International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 11 | Nov - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Voting Ensemble Framework with SMOTE for Financial Fraud Detection in Highly Imbalanced Transaction Data

Sukhdeep Singh[1], Kulwinder Singh[2]
[1]Research Scholar, Department of CSE, Bhai Maha Singh College of Engg., Sri Muktsar Sahib
[2]Professor, Department of CSE, Bhai Maha Singh College of Engg., Sri Muktsar Sahib

Abstract: The expansion of digital payment ecosystems has intensified the need for effective fraud detection mechanisms capable of identifying increasingly complex fraudulent behaviours. This research introduces a comprehensive fraud detection framework for severely imbalanced transaction datasets containing 6,362,620 records, including only 8,213 fraudulent transactions (0.1291%). The preprocessing workflow included data quality assessment, categorical transformation, normalization, and the generation of informative predictive features. Additional features such as balance differences, transaction amount ratios, and zero-balance indicators were created to improve fraud pattern identification. Stratified partitioning was adopted to preserve class distributions, while oversampling techniques were utilized to improve minority-class representation during training. A broad range of predictive approaches, spanning conventional algorithms, ensemble methods, neural architectures, and attention-based models, were comparatively analysed. It has been found that Ensemble and boosting models outperformed other approaches. The proposed Voting Ensemble + SMOTE model recorded an accuracy score 0.999953, precision 0.96641, recall 0.99817, F1-score 0.98204, ROC-AUC 0.99986, and PR-AUC 0.99873, demonstrating its effectiveness for reliable financial fraud detection with minimal false alerts.

Keywords: Financial Fraud Detection, Ensemble Learning, Voting Ensemble, Class Imbalance, Feature Engineering, Fraud Analytics, Transaction Data, ROC-AUC, PR-AUC, Imbalanced Classification.

I. INTRODUCTION

Digital financial services have expanded considerably due to the widespread adoption of electronic payment technologies with digital banking, mobile payment systems, electronic fund transfer (eFT) systems, and the online delivery of financial services contributing to an improved accessibility, convenience, and efficiency of transactions [1]. Although digitalization improves efficiency, it has also introduced new vulnerabilities that can be exploited for fraudulent purposes such as the unauthorized transfer of funds, identity theft, account compromise, transaction manipulation, and money laundering [2]. Manual fraud monitoring has become impossible with financial institutions handling millions of transactions a day, and there is a huge need for intelligent and automated fraudulent activity detection model to be in place. The datasets of monetary transaction are very skewed and make financial fraud detection difficult. In real world financial environment, the number of fraudulent transactions is a small proportion of the total transactions [3]. Traditional fraud detection methods are mostly based on pre-set rules or threshold-based detection systems. These tools are good at detecting known suspicious activity, but not so good at catching emerging trends or advanced fraud techniques [4]. With recent developments in machine learning (ML) and deep learning (DL), these methodologies have shown great promise for uncovering hidden patterns within transactions, learning intricate feature relationships, and identifying anomalous activities from large-scale financial datasets [5]. These smart techniques help to analyse characteristics of transactions, including changes in account balance, transaction type and temporal behaviour, increasing the effectiveness of the fraud detection process [6]. To solve the issue of financial fraud detection, a holistic

framework for intelligent fraud detection in financial transactions using machine learning techniques is proposed in this work. It includes data preprocessing, feature engineering, class imbalance treatment and a performance-based comparison of various machine learning algorithms. This preprocessing phase involves missing data analysis, duplicate record validation, encoding of categorical data, attribute scaling, and elimination of uninformative attributes. Moreover, some characteristics of the domains are generated, such as: difference of balance, ratio of transaction amount, zero-balance indicators and measures of balance errors, which are used to improve the discriminative power of predictive models.

II. RELATED WORK

Fraud detection is a topic that has received much discussion recently, and machine learning is the answer, as it can automatically detect patterns and anomalies in fraudulent transactions. Ensemble learning techniques can enhance the stability of the prediction by using multiple models, and deep learning techniques can learn complex nonlinear relationship from the large-scale transaction data. Also, hybrid deep learning architectures have proved to be more efficient in fraud detection, as they increase the sensitivity and minimize missed fraud cases. Ahmed et al. [1] pointed out that the merits of ensemble learning can be used to improve the fraud classification performance with the help of better precision and recall. To achieve better identification of rare fraudulent events in highly disproportionate datasets, Akouhar et al. [2] introduced a Kolmogorov–Arnold network with dynamic oversampling. Almusallam and Qayyum [5] presented a hybrid ensemble learning approach that combines feature selection and clustering techniques for real-time fraud detection. Gupta et al. [7] proposed a stacking ensemble model with feature selection which showed better predictive performance in credit card fraud detection. According to Ould-Slimane and Cedric [8] machine learning models have been shown to be more effective than traditional rule-based systems at banking fraud detection. Given that labelled fraud data are scarce, Karnavou et al. [9] investigated unsupervised learning methods for suspicious banking transactions detection. There are also some promising results from deep learning techniques. In their review, Chen et al. [10] elaborated on the recent progress of deep learning for fraud detection and pointed out data imbalance and model interpretability as two challenges. To tackle class imbalance and overfitting problems, EL Hachimi et al. [11] used LSTM networks for detecting fraud in credit card. Fanai, and Abbasimehr [12] came up with an Autoencoder-based framework to model informative transaction representations that could be used for fraud classification. The focus has been on hybrid and advanced learning models. To model both sequential and structured financial data, Chen et al. [13] merged the LSTM and Multilayer Perceptron models. Du et al. [14] proposed a heterogeneous graph learning approach to detect unknown fraud patterns and Kim et al. [15] showed that the integration of ensemble learning with deep learning could greatly improve fraud detection performances. While significant success has been obtained in previous studies, issues of high-class imbalance, feature representation, model generalization, and false alarm reduction still exist. As a result, it is essential to have a comprehensive framework that incorporates effective preprocessing, feature engineering, class balancing, and evaluation of multiple predictive models to ensure financial fraud detection with a high degree of reliability.

III. PROPOSED METHODOLOGY

The methodology goes through several steps including data collection, data preprocessing, feature scaling, and applied classifiers.

Dataset Description: A large-scale financial transaction data set with more than 6.3 million transactions, comprising of both legitimate and fraudulent ones [16] is used in the study. The data includes information like the transaction type, amount, account numbers of the sender and receiver, pre-transaction and post-transaction account balance information, along with fraud-related labels. The target is the isFraud binary classification variable which assigns a 1 to fraudulent transactions and 0 to authentic ones; isFlaggedFraud is the variable that designates the transactions flagged by the existing rule-based systems. The transaction type can be PAYMENT, TRANSFER, CASH_IN, CASH_OUT or DEBIT. The

data set is extremely skewed, with fraudulent transactions being a small percentage of the total data. Thus, suitable class imbalance processing techniques are needed to enhance detection of minority fraud cases, and the performance of the model remains stable.

Data Preprocessing: To enhance data quality, reduce data leakage, transform data attributes to suitable models for prediction and overcome the highly imbalanced nature of the financial fraud data, data preprocessing was performed. The initial data set had 6,362,620 transactions and 11 attributes. The target variable is isFraud, which has two classes – legitimate transactions (0) and fraudulent transactions (1). The final features matrix had 16 attributes after preprocessing and feature engineering (Table 1).

Table 1: Preprocessing Details

Dataset Component	Value
Original dataset shape	6,362,620 × 11
Target variable	isFraud
Normal / Fraud transaction class	0 / 1
Final feature matrix shape	6,362,620 × 16

Before feature engineering and model development, a missing value analysis was conducted on the dataset. Apart from this, the duplicates, ensuring data consistency and reliability were also checked. There were no data quality concerns that were discovered. Also, the sender and receiver account identifiers nameOrig and nameDest were dropped as they signify the origin or destination account identifier of the transaction rather than general transaction behaviour (Table 2). Keeping these characteristics may lead to data leakage and thus cause the models to remember account-specific patterns rather than fraud characteristics of transactions.

Table 2: Removed Attributes

Removed Attribute	Reason for Removal
nameOrig	Sender account identifier; may introduce memorization and data leakage.
nameDest	Receiver account identifier; does not contribute to generalized fraud behaviour.

There was significant imbalance in the target variable as seen in Table 3. Hence, the proportion of fraud was just 0.1291% of the entire data set. The presence of imbalance implies that accuracy alone will be inadequate to measure fraud prediction models.

Table 3: Target class records before balancing

Class	Meaning	Number of Records	Percentage
0	Normal transaction	6,354,407	99.8709%
1	Fraud transaction	8,213	0.1291%

Data was split into an 80:20 train-test split, stratified as illustrated in Figure 1. Both subsets contained transactions both genuine and fake and they were preserved in their original distribution using stratification.

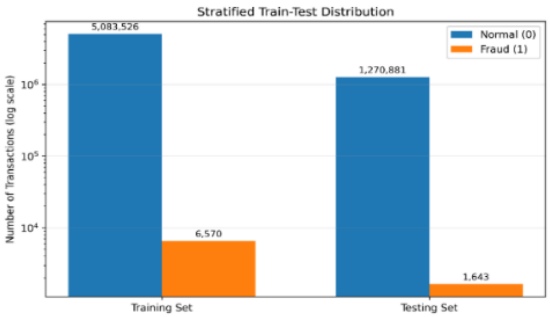


Figure 1: Stratified train-test distribution

The training data was used for model learning and class imbalance treatment while the testing data was not changed for model evaluation under the original real-world class distribution.

Feature Scaling: The dataset was first examined for missing values and duplicate records to ensure data quality and consistency. The identifier-based attributes *nameOrig* and *nameDest* were subsequently removed because they do not contribute to generalized fraud behavior and may introduce data leakage. To enhance the discriminatory power of the predictive models, several domain-specific features were engineered, including balance_diff_orig, balance_diff_dest, amount_to_oldbalance_ratio,amount_to_newbalance_ratio,orig_zero_balance_after,dest_zero_balance_before

, error_balance_orig, and error_balance_dest. These features capture balance variations, transaction-to-balance relationships, zero-balance conditions, and inconsistencies in expected account balance updates, thereby providing additional information for distinguishing fraudulent transactions from legitimate ones.Next, continuous variables were normalized through standardization to ensure comparable feature scales including transaction amount and account balances. To avoid data leakage, scaling parameters were estimated exclusively from the training subset and subsequently applied to unseen test samples. Furthermore, transaction values demonstrated substantial skewness, necessitating a transformation process. Therefore, a logarithmic transformation, log (1 + amount), was employed to improve feature distribution and model learning efficiency and analysis, as shown in Figure 2.

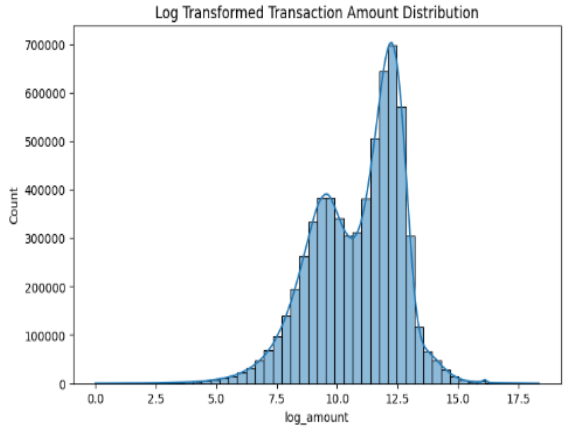


Figure 2: Log-transformed transaction amount distribution

Applied Classifiers: A comparative assessment was done to examine the performance of models such as: traditional machine learning, ensemble learning, deep learning, hybrid deep learning, and transformer-based models etc. Due to the substantial disparity between legitimate and fraudulent records, only a very small proportion of the total records, the Synthetic Minority Oversampling Technique (SMOTE) were applied to the training data to improve the representation of fraud samples and reduce model bias toward legitimate transactions.

This integration of various techniques such as SMOTE-based balancing, traditional machine learning algorithms, ensemble learning methods, deep learning architectures, hybrid networks, and transformer-based models allowed for a comprehensive assessment of the fraud detection performance. This experimental approach is highly flexible and is useful to find models that can accurately detect fraud in high dimensional financial transaction datasets with low false positive predictions.

Algorithm 1 presents the pseudocode of the proposed Voting Ensemble + SMOTE model for financial fraud detection.

Algorithm 1: Voting Ensemble with SMOTE for Financial Fraud Detection

Input: Financial transaction dataset D,
Output: Predicted class label Ypred for each transaction
1. Load the transaction dataset D.
2. Perform missing value analysis and duplicate record checking, Remove identifier-based attributes
4. Generate engineered transactions from amount and balance attributes.
5. Encode categorical transaction type using label encoding.
6. Separate independent features X and target label Y.
7. Fit StandardScaler on the training data and transform both training and testing data.
8. Apply SMOTE only on the training dataset to balance the fraud class.
9. Initialize base classifiers: Random Forest, XGBoost, LightGBM, CatBoost, and Extra Trees.
10. Train each base classifier using the balanced training dataset.
11. Obtain fraud probability predictions from each trained base classifier.
12. Compute the average fraud probability using the soft voting equation.
13. Assign the final class label using the selected decision threshold τ.
14. Select the Voting Ensemble + SMOTE model as the final proposed model.

IV. RESULTS & DISCUSSION

The classification performance of all the fraud detection models implemented is presented in this section. The patterns were tested on the original test set which had 1,272,524 transactions, 1,270,881 of which were normal transactions and 1,643 were fraud

transactions. Despite the reporting of accuracy, it is to be used in conjunction with the reporting of minority class detection measures to prevent misleading conclusions.

Table 4: Overall classification performance of implemented fraud detection models

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC	PR-AUC
Voting Ensemble	0.999953	0.96641	0.99817	0.98204	0.99986	0.99873
Random Forest	0.999948	0.96301	0.99817	0.98028	0.99938	0.99808
Stacking Ensemble	0.999945	0.96021	0.99878	0.97912	0.99986	0.99873
XGBoost	0.999944	0.96073	0.99757	0.9788	0.99984	0.99866
LightGBM + SMOTE	0.999943	0.96016	0.99757	0.97851	0.99982	0.99865
Bagging Decision Tree	0.999926	0.94795	0.99757	0.97212	0.99939	0.99715
CatBoost	0.99992	0.94358	0.99757	0.96982	0.99961	0.99871
Extra Trees	0.999815	0.87647	0.99757	0.93311	0.99938	0.99711
Decision Tree	0.999798	0.86628	0.99757	0.9273	0.99868	0.86417
AdaBoost	0.997621	0.35157	0.99757	0.51991	0.99889	0.9873
Naive Bayes	0.998357	0.38618	0.46257	0.42094	0.97157	0.31
KNN	0.995121	0.20211	0.94279	0.33287	0.97845	0.37939
Hybrid CNN-GRU+ Focal Loss	0.994843	0.19652	0.96957	0.3268	0.99905	0.86023
Logistic Regression	0.994628	0.18577	0.93427	0.30991	0.99756	0.71427
Calibrated SVM	0.9943	0.17705	0.93609	0.29777	0.99636	0.70864
ANN + Focal Loss	0.993674	0.1693	0.99817	0.2895	0.99982	0.9592
Hybrid CNN-LSTM + Focal Loss	0.991772	0.13449	0.98844	0.23677	0.99922	0.89758
Autoencoder + Classifier	0.991628	0.12663	0.93001	0.2229	0.99766	0.80324
1D CNN + Focal Loss	0.9879	0.09566	0.99026	0.17447	0.99932	0.89862
Transformer + Focal Loss	0.98638	0.06773	0.74802	0.1242	0.97167	0.53265

The results of the ensemble learning models were the best overall as seen in Table 4. Among the evaluated approaches, the Voting Ensemble enhanced with oversampling achieved the strongest balance with an accuracy 0.999953, precision 0.96641, recall 0.99817, ROC-AUC 0.99986 and PR-AUC 0.99873. It shows that the suggested voting ensemble the proposed approach correctly classified almost all the fraud cases with a relatively low false alarm rate.

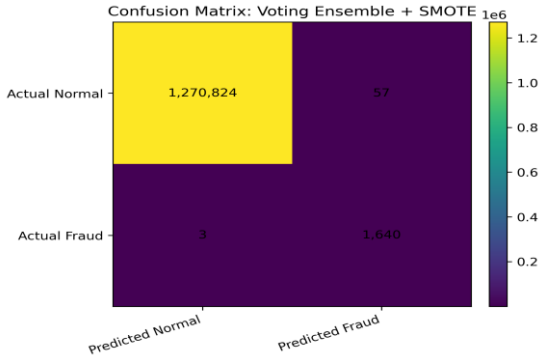


Figure 3: Confusion matrix corresponding to the proposed Voting Ensemble + SMOTE model

According to the confusion matrix (as displayed in Figure 3) experimental findings demonstrated that the model successfully classified most of both legitimate and fraudulent records. The confusion matrix revealed that only 57 legitimate transactions were mistakenly

classified as fraudulent, while merely 3 fraud cases remained undetected. These findings demonstrate the model's strong capability to identify fraudulent activities while maintaining an exceptionally low false-negative rate, which is critical for minimizing financial losses.

The parameter-wise comparison shows that the Voting Ensemble + SMOTE model achieved the best accuracy, precision and F1-score. Stacking Ensemble + SMOTE achieved the highest recall, but its precision and F1-score were slightly lower than the Voting Ensemble as shown in Table 5. Therefore, the Voting Ensemble + SMOTE model is considered the most reliable model because it provides the best overall balance across all important evaluation metrics. This section presented the overall classification results and comparative analysis of all implemented fraud detection models. The experimental results indicate that ensemble learning models outperformed conventional machine learning algorithms, neural network architectures, integrated deep learning frameworks, and attention-driven transformer models. The Voting Ensemble + SMOTE model achieved the best overall performance and are selected as the proposed model for financial fraud detection.

Table 5: Best model according to each evaluation parameter

Parameter	Best Performing Model	Value
Accuracy	Voting Ensemble + SMOTE	0.999953
Precision	Voting Ensemble + SMOTE	0.96641
Recall	Stacking Ensemble + SMOTE	0.99878
F1-score	Voting Ensemble + SMOTE	0.98204
ROC-AUC	Voting Ensemble + SMOTE / Stacking Ensemble + SMOTE	0.99986
PR-AUC	Voting Ensemble + SMOTE / Stacking Ensemble + SMOTE	0.99873

The F1-score comparison highlights the superiority of ensemble models, while precision-recall-F1 comparison shows the balance achieved by the top-performing models.

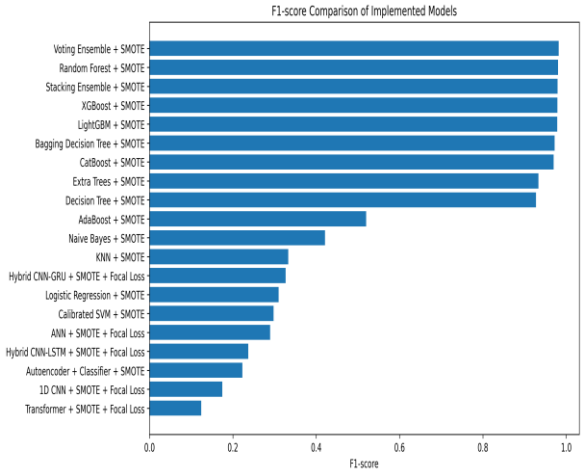


Figure 4: F1-score comparison of all implemented models

The F1-score comparison of all developed fraud detection algorithms is shown in Figure 4. The findings demonstrate that when it comes to identifying fraudulent transactions, ensemble and boosting-based methods perform noticeably better than conventional machine learning and deep learning models. Among all evaluated models, Voting Ensemble + SMOTE achieved the highest F1-score, indicating the best balance between precision and recall.

Other ensemble-based methods also achieved near-perfect F1-scores, demonstrating their strong capability in handling highly imbalanced financial transaction data. In contrast, traditional machine learning models such as Naïve Bayes, KNN, Logistic Regression, and Calibrated SVM exhibited comparatively lower performance. Similarly, deep learning and hybrid deep learning models, including ANN, 1D CNN, Transformer, Autoencoder + Classifier, CNN-LSTM, and CNN-GRU, produced lower F1-scores despite the use of SMOTE and Focal Loss. Overall, the results indicate that ensemble learning techniques demonstrate superior capability in identifying intricate fraudulent behaviours and maintaining a well-balanced combination of detection and false positive reduction, making them highly suitable for large-scale financial fraud detection applications

V. CONCLUSION

Based on the very unbalanced nature of transaction data and the increasing complexity of fraudulent actions, financial fraud detection is still a challenging area. This study proposed a comprehensive machine learning-based fraud detection framework incorporating transaction data preparation, feature construction, and synthetic minority oversampling for class distribution balancing, and comparative evaluation of multiple predictive models. A large-scale financial transaction dataset containing more than 6.3 million records was utilized to assess the effectiveness of traditional machine learning, ensemble learning, boosting, deep learning, hybrid deep learning, and transformer-based approaches. When compared to deep learning models and conventional machine learning techniques, the experimental results demonstrated that ensemble learning techniques performed the best across all evaluation metrics. The overall performance showed that the proposed Voting Ensemble + SMOTE approach achieved the best overall performance. Confusion matrix analysis also verified the efficacy of the model, in correctly recognizing a high percentage of fraudulent transactions with an acceptable false alarm rate. From the results, it has been observed that the combination of all engineered features, imbalance mitigation strategies and ensemble modelling yields significant enhancement in the fraud detection capability. The proposed framework can help financial institutions to improve their transaction monitoring, minimize financial losses and increase the security of financial operations. On-going research could involve developing real-time fraud detection algorithms, fraud analytics tools based on graphs, explainable AI methods, and adaptive learning models that can analyse evolving fraud trends in a dynamic financial landscape.

REFERENCES

[1] K. H. Ahmed, S. Axelsson, Y. Li, and A. M. Sagheer, “A credit card fraud detection approach based on ensemble machine learning classifier with hybrid data sampling,” *Mach. Learn. with Appl.*, vol. 20, p. 100675, 2025, doi: https://doi.org/10.1016/j.mlwa.2025.100675.

[2] M. Akouhar, M. Ouhssini, M. El Fatini, A. Abarda, and E. Agherrabi, “Dynamic oversampling-driven Kolmogorov–Arnold networks for credit card fraud detection: An ensemble approach to robust financial security,” *Egypt. Informatics J.*, vol. 31, p. 100712, 2025, doi: https://doi.org/10.1016/j.eij.2025.100712.

[3] Shpachuk, Vitaliy, Olena Markova, and Bogdan Adamyk. "AI-driven financial fraud: key risks and legal protections for financial institutions." *Journal of Banking Regulation* 27.1 (2026): 6.

[4] Jin, Jing, and Yongqing Zhang. "The analysis of fraud detection in financial market under machine learning." *Scientific Reports* 15.1 (2025): 29959.

[5] N. Almusallam and J. Qayyum, “A Hybrid Feature Selection and Clustering-Based Ensemble Learning Approach for Real-Time Fraud Detection in Financial Transactions,” *Comput. Mater. Contin.*, vol. 85, no. 2, pp. 3653–3687, 2025, doi: https://doi.org/10.32604/cmc.2025.067220.

[6] Khan, Talha Ahmed, et al. "Analysis and visualization of fraud detection patterns through data mining and classification using MLP and hybrid deep learning model." *Egyptian Informatics Journal* 32 (2025): 100829.

[7] R. K. Gupta *et al.*, “Enhanced framework for credit card fraud detection using robust feature selection and a stacking ensemble model approach,” *Results Eng.*, vol. 26, p. 105084, 2025, doi: https://doi.org/10.1016/j.rineng.2025.105084.

[8] C. A. Cedric and H. Ould-Slimane, “Towards effective and robust bank fraud detection thanks to machine learning,” *Procedia Comput. Sci.*, vol. 265, pp. 25–32, 2025, doi: https://doi.org/10.1016/j.procs.2025.07.152.

[9] E. Karnavou, G. Cascavilla, G. Marcelino, and Z. Geradts, “I know you’re a fraud: Uncovering illicit activity in a Greek bank transactions with unsupervised

learning," *Expert Syst. Appl.*, vol. 288, p. 128148, 2025, doi: https://doi.org/10.1016/j.eswa.2025.128148.

[10] Y. Chen, C. Zhao, Y. Xu, C. Nie, and Y. Zhang, "Deep Learning in Financial Fraud Detection: Innovations, Challenges, and Applications," *Data Sci. Manag.*, 2025, doi: https://doi.org/10.1016/j.dsm.2025.08.002.

[11] K. EL Hachimi, G. Orhanou, and L. Ballihi, "Credit card fraud detection model based on optimized long short-term memory : Imbalanced class solution and overfitting reduction," *Expert Syst. Appl.*, vol. 305, p. 130871, 2026, doi: https://doi.org/10.1016/j.eswa.2025.130871.

[12] H. Fanai and H. Abbasimehr, "A novel combined approach based on deep Autoencoder and deep classifiers for credit card fraud detection," *Expert Syst. Appl.*, vol. 217, p. 119562, 2023, doi: https://doi.org/10.1016/j.eswa.2023.119562.

[13] Z. Chen, H. Chen, Y. Liu, and Y. Shi, "Multi-label financial statement fraud detection based on long short-term memory and multilayer perceptron hybrid model," *Eng. Appl. Artif. Intell.*, vol. 170, p. 114188, 2026, doi: https://doi.org/10.1016/j.engappai.2026.114188.

[14] J. Du, C. Hu, H. Zhao, C. Liu, H. Peng, and J. Zhao, "SCALE: Adaptive Heterogeneous Graph Prototype Learning for Open-Set Unknown Financial Fraud Detection," *Expert Syst. Appl.*, p. 132749, 2026, doi: https://doi.org/10.1016/j.eswa.2026.132749.

[15] E. Kim *et al.*, "Champion-challenger analysis for credit card fraud detection: Hybrid ensemble and deep learning," *Expert Syst. Appl.*, vol. 128, pp. 214–224, 2019, doi: https://doi.org/10.1016/j.eswa.2019.03.042.

[16] Eedala, Sri Harsha. *Financial Fraud Detection Dataset*. Kaggle, 2024, https://www.kaggle.com/datasets/sriharshaeedala/financial-fraud-detection-dataset. Accessed 1 June 2026.

[17] Itoo, Fayaz, Meenakshi, and Satwinder Singh. "Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection." *International Journal of Information Technology* 13.4 (2021): 1503-1511.

[18] Yadavalli, Ramu, and Rani Polisetti. "Optimized Financial Fraud Detection using SMOTE-Enhanced Ensemble Learning with CatBoost and LightGBM." *2025 International Conference on Visual Analytics and Data Visualization (ICVADV)*. IEEE, 2025.

[19] Kumar, CH Vasanth, and Premnath Krishnamurthy. "Fraud detection in online consumer reviews using stack ensemble." *AIP Conference Proceedings*. Vol. 3369. No. 1. AIP Publishing LLC, 2026.

[20] Esmaeili, Masoumeh, Mohammad Malekinia, and Alireza Pourebrahimi. "Fraud Detection Analysis in Supplementary Health Insurance Using the LSTM Model." *Digital Transformation and Administration Innovation* 4.1 (2026): 1-9.