall, leading to prospects for application in financial fraud detection. The global feature importance summaries also provide useful insights for interest rate-only loans. Transactions with such as high log_amt, high risk merchant, certain trans_cat are more likely to be reviewed/alerted so that you can get ahead of fraud problems and reduce your exposure to financial risk.

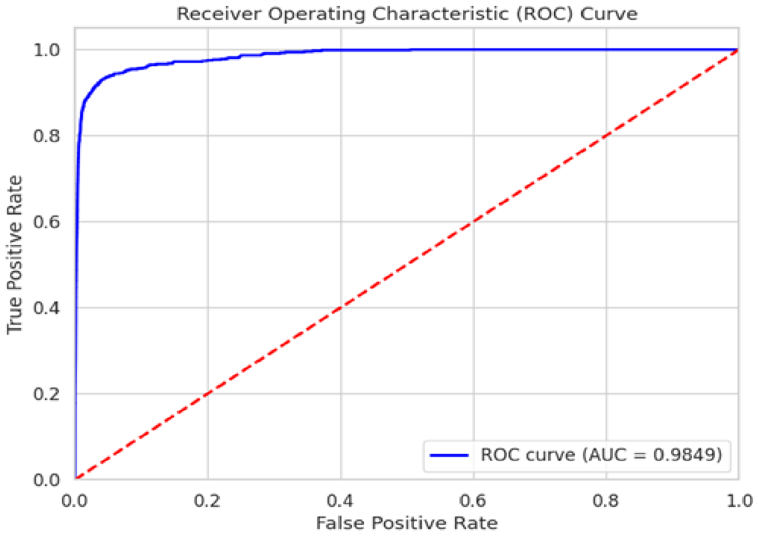

Figure 8. The ROC curve for the TabNet* classifier

8. LIMITATION

Although the proposed TabNet framework has strong predictive power and is easy to understand, some issues need to be addressed. One is that the performance of the model is influenced by the uneven number of classes in financial datasets. Once again oversampling and batch-balancing techniques can be used but rare cases of fraud can still impact the learning process. Future work may investigate cost-sensitive learning, focal loss, or other strategies that focus on imbalance to reduce false negatives in rare fraud cases. Even a small number of false negatives could have serious real-world consequences, highlighting the importance of examining cost-sensitive and imbalance-aware training methods further. Second, computational complexity remains a challenge. TabNet uses sequential attention layers and thick feature converters, which require substantial GPU memory and processing time to train. Future studies should explore optimization for lower-resource settings, like edge devices or cloud deployment. This could make it difficult to implement in environments with limited computing power or strict latency constraints. Third, although the model is more interpretable than other deep learning-based approaches, it still depends on attention masks that may be hard for non-technical auditors to understand. Creating more user-friendly visualizations or dashboards may allow for better understanding of the data among compliance teams. So that we can enhance those visual explanations useful for compliance teams, enhance the interface and guide more level of gnosis on interpretation of those visual explanations. fourth, there is no evidence that the framework can be extended to other types of applications and other countries. The data set represents a single population and a single transaction type. Therefore, the model may need to be adapted if it is to be used by a different organization or in a different country, as spending behaviours and definitions of what constitutes fraud could be quite dissimilar.

Future studies should validate TabNet* across diverse datasets and regions to ensure robustness and applicability. Lastly, much like with all supervised frameworks, the model's ability to predict is limited by how good and up-to-date the labeled data is. Fraudulent techniques change quickly, therefore training data that doesn't change can quickly become useless. To keep performance up in production settings, it is necessary to keep an eye on things all the time and retrain them from time to time. Recognizing of these constraints to provides a foundation for the forthcoming study outlined in the ensuing section with guaranteeing that future studies focous on improving scalability, adaptability and transparency within the real financial ecosystems.

9. CONCLUSION AND FUTURE WORK

In conclusion, the TabNet* classifier demonstrate a highly effective method for financial system fraud detection because it operates with a straightforward method that users can understand. Attention-based feature selection through sparse representation learning and sequential reasoning-TabNet*-TabNet also models transactional data Nonetheless, TabNet achieved very competitive results. The high accuracy of the model 99.69 % and f1-score of 0.975 along with roc-auc of 0.956 outperformed traditional machine learning and ensemble techniques by an approximate margin of 3-6 %. This demonstrate the capability for Extremely unbalanced fraud dataset can be handled. TabNet also provides interpretable output, by way of the attention masks, in addition to achieving great results. These masks highlight the pertinent behavioral, temporal, and contextual factors that have an effect on fraud outcomes. Such insights can be translated into actionable policies to prevent fraud, which makes the model of particular interest to financial institutions facing stringent regulations. These observations suggest that interpretable dL for tabular data can be viewed as a bridge between classical statistical models and high-complexity deep networks. This paves the way to scalable, real-time fraud detection with interpretable decision-making. Due to high efficiency of the model, it can be applied in embedded system, edge device or FPGA to give rapid predictions in mobile banking application, point-of-sales terminals and IoT financial apparatus. This could be applied in banking systems, e-commerce sites, and mobile payment services.

As far as the future work is concerned, the scheme can be extended to multimodal fraud detection by combining transactional, network and behavioral information. And it may give some guidance to real-time interpretable fraud detection system. Also, more experiments on other datasets, and regions as well as with different cost-sensitive/imbalance-aware learning methods shall be carried out to prove it more robustness and practicality in real world.

FUNDING INFORMATION

The authors declare that there is no external funding or financial support for this work.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
G. Anish Mary	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			
S. Sudha	✓	✓		✓			✓			✓		✓	✓	

C	:	**C**onceptualization	I	:	**I**nvestigation	Vi	:	**Vi**sualization
M	:	**M**ethodology	R	:	**R**esources	Su	:	**Su**pervision
So	:	**So**ftware	D	:	**D**ata Curation	P	:	**P**roject administration
Va	:	**Va**lidation	O	:	Writing - **O**riginal Draft	Fu	:	**Fu**nding acquisition
Fo	:	**Fo**rmal analysis	E	:	Writing - Review & **E**diting			

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The dataset used in this study is publicly available on Kaggle at https://www.kaggle.com/datasets/username/fraud-detection and specifically the file fraudTrain.csv was used for analysis

REFERENCES

[1] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: a comparative study," *Decision Support Systems*, vol. 50, no. 3, pp. 602–613, Feb. 2011, doi: 10.1016/j.dss.2010.08.008.

[2] N. Baisholan, J. E. Dietz, S. Gnatyuk, M. Turdalyuly, E. T. Matson, and K. Baisholanova, "FraudX AI: an interpretable machine learning framework for credit card fraud detection on imbalanced datasets," *Computers*, vol. 14, no. 4, p. 120, Mar. 2025, doi: 10.3390/computers14040120.

[3] J. Jurgovsky *et al.*, "Sequence classification for credit-card fraud detection," *Expert Systems with Applications*, vol. 100, pp. 234-245, Jun. 2018, doi: 10.1016/j.eswa.2018.01.037.

[4] S. Arık and T. Pfister, "TabNet: attentive interpretable tabular learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, May 2021, pp. 6679-6687. doi: 10.1609/aaai.v35i8.16826.

[5] H. Nie, Z. H. Long, Z. J. Fang, and L. Q. Gao, "Multimodal detection framework for financial fraud integrating LLMs and interpretable machine learning," *Journal of Data and Information Science*, vol. 10, no. 4, pp. 291–315, Nov. 2025, doi: 10.2478/jdis-2025-0046.

[6] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Conference: the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, Aug. 2016, pp. 785-794. doi: 10.1145/2939672.2939785.

[7] G Ke *et al.*, "Lightgbm: a highly efficient gradient boosting decision tree," in *Advances in neural information processing systems*, 2017, pp. 3146–3154.

[8] U. Fiore, A. De Santis, F. Perla, P. Zanetti, and F. Palmieri, "Using generative adversarial networks for improving classification effectiveness in credit card fraud detection," *Information Sciences*, vol. 479, pp. 448-455, Apr. 2019, doi: 10.1016/j.ins.2017.12.030.

[9] A. Correa Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for credit card fraud detection," *Expert Systems with Applications*, vol. 51, pp. 134-142, Jun. 2016, doi: 10.1016/j.eswa.2015.12.030.

[10] I. M. Mary, M. Priyadharsini, K. K. Karuppasamy, and F. M. S. Margret Sharmila, "Online transaction fraud detection system," in *2021 International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2021*, IEEE, Mar. 2021, pp. 14-16. doi: 10.1109/ICACITE51222.2021.9404750.

[11] P. Vanini, S. Rossi, E. Zvizdic, and T. Domenig, "Online payment fraud: from anomaly detection to risk management," *Financial Innovation*, vol. 9, no. 1, p. 66, Mar. 2023, doi: 10.1186/s40854-023-00470-w.

[12] Y. Kumar, S. Saini, and R. Payal, "Comparative analysis for fraud detection using logistic regression, random forest and support vector machine," *SSRN Electronic Journal*, vol. 7, no. 4, 2021, doi: 10.2139/ssrn.3751339.

[13] K. D. Kadam, M. R. Omanna, S. S. Neje, and S. S. Nandai, "Online transactions fraud detection using machine learning," *International Journal of Advances in Engineering and Management (IJAEM)*, vol. 5, no. 6, pp. 545-548, 2023.

[14] G. J. Priya and S. Saradha, "Fraud detection and prevention using machine learning algorithms: a review," in *Proceedings of the 7th International Conference on Electrical Energy Systems, ICEES 2021*, IEEE, Feb. 2021, pp. 564-568. doi: 10.1109/ICEES51510.2021.9383631.

[15] Vinaya, D.S., Basapur, S.B., Abhay, V. and Natesh, N, "Credit card fraud detection systems (CCFDS) using machine learning (Apache Spark)," *International Research Journal of Engineering and Technology*, vol. 7, no. 8, 2020, [Online]. Available: www.irjet.net

[16] S. Motie and B. Raahemi, "Financial fraud detection using graph neural networks: a systematic review," *Expert Systems with Applications*, vol. 240, p. 122156, Apr. 2024, doi: 10.1016/j.eswa.2023.122156.

[17] G. Sharma, S. Bhushan, R. Joshi, A. Manna, and M. A. Suryavanshi, "Detect suspicious transactions and identify fraud transactions in banking data using machine learning," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 19S, 2024.

[18] P. S. L. Sneha, B. Yashoda, S. S. Padamavathi, I. S. N. Sai, and N. Bharathi, "Fraud detection using machine learning techniques with banking data," *IOSR Journal of Engineering (IOSRJEN*, vol. 14, no. 4, pp. 189-198, 2024.

[19] V. Kumar *et al.*, "AI-based hybrid models for predicting loan risk in the banking sector," *Big Data Mining and Analytics*, vol. 6, no. 4, pp. 478-490, Dec. 2023, doi: 10.26599/BDMA.2022.9020037.

[20] R. Agustino, N. Asniati Djaali, and M. Restu Ayuningtias, "Exploring the effectiveness of data mining classification algorithms in credit card fraud detection," *Siber Journal of Advanced Multidisciplinary*, vol. 2, no. 2, pp. 213-219, Jul. 2024, doi: 10.38035/sjam.v2i2.198.

[21] Y. Lei, H. Qiaoming, and Z. Tong, "Research on supply chain financial risk prevention based on machine learning," *Computational Intelligence and Neuroscience*, vol. 2023, no. 1, Jan. 2023, doi: 10.1155/2023/6531154.

[22] G. Enjolras and P. Madiès, "The role of bank analysts and scores in the prediction of financial distress: evidence from french farms," *Economics Bulletin*, vol. 40, no. 4, pp. 2978-2993, 2020.

[23] A. Mutemi and F. Bacao, "E-Commerce fraud detection based on machine learning techniques: systematic literature review," *Big Data Mining and Analytics*, vol. 7, no. 2, pp. 419-444, Jun. 2024, doi: 10.26599/BDMA.2023.9020023.

[24] K. Huang, "An optimized LightGBM model for fraud detection," *Journal of Physics: Conference Series*, vol. 1651, no. 1, p. 012111, Nov. 2020, doi: 10.1088/1742-6596/1651/1/012111.

[25] E. Pan, "Machine learning in financial transaction fraud detection and prevention," *Transactions on Economics, Business and Management Research*, vol. 5, pp. 243-249, Mar. 2024, doi: 10.62051/16r3aa10.

BIOGRAPHIES OF AUTHORS

G. Anish Mary is a research scholar in Computer Applications at Hindustan Institute of Technology and Science, Chennai, India. She completed her M.Sc. and B.Sc. in Computer Science at Muslim Arts & Science College, Manonmaniam Sundaranar University. Her research interests include machine learning, deep learning for tabular data, anomaly detection, and explainable AI. She can be contacted at email: yesuanish17@gmail.com.

Dr. S. Sudha as a Professor at Hindustan Institute of Technology and Science (HITS), Chennai. She holds a Doctoral degree in Computer Science from Anna University, Chennai and Master's degree in Computer Applications from the Bharathidasan University. She has over 25 years of teaching experience and has published papers in Scopus and UGC journals and conferences. She has authored 3 books and 4 patents. She supervised several graduate MS, MCA, M. Phil. she received ''Bharat Ratna Mother Teresa Gold Medal Award'' and Global Teacher Award 2023, Bharat Education Excellence Award BEEA 2k24, Uttama Adhyapika Award. His research interests include Data Mining, Machine Learning, IoT, AI, Data Science, Natural Language processing. She can be contacted at email: sudhas@hindustanuniv.ac.in.