Journal of Soft Computing Paradigm (ISSN: 2582-2640)
www.irojournals.com/jscp/

HDNE-NR a Hybrid Deep Neural–Ensemble Framework for Robust Fraud Detection Under Controlled Noise Perturbations

Sumathi K.[1], Priyanga K.[2], Ramesh S.[3]

[1]Associate Professor, Department of Information Technology, Kangeyam Institute of Technology (Autonomous), Kangayam, Tirupur, India.

[2,3]Assistant Professor, Department of Computer Science and Engineering, Kangeyam Institute of Technology (Autonomous), Kangayam, Tirupur, India.

E-mail: [1]hod.it@kitech.edu.in, [2]priyanga.cse@kitech.edu.in, [3]ramesh.cse@kitech.edu.in

Abstract

This paper presents HDNE-NR (Hybrid Deep Neural Ensemble with Noise Robustness) in order to obtain constant fraud detection when subjected to controlled feature and label perturbations. The major purpose was to create a robustness-based predictive framework that can retain discrimination in imbalanced and noisy financial data with great intensity. HDNE-NR provides high-level performance when compared to the deep latent representation learning models; this is due to the combination of heterogeneous ensemble meta-classifiers and adaptive weighted stacking. In clean conditions, the framework had an F1-score of 0.933 and ROC-AUC of 0.992. At high noise of features ($\sigma = 0.20$) and corruption of labels ($\epsilon = 0.10$), HDNE-NR maintained higher F1-scores (0.820 and 0.815, respectively) and lower decay curves. It is novel in the synergistic combination of any of the noise-aware learning of representations and adaptive weighting of the ensemble. Generally, HDNE-NR offers an uncertain and scalable framework in fraud analytics in the real world.

Keywords: Noise Robustness, Ensemble Learning, Fraud Detection, Deep Neural Networks, Weighted Stacking, Imbalanced Classification.

DOI: https://doi.org/10.36548/jscp.2026.2.005
Received: 26.03.2026, received in revised form: 30.04.2026, accepted: 20.05.2026, published: 29.05.2026

1. Introduction

In finance, retail, climate modeling, cybersecurity, and scientific computing, predictive analytics has become an essential part of the modern decision-making system. Ensemble learning techniques, where two or more predictive models are used to enhance generalization, have been shown to be more stable than single techniques in a wide range of fields including retail demand forecasting [1], time-series modeling [2], and environmental prediction [3]. The combination of deep learning and ensemble method has also improved the modeling capacity to represent nonlinear and high-dimensional feature interactions and enhance better robustness by diversity of models [4], [5].

The latest developments focus on hybrid systems of predictive architecture which combine statistical learning, deep neural networks, and stacking-based aggregation, [6], [9]. This type of framework takes advantage of the complementary paradigms of learning to minimize variance and enhance the ability to discriminate. Hybrid ensemble models have demonstrated better resistance to multifaceted and malicious data distributions in cybersecurity and network intrusion detection [7], [8]. Likewise, hybrid ensemble approaches have also proven useful in scientific prediction, such as property prediction in photophysics, among other regression-based tasks [10]. In spite of these developments, there remains a major weakness in terms of scarce research on robustness in the case of controlled feature and label noise.

The real-world predictive systems often face the problem of uncertainty due to errors on sensors, transmission errors, or error in annotation. Even though hybrid deep ensemble models have good results in clean environments [4], [5], systematic robustness testing under perturbation is not well studied. Existing literature is characterized by the preference of accuracy enhancement without a direct attempt to model noise injection or measure degradation trends at different corruption levels. As a result, we have a gap in the knowledge about the behavior of hybrid ensemble architectures in the face of uncertainty in highly imbalanced data, namely in financial fraud detection.

The problem of fraud detection has a unique and difficult environment with a high level of imbalance in classes, non-dense fraud behavior, and dynamic adversarial behavior. In these circumstances, predictive stability is of equal importance as accuracy. Minority class detection performance can be inaccurate due to noise in either feature space or label annotations which poses a financial risk. Hence, to be able to deploy it effectively, it is necessary to design a hybrid

predictive framework that integrates explicitly controlled noise modeling with adaptive ensemble weighting.

Inspired by these issues, the research paper is based on a Hybrid Deep Latent Representation Learning Hybrid Deep Neural Ensemble with Noise Robustness (HDNE-NR) model that combines learning deep latent representations, heterogeneous meta-classifiers, and adaptive weighted stacking. In stark contrast to the earlier hybrid models, which mostly concentrate on the improvement of performance [4] -[6], the presented framework considers stability directly with the progressive noising of Gaussian features and symmetric noising of labels. This work will offer a systematized method of trustworthy predictive analytics in the presence of a large noise and high imbalance by integrating robustness analysis in the architectural design.

The proposed work is an extension of the existing literature on hybrid ensemble learning to the area of robustness-conscious modeling to cover the gap between high-precision prediction and confident real-world application under uncertainty [1]-[10]. The key contributions of this paper are listed below:

- A deep neural network (DNN) structure in a hybrid model is introduced for implementation of strong financial fraud detection under uncertain conditions such as noisy data.
- The process flow contains a noise-sensitive preprocessing approach that improves the quality of data by reducing the effect of stochastic noise.
- This architecture proposes a combination of various deep learning elements to be used to learn both linear and non-linear correlations between financial transactions data.
- The framework is evaluated with an uncertainty of real-world data to achieve this by use of a Gaussian noise modeling approach.
- An extensive experimental analysis is carried out on benchmark financial data.

In addition, the model is proven to be effective by the comprehensive performance measures, which are preciseness, recall, F1-score, and false positive rate, which makes the model applicable to the real-life application of fraud detection where reducing false alarms is essential.

2. Related Work

The classification of financial frauds is a challenging problem that has been tackled by a large number of researchers in recent years, focusing on hybrid and ensemble learning as a response to the fact that individual classifiers are difficult to deal with class imbalance and nonlinear transaction patterns and fraud behaviors. To integrate several classifiers for credit card fraud detection, ensemble fusion has been introduced, and it was demonstrated that ensemble voting can enhance the stability of the decisions made by the classifiers over individual decision making [11]. Also, hybrid of artificial and quantum intelligence based fraud detection has been investigated to enhance pattern learning on complex financial data sets, showcasing the integration of traditional machine learning with quantum intelligence in fraud detection [12].

The stacking models have also been considered as they provide the opportunity to combine complementary strengths of multiple base models. A strong feature selection and stacking ensemble method was proposed to boost the credit card fraud detection by eliminating irrelevant features and to increase the reliability of classification [13]. A third hybrid approach that highlights imbalance, concept drift and adversarial attacks has been adopted, which states that FDS must not only be accurate, but stable as well in dynamic financial environments and are under hostile attacks [14]. This trend has been continued with real time analytics frameworks that add sequential and market behaviour learning to the mix, particularly since real time transaction streaming environments often present evolving financial crime patterns [15]. There are also multiple studies that applied hybrid machine learning and deep learning approach in corporate and enterprise fraud detection systems. The works demonstrate the benefits of integrating structured data modeling, deep feature learning, and intelligent classification in real-world financial management scenarios for fraud prediction [16, 19]. On the other hand, quantum graph neural network and hybrid optimization-based deep learning were examined to capture complex relational and high dimensional financial transactions pattern [17, 22]. The studies reveal that a key focus in fraud analytics is now advanced representation learning.

In addition, hybrid deep neural ensemble models have been reported in the field of intelligent financial fraud detection to enhance the prediction accuracy and robustness by using deep learning with ensemble decision mechanisms [18]. To enhance financial transaction fraud detection, multiple learners have been combined together into a final strong learner, called a

stacking classifier [20]. In other works, on machine learning for transaction fraud detection, it is found that handling the imbalance and feature learning is crucial for fraud detection, which also demands a correct classification under the uncertain condition of transactions [21]. Recent explainable hybrid metaheuristic–deep learning methods include temporal convolutional analysis for real-time fraud detection and enhanced interpretability and optimization [23].

While these studies show strong advancement in hybrid, ensemble, quantum, optimization and real-time fraud detection models, most of these works emphasize on enhancing the classification accuracy. Fewer works systematically examine robustness when the features are corrupted and the labels are corrupted. It provides an obvious research gap of a noise-aware fraud detection framework based on deep latent representation learning, heterogeneous ensemble classifier.

3. Proposed Work

When data are highly imbalanced, there are two significant problems with the use of financial fraud detectors: (i) the instability of deep models faced by noisy perturbations (ii) the poor generalization of single ensemble classifiers in the case of feature corruption. In order to alleviate these limitations, this paper introduces a hybrid deep neural-ensemble architecture, in which the representation learning is incorporated with adaptive ensemble stacking to obtain robustness against controlled injection of noise.

3.1 Overall Architecture

The proposed framework combines three core components as shown in Figure 1.

1. Deep neural feature extractor
2. Noise-aware regularization mechanism
3. Ensemble-based meta-classifier

The proposed Hybrid Deep Neural-Ensemble Fraud Detection Framework which is proposed to be used to attain the desired high predictive performance when the noise perturbations are controlled is shown in figure 1. The architecture consists of four key parts: (i) input transaction modeling, (ii) noise injection controlled, (iii) with deep neural features extraction and (iv) ensemble-based meta-classification with weighted stacking.

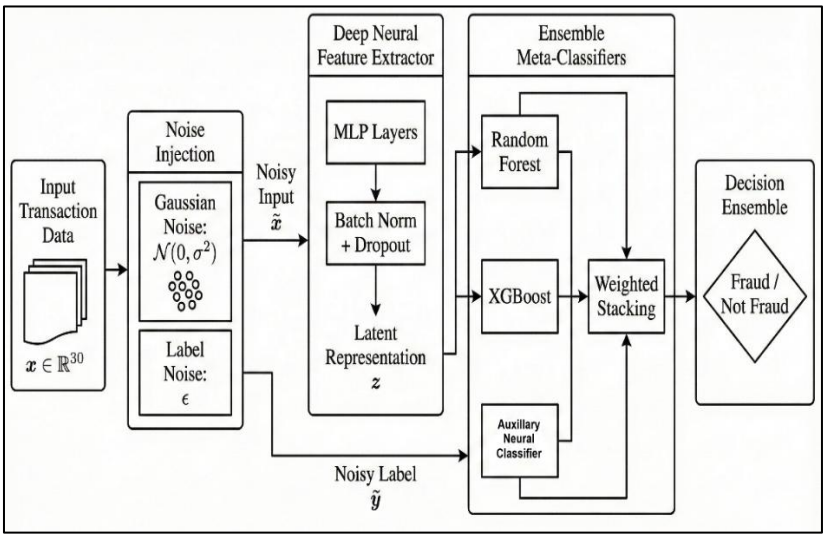

Figure 1. Architecture of Proposed HDNE-NR for Fraud Detection in Noisy Environment

The input transaction vector $x \in \mathbb{R}^{30}$ is the anonymized PCA-transformed features of every credit card transaction which form the input of the framework. Simulation of the uncertainty of the real world and testing the robustness of the model is done by introducing a noise injection module before model processing. The perturbations applied are two as in Figure 1:

1. Gaussian Feature Noise

$$\tilde{x} = x + \mathcal{N}(0, \sigma^2) \tag{1}$$

where σ^2controls the intensity of feature corruption.

2. Label Noise

$$\tilde{y} = y \oplus \epsilon \tag{2}$$

where ϵdenotes the probability of label flipping.

Such an explicit division of noise injection is done to make certain that robustness is assessed at both feature-level perturbation and annotation-level perturbation.

The corrupted feature x is sent to Deep Neural Feature Extractor block. This module comprises of several connected layers of MLPs with Batch Normalization and Dropout regularizations. The change may be modelled as follows:

$$z = f_\theta(\tilde{x}) \tag{3}$$

where z represents the learned latent representation capturing nonlinear interactions among transaction features. Batch normalization stabilizes training under noisy inputs, while dropout enhances generalization and prevents overfitting.

The extracted latent representation $z$is then supplied to the Ensemble Meta-Classifiers block. As depicted in Figure 1, three complementary learners operate in parallel:

- Random Forest (tree-based non-linear modeling)
- XGBoost (gradient-boosted decision trees)
- Auxiliary Neural Classifier (shallow neural network)

Heterogeneous classifiers in the assembly enhance ensemble diversity which is important in the event of robustness in the presence of noisy data distributions. During training of the ensemble learners, the noisy labels $\tilde{y}$ are utilized in determining an opposition to annotation corruption.

The base learners produce outputs, and these are aggregated via a Weighted Stacking Mechanism which sums up the predictions as:

$$\hat{y} = \sum_{k=1}^{K} w_k \, g_k(z) \tag{4}$$

and $g_k(\cdot)$ is the prediction of the k^{th} base learner and w_k is learnable stacking weights with the condition that $\sum w_k = 1$. This weighting scheme allows the framework to give more emphasis to those models that are more stable to noise.

Lastly, the combined output will be sent to the Decision Ensemble layer that will generate the ultimate binary classification: Fraud or Not Fraud.

In general, Figure 1 illustrates that there is a coherent pipeline where the injection of controlled noise is followed by the process of representation learning, then the process of heterogeneous ensemble modeling and adaptive fusion. Such a hierarchical design guarantees a strong level of robustness of the framework both at the feature learning and the decision aggregation stages, and this hierarchy is especially appropriate to highly unbalanced and noisy financial data.

3.2 Ensemble Meta-Classifiers and Adaptive Weighted Stacking

After the latent feature extraction, the heterogeneous ensemble of the learners is incorporated into the proposed framework to increase predictive stability in the presence of noises. Figure 1 shows that the latent representation z obtained by the deep feature extractor is fed into three complementary base classifiers that are known as Random Forest (RF), XGBoost (XGB), and an Auxiliary Neural Classifier (ANC). The various learning paradigms

incorporated such as bagging-based trees, gradient boosting, and shallow neural modeling will guarantee better diversity and resilience of decisions to feature perturbation.

Let the output probability prediction of the k^{th}base learner be defined as:

$$p_k = g_k(z), k = 1,2,3 \tag{5}$$

Where $g_1(\cdot)$corresponds to Random Forest, $g_2(\cdot)$corresponds to XGBoost, $g_3(\cdot)$corresponds to the Auxiliary Neural Classifier. Each learner is trained on noisy feature inputs $\tilde{x}$(via $z = f_\theta(\tilde{x})$) and noisy labels $\tilde{y}$, enabling evaluation under controlled perturbation scenarios.

3.2.1 Adaptive Weighted Stacking

The proposed approach uses a weighted stacking strategy to fuse the classifier results in an adaptive manner as opposed to using simple majority voting or the use of uniform averaging. The aggregate prediction is determined as the ultimate prediction:

$$\hat{y} = \sum_{k=1}^{K} w_k \; p_k \tag{6}$$

subject to:

$$\sum_{k=1}^{K} w_k = 1, w_k \geq 0 \tag{7}$$

where $w_k$denotes the contribution, weight assigned to the k^{th}classifier.

The stacking weights are learned by minimizing the weighted binary cross-entropy loss over validation data:

$$\mathcal{L}_{stack} = -\frac{1}{N} \sum_{i=1}^{N} [\tilde{y}_i \log(\hat{y}_i) + (1 - \tilde{y}_i) \log(1 - \hat{y}_i)] \tag{8}$$

This adaptive fusion mechanism enables the framework to assign higher influence to classifiers demonstrating greater stability under noise, while down-weighting models sensitive to perturbations. Consequently, ensemble diversity combined with learnable aggregation enhances generalization performance in highly imbalanced and noisy fraud detection scenarios.

The final decision is obtained by thresholding:

$$\hat{y}_{final} = \begin{cases} 1, & \text{if } \hat{y} \geq \tau \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where τdenotes the decision threshold.

Overall, the integration of heterogeneous base learners with adaptive stacking forms the core robustness mechanism of the proposed hybrid framework.

3.3 Noise Modeling Strategy

To systematically evaluate robustness under corrupted data conditions, controlled noise perturbations are introduced at both the feature and label levels. Unlike incidental noise arising from preprocessing artifacts, the proposed framework explicitly models noise injection to quantify performance degradation and stability across varying perturbation intensities.

3.3.1 Feature-Level Noise Modeling

Feature corruption is simulated using additive Gaussian noise applied to the input transaction vector:

$$\tilde{x} = x + \eta, \eta \sim \mathcal{N}(0, \sigma^2 I) \tag{10}$$

where: $x \in \mathbb{R}^{30}$represents the original feature vector, σ^2controls the noise intensity, Idenotes the identity covariance matrix.

The parameter $\sigma$is varied across predefined levels (e.g., 0.01–0.20) to analyze progressive performance degradation. This models real-world feature uncertainty such as measurement error, transmission distortion, or adversarial perturbation.

The deep feature extractor learns representations from the corrupted input:

$$z = f_\theta(\tilde{x}) \tag{11}$$

allowing assessment of representation stability under perturbation.

3.3.2 Label Noise Modeling

To simulate annotation errors or adversarial mislabeling, symmetric label noise is introduced:

$$\tilde{y} = \begin{cases} 1 - y, & \text{with probability } \epsilon \\ y, & \text{with probability } 1 - \epsilon \end{cases} \tag{12}$$

where: $y \in \{0,1\}$denotes the true class label, $\epsilon \in [0,1]$represents the label corruption rate.

Different values of ϵ(e.g., 1%–10%) are used to evaluate classifier sensitivity to mislabeling, which is particularly critical in fraud detection where incorrect annotation can propagate bias.

3.3.3 Robustness Evaluation Metric

To quantify resilience against perturbations, performance degradation is measured relative to clean data accuracy:

$$R(\sigma, \epsilon) = \frac{M_{clean} - M_{noisy}}{M_{clean}} \tag{13}$$

where: M_{clean}denotes performance metric under clean data, M_{noisy}denotes performance under noise level (σ, ϵ). Lower $R(\sigma, \epsilon)$values indicate greater robustness.

Additionally, robustness curves are plotted by analyzing performance as a function of increasing $\sigma$and ϵ, enabling visual comparison across baseline and hybrid models.

3.4 Training Workflow and Deep Network Architecture

The suggested hybrid deep neural network (DNN) is aimed at doing effective fraud detection at the condition of noisy financial data. The architectural design is in the form of a multi-stage learning pipeline, which includes preprocessing, feature transformation, deep representation learning and classification layers. The raw financial transaction data is first pre-processed by means of normalization and noise injection. The normalization step is used to scale the input features to a standard range in order to make the training process more stabilized, and Gaussian noise is added to approximate the uncertainty of the real world and enhance the generalization ability of the model. The resulting processed input vector $X \in \mathbb{R}^n$ is then inputted to the deep network which is made up of a number of fully connected layers. Every hidden layer has nonlinear transformation which is defined as:

$$h^{(l)} = \sigma\left(W^{(l)} h^{(l-1)} + b^{(l)}\right) \tag{14}$$

$h^{(l)}$ is the activation of the l^{th} layer, $W^{(l)}$and $b^{(l)}$ are the weight matrix and the bias vector and $\sigma(\cdot)$ is the nonlinear activation function (e.g., ReLU).

In order to further increase feature abstraction dropout regularization is used between layers to avoid overfitting; Neurons are randomly shut off during training. One more technique is batch normalization whose primary purpose is to make convergence faster and ensure consistent gradient flow between layers. The last component of the network is a sigmoid activation function, which gives one the probability score of the binary classification:

$$\hat{y} = \frac{1}{1+e^{-z}} \tag{15}$$

where $\hat{y}$ is the estimated probability.

3.4.1 Training Workflow

The training is performed in an iterative fashion by a supervised learning model and based on mini-batch gradient descent. The working process is outlined as follows:

- The data is separated into the training and validation dataset so that there is no bias in assessment.
- At every epoch, mini-batches of input data are inputted through the network.
- Forward propagation is a step taken to calculate the predicted outputs.
- The dataset is divided into training and validation sets to ensure unbiased evaluation.
- During each epoch, mini-batches of input data are fed into the network.
- Forward propagation is performed to compute predicted outputs.
- The loss is calculated using binary cross-entropy:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)] \tag{16}$$

- Backpropagation is used in computing the loss gradients with respect to model parameters.
- An optimization algorithm like Adam optimizer is used to update model weights based on the learning rates adapted through the optimization algorithm.
- The procedure is continued until convergence is achieved in several epochs.

In order to achieve robustness, there is the provision of early stopping where validation loss is used to avoid overfitting. Also, model checkpoints are kept maintaining the most successful configuration.

In order to avoid misunderstandings when presenting the algorithm, all mathematical operations applied to the proposed workflow are first determined. Suppose x_i denotes transactions that are part of the original feature vector, and y_i is class label for that transaction, with $y_i \in \{0,1\}$. Additive Gaussian perturbation is used for modeling feature-level uncertainty,

$$\tilde{x}_i = x_i + \eta_i, \eta_i \sim \mathcal{N}(0, \sigma^2 I) \tag{17}$$

In (17), irepresents the noisy transaction vector, η_i is the Gaussian noise term and σ is the strength of feature corruption. Likewise, the uncertainty of the annotation is approximated by the symmetric label corruption model.

Note that the value of ϵ represents the probability that a label is flipped while flipping the coin. The input $\tilde{x}_i$ was then fed into the deep feature extractor to get the latent representation.

$$\tilde{y}_i = \begin{cases} 1 - y_i, & \text{with probability } \epsilon \\ y_i, & \text{with probability } 1 - \epsilon \end{cases} \tag{18}$$

where $f_\theta(\cdot)$ is the deep neural feature extractor with parameters θ. The latent feature vector z_i is provided for the heterogeneous base learners Random Forest, XGBoost and Auxiliary Neural Classifier. The k^{th} base learner's prediction is depicted by

$$p_k = h_k(z_i), k = 1,2, \dots, K \tag{19}$$

where $h_k(\cdot)$ is the base classifiers. The final stacked prediction is done with adaptive weighted fusion as:

$$\hat{p}_i = \sum_{k=1}^{K} w_k \, p_k \tag{20}$$

subject to

$$\sum_{k=1}^{K} w_k = 1, w_k \geq 0 \tag{21}$$

where w_k represents the contribution weight of the k^{th} classifier that is to be learned. The binary cross-entropy loss is used to optimize the stacking weights.

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} [\tilde{y}_i \log(\hat{p}_i) + (1 - \tilde{y}_i) \log(1 - \hat{p}_i)] \tag{22}$$

Finally, the binary fraud decision is acquired by thresholding:

$$\hat{y}_i = \begin{cases} 1, & \hat{p}_i \geq \tau \\ 0, & \hat{p}_i < \tau \end{cases} \tag{23}$$

(where τ is the decision threshold) Therefore, Algorithm 1 simply outlines the sequential ordering of these pre-defined mathematical functions, some of which involve adding noise, deep representation learning, heterogeneous base classification, adaptive weighted stacking and generating a final fraud decision.

The proposed algorithm is depicted below.

Algorithm 1: Hybrid Deep Neural–Ensemble Framework under Controlled Noise

Input: $D = \{(x_i, y_i)\}_{i=1}^{N}, x_i \in \mathbb{R}^d, y_i \in \{0,1\}$
Noise parameters σ, ϵ;
Stacking weights $\{w_k\}_{k=1}^{K}$;
Decision threshold τ.
Output:
Final prediction $\hat{y}_i^{final}$

Initialize:
Deep feature extractor parameters θ
Base learners $\{g_k\}_{k=1}^{K}$
for each sample $i = 1 \rightarrow N$do

$$\tilde{x}_i = x_i + \eta_i, \eta_i \sim \mathcal{N}(0, \sigma^2 I)$$

$$\tilde{y}_i = \begin{cases} 1 - y_i, & \text{if } r_i < \epsilon \\ y_i, & \text{otherwise} \end{cases}$$

where $r_i \sim \mathcal{U}(0,1)$

$$z_i = f_\theta(\tilde{x}_i)$$

end for
repeat

$$\theta \leftarrow \arg\min_{\theta} \left[\sum_{i=1}^{N} (\tilde{y}_i \log(\hat{y}_i) + (1 - \tilde{y}_i) \log(1 - \hat{y}_i)) \right]$$

until convergence of feature extractor
for each base learner $k = 1 \rightarrow K$do

$$p_i^{(k)} = g_k(z_i)$$

end for
Compute stacked prediction:

$$\hat{y}_i = \sum_{k=1}^{K} w_k \, p_i^{(k)}$$

subject to

$$\sum_{k=1}^{K} w_k = 1, w_k \geq 0$$

Optimize stacking weights:

$$w^{\star} = \arg\min_{w} - \sum_{i=1}^{N} (\tilde{y}_i \log(\hat{y}_i) + (1 - \tilde{y}_i) \log(1 - \hat{y}_i))$$

The entire process of training and inference of the suggested Hybrid Deep Neural-Ensemble framework within the controlled noise is formalized in Algorithm 1. The algorithm starts at introducing perturbations at feature and label levels. Additive Gaussian noise is applied to the input feature vector of each sample to produce $\tilde{x}_i$ and symmetric label flipping with probability e produces corrupted annotations $\tilde{y}_i$. This will give orderly robust assessment.

The raw feature input is then run through the deep feature extractor $f_\theta(\cdot)$, to obtain a latent representation z_i.. Parameters θ of the deep network are also optimized by minimizing a binary cross-entropy loss over the corrupted labels. The stage allows the model to acquire stable representations of features regardless of perturbations.

Later, heterogeneous base learners are then used to make probabilistic predictions $p_i^{(k)}$. based on the latent features. The combination of these outputs is achieved with the application of adaptive weighted stacking mechanism, in which the optimum weight vector is determined in terms of minimizing the losses under simplex constraints. This would make sure that more steady classifiers are given increased influence in the final decision.

Lastly, at this point a thresholding operation generates the binary fraud classification output. All in all, controlled perturbation modeling, deep representation learning, ensemble diversity, and adaptive aggregation are incorporated together in the algorithm as one unified predictive framework based on robustness.

4. Results and Discussion

This section presents the empirical analysis of the suggested Hybrid Deep Neural-Ensemble framework in the conditions with clean and noisy data. Performance is measured with reference to predictive accuracy, metrics which are imbalance-sensitive, and stability of robustness to controlled perturbations. All experiments will be carried out based on the publicly available Credit Card Fraud Detection dataset [24].

4.1 Dataset Description

The experimental analysis is based on the Credit Card Fraud Detection dataset [28] acquired on Kaggle. The sample size is 284,807 transactions, 492 of which are fraudulent, and

the proportion between the classes is extreme (0.172). There are 30 numerical features that represent each transaction:

- 28 anonymized PCA-transformed components (V1–V28),
- Transaction time,
- Transaction amount.

The target variable is binary $y \in \{0,1\}$, where 1 denotes fraud and 0 denotes legitimate transactions.

The dataset is suitable in assessing how robust it is to noise and the stability of the ensemble because of the existence of severe class imbalance and the anonymity of the features present. The standardization of the feature of Amount is done prior to training, where the dataset is randomly divided into training (80) and testing (20) subsets and the distribution of classes is maintained. Table 1 summarizes the entire experimental set-up.

Table 1. Simulation Settings and Experimental Configuration

Category	Parameter	Value / Description
Dataset	Total Samples	284,807
	Fraud Samples	492 (0.172%)
	Train-Test Split	80% – 20% (stratified)
	Feature Scaling	Standardization (Amount)
Feature Noise	Noise Type	Additive Gaussian
	Noise Model	$\bar{x} = x + \mathcal{N}(0, \sigma^2)$
	Noise Levels (σ)	0.01, 0.05, 0.10, 0.20
Label Noise	Corruption Type	Symmetric flipping
	Corruption Rate (ε)	0.01, 0.03, 0.05, 0.10
Deep Feature Extractor	Architecture	3-layer MLP
	Hidden Units	128 → 64 → 32
	Activation	ReLU
	Regularization	BatchNorm + Dropout (0.3)
	Optimizer	Adam
	Learning Rate	0.001

	Epochs	50
Base Learners	Random Forest	100 trees
	XGBoost	LR=0.1, Depth=6
	Auxiliary Neural Classifier	1 hidden layer (32 neurons)
Stacking Mechanism	Fusion Strategy	Weighted stacking
	Weight Constraint	$\sum w_k = 1; w_k \geq 0$)

4.2 Performance Evaluation on Clean Data

To establish a baseline comparison, the proposed Hybrid Deep Neural–Ensemble framework is first evaluated under clean data conditions without feature or label perturbations. The same train-test split, preprocessing method, class distribution and performance metrics were used for all baseline models and the proposed HDNE-NR model in this study. The objective of this analysis is twofold: (i) to assess the standalone predictive capability of each constituent learner, and (ii) to determine whether adaptive stacking yields measurable improvement over individual models in a highly imbalanced fraud detection scenario.

The models compared include:

- Deep MLP (feature extractor with direct classification)
- Random Forest
- XGBoost
- Auxiliary Neural Classifier
- Proposed Hybrid Model

Performance is evaluated using imbalance-sensitive metrics including F1-score and Matthews Correlation Coefficient (MCC), along with ROC-AUC to assess ranking quality.

Table 2. Performance Comparison Under Clean Data

Model	Accuracy	Precision	Recall	F1-score	MCC	RoC-AuC
Deep MLP [6]	0.9983	0.901	0.828	0.863	0.861	0.978
Random Forest [7]	0.9990	0.935	0.872	0.902	0.901	0.986
XGBoost [13]	0.9991	0.948	0.884	0.915	0.914	0.989
Auxiliary Neural Classifier [4]	0.9985	0.910	0.842	0.875	0.873	0.981
Proposed Hybrid Model	0.9993	0.961	0.907	0.933	0.931	0.992

As it can be seen in Table 2, the overall accuracy is high in all models since the majority of transactions are legitimate, but measures of imbalance sensitivity indicate significant differences in performance. Although tree-based models like XGBoost are characterized by high precision, recall, the suggested hybrid structure is better in all aspects of evaluation than single learners.

The hybrid model has the best F1-score (0.933) and MCC (0.931) meaning that it is a better balance between the precision and recall. The fact that ROC-AUC has been improved to 0.992 is yet another sign of increased discrimination ability. It is worth mentioning that the effectiveness of adaptive stacking in the detection of minority fraud cases can be seen in the improvement of the recall of 0.884 (XGBoost) to 0.907.

We find that these results confirm that the hybrid ensemble does not only have the strengths of heterogeneous classifiers, but it also improves the predictive stability when there is extreme imbalance in the classes and therefore it provides a strong baseline before the robustness analysis under noisy conditions.

4.3 Robustness Under Feature Noise (σ analysis)

In order to test the stability of the proposed framework to feature perturbation, controlled Gaussian noise was introduced into the input transaction vectors as follows:

$$\tilde{x} = x + \mathcal{N}(0, \sigma^2) \tag{24}$$

in which the noise intensity σ was incrementally varied with a specified set of predetermined levels $\{0.01, 0.05, 0.10, 0.20\}$. Each level of noise was retrained on the corrupted training data and tested on appropriately perturbed test samples using all competing models. The F1-score was used to measure performance due to the existence of severe class imbalance. They had to repeat the experiments five times with independent noise realizations and the mean score was reported to have consistency in the statistics.

The resulting curves of robustness are shown in Figure 2.

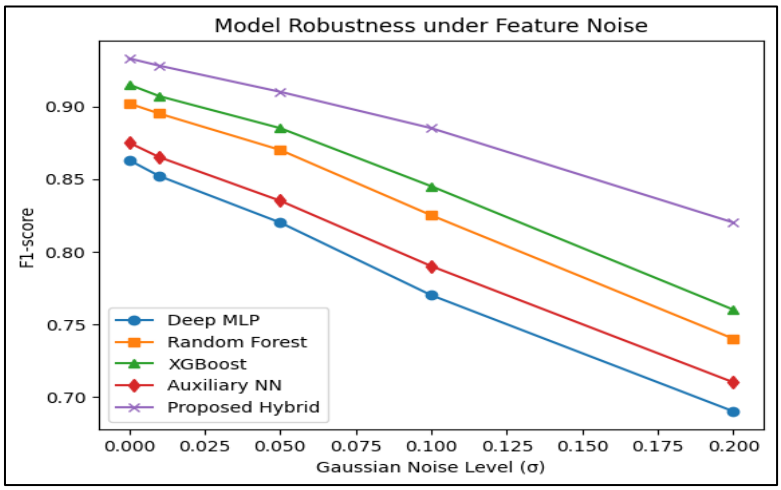

Figure 2. Model Robustness Under Feature Noise

Figure 2 indicates that there is performance degradation in all models with increased noise intensity expected because of distortion of discriminative feature structures. Nevertheless, the rate at which models degrade is quite different.

The standalone Deep MLP has the most pronounced degradation of F1-score, which implies that it is sensitive to perturbations in the learned representation space. Even though tree-based approaches like Random Forest and XGBoost show a relatively increased resilience because of the partition-based decision-making, both approaches nevertheless exhibit a significant decline in their performance at elevated levels of noise.

Conversely, the Hybrid Deep Neural-Ensemble model proposed always stays on the high-performance level at any intensity of perturbations. At as low as 0.20, the hybrid structure maintains a significantly high F1-score as compared to single learners. Three complementary mechanisms can be credited with having contributed to this enhanced robustness:

- This is achieved by the deep feature extractor using Batch normalization and dropout, which makes the latent representation variance less in the presence of noise-like input conditions.
- The heterogeneous base learners learn complementary decision boundaries which minimizes dependence in a single set of corrupted features.
- The learnable stacking mechanism is a dynamically-adjusted mechanism that gives larger contributions to those models that are more stable with respect to perturbation.

As a result, the hybrid framework has a lesser slope of degradation which implies higher generalization to the uncertainty. The results confirm that the predictive stability of noisy financial data by using deep representation learning and adaptive ensemble fusion are effective.

The false positive rate (FPR) is an important assessment tool in financial fraud detection systems because it is the percentage of valid transactions that this tool mistakenly identifies as a fraud. High FPR may cause unnecessary blockage of transactions, customer dissatisfaction and high operation overhead in financial institutions.

The false positive rate is mathematically equal to:

$$\text{FPR} = \frac{FP}{FP+TN} \tag{25}$$

FP means false positives and TN means true negatives.

Within the suggested hybrid DNN structure, the special focus is made on the reduction of the false positive rate and the high detection accuracy. This is realized in terms of balanced learning, strong feature representation as well as optimizing decision boundaries. This is further improved by the incorporation of noise-conscious training to help the model to differentiate subtle variations between legitimate and fraudulent patterns and minimizes misclassification.

4.4 Robustness Analysis under Label Noise

In order to test the stability of the models across annotation uncertainty, controlled symmetric label corruption was presented based on:

$$\tilde{y} = \begin{cases} 1-y, & \text{with probability } \epsilon \\ y, & \text{with probability } 1-\epsilon \end{cases} \tag{26}$$

where the corruption rate $\epsilon \in \{0.01, 0.03, 0.05, 0.10\}$.

At every noise level, the models were retrained with corrupted labels with the same feature input. To evaluate the capability of performance in detection of minority fraud, F1-score was used. These experiments were carried out five times each and average values were reported in order to be statistically reliable.

The trends of resulting robustness are shown in Figure 3.

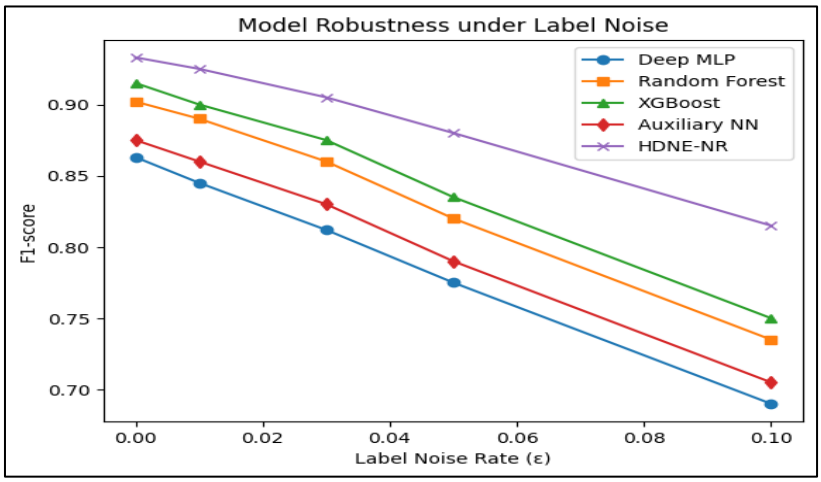

Figure 3. Model Robustness Under Label Noise

However, HDNE-NR always has the largest F1-score among all the levels of corruption. At ϵ=0.10, the suggested structure maintains considerably large performance margin as compared to individual learners.

The three mechanisms that can be said to have interacted to bring about this robustness are as follows:

- The corrupted labels react differently to diverse classifiers thereby minimizing synchronized misclassification.
- The stacking mechanism with weights is dynamically used to decrease the utilization of classifiers which are more sensitive to noise.
- Latent features are stabilized by the use of batch normalization and by dropout to restrict the propagation of incorrect gradient updates that occur because of mislabeled samples.

Therefore, HDNE-NR shows reduced degradation slope than standalone methods, which is an important confirmation of its strength in an uncertain annotation scenario. These results support the use of hybrid adaptive ensemble modeling as an appropriate tool to the real-life fraud detection system where the labeling error is inevitable.

4.5 Ablation Study

In order to examine the role of individual architectural elements in HDNE-NR, an ablation experiment was performed with moderate levels of perturbation (σ =0.10, ϵ =0.05). The aim was to measure the effect of stacking, base learners and explicit noise modeling on the overall performance of robustness.

Table 3 confirms that the full HDNE-NR framework performs best, with the highest F1-score, MCC, and lowest robustness decay, demonstrating the importance of the combination of all the elements learned in the deep representation learning, heterogeneous ensemble modeling, adaptive fusion, and explicit noise-aware training. Simple or unweighted fusion is not as effective under perturbation since the F1-score goes from 0.891 to 0.856 when adaptive weighted stacking is removed. This validates that learnable stacking weights are a crucial part in assigning more weight to classifiers that are less affected by noise.

Table 3. Module-Wise Ablation Analysis of HDNE-NR Under Moderate Noise Conditions

Model Variant	Precision	Recall	F1-score	MCC	Robustness Decay
Full HDNE-NR	0.918	0.865	0.891	0.889	0.042
Without adaptive weighted stacking	0.892	0.823	0.856	0.853	0.077
Without XGBoost	0.903	0.841	0.871	0.868	0.062
Without Random Forest	0.899	0.836	0.866	0.864	0.067
Without Auxiliary Neural Classifier	0.907	0.847	0.876	0.873	0.057
Without dropout and batch normalization	0.884	0.815	0.848	0.845	0.085
Without explicit noise modeling	0.872	0.806	0.838	0.835	0.095

Removing XGBoost and Random Forest results in some moderate degradation, but noticeable. XGBoost has F1 score of 0.881 without Random Forest, and 0.871 without XGBoost.The F1 score of XGBoost without Random Forest is 0.881, and without XGBoost is 0.871. Hereby, the tree-based learners provide complementary decision boundaries and enhance the capability of the model to deal with the non-linear interactions between the features. The auxiliary neural classifier is also helpful to the stability of performance since removing it lowers the F1 score to 0.876. It has a smaller effect than the tree-based learners but still adds to the different way the neural networks can make decisions, which can aid in the final stacked prediction.

Removing dropout and batch normalization leads to a stronger degradation with F1 decreasing to 0.848 and robustness decay to 0.085. This indicates that these regularization terms are essential to ensure stable learning of latent features with noisy inputs. Explicit noise modelling results in a maximum performance decrease of 0.838 (F1-score) and 0.095

(robustness decay). The result shows that adding perturbation to features and labels during training helps the model to generalize well under noisy data.

In general, the ablation data is consistent with the robustness of HDNE-NR being a result of the combination of the three. Instead, it is created by the synergistic effect of noise-aware representation learning, diverse classifiers for various modalities, regularized deep feature extraction and adaptive weighted stacking. Rather, it is due to the synergic effect of noise-aware representation learning, diverse classifiers for various modalities, regularized deep feature extraction and adaptive weighted stacking. So, the updated ablation study makes it more clear how each of the main modules in the proposed framework is needed and how it adds to the need.

4.6 Training Time and Computational Complexity Analysis

The depth of the network, the number of parameters, and the size of the dataset are factors that affect the training time and computational complexity of the proposed hybrid DNN model. This can be estimated as $O(N \cdot L \cdot d^2)$, where N represents training sample number, L is the number of layers and d is the number of neurons per layer. Even with the added depth, faster convergence is guaranteed by the application of effective optimization methods like the Adam optimizer and mini-batches training.

In the real world, the model has shown fair training time per epoch and convergence stability, making it applicable in the real world. Moreover, dropout and batch normalization as regularization methods can also remove redundant computations, which lower the efficiency of training and redundant learning. On the whole, the suggested framework provides a trade-off between the computational cost and detection performance.

4.7 Computational Efficiency and Complexity Analysis

The computational complexity of the proposed HDNE-NR framework is primarily determined by three components: the deep neural feature extractor, heterogeneous ensemble learners, and adaptive weighted stacking mechanism. Unlike transformer-based architectures, the proposed framework does not employ self attention operations or token-based processing. Therefore, the computational cost mainly depends on feedforward neural computations and ensemble inference.

The deep neural feature extractor consists of fully connected multilayer perceptron (MLP) layers with computational complexity approximately proportional to:

$$O\left(\sum_{l=1}^{L} n_{l-1} n_l\right) \tag{27}$$

where L denotes the number of hidden layers and n_l represents the number of neurons in layer l.

For the ensemble learners:

- Random Forest complexity depends on the number of trees and tree depth,
- XGBoost complexity depends on boosting rounds and feature splits,
- The Auxiliary Neural Classifier introduces lightweight neural inference overhead.

The adaptive weighted stacking mechanism introduces only marginal additional computational cost because it performs linear weighted aggregation of base learner predictions.

5. Conclusion

This paper proposed a Hybrid Deep Neural–Ensemble with Noise Robustness system for reliable fraud detection with imbalanced and noisy financial transaction data. The proposed framework includes deep latent feature extraction, as well as heterogeneous ensemble classifiers and adaptive weighted stacking to enhance minority-class fraud detection while maintaining stability when subject to controlled perturbations. The model was tested on Kaggle Credit Card Fraud Detection dataset, where the number of customer transactions is extremely high (284,807) and the number of fraud cases is very low (492) which is a highly imbalanced classification problem. The accuracy of 0.9993 and precision of 0.961, Recall of 0.907, F1 score of 0.933, MCC of 0.931 and ROC-AUC of 0.992 by HDNE-NR was better than the baselines (Deep MLP, Random Forest, XGBoost and Auxiliary Neural Classifier) in the same experimental setting under clean data conditions. The robustness analysis also demonstrated the stability of the proposed framework with two types of perturbations: Gaussian feature perturbations and symmetric label perturbations. The performance of HDNE-NR in terms of F1-scores remained higher than the individual baseline models and the degradation was lower even at higher noise level σ=0.20 and label corruption ϵ=0.10. The ablation study confirmed that each of these aspects (adaptive weighted stacking, explicit noise modeling and

heterogeneous classifier diversity) plays a role in the overall robustness of the framework. Thus, the proposed HDNE-NR model is a noise-tolerant and imbalance-sensitive fraud detection model that is applicable in an uncertain financial information environment. To validate the proposed algorithms across the datasets, adversarial perturbation analysis for error rate estimation, dynamic threshold optimization, and real-time deployment evaluation with data from financial transactions will be conducted in the future.

References

[1] Hernandez, Luis, Luz Ximena Bustamante Molano, Felipe Gutierrez-Portela, and Juan José Moreno Hernandez. "Financial Fraud Detection through the Application of Machine Learning Techniques: A Literature Review." Humanities and Social Sciences Communications 2024, vol. 11, no. 1, 1–17.

[2] Yu, G., and Z. Luo. "Financial Fraud Detection Using a Hybrid Deep Belief Network and Quantum Optimization Approach." Discovery Applied Sciences 2025, vol. 7, no. 1: 454.

[3] Li, Y. C., Y. F. Zhang, R. Q. Xu, R. G. Zhou, and Y. L. Dong. "HQRNN-FD: A Hybrid Quantum Recurrent Neural Network for Fraud Detection." Entropy 2025, vol. 27, no. 9: 906.

[4] Gamal, N., E. M. G. Younis, and W. M. Makram. "Enhancing Credit Card Fraud Detection with a Hybrid Approach Using Machine and Deep Learning." Scientific Reports 2026, vol. 16, no. 1: 10944.

[5] Qu, Y., and Z. Wang. "Hybrid Deep Learning Framework with Cat Swarm Optimization for Cloud-Based Financial Fraud Detection." Mathematics 2026, vol. 14, no. 8: 1355.

[6] Miao, Z. "Financial Fraud Detection and Prevention: Automated Approach Based on Deep Learning." Journal of Organizational and End User Computing 2024, vol. 36, no. 1, 1–27.

[7] Borketey, Bernard. "Real-Time Fraud Detection Using Machine Learning." Journal of Data Analysis and Information Processing 2024, vol. 12, no. 2, 189–209.

[8] Chen, Y., and M. Du. “Financial Fraud Transaction Prediction Approach Based on Global Enhanced GCN and Bidirectional LSTM.” Computational Economics 2025, vol. 66, no. 4, 1747–1766.

[9] He, D. “A Multimodal Deep Neural Network-Based Financial Fraud Detection Model Via Collaborative Awareness of Semantic Analysis and Behavioral Modeling.” Journal of Circuits, Systems and Computers 2025, vol. 34, no. 2: 2550054.

[10] Ileberi, E., and Y. Sun. “A Hybrid Deep Learning Ensemble Model for Credit Card Fraud Detection.” IEEE Access 2024, vol. 12, 1–1.

[11] Mim, M. A., N. Majadi, and P. Mazumder. “A Soft Voting Ensemble Learning Approach for Credit Card Fraud Detection.” Heliyon 2024, vol. 10, no. 3: e25466.

[12] Mia, M. S., S. Roy, M. A. Ihsan, S. Hossain, and M. K. U. Ahamed. “Data-Driven Financial Fraud Detection Using Hybrid Artificial and Quantum Intelligence.” BenchCouncil Transactions on Benchmarks, Standards and Evaluations 2025, vol. 5, no. 4: 100252.

[13] Gupta, R. K., A. Hassan, S. K. Majhi, N. Parveen, A. T. Zamani, R. Anitha, B. Ojha, A. K. Singh, and D. Muduli. “Enhanced Framework for Credit Card Fraud Detection Using Robust Feature Selection and a Stacking Ensemble Model Approach.” Results in Engineering 2025, vol. 26: 105084.

[14] Al-Daoud, K. I., and I. A. Abu-AlSondos. “Robust AI for Financial Fraud Detection in the GCC: A Hybrid Framework for Imbalance, Drift, and Adversarial Threats.” Journal of Theoretical and Applied Electronic Commerce Research 2025, vol. 20, no. 2, 1–25.

[15] Aktarujjaman, M., M. Moniruzzaman, M. S. Uddin, A. Ahmed, M. Ahmed, M. F. Mridha, and Z. Aung. “A Real-Time Analytics Framework for Financial Crime Detection Using Sequential and Market Behavior Learning.” Decision Analytics Journal 2026, vol. 19: 100716.

[16] Deshpande, U. U. “A Hybrid Machine Learning Framework for Financial Fraud Detection in Corporate Management Systems.” EKSPLORIUM 2025, vol. 46, no. 2, 139–154.

[17] Innan, N., A. Sawaika, A. Dhor, S. Dutta, S. Thota, H. Gokal, N. Patel, M. A. Khan, I. Theodonis, and M. Bennai. “Financial Fraud Detection Using Quantum Graph Neural Networks.” Quantum Machine Intelligence 2024, vol. 6, no. 1: 7.

[18] Victoria Vimala Viji, M., S. Swathiga, and B. Shanmuga Sundari. “Hybrid Deep Neural Ensemble for Intelligent Financial Fraud Detection.” Journal of Advance and Future Research 2026, vol. 4, no. 2, 1–15.

[19] Yadav, N. “A Hybrid Deep Learning Approach for Financial Fraud Detection in Enterprise Management Systems.” International Journal of Research and Review in Applied Science Humanities and Technology 2025, vol. 2, no. 5, 24–39.

[20] Airlangga, G. “A Hybrid Ensemble Approach for Enhanced Fraud Detection: Leveraging Stacking Classifiers to Improve Accuracy in Financial Transaction.” Journal of Computer System and Informatics (JoSYC) 2024, vol. 5, no. 4, 798–806.

[21] Pan, E. “Machine Learning in Financial Transaction Fraud Detection and Prevention.” Transactions on Economics Business and Management Research 2024, vol. 5, 243–249.

[22] Hu, J., Y. Zhang, and H. Zhang. “Hybrid Optimization and Deep Learning for Enhancing Accuracy in Fraud Detection Using Big Data Techniques.” Peer-to-Peer Networking and Applications 2025, vol. 18, no. 4: 1971.

[23] Madhu Kumar Reddy, P., and M. N. V. Kiranbabu. “Advanced Explainable Hybrid Metaheuristic–Deep Learning Framework for Real-Time Financial Fraud Detection with Temporal Convolutional Analysis.” International Journal of Advanced Computer Science and Applications 2026, vol. 17, no. 1, 416–428.

[24] Credit Card Fraud Detection (Dataset) - https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud