A Threshold-Tuned Hybrid Ensemble Approach for Credit Card Fraud Detection

Nitin Rathour
Department of Software Engineering
Delhi Technological University
Delhi - 42, India
rathour.nitin1522002@gmail.com

Dr. Sanjay Patidar
Department of Software Engineering
Delhi Technological University
Delhi - 42, India
sanjaypatidar@dtu.ac.in

***Abstract*—Credit card fraud detection is a chronic and significant issue faced by financial organizations, considering the increasing number and sophistication in fraudulent transactions. This work demonstrates an expert and structured methodology for fraud detection based on the commonly used Kaggle credit card transaction dataset, which is very imbalanced and representative of actual scenarios. We suggest a threshold-tuned hybrid ensemble model integrating supervised learning (Random Forest) and unsupervised anomaly detection methods (Autoencoder and Isolation Forest). Model predictions are blended via a weighted hybrid score, and dynamic threshold tuning is executed to achieve the best compromise between precision and recall. Experimental outcomes validate that our method has an accuracy of nearly 99.9%, precision of 0.96, recall of 0.81, and an F1-score of 0.88 for the fraud class. The robustness of the model is also evidenced through a ROC-AUC of 0.97, which validates its discriminative capability even in the case of extreme class imbalance. In summary, the outlined approach is a practical, understandable, and production-ready solution that considerably improves the detection of uncommon fraudulent transactions while reducing false positives.**

***Index Terms*—Hybrid ensemble, Anomaly detection, Threshold tuning, Random Forest, Autoencoder, Isolation Forest,Credit card fraud detection, Imbalanced data**

I. Introduction

The surge in online financial transactions and e-commerce has revolutionized the global economy while creating new avenues for financial crime, making credit card fraud a major and costly challenge. Industry estimates put global annual monetary losses resulting from fraudulent card transactions at over approximately 30 billion US dollars [1], impacting financial institutions and undermining consumer trust. Detecting such fraud is particularly difficult due to the rarity of fraudulent transactions and ever-evolving tactics used by fraudsters, which leads to highly imbalanced datasets and reduces the effectiveness of conventional rule-based or static machine learning systems.

Algorithms under the supervised learning categories like logistic regression, random forests, and decision trees have been popular in fraud detection for their interpretability and strong performance but require extensive, high-quality labeled data, which is costly to obtain. Moreover, severe class imbalance and the limitations of oversampling methods like SMOTE can result in overfitting and hinder generalizability. Unsupervised and techniques for anomaly detection like Isolation Forests and Autoencoders, can identify anomalies without labels or oversampling, but may have lower precision or higher false positives, especially in financial domains where reliability is critical.

To overcome these challenges, this work introduces a hybrid ensemble that bringing together the predictive power of Random Forest with the anomaly detection capabilities of Autoencoders and Isolation Forests. By employing a threshold-tuned hybrid score and deliberately avoiding synthetic oversampling, our framework maintains data authenticity and enables dynamic tuning of the decision boundary for optimal precision-recall trade-off. Evaluated on a widely-used, imbalanced public dataset, the method achieves notable gains in accuracy and reliability over single-model baselines while remaining interpretable, scalable, and suitable for real-world deployment.

II. Review of Existing Literature

The issue of pinpointing fraudulent behavior in card-based transactions has inspired numerous studies, employing machine learning alongside deep learning models, and ensemble frameworks to tackle data imbalance, evolving fraud patterns, and the need for interpretable, deployment-ready solutions.

Robust empirical baselines have been set by traditional methods. Vallarino (2025) [2] showed that properly tuned classical models can reach top benchmark performance (accuracy 98.7%, precision 94.3%, recall 91.5%), emphasizing the value of sound feature engineering and parameter tuning in skewed data. Saraf Phakatkar (2022) [3] systematically benchmarked algorithms on both actual and synthetic datasets, reporting AUCs of 98.27% and 99.3%, illustrating the trend of testing in multiple environments for robustness and comparability.

Recent work increasingly examines unsupervised and semi-supervised anomaly detection. Thakre (2024) [4] compared Isolation Forest and Autoencoder, finding precision/recall of 0.92/0.87 for the former and 0.95/0.89 for the latter, supporting their utility in imbalanced, heterogeneous settings.

Ensemble and hybrid strategies have gained traction. The Glasgow Caledonian University group (2024) [5] demonstrated that ensembles combined with resampling (SMOTE, undersampling) outperform single models in imbalanced data, but warned that synthetic generation like SMOTE may introduce overfitting and interpretability issues [6]. Deep learning's role is expanding, with Wang et al. (2023) [7] applying LSTM (sensitivity 0.975, specificity 0.990) and SHAP for interpretability. Wu and Zhang et al. (2025) [8] achieved near-perfect results (accuracy 0.9998, F1 0.9998, recall 1.0) with a Convolutional Cascade Neural Network plus SMOTE, though at the cost of complexity and interpretability.

Despite these gains, gaps remain. Most high-performing methods still rely on synthetic data creation (e.g., SMOTE), raising concerns about generalizability and validation. Few works explore threshold tuning within a supervised–unsupervised hybrid ensemble, and none have evaluated the exact combination of Random Forest, Autoencoder, and Isolation Forest—each tuned and weighted in a threshold scheme—without oversampling.

Hybrid modeling is compelling: supervised methods like Random Forest exploit labeled patterns, while unsupervised approaches (Autoencoders, Isolation Forest) adapt to changing fraud behaviors. Threshold tuning in the ensemble enables dynamic optimization of the precision–recall trade-off, critical in operational settings to balance false positives and false negatives.

Our research advances the field with a threshold-tuned hybrid ensemble uniting Random Forest's accuracy and interpretability with the anomaly detection of Autoencoder and Isolation Forest, deliberately omitting SMOTE to preserve transactional data integrity. Focus is placed on both performance and interpretability, along with practical adjustability for deployment. So far as we can determine, this is the pioneering effort to propose and evaluate this specific threshold-tuned hybrid ensemble for credit card fraud prediction, addressing both practical needs and methodological rigor.

Overview and comparison of recent techniques aimed at detecting fraudulent card based activity in the form of ensemble learning, deep architectures, and hybrid approaches along with the current work is given in TABLE I. The table also emphasizes the advantages and disadvantages of the current approaches, further highlighting the novelty and applicability of the proposed threshold-tuned hybrid ensemble framework.

III. Methodology

A. Dataset Description

Our research utilizes the Kaggle dataset focused on credit card fraud detection [9] ("mlg-ulb/creditcardfraud"), which includes 284,807 anonymized European transactions Across a span of 2 days in 2013. Only 492 transactions (0.172%) are labeled as fraud, resulting in severe class imbalance. Each sample comprises 30 features: 28 PCA-transformed components, transaction amount, and time. The output label consists of two classes namely, Fraud and Non-Fraud categories, indicating legitimate (0) or fraudulent (1) transactions. This benchmark dataset was chosen for its real-world imbalance and wide adoption in fraud detection research.

TABLE I
Comparative summary of recent explorations in fraudulent activity identification in card payments methods & the constructed hybrid ensemble system.

Reference	**Models Used**	**Data Handling**	**Key Metrics / Highlight**
Vallarino (2025) [2]	ML (not specified)	Not specified	Accuracy: 98.7%, Precision: 94.3%, Recall: 91.5%
Saraf, Phakatkar (2022) [3]	Not specified	Real & simulated	AUC: 98.27% (real), 99.3% (sim.)
Thakre (2024) [4]	IF, Autoencoder	No SMOTE	IF: Prec 0.92, Rec 0.87; AE: Prec 0.95, Rec 0.89
Glasgow Caledonian Univ. (2024) [5]	Ensemble (SVM, KNN, RF, etc.)	SMOTE, Undersampling	Ensemble best overall; outperformed individuals
Wang et al. (2023) [7]	LSTM, SHAP	Not specified	Sensitivity: 0.975, Specificity: 0.990
Wu & Zhang et al. (2025) [8]	CCNN + SMOTE	SMOTE	Accuracy: 0.9998, F1: 0.9998, Recall: 1.0
This Work (2025)	**RF, Autoencoder, IF (Hybrid)**	**No SMOTE**	**Accuracy: 99.9%, Precision: 0.96, Recall: 0.81, ROC-AUC: 0.95**

B. Data Preprocessing and Splitting

Once the dataset was imported into a pandas DataFrame, initial data exploration was conducted to ensure data integrity, identify missing data, and establish feature distributions. The "Class" column was separated as the target label, while the remaining columns formed the feature set. No explicit feature engineering, scaling, or dimensionality reduction (beyond what was already performed via PCA by the dataset creators) was applied prior to model fitting, as the focus was on evaluating ensemble and anomaly detection methods with the raw, realistic feature set.

An 80-20 split was performed, with the majority of data selected for training purposes and the remainder for testing, with stratification based on class labels and a fixed random seed of 42. All models were trained and thresholds optimized on the training portion, reserving the test set solely for final results.

C. Model Architectures

The system proposed here takes advantage of a threshold-tuned hybrid ensemble, as shown in Fig. 1, to efficiently detect fraudulent transactions within severely imbalanced credit card data. The ensemble combines several machine learning architectures with diverse strengths contributing to the detection system as a whole [10].

1) Random Forest Classifier: Random Forest classifier (utilized through scikit-learn's RandomForestClassifier) served as the principal supervised model thanks to its strength against noisy features, capability to deal with tabular and imbalanced data, and built-in interpretability via feature importance scores [11]. The classifier was created with 100 estimators (trees) and default hyperparameters (Gini impurity for split, no max tree depth, and bootstrapping on). The model was learned using

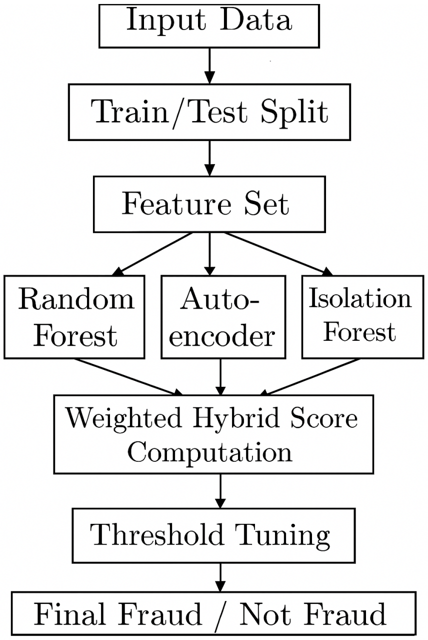

Fig. 1. Visual representation of the threshold-calibrated hybrid ensemble system for credit card transaction fraud identification.

the full set of training features and labels, to predict class probabilities for each transaction [12].

2) Autoencoder for Anomaly Detection: A symmetrical, fully-connected autoencoder unsupervised anomaly detection model functioned with the Keras and TensorFlow libraries [13]. The network structure was as follows:

- Input layer: Size same as number of features (29, not including the target).
- Encoder: Two dense layers—16 neurons with ReLU activation followed by 8 neurons with ReLU activation—to reduce the input into a reduced-dimensional encoding.
- Decoder: Symmetrical structure to decompress the input, culminating in a linear activation output structure (layer) identical in size to the input features.

Only genuine (Class=0) transactions were used to train the Autoencoder over implemented across ten epochs with groups of 64 samples per batch. The learning procedure applied the Adam optimization with objective of reducing MSE loss. During training, the reconstruction error for every transaction (MSE between input and output) was utilized as the anomaly score; the higher the score, the more it deviated from usual transaction patterns and the higher the likelihood of fraud.

3) Isolation Forest: Isolation Forest (Isolation Forest from scikit-learn) was employed as another unsupervised method of anomaly detection. It separates anomalies by randomly choosing a feature and randomly partitioning its value; points needing fewer splits are likely to be anomalies. The contamination parameter was established at 0.01, consistent with the estimated proportion of fraudulent observations. The model was then trained on the full set of training features, and an anomaly score was calculated for each transaction [14]. To scale the anomaly score to fraud probability, the negative of the decision function was utilized—thus, larger values indicated greater anomaly

D. Hybrid Ensemble Construction

Recognizing that every model identifies unique elements of fraudulent behavior, a weighted hybrid ensemble method was employed to combine their predictions [15]. The method encompassed the following steps:

- **Score Calculation**
  - For each transaction, the Random Forest model output the probability of fraud (class=1).
  - The Autoencoder yielded the reconstruction error for each transaction.
  - The Isolation Forest generated anomaly scores as described above.
- **Weighted Fusion**
  - A final hybrid fraud score was calculated for each transaction using a weighted linear combination:

    $$\text{Hybrid Score} = 0.6\times\text{RF}_{\text{scaled}}+0.2\times\text{AE}_{\text{scaled}}+0.2\times\text{IF}_{\text{scaled}}$$

  - The weights were empirically determined based on validation results, with Random Forest given highest priority due to its superior supervised performance, and the two anomaly detectors providing supporting evidence for outlier behavior.

This ensemble strategy allowed for robust and adaptive fraud detection, leveraging both labeled patterns and outlier detection [16].

E. Threshold Tuning and Optimization

Unlike most binary classifiers that use a fixed 0.5 threshold, this study employed dynamic threshold tuning to balance false positives and false negatives according to application needs. Specifically, a precision-recall curve was constructed for the hybrid score on the validation data, and the threshold that maximized the F1-score was selected as optimal. This approach is especially important in the process of identifying financial irregularities, where the cost of missing illegitimate payment activities (false negatives) is usually substantially higher than that of investigating valid transactions (false positives) [17]. The optimal threshold identified in this study was approximately 0.385, enabling a tailored trade-off between detection sensitivity and alert volume.

F. Evaluation Metrics

To critically review the impact of the proposed fraud detection framework, multiple evaluation metrics appropriate for highly imbalanced binary classification were utilized. The following notations are used throughout: TP representing True Positives, FP as False Positives, TN for True Negatives, and FN indicating False Negatives [18]. The most significant evaluation metrics are defined as follows.

Accuracy provides an overall measure of model correctness but can be misleading for imbalanced datasets (see Eq. (1)).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

Precision indicates the proportion of detected fraud cases that are genuinely fraudulent, and it gives the accuracy of positive predictions. High precision is important to avoid false alarms in real-world applications (refer to Eq. (2)).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

Recall as formulated in Eq. (3), it measures the strength of the model to uncover true illegitimate payment activities. Maximal recall is needed to minimize undetected fraud events.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

F1-Score as defined in Eq. (4), This score reflects the harmonic average combining both precision and recall metrics. This metric is especially valuable for datasets with significant class imbalance, as it balances the trade-off between missing true cases (false negatives) and generating unnecessary alerts (false alarms).

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

1) Receiver Operating Characteristic – Area Under Curve (ROC-AUC): According to Eq. (5), ROC-AUC investigates how effectively the model differentiates fraud from non-fraud at different thresholds. This is defined as the area under the ROC, which depicts how the True Positive Rate (Recall) varies in relation to the False Positive Rate.

$$\text{FPR} = \frac{FP}{FP + TN} \tag{5}$$

A superior value for ROC-AUC means the two classes are better separable.

2) Confusion Matrix: The confusion matrix presents the counts of TP, FP, TN, and FN, helping to analyze error categories and trade-offs. In fraud detection, special focus is placed on recall, F1-score, and precision for the minority (fraud) class, as these metrics best capture operational performance. ROC-AUC and the confusion matrix further clarify the balance between detecting fraud and minimizing false positives.

IV. Experimental Outcomes

Here, we present a comprehensive evaluation of the proposed threshold-tuned hybrid ensemble and its individual constituent models on the identifying credit card payment anomalies detection task. The results are reported for the highly imbalanced test set, focusing on the minority (fraud) class due to its practical significance. The evaluation leverages the metrics detailed in Section 3.6, providing a well-defined perspective on the model's effectiveness in both statistical and operational terms.

A. Quantitative Performance

The predictive power of machine learning algorithms in identifying improper activities within card payment systems must be judged on metrics that reflect both their effectiveness in recognizing instances of fraud (recall) and to avoid excessive false positives (precision). Table II compares the Hybrid Ensemble with its individual components, showing that the hybrid approach consistently attains an exemplary balance of precision and recall. Notably, the tuned ensemble delivers a substantial boost in F1-score, highlighting the advantage of combining supervised and unsupervised techniques with optimal thresholding.

TABLE II
Performance metrics for the Hybrid Ensemble and individual models on the test set (fraud class).

Model	Precision	Recall	F1-score	ROC-AUC
Hybrid Ensemble (tuned)	**0.9634**	**0.8061**	**0.8800**	**0.9700**
Autoencoder	0.1020	0.6327	0.1756	0.9550
Isolation Forest	0.9747	0.7857	0.8701	0.9580
Random Forest	0.0265	0.7959	0.0513	0.9324

The high precision and recall achieved by the hybrid model demonstrate its effectiveness in capturing fraudulent activity while keeping false alarms to a minimum, which is crucial for reducing operational costs and maintaining customer trust.

B. ROC and PR Curves

The ROC and PR curves offer a visual evaluation of the discrimination ability of the model. The area under these curves (AUC) measures the alignment of false positive and true positive rates over thresholds [19]. As seen in Fig 2, 3, 4, and 5 the hybrid ensemble (Fig. 5) has the best AUC and PR values, demonstrating its highest capacity to differentiate fraud from legitimate transactions even with extreme class imbalance [20]. For comparison, the Random Forest (Fig 2), Autoencoder (Fig 3), and Isolation Forest (Fig 4) models exhibit lower AUC and PR values, again underscoring the benefit of the proposed ensemble method.

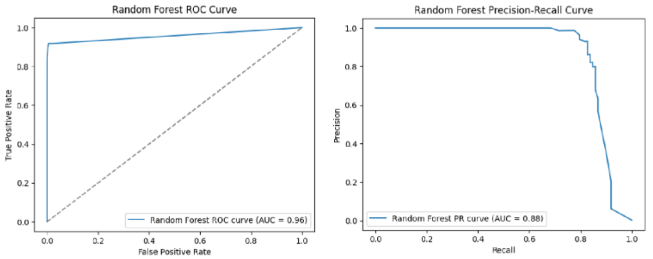

Fig. 2. ROC and PR for the Random Forest model.

C. Confusion Matrices

Confusion matrices provide a granular view of prediction outcomes, indicating the balance between correctly and incorrectly classified samples for each class. The tables below present the confusion matrices for each model, further

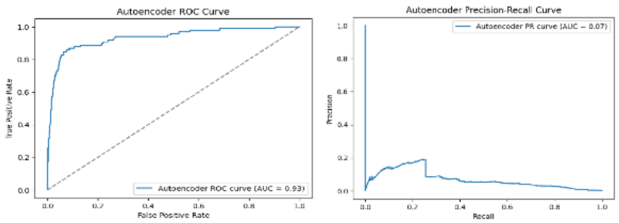

Fig. 3. ROC and PR for the Autoencoder model.

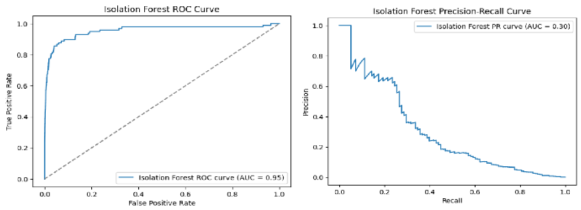

Fig. 4. ROC and PR for the Isolation Forest model.

clarifying their respective strengths and weaknesses in fraud detection, as summarized in Table III

TABLE III
CONFUSION MATRIX COMPONENTS FOR ALL MODELS ON THE TEST SET.

Model	**True Negatives**	**False Positives**	**False Negatives**	**True Positives**
Random Forest	56862	2	21	77
Autoencoder	53997	2867	20	78
Isolation Forest	56318	546	36	62
Hybrid Ensemble (tuned)	56861	3	19	79

D. *Summary*

Overall, the results confirm that the threshold-tuned hybrid ensemble not only achieves high accuracy, but also optimally balances precision and recall on the minority fraud class—outperforming each constituent model individually. The combination of supervised and unsupervised learning, along with data-driven threshold optimization, enables robust detection of rare fraudulent activities of transactions while ensuring that false positives remain minimal. These findings underscore the practical suitability within the presented framework for real-world card-based financial fraud detection deployments.

V. DISCUSSION

The experimental results underscore the efficacy and usability of the suggested threshold-tuned hybrid ensemble for credit card fraud prediction. In comparison to standalone models Random Forest, Autoencoder, and Isolation Forest, the hybrid model demonstrated the best possible compromise between precision and recall, which is critical in the case of severely skewed real-world data. Interestingly, the hybrid method achieved a considerable boost in F1-score and ROC-AUC to 0.88 and 0.97, respectively, on the minority (fraud) class.

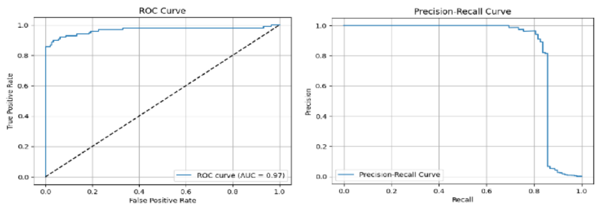

Fig. 5. ROC and PR for the Hybrid Ensemble (AUC = 0.97).

Such enhancement is largely credited to the combination of supervised and unsupervised learning modules. The ensemble model was able to effectively combine the best of every component. The outcome was significant minimization of false positives and a better capacity to identify fraudulent transactions with high confidence. Threshold tuning enabled flexible adjustment of the trade-off between missed fraud and false alarms, supporting adaptation to different operational needs. By using only real transaction data, the hybrid model improves interpretability and reliability for live deployment.

In short, the threshold-tuned ensemble hybrid surpassed all of the single-model baselines on almost all pertinent evaluation measures, affirming its promise for real-world credit card fraud detection. Its high discriminatory capability, flexibility through threshold tuning, and ease of implementation qualify it as a prime candidate to be incorporated into financial fraud detection systems.

VI. CONCLUSION AND FUTURE WORK

In brief, this paper introduces a new threshold-tuned hybrid ensemble approach combining Isolation Forest, Random Forest, and Autoencoder models to recognize suspicious activity in card-based payment systems, without relying on synthetic oversampling. The proposed approach led to significant improvements compared to single baselines, particularly attaining an F1-score of 88% as well as a ROC-AUC score of 97% on the minority (fraud) class in a real-world, extremely imbalanced dataset. These findings show not only high discriminative power but also the unique combination of high recall and high precision—qualities essential in reducing financial risk and operational expenses.

Next steps include deploying the hybrid ensemble in real-world transaction monitoring systems to assess performance and impact under practical constraints. A forthcoming journal paper will provide expanded benchmarking with additional models, ensemble strategies, explainability analysis, and focus on adaptability to evolving fraud schemes. In addition, the journal article will discuss the application of this hybrid framework to streaming data and other rare-event detection domains, including scalability, data privacy, and flexibility in dynamic environments.

The present research enhances methodologies for identifying illicit transaction activities and provides a solid basis for future academic and industrial exploration, which will be elaborated in future journal publications.

Acknowledgment

The authors express sincere gratitude to Delhi Technological University for providing the necessary infrastructure and research support for this study.

References

[1] Nilson Report, "Global Card Fraud Losses Reach $32.34 Billion," Issue 1218, February 2022, 2022, Available: https://nilsonreport.com.

[2] D. Vallarino, "Modeling adaptive fraud patterns: An agent-centric hybrid framework with moe and deep learning," *Available at SSRN 5001848*, 2024.

[3] S. Saraf and A. Phakatkar, "Detection of credit card fraud using a hybrid ensemble model," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 9, pp. 464–474, 2022.

[4] B. Thakre, "Enhancing retail fraud detection with isolation forests and autoencoders: Overcoming data limitations and regulatory challenges," *Journal of Advanced Research in Engineering and Technology*, vol. 3, no. 1, pp. 170–180, 2024. [Online]. Available: https://www.cur.org/wp-content/uploads/2024/07/NCUR-2024-Proceedings-7.31.24.pdf

[5] A. R. Khalid, N. Owoh, O. Uthmani, M. Ashawa, J. Osamor, and J. Adejoh, "Enhancing credit card fraud detection: an ensemble machine learning approach," *Big Data and Cognitive Computing*, vol. 8, no. 1, p. 6, 2024.

[6] I. M. Alkhawaldeh, M. Al-Jafari, M. H. El din Moawad, Y. J. Alabdallat, M. S. Abdelgalil, A. K. AlQurm, S. Z. Eddin, L. H. Darwish, H. K. Alsalhi, S. G. Odeh *et al.*, "Neuro-oncological research output in the middle east: A scoping review," *Cancer Treatment and Research Communications*, vol. 43, p. 100883, 2025.

[7] B. Yousefimehr and M. Ghatee, "A distribution-preserving method for resampling combined with lightgbm–lstm for sequence-wise fraud detection in credit card transactions," *Expert Systems with Applications*, 2025, article in press. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0957417424025284

[8] Y. Wu, L. Wang, H. Li, and J. Liu, "A deep learning method of credit card fraud detection based on continuous-coupled neural networks," *Mathematics*, vol. 13, no. 5, p. 819, 2025.

[9] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection data," https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud, 2015, accessed: 2025-06-28.

[10] K. Boyd, V. S. Costa, J. Davis, and C. D. Page, "Unachievable region in precision-recall space and its effect on empirical evaluation," in *Proceedings of the... International Conference on Machine Learning. International Conference on Machine Learning*, vol. 2012, 2012, p. 349.

[11] S. Gupta, S. Modgil, S. Bhattacharyya, and I. Bose, "Artificial intelligence for decision support systems in the field of operations research: review and future scope of research," *Annals of Operations Research*, vol. 308, no. 1, pp. 215–274, 2022.

[12] S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang, and C. Jiang, "Random forest for credit card fraud detection," in *2018 IEEE 15th international conference on networking, sensing and control (ICNSC)*. IEEE, 2018, pp. 1–6.

[13] A. Pumsirirat and L. Yan, "Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine.(ijacsa) international journal of advanced computer science and applications, 9 (1), 1–8, 2018.[ebook]," 2018.

[14] H. Tabrizchi and J. Razmara, "Credit card fraud detection using hybridization of isolation forest with grey wolf optimizer algorithm," *Soft Computing*, vol. 28, no. 17, pp. 10 215–10 233, 2024.

[15] F. Carcillo, Y.-A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, and G. Bontempi, "Combining unsupervised and supervised learning in credit card fraud detection," *Information sciences*, vol. 557, pp. 317–331, 2021.

[16] H. Fanai and H. Abbasimehr, "A novel combined approach based on deep autoencoder and deep classifiers for credit card fraud detection," *Expert Systems with Applications*, vol. 217, p. 119562, 2023.

[17] Q. Liu and X. Zou, "Research on trust mechanism of cooperation innovation with big data processing based on blockchain," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, pp. 1–11, 2019.

[18] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection: a realistic modeling and a novel learning strategy," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 8, pp. 3784–3797, 2017.

[19] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in *Proceedings of the 23rd international conference on Machine learning (ICML)*. ACM, 2006, pp. 233–240.

[20] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets," *PloS one*, vol. 10, no. 3, p. e0118432, 2015.