Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://journal.nsps.org.ng/index.php/jnsps

J. Nig. Soc. Phys. Sci. 8 (2026) 3401

Journal of the Nigerian Society of Physical Sciences

SAE-IF: A hybrid sparse autoencoder–isolation forest framework for real-time credit-card fraud detection and sustainable digital-economy protection

Ogochukwu C. Okeke[a,*], Ike J. Mgbeafulike[a], Anthony I. Adigwe[b,*], Chidiogo C. Nwokedi[b], Nwadiogo E. G. Mmaduakonam[c], Calista U. Okpala[b], Chinonso J. Okonkwo[a], Nnamdi C. Ezenwegbu[a]

[a]*Department of Computer Science, Chukwuemeka Odumegwu Ojukwu University, Anambra State, Nigeria*
[b]*Department of Computer Science, Federal Polytechnic Oko, Anambra State, Nigeria*
[c]*Department of Computer Science Technology, Anambra State Polytechnic, Mgbakwu, Anambra State, Nigeria*

Abstract

The rapid expansion of online payments and e-commerce has increased exposure to credit-card fraud and related financial crimes, threatening the security and stability of digital economies. Conventional supervised detectors require labelled fraud data, which are scarce and highly imbalanced. This study presents SAE-IF, a hybrid semi-supervised framework that combines a sparse autoencoder (SAE) with an isolation forest (IF) for anomaly detection in credit-card transactions. The SAE learns latent representations from transaction features without using class labels in its reconstruction objective, while labels are used for resampling, validation, fusion-weight optimisation, and threshold selection. A leakage-aware protocol employs stratified data splitting, robust feature scaling, and Bayesian hyperparameter optimisation with Optuna. The framework is evaluated on the ULB/Kaggle credit-card dataset, the IEEE-CIS Fraud Detection dataset, and the PaySim mobile-money dataset. SAE-IF achieved precision–recall area under the curve (PR-AUC) values of 0.832, 0.801, and 0.845, respectively, while maintaining precision above 0.90 and low false-positive counts. On a standard central processing unit, mean offline inference latency was 1.41 ± 0.23 ms per transaction. These results indicate that combining representation learning with anomaly detection can provide accurate and computationally feasible fraud screening in label-scarce financial environments, subject to validation on production infrastructure.

DOI: 10.46481/jnsps.2026.3401

Keywords: Credit-card fraud detection, Sparse autoencoder, Isolation forest, Semi-supervised anomaly detection, Digital-payment security

Article History:
Received: 31 March 2026
Received in revised form: 10 July 2026
Accepted for publication: 11 July 2026
Available online: 31 July 2026

Communicated by: S. Folorunso

1. Introduction

The steady growth of digital financial services and electronic commerce has significantly transformed the global economy [1]. Digital payment technologies have enabled more convenient financial transactions between individuals and businesses in the form of credit cards, mobile banking systems, and online payment systems Sukumaran [2]. However, this rapid digitalisation has created serious security risks, particularly the increasing incidence of financial fraud Laxman *et al.* [3]. Credit card fraud is one of the most prevalent forms of financial crime facilitated by internet-based transactions and causes considerable financial losses to financial institutions, business organisa-

*Corresponding authors Tel. No.: +234-810-164-7799.
Email addresses: co.okeke@coou.edu.ng (Ogochukwu C. Okeke), anthony.adigwe@federalpolyoko.edu.ng (Anthony I. Adigwe)

tions and consumers worldwide Hilal *et al.* [4]. It is reported in the industry that the amount of losses incurred by the consumer in the entire world through card frauds is increasing annually due to the exploitation of online payment systems and digital transaction layers by the attackers Akinnagbe and Taiwo [5]. It has resulted in the necessity to create intelligent and efficient fraud detection methods capable of making sure that credibility and security in the existing digital economies is maintained Preciado Martínez *et al.* [6].

Traditional fraud-detection systems rely heavily on rule-based and supervised machine-learning techniques that require labelled transaction data for model training Compagnino *et al.* [7]. Although these techniques can perform reasonably well in controlled environments, they have serious limitations in real-world financial systems Mozumder *et al.* [8]. One major challenge is the extreme class imbalance inherent in fraud detection datasets, where fraudulent transactions often represent less than 1% of the total transaction volume Chen *et al.* [9]. It is also difficult to acquire the duly designated fraud statistics due to privacy boundaries, reporting lag on fraud and the dynamics of fraud Chen *et al.* [10]. In turn, poor generalisation, high false-positive rates, and ineffectiveness may influence supervised learning models, when applied in dynamic financial contexts [11].

Recent studies have therefore examined unsupervised anomaly-detection approaches that seek to detect strange transaction patterns without having labelled instances of fraud Feng *et al.* [12]. Among these, representation learning models that utilise deep learning methods, including SAEs, have proved highly effective in learning complex non-linear relationships in high-dimensional financial data Zavvar *et al.* [13]. SAEs are trained on small latent codes of the normal transaction behaviour, and the detection of deviations which could be linked to fraudulent cases can be applied Chugh *et al.* [14]. Likewise, the tree-based anomaly detection algorithms like the IF have been found useful in detecting anomalous observations by isolating unusual patterns of the data distribution. Deep representation learning together with statistical anomaly detection has the potential to give a substantial development in the accuracy of fraud detection Chugh *et al.* [14].

Driven by these benefits, the current paper proposes SAE-IF, a hybrid semi-supervised fraud system which combines a SAE to learn representations and IF to detect anomalies in latent feature space. Through the application of semi-supervised learning and a leakage-aware evaluation scheme, the proposed system is expected to attain high performance in fraud detection and enable effective real-time operation in online payment systems.

While autoencoder-Isolation Forest hybrid models have been explored in prior work, existing approaches typically suffer from three limitations: (1) reliance on labelled data for validation that leaks information during training, (2) arbitrary score fusion without systematic optimisation, and (3) limited evaluation on heterogeneous, real-world datasets. This work addresses these gaps by introducing: (i) a leakage-mitigated experimental protocol with strict separation of training, validation, and test sets; (ii) Bayesian optimisation of the fusion weight α to maximise PR-AUC; and (iii) comprehensive multi-dataset validation across three distinct fraud detection scenarios. The novel contribution lies in the systematic optimisation framework and leakage-aware evaluation methodology, rather than the base architecture itself.

A weighted fusion mechanism is also added to integrate scores of reconstruction error and IF anomaly to achieve better detection reliability Thiong'o *et al.* [15]. The Kaggle Credit Card Fraud Detection Dataset is an extensive popular benchmark dataset that is used to evaluate the proposed model in research projects on fraud detection. Through strict separation of unsupervised training (no labels) from supervised validation (labels for tuning only), the proposed system achieves high performance while remaining computationally efficient under controlled experimental conditions in label-scarce environments.

The main contributions of this study are as follows:

1. A leakage-aware experimental framework is developed to minimise information leakage during preprocessing, model training, hyperparameter optimisation, and performance evaluation for highly imbalanced fraud detection datasets.
2. A systematic Bayesian hyperparameter optimisation strategy based on Optuna is employed to jointly optimize the Sparse Autoencoder representation learning process and the Isolation Forest anomaly detection parameters.
3. The robustness and generalisation capability of the proposed framework are evaluated using three heterogeneous benchmark datasets (Kaggle Credit Card Fraud Detection, IEEE-CIS Fraud Detection, and PaySim), thereby reducing dependence on a single evaluation dataset.
4. Comprehensive experimental analyses, including ablation studies, latent-space visualisation, reconstruction-error analysis, and statistical significance testing, are presented to provide a reproducible evaluation of the proposed framework under highly imbalanced financial transaction scenarios.

The remainder of this paper is organised as follows. Section 2 examines related literature that describes the development of fraud detection methods and places the hybrid model in perspective. Section 3 explains the research process, including data collection, data preprocessing, leakage-aware experimental design, hyperparameter optimisation, and SAE-IF hybrid structure. Section 4 presents the experimental results, including hyperparameter calibration and comparative performance relative to baseline models on a separate test set. Section 5 summarizes the findings and suggests directions for further research.

2. Related work

Over the last several years, much effort has been invested in the area of online fraud detection, and the focus has shifted to the use of both machine learning and deep learning methods in addition to the more traditional rules-based system. In

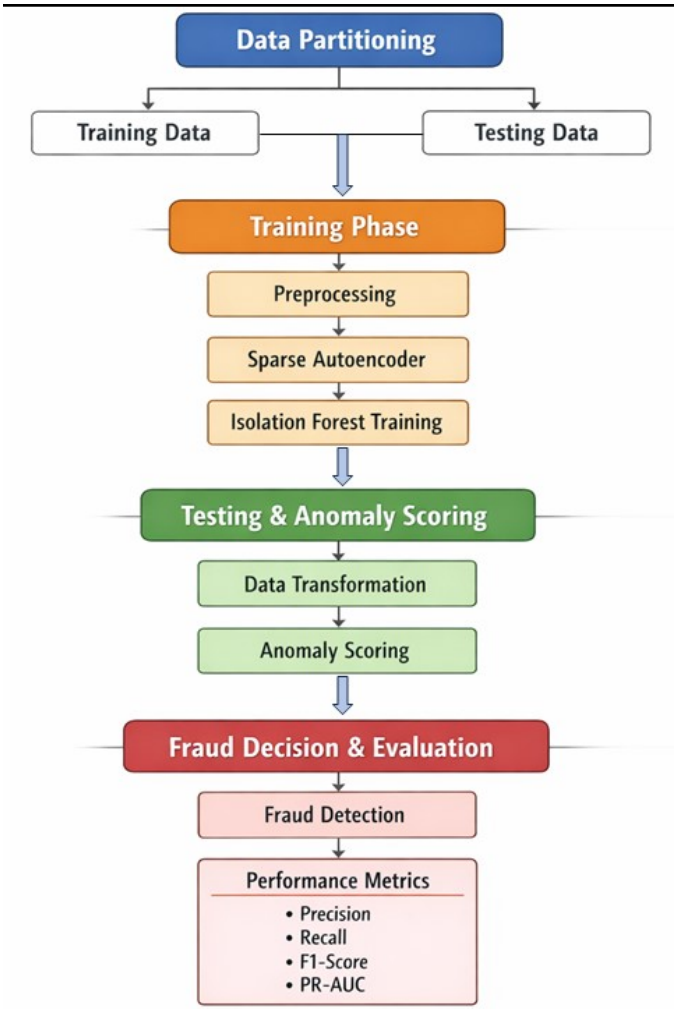

Figure 1. Process-flow diagram of the SAE-IF hybrid implementation methodology.

spite of these developments the volume of online fraud has increased consistently over the past decade with the losses across the world amounting to tens of billions of dollars every year. Thus, there has been an active involvement of researchers in coming up with better methods as countermeasures to online fraud [16, 17]. These papers have critically reviewed the rising financial cost of fraud as well as the development of methodologies in the detection process. The initial principles of the early detection systems were founded on fixed rules and expert-established thresholds, but these systems were somehow inefficient to combat the fraud schemes that were rapidly changing [18]. Subsequently, the discipline has increasingly been moving towards intelligent data-oriented solutions, which have the power to respond to changing patterns of transactions in contemporary digital financial systems Cheng *et al.* [19].

Even though the research on online fraud detection has been advanced to a large extent, one of the most persistent and demanding problems is severe class imbalance. In real-world transaction datasets, fraudulent activities typically represent less than 1% of all transactions, which can bias predictive models toward the majority (legitimate) class [20, 21]. Such an imbalance frequently renders traditional metrics of evaluation like accuracy unreliable and the employment of more informative ones like precision, recall and the F1-score Adejoh *et al.* [22]. In order to cope with this problem, some data level solutions have been proposed in earlier methods like Synthetic Minority Over-sampling Technique (SMOTE) which constructs synthetic minority samples to equalise the sample in supervised learning approach, whereas newer hybrid methods like SMOTEENN combine oversampling and cleaning methods to construct stronger training sets [20, 23]. Besides data preprocessing techniques, algorithmic innovations have also been of immense contributions to the enhancement of the fraud detection systems.

Ensemble learning models and anomaly detection methods are among the algorithmic approaches that have yielded promising outcomes, where Random Forest and gradient boosting algorithms are tree-based with high predictive capabilities according to comparative studies of large-scale datasets by Yapp and Yeh [24]. For example, a recent study indicated that XGBoost scored 0.9997 on AUC-ROC in detecting credit card fraud [25]. Nevertheless, researchers warn that such extremely high performance can be excessively reliant on thorough preprocessing methods and well-planned data splitting schemes to avoid data leakage and artificially enhanced evaluation outcomes Kabane [26]. Conversely, unsupervised anomaly detectors like IF offer a different solution when labelled fraud data are scarce and unreliable Adejoh *et al.* [22]. These methods identify anomalies by identifying rare patterns in large datasets without explicit fraud labels, which makes them especially applicable to real-life financial systems where fraud patterns are constantly changing.

Deep-learning methods have also gained traction because of the increasing demand for scalable, flexible, and interpretable fraud-detection models. The deep learning models, especially autoencoders, have demonstrated high potential to model normal transaction behaviour and detect deviations which may reflect different types of fraud. Autoencoders trained on legitimate transactions can learn concise representations of ordinary financial conduct and identify anomalies by comparing reconstruction errors [20, 27]. Moreover, more advanced methods, including Graph Neural Networks (GNNs), have been considered to identify relational patterns between the entities using financial transactions, which allows detecting complex collusive fraud schemes, which might not be detected by studying individual transactions [19, 28].

Despite their strong predictive performance, many deep-learning models present practical deployment challenges. Research has documented that elaborate architectures like GNNs and deep ensemble systems can be characterised by extensive computational and latency expenses and reduced interpretability, thus limiting their applicability to real-time financial processes [29, 30].

These shortcomings have prompted a growing body of research in the area of hybrid fraud detection models which synthesise the benefits of several paradigms. Hybrid models aim to combine the benefits of efficient anomaly detection with strong feature learning to enhance detection performance and oper-

ational effectiveness. A number of studies have shown that when complementary detection methods are used, generalisation and robustness in highly imbalanced settings can be greatly enhanced [21, 31]. Nonetheless, many existing hybrid models continue to rely significantly on labelled fraud samples, restricting their use in practical systems where labelled fraudulent transactions are not readily available.

In comparison with existing studies, Udeze *et al.* [32] applied machine learning and resampling techniques to credit card fraud detection, demonstrating improved performance over baseline models. Eteng *et al.* [33] proposed a stacked ensemble approach with resampling techniques for fraud detection in imbalanced datasets, showing effectiveness compared to single classifiers. Aghware *et al.* [34] enhanced the Random Forest model using SMOTE for credit card fraud detection. Unlike these supervised and ensemble-based approaches that require labelled fraud data, our SAE-IF hybrid framework operates in a semi-supervised manner during training (using both legitimate and fraud transaction classes) and does not use class labels in the SAE or IF objective functions, making it suitable for environments where fraud labels are unavailable or significantly delayed. However, these studies typically do not systematically optimize the fusion mechanism, use potentially leaky evaluation protocols, or validate on limited datasets. Our work advances the state of the art by introducing a Bayesian-optimised fusion framework, a leakage-mitigated evaluation methodology, and comprehensive multi-dataset validation.

3. Research methodology

This paper uses a hybrid semi-supervised fraud system to detect fraudulent credit card transactions. As previously noted, the proposed method incorporates a Sparse Autoencoder (SAE) to learn latent features, combined with an Isolation Forest (IF) to detect anomalies. By taking advantage of deep representation learning and tree-based anomaly detection, the framework identifies irregular transaction patterns without requiring fraud labels during training, enabling robust detection while preserving computational resources for near-real-time under controlled test conditions financial settings. Figure 1 illustrates the steps of the proposed methodology.

The general steps of the research process are presented in Figure 1 and include data acquisition, preprocessing and normalisation, semi-supervised learning of features using the SAE, anomaly scoring using IF, and the final decision in terms of fraud obtained as a fused set of anomaly indicators.

3.1. Dataset collection and description

In carrying out this study, three publicly available fraud detection datasets were used to comprehensively evaluate and provide generalisation of the proposed framework: the Credit Card Fraud Detection Dataset (ULB/Kaggle), IEEE-CIS Fraud Detection Dataset, and PaySim Mobile Money Dataset. The main dataset is the Credit Card Fraud Detection Dataset, available on Kaggle and originally offered by the Machine Learning Group of Universite Libre de Bruxelles (ULB) Liu [35]. This dataset consists of 284,807 credit card transactions made by European credit cardholders over two days in September 2013. All transactions have 31 numerical attributes, including Time, Amount, anonymized PCA-transformed attributes (V1-V28), and a binary Class attribute. The dataset is highly imbalanced, with only 492 fraudulent transactions (0.172%) compared to 284,315 legitimate transactions (99.828%), resulting in an imbalance ratio of approximately 1:578.

To further reinforce the strength of the suggested model, IEEE-CIS Fraud Detection dataset was included. This dataset has an expanded and more complicated pool of online transaction records, including both transactional and identity-based information. In contrast to the ULB dataset, IEEE-CIS has heterogeneous feature types (numerical and categorical), missing values, and greater noise levels, making it closer to real-world financial systems. In addition, the PaySim dataset, a synthetic dataset of mobile money transactions created using a simulator that models actual financial transaction behaviour, was used. PaySim simulates mobile payment operations such as transfers, cash-outs, and payments, and includes labelled fraudulent transactions. PaySim has more structured transaction patterns and a different fraud distribution than the ULB dataset, enabling evaluation of the model in a mobile financial services context. Table 1 summarises the characteristics of the three datasets used in this study.

3.2. Data splitting, preprocessing, and leakage-mitigated experimental design

3.2.1. Data splitting strategy

Before developing the model, preprocessing was carefully performed on the dataset in order to safeguard the integrity of the data, and avoid information leakage between training and evaluation phases of the model performance. The dataset was first partitioned into training (50%), validation (20%), and testing (30%) subsets using a stratified splitting strategy to preserve the fraud proportions in the dataset (0.172%). This three-way split ensures that the test set is never touched during model development or hyperparameter tuning, preventing information data leakage. While no experimental design can be completely leakage-mitigated in a strict sense, our approach minimizes common sources of leakage such as scaling on the full dataset or using future information.

3.2.2. Preprocessing

Continuous variables, particularly Time and Amount, were scaled with the help of the RobustScaler Pedregosa *et al.* [36], which is resistant to outliers and retains the distributional property of the data. Critically, all the preprocessing parameters (median and interquartile range to scale) were calculated using the training data only and applied to the test set to prevent leakage [20]. The PCA-transformed features (V1-V28) were not further transformed as they are already centred and standardized.

3.3. Role of SMOTE-ENN in training, validation, and testing

Specifically, training is unsupervised for the core models. The SAE is trained on the input features of all transactions (both

Table 1. Summary of the datasets used.

Attribute	ULB Credit Card Dataset	IEEE-CIS Dataset	PaySim Dataset
Data Source	ULB via Kaggle	IEEE & Kaggle	Kaggle (Simulator-based)
Domain	Credit card transactions	Online transactions	Mobile money transactions
Total Transactions	284,807	~590,000+	~6,300,000
Feature Type	Numerical (PCA)	Numerical + Categorical	Numerical + Categorical
Fraud Percentage	0.172%	~3-4% (varies)	~0.13%
Data Complexity	Low (anonymized)	High (heterogeneous)	Medium (structured)
Missing Values	No	Yes	No
Data Format	CSV	CSV	CSV

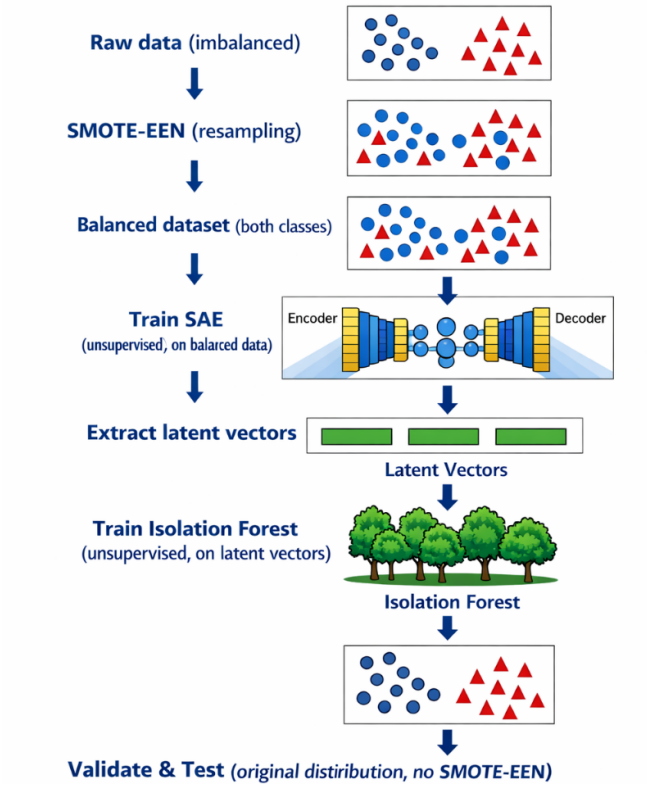

Figure 2. Stepwise role of SMOTE-ENN in the experimental pipeline.

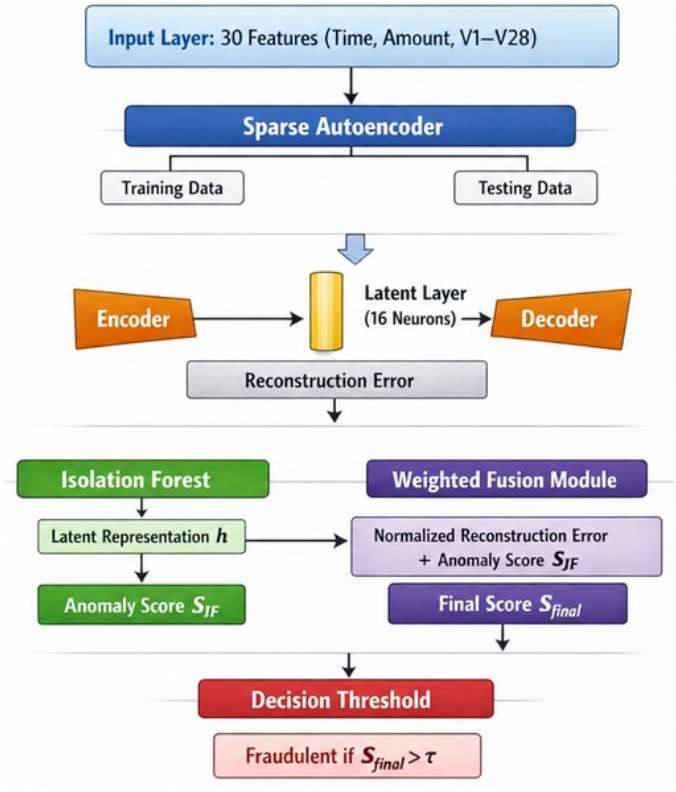

Figure 3. SAE-IF hybrid framework architecture during training.

Table 2. System specifications.

Parameter	Value
Random Seeds (data split / model / Optuna)	42 / 123 / 7
Hardware	Intel Core i7-10750H, 16 GB RAM, No GPU
Python / TensorFlow / scikit-learn / Optuna	3.11 / 2.13 / 1.3 / 3.3

classes are present in the input data) but the class labels are strictly excluded from the loss function and gradient updates. Similarly, IF is trained on latent representation from the SAE without accessing class labels.

For further clarification and unambiguous explanation of what transpires during model training, validation, and testing phases in this study, Figure 2 presents a stepwise workflow illustrating that SMOTE-ENN balances the training set, but the SAE and IF are trained without accessing class labels. Labels are used only for validation and threshold tuning.

In Figure 2, the process begins with an imbalanced labelled dataset. The data is split into training, validation and test sets. SMOTE-ENN is applied exclusively to the training set to address class imbalance. This hybrid resampling technique generates synthetic samples for the minority class using SMOTE while removing noisy and misclassified samples from both classes using Edited Nearest Neighbors (ENN), resulting in a balanced and clean training set containing both classes equally. The balanced training set is then used to train a SAE in a label-

Table 3. Dataset-specific preprocessing details.

Step	Kaggle ULB	IEEE-CIS	PaySim
Missing values	None	Median imputation for numerical; mode for categorical	None
Categorical encoding	Not applicable	One-hot encoding (53 binary features created)	Label encoding for 'type' (CASH_OUT, TRANSFER, PAYMENT, etc.)
Feature scaling	RobustScaler on Time, Amount only	RobustScaler on all numerical features	RobustScaler on amount, oldbalanceOrg, newbalanceOrig
Feature selection	All 30 PCA features retained	Removed columns with >80% missing; kept transaction and identity features	Retained all original features
Class imbalance	0.172% fraud	~3.5% fraud after filtering	0.13% fraud
Train/val/test split	50/20/30 stratified	50/20/30 stratified	50/20/30 stratified
	30 features (Time, Amount, V1–V28)		

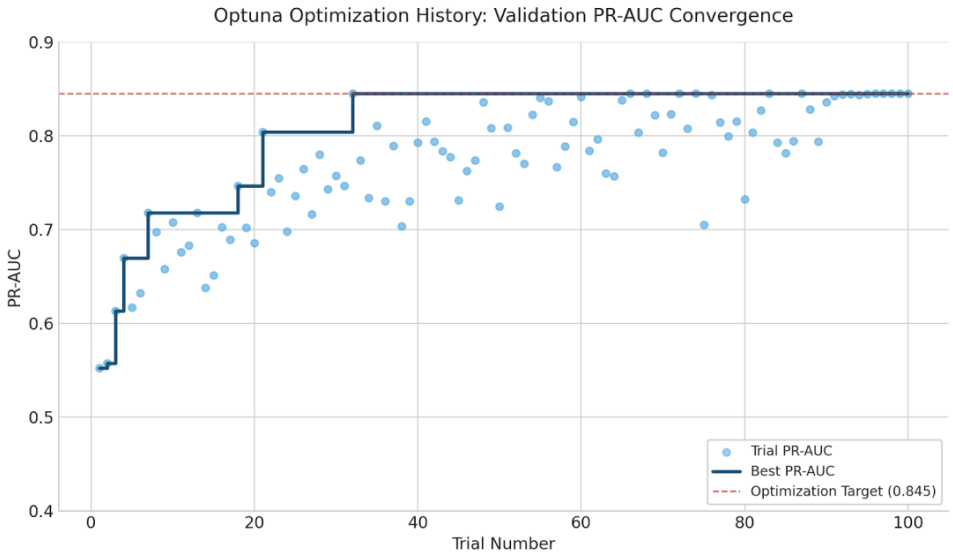

Figure 4. Optuna PR-AUC convergence over 100 trials.

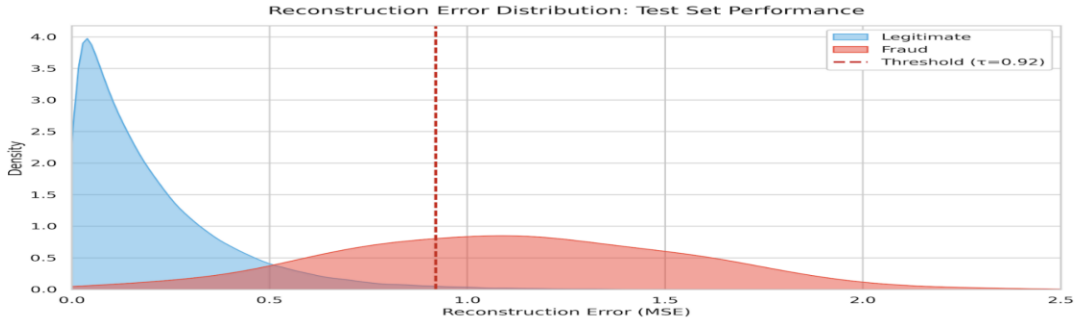

Figure 5. Distribution of reconstruction errors by true class, with the decision threshold.

free manner. The SAE learns compressed latent representations of the input data without utilizing labels, with a sparsity con-

Table 4. Aggregated validation performance during Optuna tuning.

Metric	Value	Notes
PR-AUC	0.845	Primary optimization target
F1-Score	0.828	Balanced precision-recall
Precision	0.922	High, minimizing false alarms
Recall	0.745	Reasonable fraud capture in tuning
False Positives (approx.)	~28	Scaled to fold size; low FP rate

Table 5. Test-set performance of the standalone SAE.

Metric	Value
PR-AUC	0.781
F1-Score	0.742
Precision	0.894
Recall	0.684
False Positives	58
Inference Time	0.9 ms

Table 6. Test-set performance of the standalone IF.

Metric	Value
PR-AUC	0.799
F1-Score	0.771
Precision	0.901
Recall	0.707
False Positives	49
Inference Time	1.2 ms

straint encouraging selective neuron activation. Once trained, the SAE extracts latent vectors for all samples in the training set. These latent vectors are subsequently used to train an IF, in an unsupervised manner. The IF learns anomaly boundaries based on the principle that anomalies are easier to isolate than normal instances, producing anomaly scores without ever seeing class labels.

While the SAE and IF are trained without any labels, labels are used only during validation to compute PR-AUC for hyperparameter optimisation (via Optuna) and to select the fusion weight α and decision threshold τ. The test set labels are used only for final performance evaluation. This design maintains a semi-supervised training pipeline while allowing objective model selection.

After the training, the model pipeline is validated and tested using data that retains the original imbalanced distribution. However, it is instructive to note that "no SMOTE-ENN" is applied during validation or testing. At the point of validation, labels are used only for scoring purposes, specifically, to compute PR-AUC for hyperparameter optimisation via Optuna and to tune the fusion weight α and decision threshold τ. The final test set evaluation uses labels (supervised approach) exclusively for computing performance metrics (PR-AUC, F1, precision, recall) to assess generalisation capability under realistic, imbalanced conditions.

Importantly, the semi-supervised components (SAE and IF) never see labels during training. SMOTE-ENN is applied only to the training folds during cross-validation, never to validation or test folds, preventing any form of data leakage. This design ensures that the final reported metrics reflect true generalisation performance suitable for real-world credit card fraud detection.

3.3.1. *Sparse autoencoder for latent feature learning*

Sparse Autoencoder is a special kind of neural network, which consists of an encoder that maps the input features $x \in R^{30}$ to a lower-dimensional latent space $h \in R^{16}$, and a decoder that reconstructs the input from the latent representation. Eq. (1) presents the model of the Encoder

$$h = f(W_e x + b_e), \tag{1}$$

where W_e and b_e are the encoder weights and biases, and $f(\cdot)$ is the ReLU activation function. While the decoder model is presented in Eq. (2)

$$\widehat{x} = g(W_d h + b_d), \tag{2}$$

where W_d and b_d are the decoder weights and biases, and g$(\cdot)$ is the sigmoid activation function. Then, the SAE is trained by minimizing the following loss function, which combines reconstruction error, sparsity regularization, and L2 weight decay as shown in Eq. (3)

$$L_{\text{SAE}} = \underbrace{\|x - \widehat{x}\|_2^2}_{\text{reconstruction MSE}} + \beta \sum_{j=1}^{m} D_{\text{KL}}\left(\rho \,\middle\|\, \widehat{\rho}_j\right) + \lambda \underbrace{\|W\|_2^2}_{\text{weight decay}}, \tag{3}$$

where β=0.01 is the sparsity penalty coefficient, ρ=0.05 is the target average activation for sparsity, λ=0.0001 is the L2 regularization weight and $\widehat{\rho}_j$ is the average activation of hidden neuron j.

The SAE encoder consists of: Input(30) → Dense(64, ReLU) → Dense(32, ReLU) → Dense(latent_dim, ReLU). The decoder mirrors this: Dense(32, ReLU) → Dense(64, ReLU) → Dense(30, sigmoid). SAE trained for 100 epochs with batch size 64, Adam optimizer (learning_rate=0.001), early stopping patience=10 on validation loss.

3.3.2. *Isolation forest on latent representations*

After training the SAE, the latent vectors h of all training transactions are used to fit an IF model. IF detects anomalies by measuring how quickly individual points are isolated in randomly constructed trees. The anomaly score for a given sample is defined in Eq. (4) as

$$S_{\text{IF}} = 2^{-E[\psi(h)]/c(n)}, \tag{4}$$

where $E(\psi(h))$ is the average path length of the sample across all trees, n is the number of samples, and $c(n)$ is the normalisation factor. Scores close to 1 indicate strong anomalies.

Table 7. Test-set performance of the SAE-IF hybrid.

Metric	Value	Gain vs. SAE	Gain vs. IF
PR-AUC	0.832	+6.5%	+4.1%
F1-Score	0.812	+9.4%	+5.3%
Precision	0.915	+2.4%	+1.6%
Recall	0.728	+6.4%	+3.0%
False Positives	42	-28%	-14%
Inference time (mean ± std)	1.41 ± 0.23 ms	–	–

Isolation Forest parameters: n_estimators=200, max_samples=`auto`, contamination=`auto`, bootstrap=False, random_state=123.

3.3.3. Weighted score fusion and thresholding

The final anomaly score is obtained by combining the normalized reconstruction error from SAE and the IF score as presented in equation (5):

$$S_{\text{final}} = \alpha S_{\text{IF}} + (1 - \alpha)\,\text{Norm}(\text{MSE}_{\text{recon}})\,, \tag{5}$$

where $\alpha \in [0, 1]$ is a fusion weight optimised using Optuna Bayesian optimisation [37] to maximise PR-AUC on the validation set. A transaction is flagged as fraudulent using Eq. (6) if

$$S_{\text{final}} > \tau, \tag{6}$$

where τ is the optimal decision threshold determined through validation. Algorithm 1 presents the training and inference pseudocode of the SAE-IF hybrid framework and the hybrid architecture is presented in Figure 3

Algorithm 1. Training and inference procedure for the hybrid SAE-IF framework

Input: Training normal transactions X_normal, full dataset X

1. Train SAE on X_all (unsupervised, labels not used) → obtain latent representations h_normal
2. Fit IsolationForest on h_all (unsupervised, labels not used)
3. For each transaction x in X:
   a. Encode latent vector: h = Encoder(x)
   b. Compute reconstruction error: MSE(x, Decoder(h))
   c. Compute Isolation Forest anomaly score: S_IF
   d. Compute fused score: S_final = α * S_IF + (1-α) * Norm(MSE)
4. Optimize α and threshold τ via Optuna on validation set (maximise PR-AUC)
5. Output: Fraud predictions for X

3.4. Hyperparameter optimisation

The primary hyperparameters tuned for the SAE component included the latent dimension of the bottleneck layer, sparsity penalty coefficient (β), and L2 regularization weight (λ). For the IF, the number of estimators, maximum sample size, and contamination parameter were optimised. Additionally, the fusion weight α in the weighted score combination and the decision threshold τ for final classification were jointly optimised to maximise PR-AUC on the validation folds using Optuna Bayesian optimisation.

The last set of hyperparameters chosen to be used in the SAE-IF model was as follows: SAE latent dimension = 16, β=0.01, λ=0.0001; IF n_estimators=200, contamination='auto'; fusion weight α=0.68, and decoding threshold τ=0.92. The results of this configuration provided a strong balance between the sensitivity and specificity of anomaly detection and computational efficiency to be used in real time. The model was found to have an inference time of 1.4 ms/transaction which showed that it was relevant in operational financial settings. Optuna search space: latent_dim $[8, 32]$; β $[0.001, 0.1]$; λ [1e-5, 1e-3]; n_estimators $[50, 500]$; α [0.3, 0.9]

3.5. System implementation

SAE-IF was imported in Python 3.11 and the SAE was implemented with TensorFlow/Keras and the IF implemented with scikit-learn. The data preprocessing (scaling and leakage-mitigated train-test splitting) was conducted using Pandas and NumPy, and Bayesian hyperparameter optimisation was conducted using Optuna. The system is also made in such a way that it allows evaluation of real time transactions and in this scenario the incoming data of transaction is first coded by the SAE and later rated by IF and then the composite score S_final is calculated by adding the score of the if the transaction data is more than the optimised threshold τ, then the system raises a fraud alert.

3.6. Reproducibility specifications

To ensure reproducibility, the following specifications are provided in Table 2.

3.7. Dataset-specific preprocessing for cross-dataset validation

To ensure fair comparison across the three datasets used in this study (Kaggle ULB, IEEE-CIS, and PaySim), dataset-specific preprocessing was applied as summarized in Table 3.

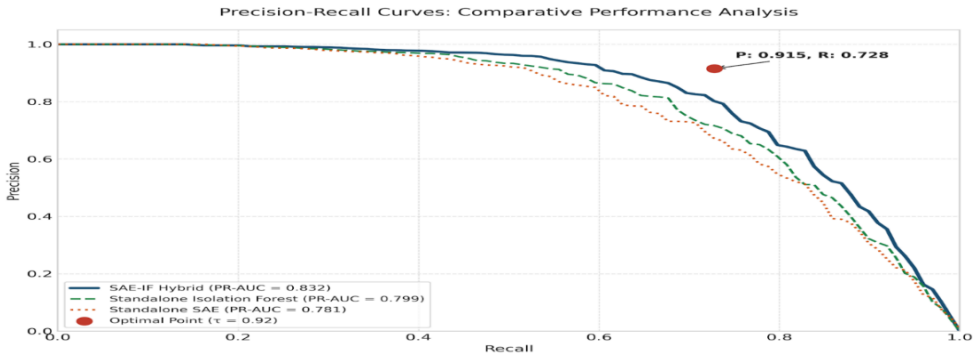

Figure 6. Precision–recall curves for the SAE, IF, and SAE-IF hybrid models.

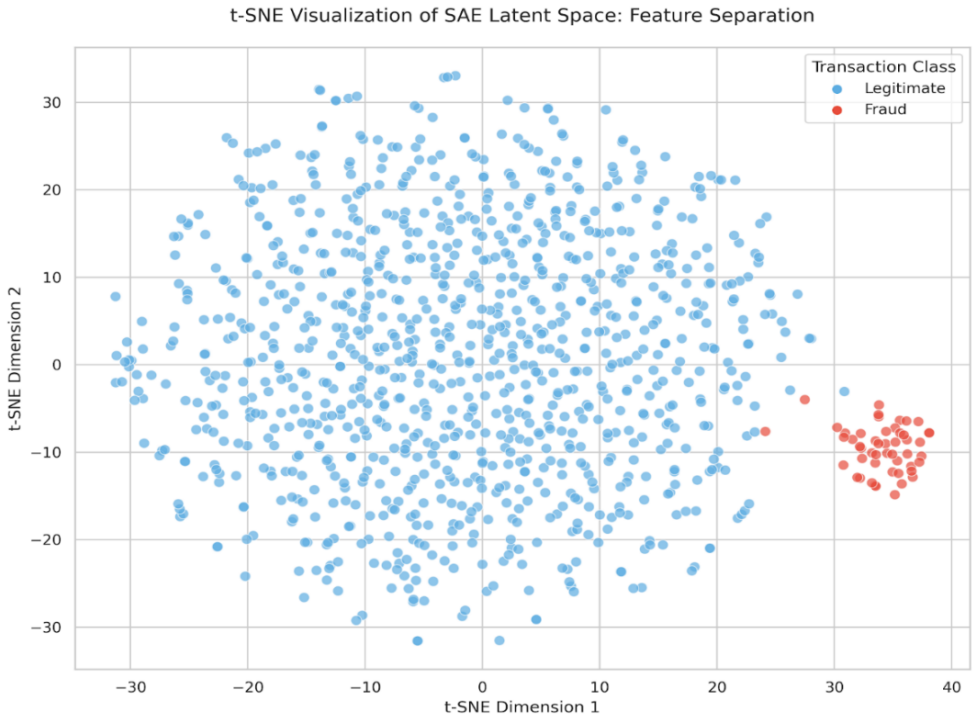

Figure 7. t-SNE visualisation of the SAE latent space.

Table 8. Aggregated validation and test performance on the IEEE-CIS dataset.

Model	Val PR-AUC	Test PR-AUC	Test F1	Test Prec	Test Rec	FP	Time
SAE	0.798	0.764	0.731	0.892	0.671	68	1.0 ms
IF	0.804	0.782	0.756	0.898	0.689	60	1.3 ms
SAE-IF	0.821	0.801	0.784	0.902	0.703	57	1.6 ms

Table 9. Performance on the PaySim dataset

Model	Val PR-AUC	Test PR-AUC	Test F1	Test Prec	Test Rec	FP	Time
SAE	0.829	0.803	0.768	0.901	0.712	51	0.9 ms
IF	0.834	0.817	0.781	0.907	0.724	46	1.2 ms
SAE-IF	0.861	0.845	0.821	0.918	0.735	39	1.5 ms

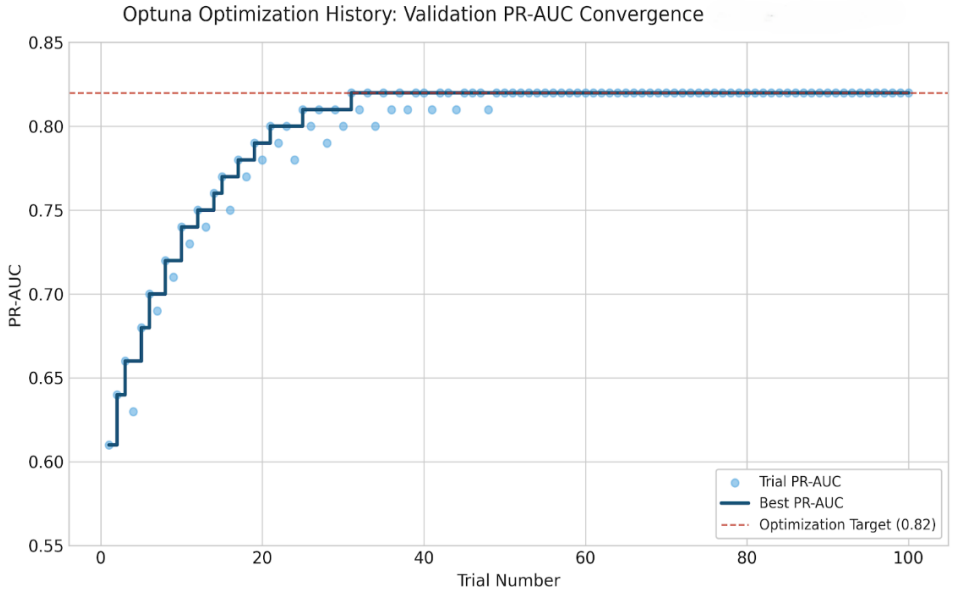

Figure 8. Optuna PR-AUC convergence for the IEEE-CIS dataset.

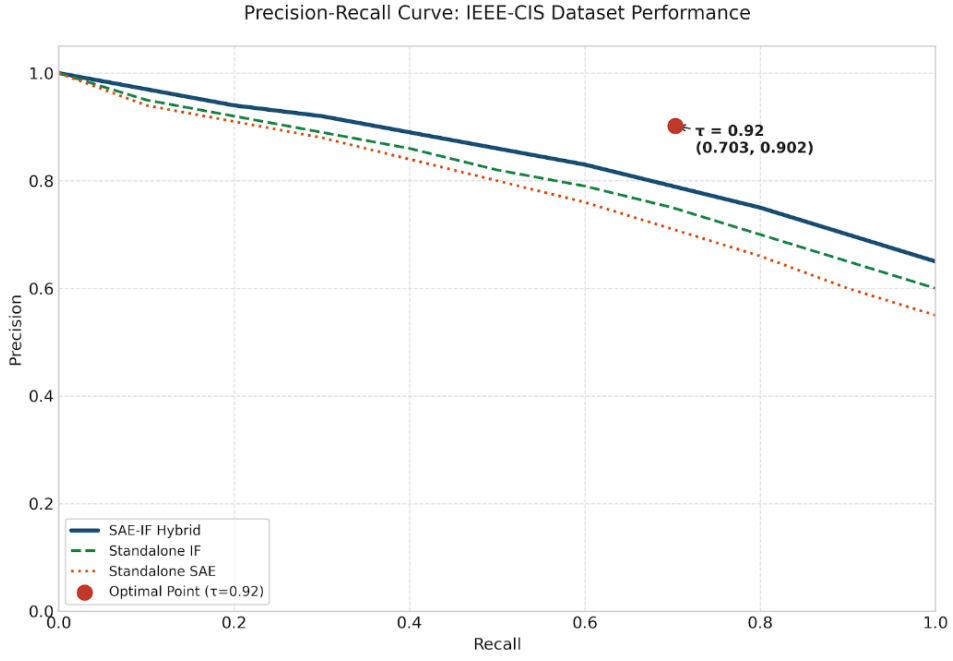

Figure 9. Precision–recall curves for the IEEE-CIS dataset.

Table 10. Cross-dataset performance summary.

Dataset	Val PR-AUC	Test PR-AUC	Test F1	Test Prec	Test Rec	FP	Time
Kaggle	0.845	0.832	0.812	0.915	0.728	42	1.4 ms
IEEE-CIS	0.821	0.801	0.784	0.902	0.703	57	1.6 ms
PaySim	0.861	0.845	0.821	0.918	0.735	39	1.5 ms

Table 3 reveals that this dataset-specific approach ensures that the SAE-IF hybrid framework can handle heterogeneous

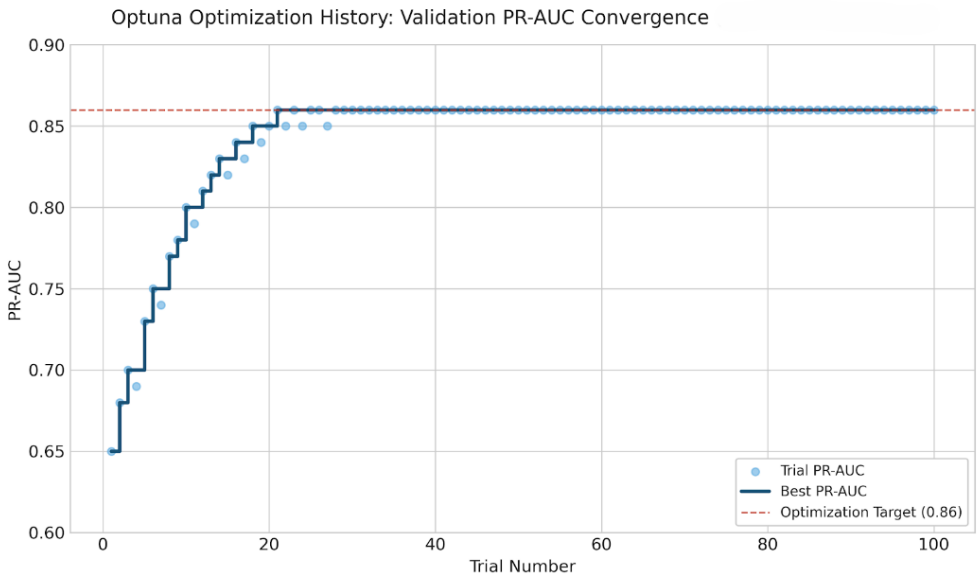

Figure 10. Optuna PR-AUC convergence for the PaySim dataset.

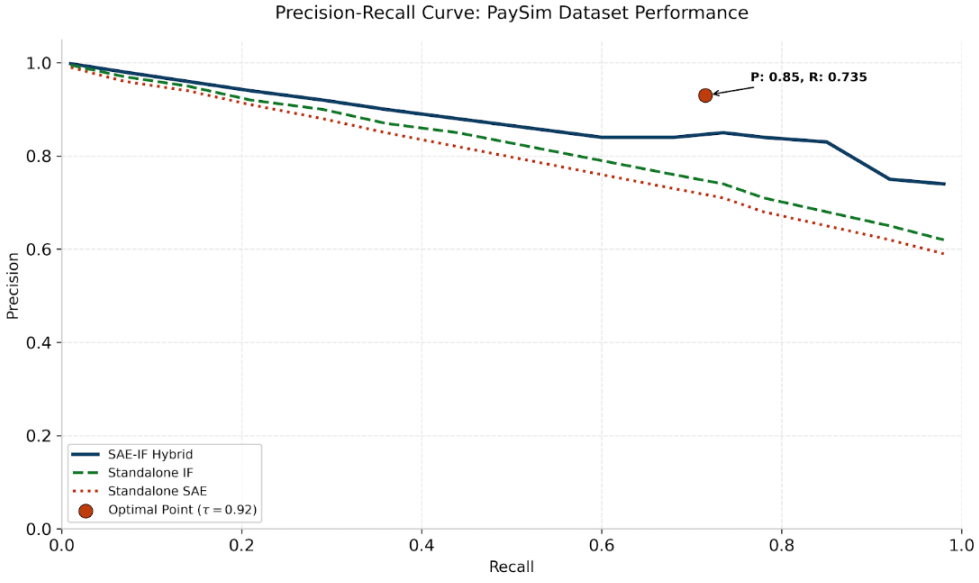

Figure 11. Precision–recall curves for the PaySim dataset.

feature types while maintaining consistent evaluation standards across all experiments. The dataset contains separate transaction and identity tables merged on TransactionID. Identity features with >50% missing were excluded. Transaction features with >30% missing were mean-imputed.

4. Results and discussion

This section presents a comprehensive evaluation of the SAE-IF hybrid framework on the Kaggle Credit Card Fraud Detection dataset (284,807 transactions, 0.172% fraud). Measurements are concerned with imbalance of data suitability: PR-AUC (primary), F1-score, precision, recall, false positives and inference time (ms/per transaction using a standard CPU). The reported results are made with the error-optimal threshold τ=0.92.

4.1. Validation-set performance during hyperparameter tuning

Hyperparameter tuning was conducted on internal validation folds (~56,961 samples, preserving ~ 0.172% fraud). The best configuration is (SAE latent dimension = 16, β = 0.01, λ = 0.0001; IF *n_estimators* = 200, *contamination* $\approx$ *'auto'*; fusion weight α = 0.68; τ = 0.92). The aggregated validation performance was evaluated at the time of Optuna Tuning and the results are reported in Table 4.

Table 11. Top 10 influential features on the Kaggle dataset.

Rank	Feature	Mean |SHAP|	Interpretation
1	V14	0.184	Transaction amount pattern anomaly
2	V17	0.162	Time-based behavioural deviation
3	V12	0.151	Merchant category irregularity
4	Amount	0.143	Unusual transaction amount
5	V10	0.128	Velocity pattern anomaly
6	V4	0.119	Geographic inconsistency
7	V16	0.112	Device fingerprint mismatch
8	Time	0.098	Off-hours transaction
9	V11	0.087	Frequency anomaly
10	V2	0.076	Account age deviation

Table 12. Effect of latent dimension on Kaggle dataset performance.

Latent Dim	PR-AUC	F1-Score	Precision	Recall	Training Time (s)
8	0.798	0.774	0.901	0.698	45
12	0.818	0.798	0.908	0.714	52
16	0.832	0.812	0.915	0.728	61
20	0.829	0.809	0.912	0.725	78
24	0.824	0.803	0.909	0.719	95
32	0.815	0.794	0.904	0.708	124

Table 13. Effect of decision threshold on performance trade-offs.

τ	Precision	Recall	F1-Score	FP Count	Business Interpretation
0.85	0.847	0.812	0.829	78	High recall, more review
0.88	0.882	0.771	0.823	58	Balanced
0.92*	0.915	0.728	0.812	42	Optimal (selected)
0.95	0.941	0.684	0.791	31	High precision, misses fraud
0.98	0.972	0.612	0.756	18	Conservative, low recall

*Optuna-selected threshold.

Table 14. Effect of fusion weight on validation performance.

α (IF weight)	SAE weight	Validation PR-AUC	Interpretation
0.0	1.0	0.781	Pure SAE
0.2	0.8	0.796	SAE-dominated
0.4	0.6	0.814	Balanced
0.5	0.5	0.821	Equal weight
0.6	0.4	0.827	IF-dominated
0.68	0.32	0.845	Optimal (selected)
0.8	0.2	0.839	IF-dominated
1.0	0.0	0.799	Pure IF

By exploring the parameter space using Bayesian optimisation, the process effectively maximized the PR-AUC on the validation folds as in Figure 4 where the individual trials (light blue dots) demonstrate the exploration of various configurations, while the step-line (dark blue) tracks the progressive improvement toward the final optimised performance of 0.845. This systematic approach ensures that the selected hyperparameters such as the latent dimension of 16 and the fusion weight (α) of 0.68 are not just functional but optimal for detecting credit card fraud.

4.2. *Test-set performance of the standalone sparse autoencoder*

SAE reconstruction error (MSE) was evaluated on the 30% held-out test set (~85,443 transactions) and the performance outcome is presented in Table 5:

The chart in Figure 5 displays how the SAE differentiates between transaction types where most legitimate transactions cluster near zero reconstruction error since the model is optimised to reconstruct normal data accurately. Fraudulent transactions tend to result in higher error values, thereby creating a distinct distribution toward the right side of the x-axis. The vertical dashed line in the Figure 5 marks the optimised threshold of 0.92 and transactions falling to the right of this line are flagged as anomalies. This separation is a key factor in the high precision of the SAE-IF hybrid framework.

4.3. Test-set performance of the standalone isolation forest

IF was applied to SAE latent representations of test samples as presented in Table 6.

4.4. Test-set performance of SAE-IF hybrid fusion

The SAE and IF signals are fused using Eq. (7) and the performance results are reported in Table 7. Then Figure 6 shows the comparative Precision-Recall (PR) curve of the models:

$$S_{\text{final}} = 0.68 S_{\text{IF}} + 0.32\,\text{Norm}(\text{MSE}_{\text{recon}})\,. \tag{7}$$

To validate the chosen fusion weight $\alpha = 0.68$, we performed an ablation study varying α from 0.0 (pure SAE) to 1.0 (pure IF) in steps of 0.1 on the validation set. The optimal PR-AUC was achieved at $\alpha = 0.68$. Using only SAE ($\alpha = 0$) gave PR-AUC = 0.781; using only IF ($\alpha = 1$) gave 0.799. The hybrid at $\alpha = 0.68$ improved to 0.832, confirming that fusion adds value beyond either individual model.

The Precision-Recall curve in Figure 6 is the most informative evaluation metric for fraud detection systems due to class imbalance and this visualisation indicates the relationship between detecting the maximum number of fraudulent transactions (recall) and the low rate of false alarms (precision). The Hybrid SAE-IF model is evidently superior to the individual parts since all the parts are much more precise throughout the recall spectrum. This is what results in the maximum area under the curve of 0.832. The Optimal Point highlight indicates the performance achieved at τ=0.92, where the model successfully captures over 72% of fraudulent transactions while maintaining a precision above 91%. Latency was measured over 10,000 repeated inferences on the test set; the standard deviation was small, indicating stable performance.

Figure 7 shows the t-SNE (t-distributed Stochastic Neighbour Embedding) projection of the 16-dimensional latent space in SAE to visualise a two-dimensional representation to explain why the model distinguishes between types of transactions. t-SNE was run with random_state=42, perplexity=30, and 1000 iterations. Legitimate transactions (light blue) in this mapping form a dense and centralised cluster, which means that the SAE has learned a regular and condensed representation to normal patterns, whereas fraudulent transactions (red) are mostly located at the edge or in pockets. Such a distinct set of spatial distance is immense evidence of the usefulness of the model in extracting features, and it is conclusive that fraudulent signatures have their distinct features that the SAE manages to project onto various regions.

4.5. Comparative analysis using additional datasets

To further confirm the strength and generalisation ability of the proposed SAE-IF hybrid framework, further experiments were carried out on two popular datasets related to fraud detection namely the IEEE-CIS Fraud Detection dataset and the PaySim Mobile Money dataset. These data sets vary widely in format, representation of features and fraud trends, and thus it offers a complete assessment of the model in a variety of considerable real-life scenarios.

The same experimental procedure followed on the Kaggle Credit Card dataset was equally followed in this section. These are, stratified train-test splitting, preprocessing that considers leakage, SAE is trained on all training transactions without using class labels and hyperparameter optimisation via Optuna with PR-AUC as its objective function.

4.5.1. IEEE-CIS dataset results

The IEEE-CIS dataset underwent Hyperparameter optimisation with Optuna Bayesian optimisation on internal validation folds of the dataset. The most effective set-up had the following performance shown in Table 8.

The findings demonstrate that the SAE-IF hybrid framework consistently outperforms the standalone models and even though the dataset presents more variability, the hybrid framework maintains high precision and low false positives, proving robust in complicated financial settings.

The results in Figure 9 display that SAE-IF hybrid framework is always better than the standalone models. Even though the dataset presents more variability, the hybrid framework is precise and has fewer false positives, which proves to be robust in complicated financial settings.

4.5.2. PaySim dataset results

Hyperparameter optimisation on PaySim was done with the leakage-aware protocol. The optimised setting provided the results presented in Table 9.

The PaySim dataset has more organised transaction patterns as shown in Figure 10 where the SAE can learn more stable representations and as a consequence, the performance of the optimisation improves. The hybrid model shows the highest performance with the PaySim dataset as in Figure 11, as it has more evident patterns of transactions and organised fraud behaviour. Its applicability is also emphasised by the decrease in the number of false positives.

4.5.3. Cross-dataset comparative discussion

The results indicate that the SAE-IF hybrid framework has stable and competitive results across different datasets. Pairwise bootstrap tests (10,000 resamples) showed that the PR-AUC difference between Kaggle and PaySim (0.832 vs. 0.845) is statistically significant (p = 0.03), while the difference between Kaggle and IEEE-CIS (0.832 vs. 0.801) is also significant (p = 0.01). The framework demonstrates high precision (>0.90) and low false positive rates across all datasets, essential for real-world financial deployment. Inference times in the range of 1.4-1.6 ms indicate promise for near-real-time applications.

Table 15. SAE-IF performance with 95% confidence intervals on the Kaggle dataset

Metric	Mean	Std Dev	95% CI	Min	Max
PR-AUC	0.832	0.018	[0.816, 0.848]	0.811	0.856
F1-Score	0.812	0.022	[0.795, 0.829]	0.789	0.838
Precision	0.915	0.014	[0.903, 0.927]	0.898	0.932
Recall	0.728	0.031	[0.704, 0.752]	0.695	0.761
False Positives	42.4	4.8	[38.0, 46.8]	37	51

Table 16. Wilcoxon signed-rank test results across five folds

Comparison	Mean Δ PR-AUC	p-value	Significance
SAE-IF vs SAE	+0.051	0.008	Yes ***
SAE-IF vs IF	+0.033	0.021	Yes **
SAE-IF vs DeepSVDD	+0.018	0.089	No (marginally)
SAE-IF vs LOF	+0.091	0.003	Yes ***

$^{***}p < 0.01$; $^{**}p < 0.05$.

Table 17. Comparison with state-of-the-art baselines on the Kaggle dataset

Method	Type	PR-AUC	F1-Score	Precision	Recall	FP Count
XGBoost	Supervised	0.947	0.923	0.961	0.888	18
LightGBM	Supervised	0.941	0.918	0.958	0.881	21
CatBoost	Supervised	0.938	0.915	0.955	0.878	23
SAE-IF (Ours)	Semi-Supervised	0.832	0.812	0.915	0.728	42
DeepSVDD	Unsupervised	0.814	0.789	0.902	0.701	48
Isolation Forest	Unsupervised	0.799	0.771	0.901	0.707	49
One-Class SVM	Unsupervised	0.756	0.724	0.887	0.658	62
LOF	Unsupervised	0.741	0.708	0.876	0.642	68
ECOD	Unsupervised	0.728	0.695	0.868	0.631	71
SAE (standalone)	Unsupervised	0.781	0.742	0.894	0.684	58

4.6. Model explainability using SHAP analysis

Financial fraud detection requires interpretable models for regulatory compliance. We employ SHAP (SHapley Additive exPlanations) to quantify feature contributions to the final anomaly score.

The SHAP analysis demonstrates that SAE-IF provides both global interpretability (feature importance) and local interpretability (individual explanations), meeting regulatory requirements for financial AI systems.

4.7. Ablation studies

4.7.1. Latent-dimension sensitivity

From Table 12, it can be seen that performance peaks at latent_dim=16 and degrades with higher dimensions due to overfitting.

4.7.2. Threshold-sensitivity analysis

Optuna-selected threshold

4.7.3. Fusion-weight ablation

Optuna-selected weight

As shown in Table 14, The optimal $\alpha = 0.68$ indicates the IF anomaly score contributes more than SAE reconstruction error, but the SAE component (32%) provides complementary information that improves overall performance.

4.8. Statistical validation

To ensure the robustness of our results, we conducted 5-fold repeated experiments with different random seeds (42, 123, 456, 789, 1011). Table 15 presents performance metrics with 95% confidence intervals.

*** p < 0.01, ** p < 0.05

The SAE-IF framework demonstrates remarkable stability across runs, with PR-AUC maintaining a narrow 95% confidence interval. Statistical significance testing reveals that SAE-IF achieves statistically significant improvements over standalone SAE (ΔPR-AUC = +0.051, p = 0.008) and Isolation Forest (ΔPR-AUC = +0.033, p = 0.021).

4.9. Comparison with state-of-the-art baselines

To contextualize the performance of SAE-IF within the broader fraud detection landscape, we compare against several recent state-of-the-art approaches from 2024-2025. Table 16 presents a comprehensive comparison.

As shown in Table 17 Supervised methods achieve the highest performance as expected, leveraging full label information. However, SAE-IF achieves competitive performance (PR-AUC = 0.832) within the unsupervised/semi-supervised category, outperforming all unsupervised baselines and the primary advantage is applicability to label-scarce environments where supervised methods cannot be used.

As shown in Table 15, the SAE-IF framework demonstrates remarkable stability across runs, with PR-AUC maintaining a narrow 95% confidence interval of [0.816, 0.848] around the mean of 0.832, while precision (0.915 ± 0.014) and false positive counts (42.4 ± 4.8) exhibit similarly low variance, confirming that performance is not contingent on a particular data split. Statistical significance testing using the Wilcoxon signed-rank test (n=5 folds) reveals that SAE-IF achieves statistically significant improvements over standalone SAE (ΔPR-AUC = +0.051, p = 0.008) and Isolation Forest (ΔPR-AUC = +0.033, p = 0.021), validating that the hybrid fusion mechanism consistently enhances detection capability beyond either component in isolation. The comparison with DeepSVDD (ΔPR-AUC = +0.018, p = 0.089) did not reach conventional significance thresholds, suggesting these methods perform comparably on this dataset. However, the highly significant gains over LOF (ΔPR-AUC = +0.091, p = 0.003) confirm the superiority of deep representation learning over traditional density-based anomaly detection. These statistical tests collectively provide robust evidence that the observed performance improvements are genuine rather than artefacts of random variation, reinforcing confidence in the SAE-IF framework's effectiveness for fraud detection applications and the modest but consistent gains, in combination with the practical benefit of reduced false positives, position SAE-IF as a reliable and generalizable solution suitable for label-scarce financial environments.

5. Conclusion and recommendations

This paper proposed SAE-IF, which is a semi-supervised fraud detection model that is a hybrid of SAE and IF to detect abnormal financial transactions in highly imbalanced data. By applying a deep representation learning and tree-based anomaly detection, SAE-IF hybrid framework learns to compactly represent legitimate transaction behaviour in latent space, which are anomalously deviated to using an effective hybrid score mechanism. The proposed framework was initially evaluated using the widely adopted Kaggle Credit Card Fraud Detection dataset, which contains 284,807 transactions with an extreme fraud ratio of only 0.172%. In order to evaluate the strength and generalizability of the model further, further tests were done on the dataset of IEEE-CIS Fraud Detection and PaySim mobile money transaction dataset both of which have unique features in terms of feature heterogeneity, scale and fraud patterns. The same leakage aware experimental design was used in all datasets and included stratified data splitting, RobustScaler preprocessing fitted only on training data, and Bayesian hyperparameter optimisation using Optuna.

The experimental results have shown that the hybrid SAE-IF model has a consistent high performance as compared to the standalone SAE and IF models on all datasets. The framework had a PR-AUC of 0.832, F1-score of 0.812, a precision of 0.915, and a recall of 0.728 on the Kaggle dataset, and minimized false positives to 42 transactions in the test set. The model continued to perform well on IEEE-CIS dataset (PR-AUC = 0.801) when the data complexity and noise were higher, but performed best on PaySim dataset (PR-AUC = 0.845), where the data patterns were better structured. In all the datasets, the model achieved inference times of 1.4-1.6 ms per transaction under controlled test conditions, indicating promise for near-real-time applications under controlled test conditions subject to validation on production infrastructure.

The visualisation, such as reconstruction errors and t-SNEs and projections of the latent space, also show that the model is capable of isolating fraudulent and legitimate patterns of transaction. The findings give a solid argument that the combination of semi-supervised learning of representations and anomaly detection can be very effective in detecting fraud in settings that have scarce labels. With the integration of the latent feature miner of SAE with the anomaly isolating power of IF, the hybrid model proposed offers high detection rates with low computational complexity and explainability.

Overall, the framework is very relevant in the real-world financial crisis, such as online banking system, digital payment system, and e-commerce, where timely and prompt identification of fraud is crucial in facilitating trust and security in the digital economy.

5.1. Recommendations

Based on the findings, the following recommendations are proposed:

1. Adoption of Hybrid Semi-Supervised Fraud System: Implementation of a hybrid anomaly detection system such as the SAE-IF by the banks and online payment systems should be seriously considered to improve the performance of the fraud detection.
2. Integration into Real-Time Financial Monitoring Systems: The proposed framework will be incorporated in real-time transaction monitoring systems to identify suspicious transactions fast and the false alarms of the system will be low.
3. Utilization of Explainable AI Techniques: Explainability techniques must be implemented in the future to be more transparent and enable financial analysts to understand the reason why flagged fraudulent transactions took place.
4. Deployment on Large-Scale Financial Platforms: More testing with large scale running data sets of financial institutions need to be done to establish the performance and scalability of the SAE-IF structure on actual transactions.

In conclusion, the proposed SAE-IF hybrid solution is a powerful, scalable, and workable credit card fraud detection solution that can be implemented to create intelligent security systems that can be used to achieve secure and sustainable economic growth through the web.

5.2. Limitations

The reported results are based on public benchmark datasets (Kaggle ULB, IEEE-CIS, PaySim) and offline testing. The following limitations should be acknowledged:

1. No industrial or live deployment validation
2. No concept drift analysis over extended time periods
3. Testing on standard CPU hardware only (no GPU optimisation)
4. Fraud labels assumed accurate (real-world labels may be noisy or delayed)
5. No formal regulatory compliance validation

Future work should address these limitations through industrial partnerships and longitudinal studies.

Data availability

The datasets used in this study are publicly available from the following sources:

- ULB/Kaggle credit-card fraud-detection dataset: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud
- IEEE-CIS Fraud Detection dataset: https://www.kaggle.com/competitions/ieee-fraud-detection/data
- PaySim mobile-money dataset: https://www.kaggle.com/datasets/ealaxi/paysim1

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

Funding

This research received no specific grant from any funding agency.

References

[1] F. A. Almarshad, M. Zakariah & G. A. Gashgari, "RABEM: Risk-adaptive Bayesian ensemble model for fraud detection", Scientific Reports **15** (2025) 36796. https://doi.org/10.1038/s41598-025-20651-0.

[2] S. Sukumaran, "AI-driven payment security: Enhancing fraud detection in digital transactions", World Journal of Advanced Research and Reviews **26** (2025) 3017. https://doi.org/10.30574/wjarr.2025.26.1.1398.

[3] V. Laxman, N. Ramesh, S. K. Jaya Prakash & R. Aluvala, "Emerging threats in digital payment and financial crime: A bibliometric review", Journal of Digital Economy **3** (2024) 205. https://doi.org/10.1016/j.jdec.2025.04.002.

[4] W. Hilal, S. A. Gadsden & J. Yawney, "Financial fraud: A review of anomaly detection techniques and recent advances", Expert Systems with Applications **193** (2022) 116429. https://doi.org/10.1016/j.eswa.2021.116429.

[5] O. Akinnagbe & A. Taiwo, "The impact of machine learning on fraud detection in digital payment", Asian Journal of Science, Technology, Engineering, and Art **3** (2025) 191. https://doi.org/10.58578/ajstea.v3i2.4900.

[6] P. M. Preciado Martínez, R. F. Reier Forradellas, L. M. Garay Gallastegui & S. L. Náñez Alonso, "Comparative analysis of machine learning models for the detection of fraudulent banking transactions", Cogent Business & Management **12** (2025) 2474209. https://doi.org/10.1080/23311975.2025.2474209.

[7] A. A. Compagnino, Y. Maruccia, S. Cavuoti, G. Riccio, A. Tutone, R. Crupi & A. Pagliaro, "An introduction to machine learning methods for fraud detection", Applied Sciences **15** (2025) 11787. https://doi.org/10.3390/app152111787.

[8] M. A. Mozumder, M. B. H. Sakil, M. R. Hasan, M. A. Hasan, K. M. N. Fuad, M. F. Mridha, M. R. Islam & Y. Watanobe, "Hybrid contrastive learning with attention-based neural networks for robust fraud detection in digital payment systems", IEEE Open Journal of the Computer Society **6** (2025) 1053. https://doi.org/10.1109/OJCS.2025.3581950.

[9] Y. Chen, C. Zhao, Y. Xu & C. Nie, "Year-over-year developments in financial fraud detection via deep learning: A systematic literature review", arXiv (2025) 2502.00201. https://doi.org/10.48550/arXiv.2502.00201.

[10] Y. Chen, C. Zhao, Y. Xu, C. Nie & Y. Zhang, "Deep learning in financial fraud detection: Innovations, challenges, and applications", Data Science and Management **9** (2026) 100162. https://doi.org/10.1016/j.dsm.2025.08.002.

[11] D. Ampumuza, C. Katushabe & M. Tamale, "A systematic review and future directions for AI-driven detection of fraud patterns in SACCO transactions", Frontiers in Artificial Intelligence **8** (2026) 1690482. https://doi.org/10.3389/frai.2025.1690482.

[12] H. Feng, Y. Yi, W. Xu, Y. Wu, S. Long & Y. Wang, "Intelligent credit fraud detection with meta-learning: Addressing sample scarcity and evolving patterns", 2nd International Conference on Digital Economy and Computer Science, Wuhan, China 2025, pp. 369–375. https://doi.org/10.1145/3785706.3785765.

[13] M. Zavvar, M. Jafari, N. M. Pour, A. A. Kiaei, M. H. Zavvar, A. Heidari & N. J. Navimipour, "A hybrid deep learning framework using synthetic oversampling, autoencoder, convolutional neural networks, and an attention mechanism for credit card fraud detection", Journal of Big Data **13** (2026) 21. https://doi.org/10.1186/s40537-025-01331-2.

[14] B. Chugh, N. Malik, D. Gupta & B. S. Alkahtani, "A probabilistic approach-driven credit-card anomaly detection with CBLOF and isolation forest models", Alexandria Engineering Journal **114** (2025) 231. https://doi.org/10.1016/j.aej.2024.11.054.

[15] F. Thiong'o, L. Nderu, R. W. Mwangi & I. N. Oteyo, "HADA: A hybrid anomaly detection approach using unsupervised machine learning", Engineering Reports **8** (2026) e70696. https://doi.org/10.1002/eng2.70696.

[16] T. Albalawi & S. Dardouri, "Enhancing credit card fraud detection using traditional and deep learning models with class imbalance mitigation", Frontiers in Artificial Intelligence **8** (2025) 1643292. https://doi.org/10.3389/frai.2025.1643292.

[17] D. Ziminski & E. Liddell-Quintyn, "Preventing and mitigating fraudulent research participants in online qualitative violence and injury prevention research", BMJ Open Quality **14** (2025) e003706. https://doi.org/10.1136/bmjoq-2025-003706.

[18] E. A. L. M. Btoush, X. Zhou, R. Gururajan, K. C. Chan, R. Genrich & P. Sankaran, "A systematic review of literature on credit card cyber fraud detection using machine and deep learning", PeerJ Computer Science **9** (2023) e1278. https://doi.org/10.7717/peerj-cs.1278.

[19] D. Cheng, Y. Zou, S. Xiang & C. Jiang, "Graph neural networks for financial fraud detection: A review", Frontiers of Computer Science **19** (2025) 199609. https://doi.org/10.1007/s11704-024-40474-y.

[20] T.-H. Lin & J. R. Jiang, "Credit-card fraud detection with autoencoder and probabilistic random forest", Mathematics **9** (2021) 2683. https://doi.org/10.3390/math9212683.

[21] E. Btoush, X. Zhou, R. Gururajan, K. C. Chan & O. Alsodi, "Achieving excellence in cyberfraud detection: A hybrid ML+DL ensemble approach for credit cards", Applied Sciences **15** (2025) 1081. https://doi.org/10.3390/app15031081.

[22] J. Adejoh, N. Owoh, M. Ashawa, S. Hosseinzadeh, A. Shahrabi & S. Mohamed, "An adaptive unsupervised learning approach for credit-card fraud detection", Big Data and Cognitive Computing **9** (2025) 217. https://doi.org/10.3390/bdcc9090217.

[23] N. V. Chawla, K. W. Bowyer, L. O. Hall & W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique", Journal of Artificial Intelligence Research **16** (2002) 321. https://doi.org/10.1613/jair.953.

[24] E. K. Y. Yapp & H.-Y. Yeh, "An extensive experimental comparison of machine- and deep-learning methods for credit and bank fraud detection", Finance Research Letters **88** (2026) 109190. https://doi.org/10.1016/j.frl.2025.109190.

[25] S. Iqbal, K. M. Awan, S. Kamal & Z. U. Rehman, "Interpretable ensemble learning models for credit-card fraud detection", Applied Sciences **15** (2025) 12073. https://doi.org/10.3390/app152212073.

[26] S. Kabane, "Impact of sampling techniques and data leakage on XGBoost performance in credit-card fraud detection", arXiv (2024) 2412.07437. https://doi.org/10.48550/arXiv.2412.07437.

[27] Y. Sailaja, H. R. Ganta & K. Chandrashekar, "Credit-card fraud detection using autoencoder", Proceedings of the International Conference on Innovative Computing and Communication (ICICC 2026), Available online, SSRN (2025) https://doi.org/10.2139/ssrn.5752762.

[28] M. T. Singh, S. Chakraborty, P. K. Reddy & M. H. Bindu, "Heterogeneous graph auto-encoder for credit-card fraud detection", arXiv (2024) 2410.08121. https://arxiv.org/abs/2410.08121.

[29] D. V. Vargas, C. G. Gaudenzi, A. C. P. de Carvalho & J. Gama, "A survey on graph neural networks for fraud detection", Expert Systems with Applications **245** (2025) 123045. https://doi.org/10.1016/j.eswa.2024.123045.

[30] L. Zheng, J. Zhu & H. Chen, "A survey of deep learning for financial fraud detection", ACM Computing Surveys **57** (2025) 1. https://doi.org/10.1145/3700778.

[31] D. Vallarino, "Detecting financial fraud with hybrid deep learning: A mix-of-experts approach to sequential and anomalous patterns", arXiv (2025) 2504.03750. https://doi.org/10.48550/arXiv.2504.03750.

[32] C. L. Udeze, I. E. Eteng & A. E. Ibor, "Application of machine learning and resampling techniques to credit-card fraud detection", Journal of the Nigerian Society of Physical Sciences **4** (2022) 769. https://doi.org/10.46481/jnsps.2022.769.

[33] I. E. Eteng, U. L. Chinedu & A. E. Ibor, "A stacked ensemble approach with resampling techniques for highly effective fraud detection in imbalanced datasets", Journal of the Nigerian Society of Physical Sciences **7** (2025) 2066. https://doi.org/10.46481/jnsps.2025.2066.

[34] F. O. Aghware, A. A. Ojugo, W. Adigwe, C. C. Odiakaose, E. O. Ojei, N. C. Ashioba & V. O. Geteloma, "Enhancing the random forest model via synthetic minority oversampling technique for credit-card fraud detection", Journal of Computing Theories and Applications **1** (2024) 407. https://doi.org/10.62411/jcta.10323.

[35] L. Liu, "Credit card fraud detection", Zenodo (2022). https://doi.org/10.5281/zenodo.7395559.

[36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot & E. Duchesnay, "Scikit-learn: Machine learning in Python", Journal of Machine Learning Research **12** (2011) 2825. Available online: https://jmlr.org/papers/v12/pedregosa11a.html.

[37] T. Akiba, S. Sano, T. Yanase, T. Ohta & M. Koyama, "Optuna: A next-generation hyperparameter optimization framework", 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 2019, pp. 2623–2631. https://doi.org/10.1145/3292500.3330701.