

“Deep Autoencoder-Based Anomaly Detection for Financial Fraud Transactions”

Vaishnavi T, vaishnavi.mythili@gmail.com

Abstract

The growing adoption of digital payment platforms has led to a rise in financial fraud, creating a strong need for intelligent and flexible fraud detection techniques. Traditional rule-driven methods and supervised machine learning models frequently struggle to detect new or evolving fraud patterns because they rely heavily on labelled datasets. This study proposes an unsupervised anomaly detection approach using a deep autoencoder model to detect suspicious financial transactions. The framework is trained only with genuine transaction records so that it can capture the normal behaviour of users through compressed latent features. Fraudulent activities are detected by measuring reconstruction errors and applying a threshold determined through percentile-based analysis. Tests conducted on a real credit card transaction dataset show that the proposed method provides high recall and effective classification performance, indicating its potential for practical use in real-time financial fraud monitoring systems.

Keywords

Anomaly Detection, Autoencoder, Fraud Detection, Deep Learning, Financial Transactions

I. Introduction

The widespread use of online banking, mobile wallets, and digital payment technologies has greatly reshaped the financial sector. Although these advancements have improved convenience and accessibility, they have also increased the risk of financial fraud, leading to major economic losses worldwide. Detecting fraudulent transactions remains a difficult task because fraudulent activities occur rarely compared to legitimate transactions, and fraud techniques continuously evolve over time.

Conventional fraud detection methods mainly depend on predefined rules or supervised machine learning algorithms. These approaches require large amounts of labelled transaction data and often fail to identify new or sophisticated fraud patterns. To overcome these limitations, unsupervised learning methods have gained attention as they focus on learning the characteristics of normal transaction behaviour and identifying abnormal activities without requiring labelled fraud examples.

This research proposes a fraud detection framework based on a deep autoencoder model for anomaly detection in financial transactions. The suggested approach is trained only on genuine transaction records, enabling it to recognize deviations from normal behaviour and effectively detect unknown fraudulent activities.

II. Related Work

Financial fraud detection has been explored using several computational and analytical methods. Initial studies primarily employed statistical techniques to detect unusual transaction patterns. While these methods were useful for smaller datasets, they often struggled with scalability and performance in large-scale financial systems.

Later, machine learning algorithms such as decision trees, support vector machines, and ensemble models improved fraud detection accuracy. However, these supervised approaches heavily depend on labelled datasets and may not adapt effectively to newly emerging fraud strategies.

In recent years, deep learning techniques have shown promising results because of their capability to model complex and non-linear relationships within transaction data. Among these approaches, autoencoders have become widely recognized for anomaly detection tasks. By learning compressed representations of legitimate data, autoencoders can identify abnormal patterns based on reconstruction differences, eliminating the need for explicit fraud labels.

III. Methodology

A. Autoencoder Architecture

The proposed framework is built using a deep autoencoder network composed of the following components:

- **Input Layer** - The input layer receives the original transaction data and passes it into the neural network. Each transaction is represented as a feature vector containing the relevant attributes of the financial record.
- **Encoder** - The encoder is responsible for transforming the input data into a lower-dimensional representation. It contains multiple neural network layers that progressively reduce the dimensionality of the data through linear operations and non-linear activation functions such as ReLU or sigmoid.

The encoder mainly includes:

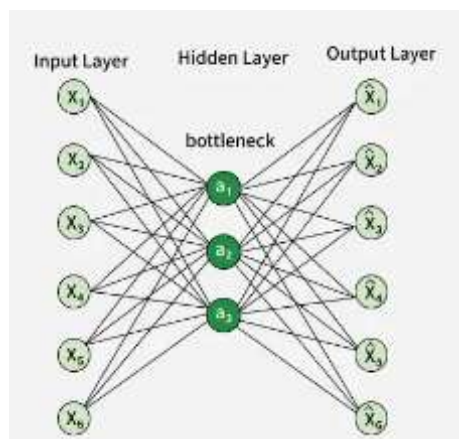
1. **Linear Transformation**: The input features are multiplied by weight matrices, and bias values are added to generate transformed outputs..
2. **Activation Function**: Non-linear activation functions help the network learn complex relationships and hidden structures within the data.

- **Bottleneck Layer** - The bottleneck layer represents the compressed form of the input data, commonly known as the latent representation. This layer preserves the most significant features of the original transaction while reducing redundant information. Two common Configurations include:

1. **Undercomplete Autoencoder**: The latent layer contains fewer neurons than the input layer, forcing the model to learn compact and meaningful representations.
2. **Overcomplete Autoencoder**: The latent layer contains more neurons than the input, allowing the network to capture additional complex patterns from the data.

- **Decoder** - The decoder reconstructs the original input data from the compressed latent representation. It performs operations similar to the encoder but in reverse order, gradually restoring the data to its initial dimensionality.

- **Output Layer** - The output layer generates the reconstructed transaction data. The reconstructed output is then compared with the original input to measure the reconstruction error, which is used to identify anomalies and potential fraudulent transactions.



B. Working Principle

The proposed autoencoder model is trained solely using legitimate financial transactions so that it can learn the normal behavioural patterns present in the dataset. During the testing or inference stage, every incoming transaction is processed through the trained network to generate a reconstructed version of the input.

- **Normal Transactions → Low Reconstruction Error**
- **Fraudulent Transactions → High Reconstruction Error**

Since the model is optimized to reconstruct genuine transactions accurately, normal records generally produce very small reconstruction errors. On the other hand, fraudulent transactions differ significantly from the learned patterns of legitimate behaviour, making them difficult for the autoencoder to reconstruct correctly. As a result, such abnormal transactions generate comparatively high reconstruction errors. This variation in reconstruction loss serves as the primary criterion for distinguishing suspicious activities from normal transactions.

C. Algorithm

The proposed anomaly detection system employs a deep autoencoder network to learn the behavioural characteristics of genuine financial transactions. The complete workflow includes data preprocessing, model training, anomaly score generation, adaptive threshold determination, and transaction classification.

A. Data Preprocessing and Normalization

Initially, the raw transaction dataset undergoes preprocessing to improve data consistency and quality. Unnecessary attributes are removed, while numerical features are standardized so that they possess zero mean and unit variance. This normalization process is important because features with different scales may negatively influence neural network training and convergence performance.

B. Training on Non-Fraudulent Transactions

To ensure that the model captures only normal transaction behaviour, the autoencoder is trained exclusively on genuine transaction samples. Let X_n represent the collection of non-fraudulent transactions. The network learns by minimizing the difference between the original input and its reconstructed output. Through this process, the model acquires the underlying structure and distribution of legitimate transaction patterns.

C. Anomaly Scoring Using Reconstruction Error

To ensure that the model captures only normal transaction behaviour, the autoencoder is trained exclusively on genuine transaction samples. Let X_n represent the collection of non-fraudulent transactions. The network learns by minimizing the difference between the original input and its reconstructed output. Through this process, the model acquires the underlying structure and distribution of legitimate transaction patterns.

$$e(x) = \frac{1}{n} \sum_{j=1}^n (x_j - \hat{x}_j)^2$$

Transactions that closely match learned normal behaviour generate low reconstruction errors, whereas anomalous or fraudulent transactions produce larger error values.

D. Adaptive Threshold Selection

Rather than selecting a fixed threshold manually, the proposed framework uses a percentile-based strategy to determine the anomaly cutoff dynamically.

$$\tau = \text{Percentile}_{95}(e(x))$$

This adaptive approach improves the robustness of the model by allowing the threshold to adjust according to the distribution of reconstruction errors within the dataset.

E. Classification Strategy

The final classification decision is made by comparing the reconstruction error of a transaction with the selected threshold value.

$$y(x) = \begin{cases} 0, & e(x) \leq \tau \\ 1, & e(x) > \tau \end{cases}$$

If the reconstruction error is below the threshold, the transaction is considered normal; otherwise, it is classified as fraudulent. This approach enables the detection of previously unseen fraud patterns without relying on labelled anomaly data.

IV. Dataset

The proposed framework is evaluated using a publicly accessible credit card transaction dataset containing anonymized attributes generated through dimensionality reduction methods. The dataset reflects real-world financial scenarios by exhibiting a highly imbalanced distribution, where legitimate transactions greatly outnumber fraudulent ones.

V. Implementation Details

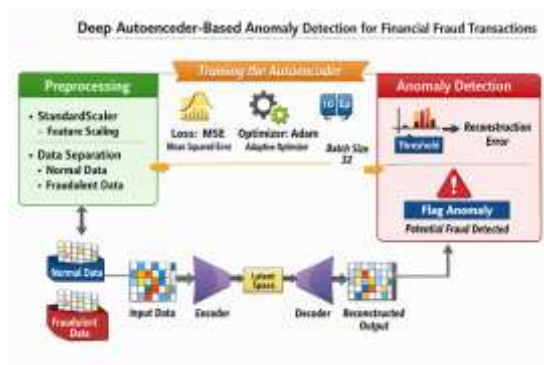
The proposed model is developed using Python along with deep learning and machine learning libraries such as TensorFlow and Scikit-learn.

A. Preprocessing

- Feature scaling - The StandardScaler technique is applied to normalize input features by transforming them to zero mean and unit variance.
- Data Separation- The dataset is divided into normal and fraudulent transaction groups for training and evaluation purposes.

B. Training

- Loss function: Mean Squared Error (MSE) is used to measure the reconstruction difference between input and output vectors.
- Optimizer: The Adam optimization algorithm is utilized for efficient and adaptive gradient-based learning.
- Epochs: The model is trained for 10 epochs, representing complete iterations over the training dataset.
- Batch size: A batch size of 32 is used, indicating the number of samples processed before updating model parameters.



VI. Results

A. Evaluation Metrics

The model is evaluated using:

- Precision
- Recall
- F1-score
- ROC-AUC

Scenario	Precision	Recall	F1-score	ROC-AUC	Interpretation
High Precision, Low Recall	High (few false positives)	Low (many frauds missed)	Moderate	Usually good	Customers trust the system (few false alarms), but fraudsters slip through.
Low Precision, High Recall	Low (many false alarms)	High (most frauds caught)	Moderate	Usually good	Fraud is caught, but many legitimate transactions are wrongly flagged.
Balanced Precision & Recall	Moderate	Moderate	High	Good	Best trade-off: system catches fraud while minimizing false alarms.
High ROC-AUC	Varies	Varies	Varies	High (close to 1)	Model distinguishes fraud vs. normal well across thresholds.
Low ROC-AUC	Varies	Varies	Varies	Low (close to 0.5)	Model struggles to separate fraud from normal transactions.

OUTPUT :

```
Epoch 1/10
7997/7997 — 24s 3ms/step - loss: 0.7280 - val_loss: 0.6849
Epoch 2/10
7997/7997 — 23s 3ms/step - loss: 0.6312 - val_loss: 0.6481
Epoch 3/10
7997/7997 — 24s 3ms/step - loss: 0.5944 - val_loss: 0.6153
Epoch 4/10
7997/7997 — 25s 3ms/step - loss: 0.5795 - val_loss: 0.6047
Epoch 5/10
7997/7997 — 25s 3ms/step - loss: 0.5780 - val_loss: 0.5944
Epoch 6/10
7997/7997 — 23s 3ms/step - loss: 0.5625 - val_loss: 0.5938
Epoch 7/10
7997/7997 — 21s 3ms/step - loss: 0.5577 - val_loss: 0.5872
Epoch 8/10
7997/7997 — 23s 3ms/step - loss: 0.5537 - val_loss: 0.5763
Epoch 9/10
7997/7997 — 22s 3ms/step - loss: 0.5489 - val_loss: 0.5807
Epoch 10/10
7997/7997 — 23s 3ms/step - loss: 0.5426 - val_loss: 0.5754
```

B. Performance Analysis

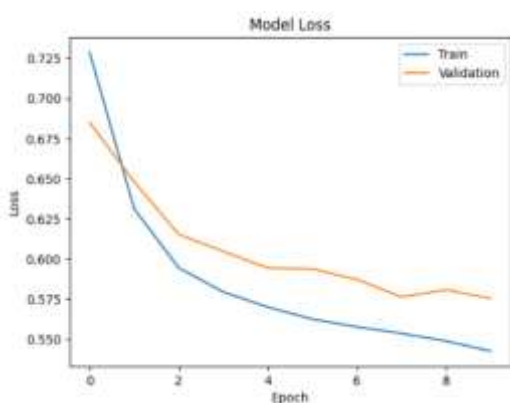
- High Recall:** The proposed model achieves high recall performance, ensuring that the majority of fraudulent transactions are successfully identified and minimizing the chances of undetected fraud.
- Strong Discriminative Performance:** The Receiver Operating Characteristic (ROC) curve demonstrates the model's strong ability to effectively distinguish between legitimate and fraudulent transactions.
- Reduced False Negatives:** The framework produces a low false negative rate, making it highly suitable for real-world financial systems where missing fraudulent transactions can result in significant financial losses.



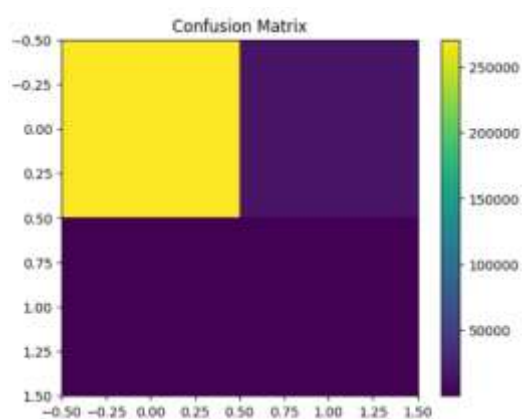
C. Visualization

The following graphs are used:

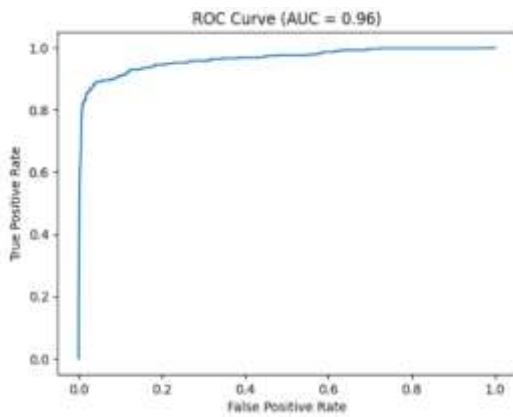
- Training vs Validation Loss Curve



- Confusion Matrix



- ROC Curve



VII. Conclusion

This study proposes an efficient and scalable deep learning framework for anomaly detection in financial transactions. By utilizing an unsupervised learning strategy, the model can successfully detect fraudulent activities without the need for labelled training data. Experimental results demonstrate the effectiveness of the proposed approach and its suitability for deployment in practical financial fraud detection environments.

Future enhancements of this work may include:

- **Hybrid Detection Models:** Combining autoencoder-based anomaly detection with supervised learning techniques to further improve detection accuracy.
- **Real-Time Fraud Monitoring:** Extending the framework to support real-time transaction analysis and streaming fraud detection systems.
- **Graph-Based Fraud Analysis:** Integrating graph learning methods to capture hidden relationships and complex fraud networks among transactions.

References

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
2. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), 1–58.
3. West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. *IEEE Systems Journal*, 11(4), 1–12.
4. Kaggle. (n.d.). *Credit card fraud detection dataset*. Retrieved from <https://www.kaggle.com/datasets>
5. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P. A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11, 3371–3408.