Available online at www.HighTechJournal.org

ISSN: 2723-9535

HighTech and Innovation Journal

Vol. 5, No. 1, March, 2024

Real-Time Online Banking Fraud Detection Model by Unsupervised Learning Fusion

Hanae Abbassi [1*], Saida El Mendili [1], Youssef Gahi [1]

[1] *Laboratory of Engineering Sciences, National School of Applied Sciences, Ibn Tofail University, Kenitra 14000, Morocco.*

Received 28 November 2023; Revised 12 February 2024; Accepted 17 February 2024; Published 01 March 2024

Abstract

Digital trades and payments are becoming increasingly popular, as they typically entail monetary transactions. This not only makes electronic transactions more convenient for the end customer, but it also raises the likelihood of fraud. An adequate fraud detection system with a cutting-edge model is critical to minimizing fraud costs. Identifying fraud at the ideal time entails establishing and setting up ubiquitous systems to consume and analyze massive amounts of streaming data. Recent advances in data analytics methods and introducing open-source technology for big data storage and processing opened new options for detecting fraud. This study aims to tackle this critical issue by providing a newly real-time e-transaction fraud detection schema that consolidates the advantages of both unsupervised learners, including autoencoder and extended isolation forests, with cutting-edge big data gadgets such as Spark streaming and sparkling water. It addresses the shortage of non-fraudulent instances and handles the excessive dimension of the set of features. On two real-world transactional datasets, we assess our suggested technique. Compared with other current fraud identification systems, our methodology delivers an elevated accuracy yield of 99%. Furthermore, it outperforms state-of-the-art approaches in reliably identifying fraudulent samples.

Keywords: Online Fraud Detection; Big Data Analytics; Autoencoder; Extended Isolation Forest; Real-Time Detection.

1. Introduction

Payment methods have been wholly improved over the years due to the advancement of technologies, massive data analytics, and machine learning. Because of the prevalence of mobile payments, criminals now have more options for committing online transaction fraud, including account takeover, chargebacks, money laundering, etc. Currently, fraud methods are distinguished by technological sophistication, hiding, and cross-regional offenses [1]. The matter's deployment processes are becoming increasingly veiled, approaches are continually being updated, and the danger of fraud rapidly progresses to the company's applicant process. This led to an enormous number of assets failing and having some influence on the financial equilibrium. As a result, avoiding and identifying fraudulent transactions remains a hotly debated research topic.

Earlier strategies of fraud detection relied on computational and signature-based tactics. These techniques were irrelevant in investigating the complexities of detecting fraud [2]. Moreover, computational approaches may produce an excessive number of false positives, misidentifying normal actions as illegal, resulting in operational shortcomings and consumer disappointment. As a result, we require a practical theft and security risk prediction engine that surpasses the

* Corresponding author: hanae.abbassi@uit.ac.ma

http://dx.doi.org/10.28991/HIJ-2024-05-01-014

shortcomings of existing approaches. Data mining and machine learning (ML) have garnered academics worldwide since heuristic strategies serve for rough results.

In contrast, ML techniques are employed for precise decisions -One well-known approach for classifying online transactions is to differentiate fraud with typical labels in the training set, often called supervised learning [3]. Common approaches in this subject comprise logistic regression [4], k-nearest neighbors [5], support vector machines [6], and decision trees [7]. This strategy uses labeled past transactions to create a predictive fraud algorithm that predicts the chance of every new transaction being fraudulent. However, there are hurdles to overcoming adaptive financial transaction treatment, notably shifting client behavior that must be handled to maintain legitimate operations. Due to these shifts and difficulties, e-banks and digital payment system providers are quickly modernizing how they process payments, which may cause security vulnerabilities. Hence, businesses must maintain solid and up-to-date real-time transactional fraud detection procedures.

This study aims to identify fraud in real-time in digital banking. In this regard, fraud detection architecture seeks to assess the threat inside each item in terms of fraud possibility in real-time. The fintech institutions can subsequently authorize, deny, or mandate the end user to provide a particular authentication once the transaction has been completed. To cope with these shortcomings, we put forward a new real-time e-transactional fraud detection scheme aimed at creating a digital banking fraud detection system that draws on the most relevant extensive data analysis techniques (such as spark streaming and sparkling water) alongside the autoencoder, which is a deep learning model, and an unsupervised learner, namely the extended isolation forest. A feature engineering technique founded on rule-based analysis is provided to create variable features to feed the fraud detection model. Afterward, a deep learning model with an unsupervised approach is included in the fraud detection process to enhance fraud detection speed and accuracy. We perform empirical tests using real-world datasets to evaluate the system's efficacy. The experiments' findings demonstrate the suggested technique's effectiveness as a viable tool for detecting e-transaction fraud.

Our study's primary contributions are as follows:

- Initially, we present a rule-based feature engineering approach for improving variable features for detecting fraudulent online transaction behavior.
- Secondly, we use the autoencoder to identify suspicious transactions in the data set and separate the malicious and regular transactions; then, we use the extended isolation forest to more accurately model transaction behaviors and obtain superior fraud detection performances.
- Thirdly, we conduct a practical evaluation of our fraud detection system and study its effectiveness on massive real-world datasets.
- Lastly, we compare our model to cutting-edge approaches.

The findings suggest that digital payment system providers may use the proposed approach to easily detect fraudulent transactions amid vast transactions, protecting consumers' interests while reducing false positive and negative rates. The technique can fulfill stringent operational time constraints while maximizing relevant prediction performance requirements.

The rest of the article is structured as follows. The digital banking fraud topology is provided in Section 2. The third section outlines our research methodology. Section 4 covers the tactics for detecting online banking fraud. Section 5 addresses a critical examination of the related work. Section 6 highlights our proposed work, whereas Section 7 focuses on experimental results and evaluation. Section 8 discusses our findings and compares them to current research. Finally, Section 9 summarizes the article and suggests future scopes.

2. Digital Banking Fraud Types

Banking fraud is a broad term that refers to any unlawful act involving a bank account. It entails seizing somebody's bank account, creating a banking account using someone else's identifier, or persuading them to give away funds beyond their approval [8]. Banking fraud can be classified as domestic fraud, which means that someone close to you or anybody else may have access to your personal information regarding your account. And sophisticated fraud, in which a fraudster purports to provide genuine technological assistance. Figure 1 highlights the digital banking fraud topology.

Phony operations cause financial losses and negatively impact businesses' reputations. As a side effect of these considerations, firms and researchers have developed a significant interest in identifying fraud, presenting a variety of theories built upon predictive modeling and data analytics methodologies. The upcoming section will examine various intriguing models for detecting digital banking fraud.

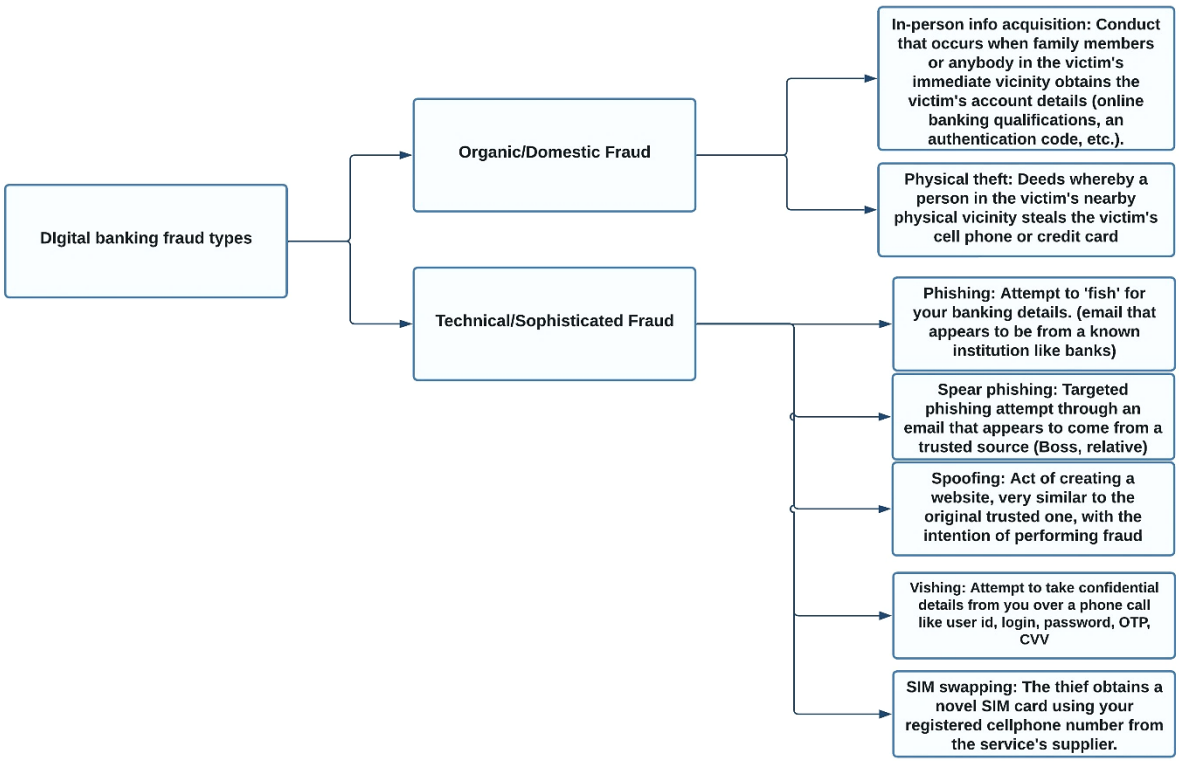

Figure 1. Banking fraud topology

3. Research Methodology

As shown in Figure 2, the three-phase process used by Sánchez-Aguayo et al. [9] was followed. Ali et al.'s [10] systematic strategy for literature review was utilized to gather and consolidate pertinent and current research tackling banking fraud detection.

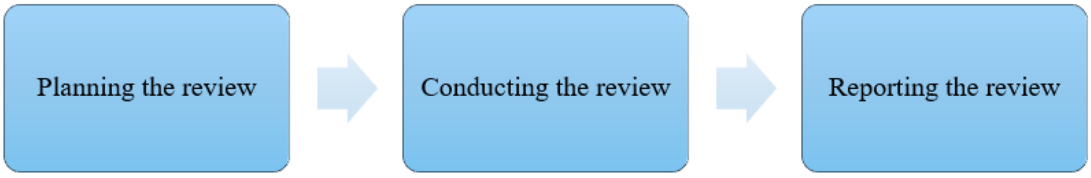

Figure 2. Related work research methodology

3.1. Review Planning

Determining the study field, creating review objectives, and establishing the research scope are all part of the first step. The present study focuses on the field of "banking fraud detection" with two primary objectives:

- Recognizing existing trends,
- Showcasing recent research that offers innovative methods for detecting banking fraud.

Studies using Machine Learning, Deep Learning, and significant data analytics approaches are included in the scope.

3.2. Conducting the Review

Conducting the review comes next, following review planning. Finding scientific datasets with publications relevant to our study area is the first stage in this process. ACM Digital Library, IEEE Xplore, Springer, Scopus, Web of Science, and ScienceDirect are the six primary online scientific datasets that were chosen. Setting criteria for which publications to keep or omit from our investigation was the first step in the research procedure. We established three requirements for the article: it must be written in English, published between 2019 and 2023, and published in a peer-reviewed academic publication.

Specific search phrases related to the research questions of this study were created, such as "What common banking frauds are handled depending on the ML approaches?" "Which well-liked machine learning techniques are used to identify financial fraud?" and "What assessment metrics are used to detect banking fraud?" This was done to obtain a more efficient and thorough search strategy, which entails combining search phrases that are pertinent to the research

questions using Boolean terms like "OR" or "AND." The following search phrases were utilized in this study: "machine learning" OR "big data analytics," "online transaction," and "digital banking fraud."

3.3. Reporting the Review

In this final phase, conclusions are drawn from the literature review, and the key outcomes are reported. The literature examination reveals a shortage of work identifying new systems to detect fraud in online banking that meet customers' needs with behavioral changes. This leads us to analyze the existing research to clarify the need and contribute significantly to this field. The next part summarizes the key conclusions of our related work and discusses the publications chosen for our investigation.

4. Online Banking Fraud Detection Tactics

With the advancement of FinTech, fraudulent digital transaction detection has become a prominent study area for academia and industry. It covers the detection of credit card fraud, mobile wallet fraud, e-commerce transaction fraud, and more.

Conventional rule-based and statistical techniques were frequently utilized to handle the issue of banking fraud detection before the development of machine learning and data-driven methodologies. Because rules are built upon proven trends [11], these systems are limited to identifying proven fraudulent trends; they cannot detect unexpected or emergent patterns. A deceptive transaction requires around 72 hours to be identified overall [12]. As a result, we undertook a literature analysis using the methods of research outlined in the preceding section to examine all accessible studies that contributed to the continuing interactions on online transaction fraud detection. For our review, we picked 15 papers. This section tackles many crucial research subjects pertinent to our work.

The potential of using machine learning algorithms, including classification, to spot fraudulent transactions has been addressed [13]. The authors have employed a variety of supervised learners, including random forests (RF), logistic regression (LR), support vector machines (SVM), and artificial neural networks (ANN), on a transactional dataset. The findings revealed that ANN performed well, with an F1 score of 0.91. Regarding banking transaction fraud detection context, Mytnyk et al. [14] have compared seven machine learning models: RF, k-nearest neighbors (KNN), LR, stochastic gradient descent (SGD), decision tree (DT), naive Bayes (NB), and SVM on a transactional dataset. According to the findings of the various methods, the LR works better, yielding a final AUC value of around 94.6%. Also, the AUC of the SGD is the best, with 95.4%.

Along with other kinds of e-transaction fraud, Mutemi et al. [15] suggested an automatic fraud identification system based on a vast transactional database from an online marketplace portal to identify potential fraud in the organized retail offense area. LR, DT, SVM, KNN, RF, Gaussian Naive Bayes (GNB), and Gradient Boosting (GB) comprise the seven supervised machine learning algorithms. Even though the GNB approach has the best recall value of 95.4% for all models examined, it fails to identify real positive cases and offers the least accuracy, with 40%. Similarly, in Field [16], we have proposed a unique technique for identifying Ponzi schemes over Ethereum by employing three algorithms: RF, ANN, and KNN. Over twenty thousand samples relating to Ethereum interaction channels were collected and prepared by Kaggle to train the models. Upon going through and contrasting each of the models, it was determined that RF performed the best, achieving an accuracy, an average score, and a total rating of 94%, 88.33%, and 96.6%, respectively.

Furthermore, Kodate et al. [17] examined interactions between users within a digital consumer-to-consumer market where a direct side links a supplier and the matching purchaser of a transaction. They used the RF classifier via user network attributes on a Japanese market dataset. The results demonstrated that the suggested technique differentiated distinct categories of phony users from regular users, with an AUC of roughly 91%–98%.

Emerging technologies in the arena of detecting fraud fitted investigators with more gadgets. Thoroughly, we have tackled fraud and abnormalities in the Bitcoin chain [18]. They suggested a safe fraud detection methodology made up of blockchain, using two algorithms for transaction classification: XGBoost and RF. The findings from simulations demonstrate that the proposed approach can detect transaction fraud and is immune to double-spending and Sybil threats. Following the same logic, Ren et al. [19] have presented a gradient-boosting decision tree-based fraud prevention integrating blockchain Tech (GBDT-APBT), which takes the prevention of fraud transaction algorithms as a collection of classifier weaknesses and then constructs the classifiers to determine whether a transaction is suspect. The private data of every consumer has been trained offline on the closest blockchain node; after that, the trained model is instantly loaded into the cloud, and the ultimate consensual model is established by vote. GBDT-APBT exhibits superior performance and efficacy in identifying malignities and proposes an intriguing approach regarding transaction safety in terms of accuracy in detection.

Other studies have addressed the class imbalance fraud detection problem [20]. I have explored a range of data augmentation strategies for identifying credit card fraud on an unbalanced dataset. The efficacy of the augmentation strategies is then evaluated using a variety of primary classification methodologies (such as SMOTE, ADASYN, B-

SMOTE, CGAN, Vanilla GAN, WS GAN, SDG GAN, NS GAN, and LS GAN). Compared to other augmentation techniques, these findings reveal that B-SMOTE, K-CGAN, and SMOTE have the greatest Precision and Recall. K-CGAN has the most excellent F1 Score and Accuracy among them. On the other side [21–23], we have developed an automatic rule-generating system to identify fraud systems that employ dispersed tree-based models involving DT, RF, and Gradient Boosting, with the parts of the expert rules serving as model attributes. They tested the proposed method using a bank's card transactions. The rules developed utilizing this system were revealed to be satisfying and practical, with a measurable commercial impact.

Because of its superior efficacy across numerous classification tasks, the deep learning method has been applied to fraud recognition in digital transactions. The authors have suggested an effective method for detecting credit card fraud that uses long short-term memory (LSTM) as the basis layer in the AdaBoost methodology. It will also include a combination of data resampling approaches, namely the synthetic minority oversampling technique with edited nearest neighbor (SMOTE-ENN). The suggested method's efficacy is proven by employing accessible real-world transaction sets. Research results reveal that the suggested LSTM ensemble outperforms benchmarked methods like SVM, MLP, DT, and conventional AdaBoost, with sensitivity and specificity of 99.6% and 99.8%, respectively.

Moreover, detecting fraudulent financial transactions has suggested novel features of engineering architecture and subsequently created an autoencoder in a real-life transactional database. The dataset was divided into three sets: original data, produced features, and picked successful features. The outcomes show that the autoencoder with the specified characteristics performs much better with the new framework than the data as it is. Using the same architecture to detect suspicious transactions [24], we have presented a new multiple-step deep learning architecture that combines a technique for selecting features based on an Autoencoder algorithm with a deep convolutional neural network (CNN). A broad collection of experimental cases demonstrates the proposed scheme's performance in contrast with SVM and CNN.

Recently, the area of fraud detection in digital transactions has begun to employ detection methods that combine the benefits of both machine learning and deep learning. Researchers have offered a hybrid approach to identifying bank fraud detection involving an autoencoder with probabilistic LightGBM (AED-LGB) [25, 26]. Primarily, the autoencoder separates low-dimensional characteristic data from incredibly dimensional banking feature data. The data is then resampled using the SMOTE technique before the features obtained are loaded into LightGBM. The findings show that the AED-LGB performs better with unbalanced data and does not enhance with resampling. In addition, when AED-LGB is compared to KNN and LightGBM, the ACC improves by 2% overall. The authors have also employed the Artificial Bee Colony with Recurrent Neural Network ABC-RNN to divide up fraud behavior and contrast it to current techniques by emphasizing fraud circumstances that can't be discovered through supervised learning. The results show that the accuracy of the suggested model is superior and the amount of training error is low. In another study by Berhane et al. [27], a hybrid CNN-SVM technique for identifying fraud in credit card transactions was established in this paper, which was evaluated against a real-life available transactional database. Based on the experimental outcomes, the CNN-SVM approach provided classification efficacy with precision, accuracy, and F1-score of 90.50%, 91.08%, and 90.41.

Despite the effectiveness of these approaches in fraud detection, they remain limited. The following section critically discusses these methods and presents our research motivation.

5. Summary of Research

By assessing the above studies, researchers have provided numerous methods for identifying online banking fraud, including the LGBM method and supervised ML techniques such as DT, RF, DT, NB, and KNN. Most of the methods investigated have proven advantageous in the procedure, yet their detection accuracy falls as the input variables rise, and they are susceptible to overfitting [28]. A further drawback of this method is that a single technique might fail to yield reliable results or generalize adequately for novel or unexplored information [2]. Furthermore, they're costly to build and require enormous computing power [29] to get quicker and better outcomes when detecting fraud.

The inspiration for this work stems from the reason that previous methodologies were unable to investigate the full scope of identifying fraud. Machine learning can learn from past information to uncover previously undiscovered fraud, making it an essential tool. Consequently, we require a practical fraud identification approach that tackles the deficiencies of existing methods. Our contribution differs significantly from cutting-edge techniques in a few significant regards. At the same time, we offer an innovative approach incorporating the advantages of rule-based, autoencoder-extended isolation forests and the most relevant big data engines. Through this integration, we can provide a complete and more flexible framework for identifying fraud, which helps us transcend the constraints of current methods.

Furthermore, this approach is designed to cope with the changing fraud scene, where illicit behaviors are becoming more dynamic and complex. Our system provides a proactive offense toward financial harm. It maintains the reliability of electronic transactions by utilizing the most modern advances in extensive data analysis and real-time data processing capacity, which allow immediate identification and reaction to fraud.

Specifically, our suggested approach excels in the following aspects:

- Both autoencoders and extended isolation forests are unsupervised machine learning algorithms that enable the identification of fraudulent activity without the need for labeled fraud events.
- Autoencoders are good at recognizing non-linear correlations in datasets [30]. The proposed strategy uses this power to identify complicated patterns and trends that standard techniques may need help detecting.
- The extended isolation forest, noted for its tolerance to outliers, is less susceptible to their impact than other techniques [31]. This resilience is especially useful in fraud detection because illicit transactions are frequently regarded as outliers. Furthermore, it enables interpretability within aberrant ratings, which helps rate cases according to their divergence from the norm. This interpretability helps comprehend system decisions and investigate possibly unauthorized transactions.
- Transaction fraud datasets are often very imbalanced, with a low ratio of phony transactions. The proposed model can deal with imbalanced sets without requiring considerable preparation or synthetic data processing.
- It strikes a harmonious blend of interpretability, adaptability, and efficiency, rendering it highly suitable for the real-time detection of suspicious transactions.

The details of the proposed approach will be highlighted in the upcoming section.

6. Digital Transaction Fraud Detection Framework

Before discussing our proposed strategy, we'll review the autoencoder and extended isolation forest basics.

6.1. AutoEncoder

The architecture of the Autoencoder resembles that of a neural network based on feed-forward reasoning with identical input and output properties [32]. As seen in Figure 3, the autoencoder merely transfers the input to the output. Throwing a constraint on the neural network may complicate an otherwise simple neural network. Limiting the number of neurons among *n* inputs and outputs is an example of a symbolic restriction. This limitation offers two benefits. For starters, Field would prohibit autoencoders from simply duplicating inputs to production. In addition, it would help me learn how to portray data better. This basic autoencoder is specified as an undercomplete autoencoder, the autoencoder's conceptual representation [33, 34]. The encoder and decoder are independent components of the autoencoder.

The encoder converts the given input X to a concealed representation Z. If Wφ and Bφ are the encoder layer biases and weighs, then concealed representation Z may be written $BZ = f_E\,(W_\varphi \times X + B_\varphi)$.

The activation function of the encoder is denoted by f_E.

The decoder: converts Z to its original data X reconstructing X'. If W_θ and B_θ are the encoder layer's weights as well as biases, the concealed structure Z may be expressed as: $X' = f_D(W_\theta \times Z + B_\theta)$.

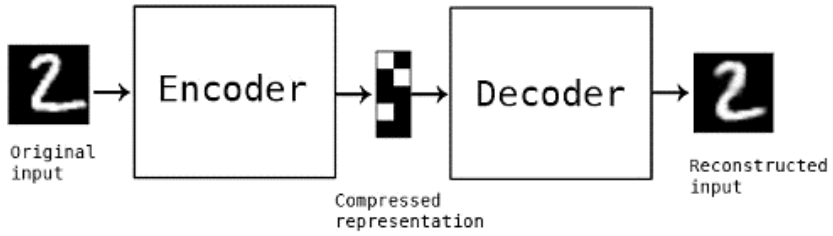

Figure 3. Architecture of an under complete autoencoder

6.2. Extended Isolation Forest

The Extended Isolation Forest methodology generalizes the Isolation Forest method. The predecessor Isolation Forest technique introduces a new type of identification despite exhibiting bias due to tree branching [35, 36]. The bias is mitigated by modifying the branching in the method, and the original approach is reduced to a particular instance.

The bias arises because branching is characterized by its resemblance to a binary search tree (BST). The attribute and its value are picked at each branching instance, which causes bias owing to the branching point sitting perpendicular to one of the vectors. Each branch node must have an arbitrary slope defined in the broader scenario. Rather than picking the attribute and value, it chooses an arbitrary gradient n and apex p to generate the branching split. The gradient can be created using an N (0,1) Gaussian distribution, and the point of intersection may be obtained using an even distribution with limits determined by the data to be divided [31, 37]. For an instance of point x, the branching requirements to determine data dividing are as follows: $(x - p) \times n \leq 0$.

6.3. Proposed Fraud Detection Architecture

This part proposes a real-time robust framework for catching fraudulent transactions in digital banking through a hybrid approach that takes advantage of big data engines, deep learning, and unsupervised learning, such as autoencoders and extended isolation forests, respectively. This enhances the ability to detect and oversee extremely complicated digital transaction fraud scenarios. Figure 4 presents our high-level proposed framework.

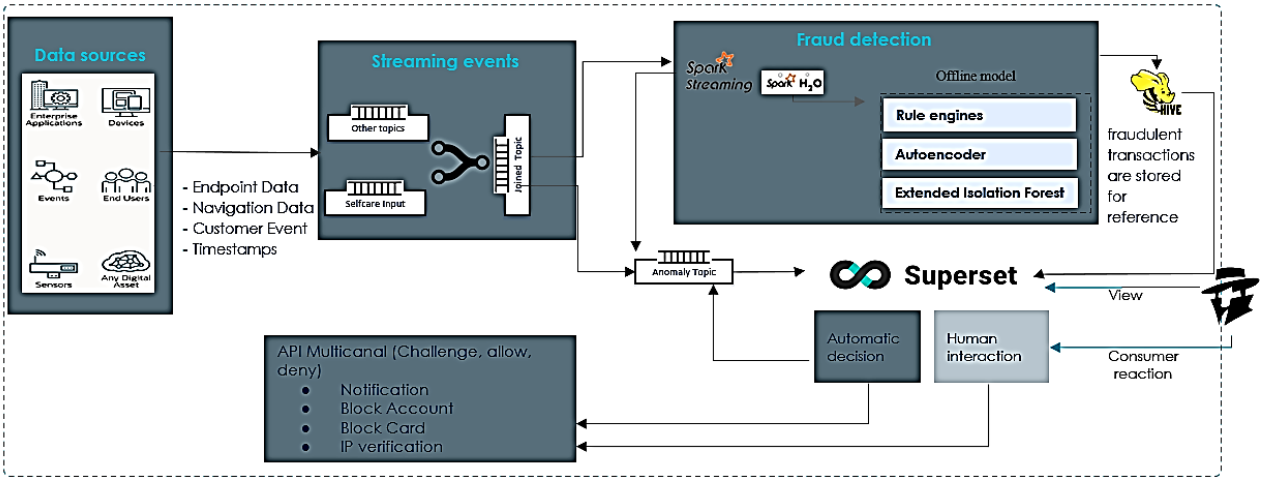

Figure 4. High level architecture

When proceeding with online payments, transactions must be checked and allowed only upon determining whether the consumer is a legitimate user or a fraudster. These transactions are fed through a sophisticated motor leveraging Apache Kafka and Spark streaming for detection. Reliable transactions can be wrapped up; however, phony ones cannot. These operations are recorded in the set for reference later.

The primary focus of this study is the Offline Model Training aspect. The training repository is generated via analysis of historical transactions in the hive data repository throughout a specific period. The predictive algorithm's feature properties are derived by rule-based feature engineering. Following that, an unsupervised hybrid model called AE-EIF (AutoEncoder-Extended Isolation Forest) is constructed to identify unlawful digital transactions. Following training, the model is fed into the online Fraud Detection component, which identifies fraud in real-time by calculating online transaction suspicion scores. Figure 5 shows our hybrid model architecture.

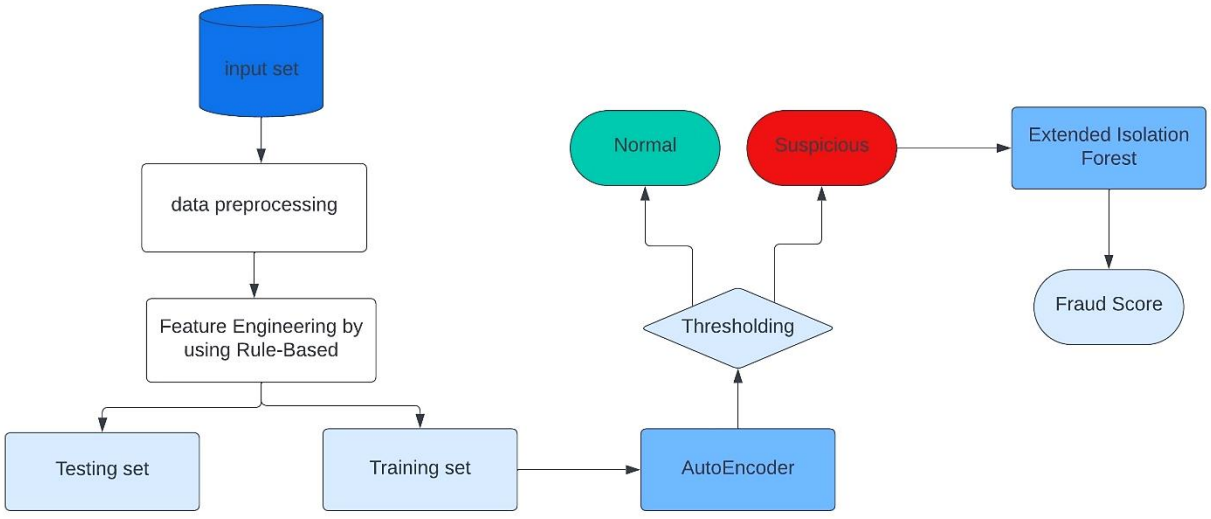

Figure 5. Flowchart of AE-EIF model

6.3.1. Architecture Work streams

Within our work, we apply a pair of steps for detecting questionable transactions. The outcome of the first step is used as the input for the following step. The training set is loaded onto the autoencoder in step 1. AE detects suspicious transactions and categorizes transactional data into two groups: aberrant and regular. The Extended Isolation Forest searches for these misfit observations in step 2. The primary phases of the suggested strategy are as follows:

A. In the initial stages, we begin by setting up an ensemble of rule-based features that will reveal transactions as possibly illegitimate if they display odd behavior according to criteria related to different transaction parameters, such as transaction amount, frequency, location, and others.
B. Establish our dataset for autoencoder training, incorporating these newly built rule-based attributes. The autoencoder picks up to encode and decode transactions while capturing their underlying trends and adding rule-based characteristics. The AE architecture consists of the following components:
    - Takes in the rules and properties created during the rule-based feature engineering process.
    - Minimize the data dimension and retain latent patterns.
    - A compressed form of the data is included.
    - Reassembles the data from its hidden representation.
    - Returns the rebuilt data.
C. A suspicious threshold will be established throughout the training phase.
D. For each transaction, the features and rules are put via the learned AutoEncoder. The discrepancy between the input data and the AutoEncoder outcome (reconstruction error) is used to calculate the fraud score. The greater the reconstruction mistake, the more suspicious the transaction is.
E. Contrast the fraud score from D with the suspicious threshold. If the rating exceeds the threshold, the sample is considered suspicious. This means that the transaction must be investigated via the EIF learner. Otherwise, the prediction operation is aborted, and the sample is deemed normal.
F. We will explore the suspect samples mentioned in E using the Extended Isolation Forest. Then, we will compute a final fraud score for each sample.
G. Finally, the model decides based on the final fraud ratings generated in the preceding phase. If the final fraud score surpasses a particular threshold, the sample is classified as an attempt to defraud. If the score is less than this, the transaction is deemed regular.

6.3.2. AE-EIF-Based Fraud Detection

Regarding the application's environment under consideration, where the suggested system must identify suspicious transactions in real time, the accuracy score for the fraud detection method must be maximized. Low accuracy implies a high number of false-positives. This might cause higher latency while analyzing and responding to actual suspects in the monitored transactions. As a result, even if the recall index is penalized in some circumstances, it is vital to adopt an oriented algorithm to maximize accuracy. A new hybrid method combining Deep Learning and the Extended Isolation Forest is suggested to achieve these objectives. Figure 6 presents the training and detection stages of the proposed model.

First, a set of rules based on fraud detection heuristics is defined. These rules include conditions related to various transaction attributes, such as transaction amount, frequency, location, and other relevant features. The defined rules are applied to the transactional dataset, creating binary features (0 or 1) for each transaction. These binary features indicate whether a transaction satisfies the defined rules. For example, a "high amount" binary feature is set to 1 if the transaction amount exceeds a specific threshold and 0 otherwise. These rule-based features serve as additional information that helps the model identify potentially fraudulent transactions.

After that, the Autoencoder was trained, generating three new features for each sample analyzed, which are as follows:

- A: A compressed set computed during the initial step.
- B: Stands for Euclid's distance concerning the inspected Ω and the reconstructed set Y, whereas
- C: Represents the degree to which similarities exist throughout the two samples.

Within this manner, all of the sample's Ω to be studied are outlined by a total of n characteristics present in Ω as well as three vector attributes denoted by μ, whereby μ = [A, B, C]. The EIF algorithm's forest of trees must be built in the following stage of the training procedure, with each of the samples of the training set including the characteristics of Ω beside those of μ.

By combining the speed of the AE with the accuracy of the Extended Isolation Forest method, the proposed approach intends to improve the reliability of the fraud detection system. The AE method is used in the first stage to forecast the suspicious score on Ω. If Ω's score exceeds the fraud threshold set during training, it is deemed suspect and must be examined using the EIF. Instead, the forecasting procedure is terminated, and the examined sample is considered usual. The EIF algorithm analyses the questionable samples and determines the final abnormal behavior score. The EIF method, for instance, involves more than the characteristics characterizing the set Ω; additionally, the attributes of μ taken from the set rebuilt by the Autoencoder as input.

The EIF examines the questionable samples in the suggested method; therefore, the forecasting speed stays the same as that of AE despite ordinary samples. Extended Isolation Forest then analyzes the suspect samples, which may detect some of the erroneous positives and improve the solution's reliability for fraud identification.

The following section will concentrate on the data set employed and the H2O AE-EIF model's implementation. We will reveal the strategy, crucial findings, and evaluation measures for both of these components.

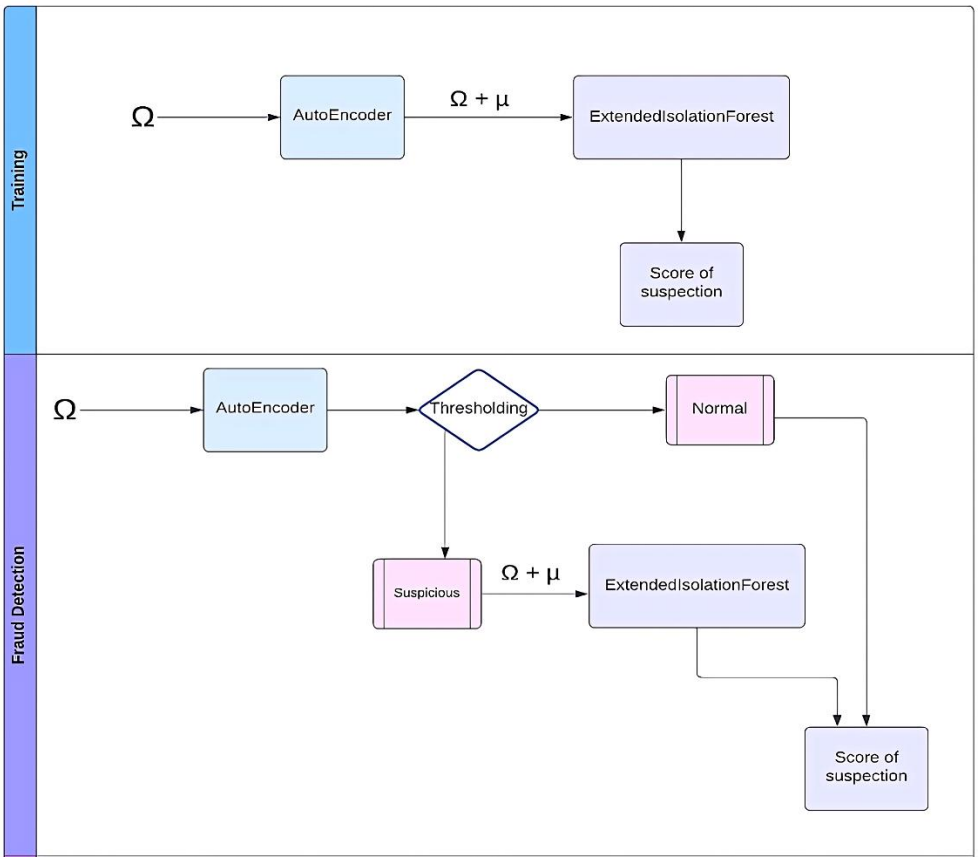

Figure 6. Training and Detection stages of AE-EIF model

7. Experimental Results and Analysis

This section describes the data sets, experimental setup, and evaluation measures used in our work. It then presents the results of the proposed approach's experiments.

7.1. Datasets

The data sets serve as training and examination of offered methodologies, making them critical in driving research. This subsection details two different data sets employed in our suggested framework tests.

Dataset 1: The database analyzed for our research contains digital transactions performed by European cardholders through Kaggle spanning two days; it includes 284,807 transactions, 492 being fraudulent. Moreover, as shown in Table 1, this data set comprises 31 feature input parameters that determine the outcome of a principal component analysis upgrade. The choice of this database is supported by its wide application as standard benchmark data in transaction fraud detection studies, allowing for a thorough review and contrast of our suggested approach against existing state-of-the-art methodologies.

Table 1. Features of Dataset 1

Attribute	Description
From V1, …., to V28	Encrypted cardholder information
Time	The time when a transaction occurred
Amount	The overall number of transactions
Class	Determines if a transaction is genuine or not using binary values '1' as well as '0'

Dataset 2: This second database comprises 594,643 transactions performed over 180 replicated days, 7200 of which are judged suspicious. This fictitious database is generated to detect fraudulent transactions through the BankSim app, a simulation setup mainly built to simulate fraudulent records. Table 2 displays all dataset properties.

Table 2. Features of Dataset 2

Attribute	Description
Step	The date when a transaction occurred
Customer ID	A unique code that identifies the customer's account that is engaged in the transaction.
Account Zip Code	The customer's related zip code.
Merchant ID	A code that identifies the merchant who is conducting the transaction.
Merchant Zip Code	The merchant's zip code.
Purchase Category	A variable of sorts that indicates the category of product or service bought.
Purchase Amount	The entire amount spent on the transaction.
Age Category	A categorical variable that assigns the consumer to one of eight different age categories.
Gender	A type of variable presenting the client's gender.
Status of Fraud	A binary number denotes whether or not the transaction was unlawful.

7.2. Experimental Setup

All of the experiments were carried out using a device equipped with an Intel Core i7-11800H CPU, 16 GB of RAM, and an Nvidia RTX 3050 GPU. In this subsection, we will highlight the implementation of our proposed model.

7.2.1. Feature Engineering

As asserted before, we establish a set of rules as feature engineering to profile user behavior in our data analysis. Rule-based feature engineering's rationale originates from its capacity to identify significant characteristics transparently and effectively from raw sets. App record information representing the number of transactions, frequency range, geolocation, and other pertinent features was used to construct them. The key features that were employed for training the model are summed up as follows:

- Any attempts to enroll new customers via a specific device.
- Sign-in tries that were both successful and unsuccessful.
- The day and time of the last login.
- Amount of the most recent transaction.
- Frequency of transactions.
- The latest timestamp for the device ID.
- Customer IP address.

The training step was provided once the features were extracted. The following subsection tackles the training of models.

7.2.2. Model Training

Once the features' engineering step was done, the suggested model was trained over the Sparkling Water package to cope with the intricacies and scale of real-world transactions. In the algorithm's training phase, we have generated a set comprising just the transactions judged normal. This strategy has multiple pluses. Starting with training AE just on a standard set gets rid of the unbalanced class problem. Furthermore, it enables the model to record regular transactions while discarding suspicious ones, rendering our technique more practical for real-time applications in which lawful and suspicious transactions must be decided instantaneously. The training set is first separated into 70% only to train the AE, and 30% of the training set is used to determine the fraud threshold and train the EIF.

The finest results throughout the training process were achieved via epoch values of 100 and 110 on data 1 and 2, respectively. The AE employs the Rectified Linear Unit activating function through both data, with these layers, including an input layer made up of 38 nodes containing the attributes of the data gathered; a trio of hidden layers of 14, 7, and 14 nodes; and a layer for output of 38 nodes containing the reestablished attributes of the original input data, regarding both datasets. In addition, for each data collection, a forest of 200 trees and a tree depth of 18 were used. The effectiveness of the approach with these settings is revealed in the next subsection via classification measures.

7.3. Evaluation Measures

The proposed technique is evaluated using various metrics, including accuracy, recall, precision, F1-F2 measures, and MCC. These measurements are often used to evaluate fraud detection strategies, as they offer an in-depth evaluation of the model's effectiveness. Accuracy is defined as the proportion of correctly classified occurrences among all cases. It is calculated using the formula that follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

Precision can be defined as the number of actual positive events compared to the total positive events categorized per the algorithm. It is obtained via this formula:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

Recall is the proportion of accurately detected affirmative instances to the total number of favorable cases in the dataset. The following formula calculates it:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

The F-score represents a statistical mean that employs a harmonic average to incorporate recall and precision. The following formulas calculate it:

$$\text{F1 score} = 2 \times \frac{precision \times recall}{precision+recall} \quad (4)$$

$$\text{F2 score} = 5 \times \frac{precision \times recall}{(4\times(precision))+recall} \quad (5)$$

When the goal variable is unbalanced, MCC (Matthews Correlation Coefficient) is an appropriate scorer to utilize instead of Accuracy. The MCC score goes (-1-1), with -1 indicating a model that forecasts an inverse category of the actual value, 0 indicating a learner that performs no more effectively than randomized speculating, and 1 indicating a flawless learner. It is obtained via this formula:

$$\text{MCC} = \frac{(TP \times TN)-(FP\times \text{FN})}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (6)$$

When assessing fraud detection programs, financiers confront multiple hurdles, including false positive and negative rates. A false positive is when the Fraud Detection System classifies operations as malicious when they are regular behaviors. Despite these examples involving categorization mistakes, they weren't responsible for significant losses for FinTechs. On the other hand, false negatives are occasions where transactions are deemed as regular yet fraudulent, resulting in considerable expenses for FinTechs and a drop in customer retention. As a result, we will be more intrigued by the measures listed above that provide reliability within fraud scenario categorization since they are the most relevant assessment metrics in this area for determining the efficacy of our suggested approach. Additionally, MCC is excellent for evaluating the efficiency of models in the setting of unbalanced datasets, providing insights on predictive capabilities.

7.4. Results

Based on the datasets described previously, we verified our suggested approach to detecting fraudulent transactions. Python with the sparkling water engine was utilized to experiment's findings. In unsupervised model testing, the fraud identification technique can categorize the analyzed transactions into regular (0) and suspicious (1) transactions. Because the algorithms EIF, AE, and AE-EIF output a fraud score for each sample analyzed, a threshold must be defined to distinguish between regular and abnormal values. For the experiments described in this part, because the proportion of aberrations in the test dataset is known, the aberration threshold is set to a number that permits the proportion of higher scores to be isolated. The databases have been analyzed only with numerical attributes. Using the criteria above, we compared the efficiency of our model to that of a single AE and single EIF. Figures 7 and 8 summarize the results obtained from Dataset 1 and Dataset 2, respectively, when the training stage is finalized.

The experiments reveal that our Model AE-EIF, singly outperforms the autoencoder and Extended Isolation Forest. The learners produce good outcomes independently, yet when combined, they acquire superior accuracy and F scoring. As such, we deduce that our suggested approach may effectively reduce the number of false alarms and identify uncommon illegal activities, which are crucial in the real world for banks and other financial institutions. As for the MCC metric, AE-EIF outperformed the different techniques, demonstrating a well-balanced effectiveness across genuine and fraudulent classes. Conversely, the EIF performed highly only in precision, F score, and Recall. Moreover, the AE-EIF provides a higher recall than the single AE and EIF, lowering the occurrence of false negatives.

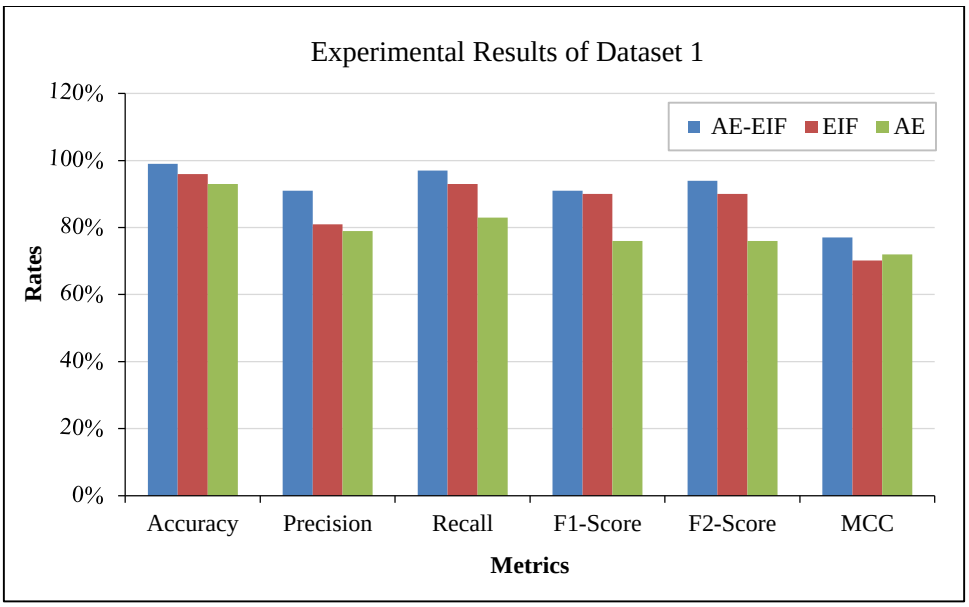

Figure 7. Experimental outcomes of dataset 1

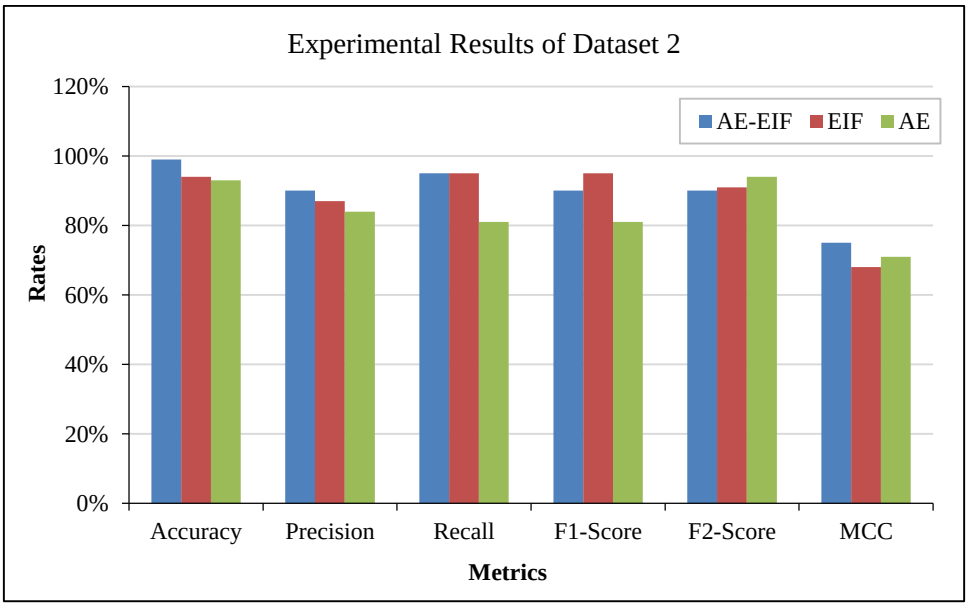

Figure 8. Experimental outcomes of dataset 2

F1-Score is a classification issue metric that equips recall and precision equitably. In Figures 7 and 8, the F1-Score test outcomes of AE-EIF range from 90% to 91%, greater than those of the two individual learners. This demonstrates that our proposed model is more consistent with overall efficacy and has superior classification impacts. When identifying fraudulent digital behaviors, it is vital to discover as many scams as possible to avert significant losses. Figures 7 and 8 illustrate the F2 Score test outcomes, with the recall outweighing the accuracy. The F2-Score findings for AE-EIF are greater than 90% in both datasets.

Furthermore, in the next section, the experimental findings are contrasted with state-of-the-art fraud detection techniques to evaluate their validity.

8. Discussion

Because of its importance in today's cyber environment, real-time transactional fraud detection is a hotly debated issue. Researchers have used many powerful and sophisticated algorithms based on machine learning to identify fraudulent activity. In this part, we compare our approach's performance over the European cardholder database (Dataset 1) to that of other cutting-edge transactional fraud detection approaches based on predictive techniques. The accuracy, precision, F1-Score, and recall measurement results of every method are shown in Table 3.

Table 3. Comparison between AE-EIF and state-of-the-art Methods

Reference	Technique	Accuracy	Precision	Recall	F1-Score
Karthikeyan et al. (2023) [26]	CNN-SVM	0.9	0.91	-	0.9
Afriyie et al. (2023) [38]	Decision Tree	0.92	0.05	0.93	0.09
Alfaiz & Fati (2022) [39]	K-Nearest Neighbors with CatBoost	0.97	-	0.95	0. 87
Kolli & Tatavarthi (2021) [40]	Harris water optimization- RNN	0. 91	0.99	0.76	-
Sadgali et al. (2021) [41]	Bidirectional Gated Recurrent Units	0.97	-	0.97	-
Present Model	AE-EIF	0.99	0.91	0.97	0.91

In keeping with the outcomes of this study and present state-of-the-art technologies for identifying fraud. For example, our framework obtains an accuracy of 99%, outperforming the CNN-SVM [27], Harris water optimization-RNN [40], and Decision Tree [39] models, producing 90% to 97% accuracy. Moreover, our model has superior precision, recall, and F1-Score ratings of 91%, 97%, and 91%, surpassing the Bidirectional Gated Recurrent Units [41], Decision Tree [38], and K-Nearest Neighbors with CatBoost [39] on these measures. Indeed, although some strategies obtained excellent precision results, they compromised other criteria, including recall, emphasizing the balanced effectiveness of our AE-EIF model.

Furthermore, the model's higher accuracy and precision, compared with existing approaches, indicate its usefulness in detecting illicit transactions while minimizing instances of false positives. Ultimately, this analysis highlights the resilience and usefulness of the suggested AE-EIF model in identifying digital transaction fraud, demonstrating its ability to beat cutting-edge methodologies and improve fraud detection skills in the contemporary digital world..

9. Conclusion

The present study provides a novel online banking fraud detection framework. Feature engineering techniques were utilized to create feature variables representing the behavior characteristics that fraud detection learners needed. On top of that, a hybrid approach that combines the strengths of autoencoder deep learning and extended isolation forest approaches was implemented to improve fraud detection modeling. Two real-world datasets were evaluated for the proposed model's performance with the singles autoencoder and extended isolation forest. A comparison of the AE-EIF-based suspicion detection performance to other powerful machine-learning algorithms revealed a pioneering approach to online banking fraud detection.

To the best of current knowledge, this study represents the first attempt to integrate the Autoencoder with the Extended Isolation Forest for detecting fraudulent transactions in digital banking. The practical impact of this study lies in the potential for digital banking providers to utilize the proposed approach for effective and efficient real-time detection of fraudulent transactions, thereby safeguarding consumer interests and reducing financial losses from fraud and compliance expenses. However, the study is limited by real-world implementation challenges related to data privacy. Additionally, the computing cost of the suggested framework was not assessed. Future research will focus on investigating the computing requirements for real-time online banking fraud detection and exploring the use of advanced AI algorithms and their combinations for fraud detection.

10. Declarations

10.1. Author Contributions

Conceptualization, H.A.; methodology, H.A.; software, H.A.; validation, Y.G. and S.E.M.; formal analysis, H.A.; investigation, H.A., Y.G., and S.E.M; resources, H.A., Y.G., and S.E.M; data curation, H.A.; writing—original draft preparation, H.A.; writing—review and editing, S.E.M. and Y.G.; visualization, H.A.; supervision, S.E.M. and Y.G.; project administration, Y.G. and S.E.M. All authors have read and agreed to the published version of the manuscript.

10.2. Data Availability Statement

Data sharing is not applicable to this article.

10.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

10.4. Institutional Review Board Statement

Not applicable.

10.5. Informed Consent Statement

Not applicable.

10.6. 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 paper.

11. References

[1] Jiang, S., Dong, R., Wang, J., & Xia, M. (2023). Credit Card Fraud Detection Based on Unsupervised Attentional Anomaly Detection Network. Systems, 11(6), 305. doi:10.3390/systems11060305.

[2] Nayyer, N., Javaid, N., Akbar, M., Aldegheishem, A., Alrajeh, N., & Jamil, M. (2023). A New Framework for Fraud Detection in Bitcoin Transactions through Ensemble Stacking Model in Smart Cities. IEEE Access, 11, 90916–90938. doi:10.1109/ACCESS.2023.3308298.

[3] Chang, V., Doan, L. M. T., Di Stefano, A., Sun, Z., & Fortino, G. (2022). Digital payment fraud detection methods in digital ages and Industry 4.0. Computers and Electrical Engineering, 100, 107734. doi:10.1016/j.compeleceng.2022.107734.

[4] Roseline, J. F., Naidu, G. B. S. R., Pandi, V. S., alias Rajasree, S. A., & Mageswari, N. (2022). Autonomous credit card fraud detection using machine learning approach☆. Computers and Electrical Engineering, 102, 108132. doi:10.1016/j.compeleceng.2022.108132.

[5] Habibpour, M., Gharoun, H., Mehdipour, M., Tajally, A., Asgharnezhad, H., Shamsi, A., ... & Nahavandi, S. (2023). Uncertainty-aware credit card fraud detection using deep learning. Engineering Applications of Artificial Intelligence, 123, 106248. doi:10.1016/j.engappai.2023.106248.

[6] Khan, S., Alourani, A., Mishra, B., Ali, A., & Kamal, M. (2022). Developing a Credit Card Fraud Detection Model using Machine Learning Approaches. International Journal of Advanced Computer Science and Applications, 13(3), 411–418. doi:10.14569/IJACSA.2022.0130350.

[7] Bin Sulaiman, R., Schetinin, V., & Sant, P. (2022). Review of Machine Learning Approach on Credit Card Fraud Detection. Human-Centric Intelligent Systems, 2(1–2), 55–68. doi:10.1007/s44230-022-00004-0.

[8] Moreira, M. Â. L., De Souza Rocha Junior, C., Silva, D. F. D. L., De Castro Junior, M. A. P., De Araújo Costa, I. P., Gomes, C. F. S., & Dos Santos, M. (2022). Exploratory analysis and implementation of machine learning techniques for predictive assessment of fraud in banking systems. Procedia Computer Science, 214(C), 117–124. doi:10.1016/j.procs.2022.11.156.

[9] Sánchez-Aguayo, M., Urquiza-Aguiar, L., & Estrada-Jiménez, J. (2021). Fraud detection using the fraud triangle theory and data mining techniques: A literature review. Computers, 10(10), 121. doi:10.3390/computers10100121.

[10] Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review. Applied Sciences (Switzerland), 12(19), 9637. doi:10.3390/app12199637.

[11] Kanika, Singla, J., Bashir, A. K., Nam, Y., Hasan, N. U. I., & Tariq, U. (2022). Handling class imbalance in online transaction fraud detection. Computers, Materials and Continua, 70(2), 2861-2877. doi:10.32604/cmc.2022.019990.

[12] Adewumi, A. O., & Akinyelu, A. A. (2017). A survey of machine-learning and nature-inspired based credit card fraud detection techniques. International Journal of System Assurance Engineering and Management, 8(S2), 937–953. doi:10.1007/s13198-016-0551-y.

[13] Sharma, P., Banerjee, S., Tiwari, D., & Patni, J. C. (2021). Machine learning model for credit card fraud detection-A comparative analysis. International Arab Journal of Information Technology, 18(6), 789–796. doi:10.34028/iajit/18/6/6.

[14] Mytnyk, B., Tkachyk, O., Shakhovska, N., Fedushko, S., & Syerov, Y. (2023). Application of Artificial Intelligence for Fraudulent Banking Operations Recognition. Big Data and Cognitive Computing, 7(2), 93. doi:10.3390/bdcc7020093.

[15] Mutemi, A., & Bacao, F. (2023). A numeric-based machine learning design for detecting organized retail fraud in digital marketplaces. Scientific Reports, 13(1), 12499. doi:10.1038/s41598-023-38304-5.

[16] Jacinta, O. I., Omolara, A. E., Alawida, M., Abiodun, O. I., & Alabdultif, A. (2023). Detection of Ponzi scheme on Ethereum using machine learning algorithms. Scientific Reports, 13(1), 18403. doi:10.1038/s41598-023-45275-0.

[17] Kodate, S., Chiba, R., Kimura, S., & Masuda, N. (2020). Detecting problematic transactions in a consumer-to-consumer e-commerce network. Applied Network Science, 5(1), 90. doi:10.1007/s41109-020-00330-x.

[18] Ashfaq, T., Khalid, R., Yahaya, A. S., Aslam, S., Azar, A. T., Alsafari, S., & Hameed, I. A. (2022). A Machine Learning and Blockchain Based Efficient Fraud Detection Mechanism. Sensors, 22(19), 7162. doi:10.3390/s22197162.

[19] Ren, Y., Ren, Y., Tian, H., Song, W., & Yang, Y. (2023). Improving transaction safety via anti-fraud protection based on blockchain. Connection Science, 35(1), 2163983. doi:10.1080/09540091.2022.2163983.

[20] Strelcenia, E., & Prakoonwit, S. (2023). Improving Classification Performance in Credit Card Fraud Detection by Using New Data Augmentation. AI (Switzerland), 4(1), 172–198. doi:10.3390/ai4010008.

[21] Vorobyev, I., & Krivitskaya, A. (2022). Reducing false positives in bank anti-fraud systems based on rule induction in distributed tree-based models. Computers & Security, 120, 102786. doi:10.1016/j.cose.2022.102786.

[22] Esenogho, E., Mienye, I. D., Swart, T. G., Aruleba, K., & Obaido, G. (2022). A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection. IEEE Access, 10, 16400–16407. doi:10.1109/ACCESS.2022.3148298.

[23] Ikeda, C., Ouazzane, K., Yu, Q., & Hubenova, S. (2021). New Feature Engineering Framework for Deep Learning in Financial Fraud Detection. International Journal of Advanced Computer Science and Applications, 12(12), 10–21. doi:10.14569/IJACSA.2021.0121202.

[24] Zioviris, G., Kolomvatsos, K., & Stamoulis, G. (2021). On the Use of a Sequential Deep Learning Scheme for Financial Fraud Detection. Intelligent Computing - Proceedings of the 2021 Computing Conference, 507–523. doi:10.1007/978-3-030-80126-7_37.

[25] Du, H., Lv, L., Guo, A., & Wang, H. (2023). AutoEncoder and LightGBM for Credit Card Fraud Detection Problems. Symmetry, 15(4), 870. doi:10.3390/sym15040870.

[26] Karthikeyan, T., Govindarajan, M., & Vijayakumar, V. (2023). Intelligent Financial Fraud Detection Using Artificial Bee Colony Optimization Based Recurrent Neural Network. Intelligent Automation and Soft Computing, 37(2), 1483–1498. doi:10.32604/iasc.2023.037606.

[27] Berhane, T., Melese, T., Walelign, A., & Mohammed, A. (2023). A Hybrid Convolutional Neural Network and Support Vector Machine-Based Credit Card Fraud Detection Model. Mathematical Problems in Engineering, 2023, 1–10. doi:10.1155/2023/8134627.

[28] Al Smadi, B., & Min, M. (2020). A Critical review of Credit Card Fraud Detection Techniques. 2020 11th IEEE Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2020, 0732–0736. doi:10.1109/UEMCON51285.2020.9298075.

[29] Hanae, A., Abdellah, B., Saida, E., & Youssef, G. (2023). End-to-End Real-time Architecture for Fraud Detection in Online Digital Transactions. International Journal of Advanced Computer Science and Applications, 14(6), 749–757. doi:10.14569/IJACSA.2023.0140680.

[30] Chen, S., & Guo, W. (2023). Auto-Encoders in Deep Learning—A Review with New Perspectives. Mathematics, 11(8), 1777. doi:10.3390/math11081777.

[31] Hariri, S., Kind, M. C., & Brunner, R. J. (2021). Extended Isolation Forest. IEEE Transactions on Knowledge and Data Engineering, 33(4), 1479–1489. doi:10.1109/TKDE.2019.2947676.

[32] Misra, S., Thakur, S., Ghosh, M., & Saha, S. K. (2020). An Autoencoder Based Model for Detecting Fraudulent Credit Card Transaction. Procedia Computer Science, 167, 254–262. doi:10.1016/j.procs.2020.03.219.

[33] Cheon, M. J., Lee, D. H., Joo, H. S., & Lee, O. (2021). Deep learning based hybrid approach of detecting fraudulent transactions. Journal of Theoretical and Applied Information Technology, 99(16), 4044–4054.

[34] Chen, Z., Yeo, C. K., Lee, B. S., & Lau, C. T. (2018). Autoencoder-based network anomaly detection. Wireless Telecommunications Symposium, 2018-April, 1–5. doi:10.1109/WTS.2018.8363930.

[35] Chen, X., Xu, W., Wang, S., Li, Y., & Lin, Z. (2022). An Anomaly Detection Scheme with K-means aided Extended Isolation Forest in RSS-based Wireless Positioning System. IEEE Wireless Communications and Networking Conference, WCNC, 2022-April, 1910–1915. doi:10.1109/WCNC51071.2022.9771602.

[36] de Santis, R. B., & Costa, M. A. (2020). Extended isolation forests for fault detection in small hydroelectric plants. Sustainability (Switzerland), 12(16), 6421. doi:10.3390/SU12166421.

[37] Zheng, F., Bonnet, S., Villeneuve, E., Doron, M., Lepecq, A., & Forbes, F. (2020). Unannounced Meal Detection for Artificial Pancreas Systems Using Extended Isolation Forest. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2020-July, 5892–5895. doi:10.1109/EMBC44109.2020.9176856.

[38] Afriyie, J. K., Tawiah, K., Pels, W. A., Addai-Henne, S., Dwamena, H. A., Owiredu, E. O., Ayeh, S. A., & Eshun, J. (2023). A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. Decision Analytics Journal, 6, 100163. doi:10.1016/j.dajour.2023.100163.

[39] Alfaiz, N. S., & Fati, S. M. (2022). Enhanced Credit Card Fraud Detection Model Using Machine Learning. Electronics (Switzerland), 11(4), 662. doi:10.3390/electronics11040662.

[40] Kolli, C. S., & Tatavarthi, U. D. (2020). Fraud detection in bank transaction with wrapper model and Harris water optimization-based deep recurrent neural network. Kybernetes, 50(6), 1731–1750. doi:10.1108/K-04-2020-0239.

[41] Sadgali, I., Sael, N., & Benabbou, F. (2021). Bidirectional gated recurrent unit for improving classification in credit card fraud detection. Indonesian Journal of Electrical Engineering and Computer Science, 21(3), 1704–1712. doi:10.11591/ijeecs.v21.i3.pp1704-1712.