IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 2 November 2023, accepted 19 November 2023, date of publication 4 December 2023,
date of current version 12 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3339226

RESEARCH ARTICLE

Online Payment Fraud Detection Model Using Machine Learning Techniques

ABDULWAHAB ALI ALMAZROI[1] AND NASIR AYUB[2], (Student Member, IEEE)
[1]Department of Information Technology, College of Computing and Information Technology at Khulais, University of Jeddah, Jeddah 21959, Saudi Arabia
[2]Department of Creative Technologies, Air University, Islamabad 44000, Pakistan

Corresponding author: Abdulwahab Ali Almazroi (Aalmazroi@uj.edu.sa)

This work was supported by the Deputyship for Research and Innovation, Ministry of Education, Saudi Arabia, under Project MoE-IF-UJ-22-4100409-9.

ABSTRACT In a world where wireless communications are critical for transferring massive quantities of data while protecting against interference, the growing possibility of financial fraud has become a significant concern. The ResNeXt-embedded Gated Recurrent Unit (GRU) model (RXT) is a unique artificial intelligence approach precisely created for real-time financial transaction data processing. Motivated by the need to address the rising threat of financial fraud, which poses major risks to financial institutions and customers, our artificial intelligence technique takes a systematic approach. We commence the process with artificial intelligence data input and preprocessing, mitigating data imbalance using the SMOTE. Feature extraction uses an artificial intelligence ensemble approach that combines autoencoders and ResNet (EARN) to reveal critical data patterns, while feature engineering further enhances the model's discriminative capabilities. The core of our artificial intelligence classification task lies in the RXT model, fine-tuned with hyperparameters using the Jaya optimization algorithm (RXT-J). Our artificial intelligence model undergoes comprehensive evaluation on three authentic financial transaction datasets, consistently outperforming existing algorithms by a substantial margin of 10% to 18% across various evaluation metrics while maintaining impressive computational efficiency. This pioneering artificial intelligence research represents a significant advancement in the ongoing battle against financial fraud, promising heightened security and optimized efficiency in financial transactions. In defense against wireless communication interference, our artificial intelligence work aims to strengthen security, data availability, reliability, and stability against cyber warfare attacks within the financial industry.

INDEX TERMS Financial transaction fraud, deep learning, fraud defense mechanism, detection, optimization methods, classification, ResNeXt, cyber attacks.

I. INTRODUCTION

With the expansion of e-commerce and the widespread adoption of online payment methods, fraudulent activities have seen a noticeable surge. Credible reports indicate a stark and rapid increase in financial losses attributed to credit and debit card fraud between 2020 and 2022 [1]. What's particularly striking is that while unauthorized purchases and the use of counterfeit credit cards make up a relatively small portion, approximately 12-17%, of the total reported fraud cases, they account for a disproportionately large share, ranging from 75% to 80%, of the overall financial losses. In light of these critical issues, private businesses and government organizations have substantially increased their funding for research and development projects. Their primary objective is to create more resilient and effective systems for detecting fraudulent activities. Implementing automated fraud detection systems has become essential for financial institutions that oversee credit card issuance and online transaction management. These systems not only help reduce financial losses but also play a crucial role in enhancing customer faith and assurance. Innovative big

The associate editor coordinating the review of this manuscript and approving it for publication was Tyson Brooks.

data and artificial intelligence possibilities have opened up, giving intriguing potentials, particularly in utilizing powerful machine learning algorithms to combat financial crime. Modern fraud detection systems, aided by cutting-edge data analysis and advanced machine/deep learning algorithms, have demonstrated extraordinary efficacy [2]. Typically, these algorithms are trained on large datasets of labeled transactions, allowing them to differentiate between regular and fraudulent activity. The ultimate result is the development of binary classification models capable of distinguishing between valid and fraudulent transactions. Detecting fraudulent transactions using classification algorithms is a difficult task that requires constant innovation and flexibility. In the same way, innovation assures security, data availability, dependability, and resilience against cyber warfare assaults in the fight against wireless communication interference, the financial industry must continually innovate to stay ahead in the struggle against financial crime.

Firstly, there's the issue of an imbalanced dataset, where the number of fraudulent transactions is significantly lower compared to legitimate ones. Secondly, there is the issue of cost sensitivity, where misclassifying fraudulent and normal transactions carries dissimilar costs, potentially having severe consequences. Additionally, transactions exhibit temporal dependence, necessitating consideration of their temporal relationships. Moreover, concept drift exists over time, which means that class conditional distributions can evolve, mandating periodic updates to the classifier. Lastly, managing the dimensionality of the feature space is a significant challenge, demanding sophisticated preprocessing techniques [3].

A thorough examination of existing research [4] found that when it comes to artificial neural networks, supervised learning techniques, logistic regression, and decision trees are the most commonly employed methods. Researchers have been drawn to the remarkable achievements of deep learning in various domains like computer vision, translation, speech recognition, and complex time series forecasting. As a result, several studies have started to utilize recurrent neural network variants, such as GRU, to develop credit card transaction fraud detection systems, taking inspiration from the success witnessed in those areas.

LSTM and GRU represent recurrent neural networks (RNNs) aimed at mitigating the issues of gradient vanishing and exploding gradients in RNNs [5]. These architectures are designed explicitly for capturing temporal patterns within sequential data. Deep learning models are attractive because of the valuable information from unprocessed data. In many research areas, neural networks and deep learning methods have consistently shown better results than conventional algorithms. However, it's important to note that these models haven't been widely used in fraud detection of credit card fraud.

The key to building effective and accurate fraud detection systems is to transform the input data from a fraud dataset into a simplified, lower-dimensional form [6]. This lower-dimensional representation is achieved using representation learning techniques, yielding a more detailed and informative depiction of the data. Prominently featured in this study, Autoencoders have gained prominence as practical tools within the repertoire of representation learning techniques. Their allure lies in their capacity to unveil latent patterns within input data before subjecting them to classification. An Autoencoder consists of two key components: an encoder and a decoder. The decoder attempts to reconstruct the original inputs using the condensed representation that the encoder has created from the input data.

To guide our investigation, we formulate three core research questions to drive our study:

- How can the proposed ResNeXt-embedded Gated Recurrent Unit (GRU) model (RXT) contribute to enhancing the accuracy and efficiency of financial fraud detection in real-time transaction data analysis, particularly in the context of credit card fraud?
- What specific techniques, such as Synthetic Minority Over-sampling Technique (SMOTE), ensemble feature extraction (EARN), and hyperparameter fine-tuning with the Jaya optimization algorithm (RXT-J), are employed in the proposed system model to address the challenges of data imbalance, feature extraction, and model scalability in financial fraud detection?
- In what ways does the proposed research represent a significant advancement in combating financial fraud, and what are the tangible benefits it offers, such as delicate security and operational efficiency, for financial institutions and consumers?

This study introduces an innovative framework that amalgamates feature vectors from the Ensemble Autoencoder and ResNet (EARN) and processes them through a unified learning ensembler. EARN offers the advantage of effectively capturing both high-dimensional and low-dimensional features in financial transaction data. Following the EARN stage, we conduct feature engineering, which includes K-Means Clustering, Principal Component Analysis (PCA), Silhouette scoring, and the Isolation Forest Model (IFM). Subsequently, we leverage the ResNeXt-GRU (RXT) for the classification task. To ensure scalability across diverse datasets in terms of type and size, we meticulously fine-tune the hyperparameters of RXT using the Jaya Algorithm (JA). The resulting RXT-J model can handle varying data sizes, including big data. Our model provides a significant economic benefit by swiftly and accurately identifying transaction fraud, making it a valuable tool for the financial sector. The key contribution of this study is described as:

- EARN Ensembler: Our study introduces the EARN Ensembler, a sophisticated feature extraction tool that adeptly captures both high-dimensional and low-dimensional aspects of financial transaction data. By seamlessly combining the strengths of various

feature extraction methods, this innovative framework offers versatility and effectiveness in revealing essential data patterns.

- Ensemble Learning Feature Vector: Within the EARN Ensembler, we present the concept of an ensemble learning feature vector. This approach leverages the collective learning from multiple feature extraction methods, harnessing their strengths to create a unified, comprehensive feature representation. This vector becomes a powerful asset in enhancing the model's ability to discern meaningful patterns in financial transaction data.
- RXT-J Classification Model: Our study introduces the RXT-J classification model, designed to navigate big data's complexities while maintaining exceptional accuracy in distinguishing between normal and fraudulent financial transactions. Leveraging advanced techniques and fine-tuned hyperparameters through the Jaya Algorithm (JA), this model sets a high standard for classification performance.
- Real-time Performance: Our model excels in real-time scenarios, delivering rapid and precise results crucial for timely fraud detection. Notably, it exhibits remarkable resilience when faced with data size scalability challenges, ensuring consistent and reliable performance across a diverse range of datasets. This capability makes it a valuable tool for real-world financial applications, where timely and accurate fraud detection is paramount.

II. BACKGROUND AND RELATED WORK

This section delves into the extensive studies conducted on financial Transaction fraud detection providing a detailed exploration of their findings and limitations.

A. BACKGROUND

As previously discussed, in the era of widespread utilization of e-commerce platforms and the consequential reliance on electronic payment systems, implementing an efficient fraud detection system assumes paramount significance in mitigating potential losses. The significant rise in the volume of credit card transactions processed through electronic payment systems provides a rich data source that can be effectively utilized to develop a fraud detection system driven by data analysis [5]. These datasets related to credit card transactions contain diverse features that can be incorporated into machine learning models, including transaction details, cardholder data, and transaction histories. The datasets employed for fraud detection exhibit the following unique characteristics: Absence of Public Data Sets: While credit card transaction data are abundant, there is a significant lack of publicly accessible datasets for researchers to use in their experiments. This scarcity of data stems from stringent constraints on disclosing information within this domain, preventing operators from divulging insights into their business activities. Even anonymized data releases encounter resistance from many financial institutions [6]. Cost Sensitivity: The nature of credit card fraud detection inherently embodies cost sensitivity, as false positives incur higher costs than false negatives. When a legitimate transaction is incorrectly labeled as fraudulent, it entails administrative expenses for the financial institution. Conversely, failure to detect fraud results in losing the transaction's value [7]. Data Imbalance: In the context of fraud detection, datasets often contain many legitimate transactions, and only a small proportion of fraudulent ones are used to train the model. This skewed distribution of data poses a difficulty in achieving accurate classification results. Some Machine Learning (ML) approaches necessitate including genuine and fraudulent instances during model training. Addressing this challenge often involves preprocessing the dataset to balance the data artificially [8]. Strategies to rectify data imbalance can be executed through oversampling or undersampling methods. Oversampling techniques include increasing the representation of fraudulent instances in a dataset to restore data balance, often duplicating them. On the other hand, undersampling methods achieve data balance by reducing the number of normal instances. Numerous research studies have shown that employing these strategies can improve the performance of modeling techniques. Notably, oversampling methods tend to outperform undersampling procedures in this regard. Feature space dimensionality: Credit card transaction databases encompass a diverse array of features, spanning transaction-specific properties, cardholder details, and transaction histories. Consequently, [9] employing dimensionality reduction techniques is vital to optimize the effectiveness of learning methods. Past research has used approaches like PCA and Neighborhood Components Analysis (NCA). However, deep representation techniques are imperative to attain a rich data representation from the input data [10].

B. RELATED WORK

Extensive fraudulent activities within the corporate and global financial sectors have resulted in substantial investment losses, substantial legal expenses, and the complete upheaval of the entire organization [25]. Academic researchers and investors have proposed innovative technological solutions and devoted significant efforts to combat these fraudulent activities within the financial industry. Safeguarding financial operations has become increasingly dependent on fraud detection systems. There has been a substantial surge in research efforts over the past decade focusing on fraud detection. Researchers have leveraged numerous machine learning and deep learning algorithms to predict and identify fraudulent financial activities. For instance, credit card fraud detection has garnered considerable attention with the application of neural networks [13], [14], [26]. In a real-world dataset, analysis focused on credit card fraud detection, it was found that logistic regression and artificial neural networks (ANN) exhibited comparable performance when trained with the training data derived from the empirical

TABLE 1. Summary of research contributions and techniques in financial transactions fraud detection.

Ref	Research Contribution	Applied Techniques	Data Dimension Reduction	Reduction Method Employed	Data Regularization Handling
[8]	Pioneering a novel ensemble method for credit card fraud prediction	LSTM, GRU, Ensemble of Recurrent Neural Networks (RNNs)	Not Applied	Not Applied	Employed L1 and L2 regularization to prevent overfitting.
[9]	Using RNNs for sequence classification. Applying traditional feature engineering techniques.	LSTM and RF	Not Applied	Not Applied	Applied dropout regularization to mitigate overfitting in RNN models.
[10]	Introducing a method for generating fake fraud data using GANs and comparing it to SMOTE.	Artificial Neural Network (ANN) classifier	Not Applied	Not Applied	Employed synthetic data regularization techniques to balance class distribution.
[11]	Designing a system to split data, train separate models on each split, and merge them into an ensemble.	ANN	Not Applied	Not Applied	Employed cross-validation for regularization during model training.
[12]	Designing a sliding window-based algorithm to downsample data.	ANN, LSTM, Field (LSTM-CRF)	Not Applied	Not Applied	Utilized data augmentation techniques for regularization.
[13]	Proposing a method involving feature selection, dimension reduction, and application to two deep learning models.	LSTM, Incorporating an Attention Mechanism	Employed	PCA and t-SNE	Implemented early stopping to prevent overfitting in deep models.
[14]	Using GANs and VAEs to handle imbalanced datasets and comparing their effectiveness with SMOTE	Artificial Neural Network (ANN)	Not Applied	Not Applied	Employed GAN and VAE for data regularization and balancing.
[15]	Capturing customer spending patterns and incorporating transaction timestamps for improved classification	Decision Tree, Logistic Regression, Random Forest	Not Applied	Not Applied	Utilized class-weighted regularization to account for class imbalance.
[16]	Leveraging dimension reduction techniques for data preparation before applying the KNN	K-Nearest Neighbors (KNN)	Employed	PCA and NCA	Implemented instance-based regularization in KNN.
[17]	Offering cutting-edge methods that entirely depend on authentic transaction data and employ a Discrete Wavelet Trans and Discrete Fourier Analysis (DFT) to identify credit card fraud before it happens.	Two proactive approaches based on DWT and DFT	Not Applied	Not Applied	Applied time-based regularization for anomaly detection.
[18]	Building an advanced fraud detection model that uses invariant diversity to identify patterns in the device characteristics used during transactions.	Regression	Not Applied	Not Applied	Utilized regularization techniques to prevent overfitting in the regression model.
[19]	Assessing how well filter and wrapper feature selection techniques improve the performance of a classifier.	Decision tree, Random Forest, AdaBoost	Feature selection based on Information gain and a wrapper feature selection method	Not Applied	Utilized regularization techniques to control the complexity of decision tree models.
[12], [20]	Suggest a suitable model for dealing with extensive datasets by integrating a subsampling technique within a stochastic step-by-step framework.	Stochastic stagewise approach	Not Applied	Not Applied	Implemented regularization in the stochastic stagewise framework to prevent overfitting.
[21]	Presenting a cutting-edge neural network ensemble to boost accuracy and resilience in credit card fraud detection.	Neural Network Ensemble	Not Applied	Not Applied	Employed dropout regularization in neural networks to improve generalization.
[22]	introducing a self-organizing map (SOM)-based anomaly-based credit card fraudulent transactions system.	Self-Organizing Maps (SOMs)	Employed	Not Applied	Utilized SOMs with regularization to control map size and adapt to data distribution.
[23]	Proposing a hybrid approach combining decision tree classification with genetic algorithm feature selection for credit card fraud detection	Decision Tree, Genetic Algorithm	Employed	Not Applied	Implemented regularization within the genetic algorithm to guide feature selection.
[24]	Introducing a credit card fraud detection system using a combination of KNN, RF	KNN, Random Forest (RF)	Not Applied	Not Applied	Employed regularization techniques within RF to control tree depth.

observations [27]. However, ANN outperformed logistic regression when evaluating their performance on test data [15], [28]. It is crucial to highlight that for the ANN to deliver optimal results, it must be trained on real-time datasets. Task classification within ANN may not effectively identify abnormal behaviors or detect fraud if the model lacks adequate training.

To improve the accuracy of detecting credit card fraud, logistic regression is used to identify essential factors while fine-tuning the approach for capturing transactions [16], [29]. Real-time credit card transaction data can benefit existing credit card fraud detection methodologies. This dataset enables the identification of credit card fraud based on various classification parameters, such as location, product

type, transaction type, and others. However, an essential limitation of the logistic regression algorithm is its inclination to produce results that fall into only one category when used in fraud detection. Furthermore, this algorithm is vulnerable to the problem of overfitting. To improve fraud detection, various alternative approaches and techniques have been explored. Applying a Bayesian network, researchers are exploring the effectiveness of a hidden Markov model for fraud detection alongside an artificial immune system and a support vector machine. They are also exploring incorporating a fuzzy neural network and fuzzy Darwinian system in this particular scenario [30]. Furthermore, they are investigating using a genetic algorithm and conducting a k-nearest neighbor analysis. These various models are being assessed by examining cost, speed, and accuracy, with comparisons being drawn. These strategies attempt to boost fraud detection in the financial industry, protecting against fraudulent activities by pursuing greater security and resilience against interference in wireless communications.

In terms of speed of processing, the Markov model has exhibited remarkable performance, while for accuracy, the optimal choice has been the fuzzy Darwinian method [31]. To address the classification challenge with high precision and effectively manage the variable costs associated with misclassification, they are developing an efficient algorithm combining multiple approaches. Despite its impressive performance, the fuzzy Darwinian approach suffers from drawbacks like high expenses and sluggish detection speed [19], [32]. Although the hidden Markov model exhibits swift processing speed, it faces challenges related to low accuracy, high costs, and difficulties in managing large datasets, rendering it non-scalable.

Author in study [34] addresses the crucial challenge of credit card default prediction in the financial sector. The proposed approach, utilizing a stacked sparse autoencoder (SSAE) for feature learning, shows promise in enhancing predictive performance. However, it's important to consider that this method may require significant computational resources and its performance can vary depending on the dataset. Despite these limitations, the SSAE-based approach outperforms traditional methods, suggesting its potential for improving risk management in credit card usage.

C. PROBLEM STATEMENT

Accurately identifying fraudulent transactions remains a persistent challenge in credit card fraud detection, exacerbated by several pressing issues. First, the prevalence of imbalanced datasets, stemming from the rarity of credit card fraud, presents a fundamental obstacle, affecting the efficacy of traditional classification models [13], [16]. Second, the dynamic nature of fraudulent activities results in concept drift, necessitating the development of models capable of adapting to evolving patterns in transaction data over time [26], [27]. Additionally, ensuring the interpretability of fraud detection models is paramount for financial institutions, compelling the need for models that can provide meaningful explanations for their decisions [29]. Scalability issues, driven by the increasing volume of transactions, underscore the demand for efficient algorithms and hardware solutions to ensure real-time fraud detection [27], [28], [29]. Lastly, the persistent threat of adversarial attacks highlights the importance of robust models capable of withstanding manipulation attempts [33], [35]. To tackle these multifaceted challenges, this research investigates the effectiveness of the Ensemble AutoEncoder with ResNet (EARN) model in dimensionality reduction while preserving essential features in credit card transaction data. Through this investigation, we aim to select the most relevant high-dimensional and low-dimensional features for improved handling of financial transactions. This selection enhances the training process of our proposed classification method, RXT-J.

III. PROPOSED SYSTEM MODEL

This section presents the conceptual framework designed to identify and flag fraudulent transactions within financial transactions. Addressing the critical need for robust and accurate fraud detection mechanisms, our model utilizes a combination of ensembler techniques to mitigate the challenges associated with credit card fraud. This comprehensive approach encompasses feature engineering, dimensionality reduction, and a novel classification method, ensuring the effective identification of fraudulent activities while maintaining operational efficiency. The proposed system model is designed to enhance fraud detection accuracy and address the evolving nature of fraudulent behavior and the interpretability requirements crucial for financial institutions. Figure 1 depicts the model being proposed along with the associated stages.

Initially, we have taken dataset features as input. We have applied the preprocessing step, including filling in missing values, normalization, etc. Given the class imbalance in the dataset, where fraudulent activity accounts for only 3.27% of all transactions, addressing this imbalance is crucial for robust model training. To mitigate class imbalance, we commence with data preprocessing using the Synthetic Minority Over-sampling Technique (SMOTE) algorithm. SMOTE is applied to balance the class distribution in the dataset, generating synthetic samples for minority classes and thus enhancing the overall dataset balance. After this critical preprocessing step, we proceed to the feature extraction phase, employing the Ensemble AutoEncoder with ResNet (EARN) model.

The EARN model plays a pivotal role in capturing both high-dimensional and low-dimensional features, providing a comprehensive approach to handling the varying degrees of fraudulent fluctuations in the data. During the feature extraction phase, we deploy multiple autoencoder and ResNet models, each configured with different architectural complexities and layer configurations. Within the EARN model, these autoencoders are trained unsupervised,

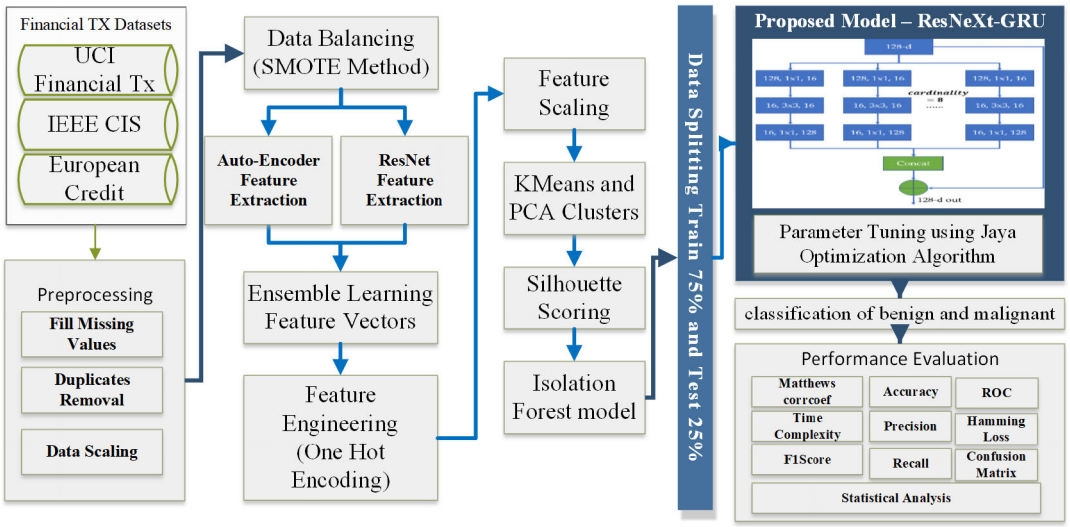

FIGURE 1. Proposed system model of fraud financial transaction detection.

focusing on reducing data dimensionality while capturing intrinsic features. Training them individually involves using reconstruction loss functions, such as mean squared error (MSE), which facilitate feature extraction from the encoder part of each autoencoder. This phase operates independently of class labels, concentrating solely on extracting latent features essential for subsequent classification. In the feature fusion / Ensembler Learner stage, we combine the features obtained from the autoencoders with those from the ResNet model, creating a hybrid feature representation.

Subsequently, a classification layer RXT-J model with appropriate layers and skip connections is implemented on top of this concatenated feature vector. This classifier learns to map the combined features to class labels, resulting in the model's classification ability. Furthermore, we employ the novel ensemble learning method to enhance the model's performance and robustness; EARN models are constructed, introducing variations in autoencoder architectures, ResNet configurations, and initialization seeds. This ensemble approach mitigates overfitting and enhances the model's generalization capabilities. Hyrbid RXT-J method is then applied to consolidate the predictions from the individual EARN models, yielding the final classification outcome.

A. DATASET DESCRIPTION

In our article, we have implemented our proposed model on three distinct datasets: the Paysim Financial credit dataset [36], the IEEE CIS Fraud Dataset [37], and the European transaction dataset [38]. The IEEE-CIS Fraud Detection dataset is frequently used in fraud detection research and performance assessment. Originally introduced as part of a Kaggle competition, this dataset primarily revolves around transaction-related data. Its primary aim is

TABLE 2. Summary of datasets for financial transaction fraud detection.

Dataset Name	Source	No. of Features	No. of Instances	Imbalanced Classes
IEEE-CIS Fraud Detection Dataset	Kaggle	433	590,000	Exhibits class imbalance, fewer fraudulent transactions
PaySim Synthetic Dataset	GitHub	11	600,000	More class imbalance
UCI Credit Card Dataset	UCI Repository	23	30,000	Exhibits some class imbalance

a binary classification task to distinguish between fraudulent transactions (class 1) and legitimate ones (class 0). The second dataset we utilized is the Paysim Dataset. The PaySim Synthetic dataset is a fabricated financial transaction dataset commonly used for research and experimentation in fraud detection and financial analytics. This dataset was specifically generated to mimic mobile money transactions and encompasses various features that capture different facets of these transactions. Lastly, our third dataset is the UCI Credit Card Dataset, sourced from the UCI repository. The details about the datasets are provided in Table 2.

B. PREPROCESSING

The initial and crucial phase in preparing datasets for analyzing and modeling financial fraud involves essential preprocessing steps. These steps entail handling missing data, removing duplicate records, and standardizing data scaling. Through these actions, we aim to cleanse and structure the dataset, establishing a robust basis for precise analysis and efficient modeling to identify possible cases of financial fraud.

Addressing Missing Data: Managing missing values entails handling the gaps in our dataset that arise due to

the absence or incompleteness of information. A widely employed technique for this purpose is mean imputation. In mean imputation, we compute the average of the known data points within a specific feature and utilize this calculated mean to substitute the missing values. The process can be summarized as in Equation 1 [39]:

$$X_{\text{imputed}} = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{1}$$

X_{imputed} signifies the values estimated to fill the gaps, X_i denotes the values initially observed, and N represents the total count of values not missing.

Eliminating Duplicate Records: Removing duplicate records involves ensuring that our dataset consists of distinct data points, preventing any bias caused by duplications, and maintaining the accuracy and reliability of our data. This process involves examining each record and keeping only the unique ones by comparing them to all other records.

$$D_{\text{unique}} = \{d_i \in D : \text{No identical record in } D \text{ matches } d_i\} \tag{2}$$

D_{unique} refers to the dataset with all duplicate records removed. ''di'' represents a single, distinct record within this dataset, and ''D'' represents the initial/original dataset that may contain duplicate records.

Data Scaling: Data scaling is a crucial step in data preprocessing that focuses on bringing all numerical features to a uniform and easily comparable scale. There are two common methods for achieving this: standardization and min-max scaling. Standardization is a process that transforms a particular feature, denoted as X, in a way that it possesses an average (or mean) value of 0 and a consistent spread represented by a standard deviation of 1. This transformation enables us to compare feature X effectively with other features within a dataset [39] as in Equation 3;

$$X_{\text{scaled}} = \frac{X - \mu}{\sigma} \tag{3}$$

where X_{scaled} represents the adjusted feature, X denotes the initial feature, μ stands for the average, and σ signifies the standard deviation. Meanwhile, min-max scaling typically alters a feature X to fit within a specified range [0, 1] [39].

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{4}$$

X_{scaled} represents a feature that has been normalized or rescaled based on its size, taking into account the original feature (X), its minimum value ($X_{\min}$), and its maximum value ($X_{\max}$).

C. DATA BALANCING USING THE SMOTE METHOD

In fraud detection within financial transactions, a common challenge arises due to the significant disparity in the number of legitimate transactions compared to fraudulent ones in our dataset. This class imbalance poses a unique hurdle when constructing effective fraud detection models.

While these models tend to perform well in identifying non-fraudulent transactions, they often struggle with accurately detecting the minority class, which represents instances of fraud. To address this issue, specialized techniques are required. One such technique that we discuss in this article is undersampling. However, instead of removing instances from the majority class, we employ a more advanced approach, the Synthetic Minority Over-sampling Technique (SMOTE) [40]. The primary objective of SMOTE is to generate synthetic data points for the minority class in a way that retains the underlying patterns and relationships within the data.

SMOTE accomplishes this by creating synthetic instances that lie between a minority class instance (referred to as m1) and its closest neighbors in the feature space (denoted as x01 and x02). To introduce an element of randomness and variability into the process, we incorporate a random value ranging from 0 to 1, denoted as random(0, 1). we take a random value and add it to the differences between the characteristics of the original data point and its closest neighbors. This creates new data points that help connect the underrepresented minority class with its neighboring data, making the minority class more prominent and better represented in the overall dataset as in Equation 5 [40].

$$(M1; M2) = (m1; m2) + \text{random}(0; 1) \cdot (x01 - x1; x02 - x2) \tag{5}$$

Utilizing Random(0, 1) generates a random value ranging from 0 to 1. We compute the disparity between the feature values of the given instance and those of its closest neighbors, expressed as (x01 - x1; x02 - x2). This procedure is reiterated several times to produce multiple fabricated instances for the underrepresented class. Implementing the SMOTE algorithm within our fraud detection framework yields an equitable distribution between the two categories: the quantity of authentic (non-fraudulent) transactions and that of deceitful transactions. This method effectively tackles the class imbalance issue, enhancing the model's capacity to identify fraudulent activities without compromising its performance in authentic transactions.

D. ENSEMBLE LEARNING OF FEATURES VECTORS

We explore the ensemble learning approach we use to leverage the benefits of feature extraction using autoencoders and ResNet models with balanced data. Combining these feature extraction techniques, we aim to strengthen our ensemble model's resilience and discriminative capabilities, enhancing its performance across different tasks.

Feature Extraction with Autoencoder: Autoencoders have demonstrated remarkable effectiveness in extracting valuable representations from input data, particularly when confronted with datasets characterized by a balanced distribution of samples. This unique ability equips them to uncover intricate patterns and subtle nuances that might remain obscured when dealing with imbalanced datasets. In our ensemble-based

approach, we leverage autoencoders to extract features from the well-balanced dataset. These autoencoders are trained to reconstruct the input data, with their central encoding layer, often termed the bottleneck layer, serving as a concise and informative representation of the input data.

In this particular context, we harness the capabilities of autoencoders to derive these meaningful features from the balanced dataset. The encoder function, denoted as E(input), is central in transforming the input data x into a condensed representation. This resultant encoder output, referred to as ensmb_auto, is obtained in the following manner:

$$\text{ensmb\_auto} = E(\text{input}) \tag{6}$$

Feature Extraction with ResNet: Integrating ResNet architectures into our approach significantly enhances our feature extraction strategy. ResNet models have established their prowess in capturing complex hierarchical features, and their effectiveness in various computer vision tasks is well-documented. When dealing with a well-balanced dataset, ResNet models shine in the task of distilling crucial high-level features that play a pivotal role in distinguishing between different classes.

ResNet models are essentially deep neural networks with skip connections, allowing them to capture and represent intricate data patterns efficiently. In our ensemble methodology, we use pre-trained ResNet models, fine-tuned on our balanced dataset, to extract discriminative features from the input data. These features are removed from the final convolutional or fully connected layers of the ResNet models, presenting us with an additional set of feature vectors for our ensemble.

In combination with our autoencoders, integrating ResNet architectures for feature extraction adds a valuable dimension to our approach. Denoted as 'F(x),' ResNet excels at capturing intricate patterns and hierarchically organized features. The feature vector obtained from ResNet, which we refer to as 'f_resnet,' is generated from the network's ultimate convolutional or fully connected layers.

$$f_{\text{resnet}} = F(x) \tag{7}$$

Ensembler Integration: Our approach integrates feature vectors from autoencoders and ResNet models into a unified representation. This combined representation effectively captures a comprehensive range of features drawn from a balanced dataset, spanning intricate details and high-level characteristics. This fusion process substantially bolsters the overall discriminative capability of our ensemble model. We use concatenation and weighted averaging to create the ensemble feature vector for each data point. In the case of concatenation, we straightforwardly merge the two feature vectors, forming a singular ensemble feature vector. These ensemble feature vectors are then employed as inputs for subsequent tasks, such as classification or anomaly detection, leveraging their enriched feature set.

$$\text{ensemble\_feature} = [e_{\text{auto}}, f_{\text{resnet}}] \tag{8}$$

When performing weighted averaging, we allocate specific weights (w_{auto} and w_{resnet}) to the feature vectors obtained from autoencoders and ResNet. This results in an ensemble feature vector calculated by blending the feature vectors using these assigned weights.

$$\text{ensemble\_feature} = w_{\text{auto}} \cdot e_{\text{auto}} + w_{\text{resnet}} \cdot f_{\text{resnet}} \tag{9}$$

This combination of feature vectors encompasses a comprehensive range of data, merging intricate specifics obtained through autoencoders with the broader, top-level characteristics ResNet identifies. This merging elevates the model's ability to distinguish patterns, rendering it highly effective for tasks that require analyzing well-balanced datasets.

E. FEATURE ENGINEERING AND FEATURE PROCESSING

After obtaining feature vectors from ensemble learning, enhancing their quality further and preparing them for downstream machine learning tasks is essential. This post-processing phase includes the following steps:

One Hot Encoding: One Hot Encoding handles categorical variables within the feature vectors. By converting these categories into binary vectors, we make them compatible with machine learning algorithms and ensure they contribute effectively to the model's performance.

PCA: PCA is a technique used to simplify data by reducing its complexity while preserving important patterns and relationships. It accomplishes this by transforming a data point 'X' into its principal components through a specific mathematical formula.

$$\text{PCA} = X \cdot V \tag{10}$$

When performing weighted averaging, we allocate specific weights (w_{auto} and w_{resnet}) to the feature vectors obtained from autoencoders and ResNet. This results in an ensemble feature vector calculated by blending the feature vectors using these assigned weights.

Silhouette Scoring: The Silhouette Score is a measure employed to assess the effectiveness of cluster formation in clustering analyses, like KMeans clustering. It quantifies the extent to which an item within a cluster resembles the other items in the same cluster (cohesion) in comparison to items in different clusters (separation) [41]. The Silhouette Score is on a scale from -1 to 1, where:

- A high positive silhouette score, close to 1, indicates that a data point fits its cluster well and is not very similar to data points in other clusters. It suggests that the clustering is effective.
- A silhouette score around 0 suggests a data point near the boundary between two neighboring clusters. In other words, it's not associated with one cluster or another.
- A silhouette score approaching -1 indicates the possibility that a data point has been assigned to an incorrect cluster, as it seems more similar to points in another cluster.

The silhouette score is useful for evaluating how well clustering algorithms organize data points. It aids in deciding

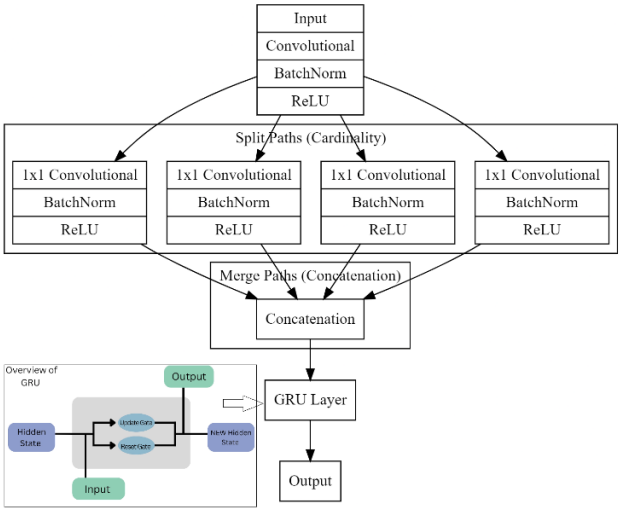

FIGURE 2. ReNeXt-GRU model.

the right number of clusters and gauges how effectively the algorithm groups similar data points together while maintaining separation from other clusters.

IFM: Anomaly detection is efficiently handled by the IFM, effectively identifying and isolating outliers within a dataset [42]. Unlike traditional methods that focus on modeling normal data points, the isolation forest takes a different approach. It isolates anomalies by building an ensemble of decision trees. Anomalies can be identified by needing only a small number of divisions in a tree-like structure to distinguish them from most data points. The fundamental concept of the isolation forest is that anomalies are infrequent and should stand out distinctly from most of the data with minimal effort. This model can be advantageous when dealing with datasets with imbalanced classes or focusing on detecting unusual or fraudulent behavior within a dataset. Trained on the data, the IFM assigns an anomaly score to each data point during the testing phase.

In fields such as cybersecurity, fraud detection, and quality control, data points that receive higher anomaly scores are typically regarded as more probable outliers or anomalies. This characteristic makes the model useful for identifying unusual or irregular data points.

F. CLASSIFICATION USING RXT-J

The ResNeXt-GRU model represents an advanced and versatile architecture designed to classify financial transaction data, particularly in detecting fraudulent activities. This model amalgamates the robust feature extraction capabilities of the ResNeXt architecture with the sequential learning and contextual understanding existing by GRUs. Figure 2 shows the ResNeXt GRU Model's internal structure.

Feature Extraction using ResNeXt: The classification begins with the initial input of raw financial transaction data. This data undergoes a series of transformations within the ResNeXt component. The modeling of this feature extraction stage presents For each path i within the ResNeXt block:

- Convolutional Layer: $Conv_i = \text{Convolutional Layer}_i(x)$
- Batch Normalization: $BN_i = \text{Batch Normalization}_i(Conv_i)$
- ReLU Activation: $ReLU_i = \text{ReLU}(BN_i)$

x denotes the input financial transaction data, and $Conv_i$, BN_i, and $ReLU_i$ represent the output of the convolutional layer, the result after batch normalization, and the activation function (ReLU) applied to the batch-normalized output for path i, respectively.

Path Cardinality and Concatenation: Strategically leveraging cardinality-based segmentation, the data is divided into multiple paths to augment the model's capacity for capturing a wide range of features. Independently, each path processes the data through the ResNeXt architecture. These path outputs, now enriched with distinct information, are concatenated to form a comprehensive feature representation:

$$\text{Concatenation} = [RFU1, RFU2, RFU3, RFU4] \quad (11)$$

The Concatenation operation combines the outputs ($ReLU_i$) from all paths (in this example, four paths) into a unified representation.

1) SEQUENTIAL MODELING WITH GRU

The concatenated features are channeled through a GRU layer. This step introduces a temporal modeling aspect, allowing the model to understand the sequential dependencies and evolving patterns within the financial transactions. The mathematical formulation of GRU operations is as follows [43]:

$$r = \sigma(W_r \cdot [\text{Concatenation}, h_{t-1}] + b_r) \quad (12)$$

$$z = \sigma(W_z \cdot [\text{Concatenation}, h_{t-1}] + b_z) \quad (13)$$

$$\tilde{h}_t = \tanh(W \cdot [\text{Concatenation}, r \cdot h_{t-1}] + b) \quad (14)$$

$$h_t = (1 - z) \cdot h_{t-1} + z \cdot \tilde{h}_t \quad (15)$$

These equations describe how the GRU layer functions. In this context, 'ht' stands for the hidden state at a specific time 't,' and we use different gates, denoted as 'r' and 'z,' to manage how information flows and the gating mechanisms operate.

2) TUNNING WITH JA

Hyperparameters are parameters not learned during training but set before training. They profoundly impact the model's performance, convergence, and generalization. In the context of our ResNeXt-GRU model, the following hyperparameters are obtained: The Jaya Algorithm is a population-based optimization technique inspired by improvement and collaboration within a population [44]. Its objective is to iteratively explore and update hyperparameter values to identify the optimal combination. The stages described in Algorithm 1 are followed during the optimization process, which is guided by an objective function that evaluates the efficacy of the model on a validation data.

TABLE 3. Hyperparameters and obtained values.

Hyperparameter	Obtained Value
Batch Size	32
GRU Hidden Units	128
ResNeXt Block Config	4 blocks, depth 32, width 4
Dropout Rate	0.3
Learning Rate	0.001
Weight Decay	0.0001
Epochs	50
Jaya Algorithm Parameters	
Population Size	10
Convergence Threshold	1e-5
Exploration Range	[0.1, 0.9]

Algorithm 1 Hyperparameter optimization using the jaya algorithm

Input:
Population P of hyperparameter sets
Objective function f(solution)
Exploration range $[L, U]$ for each hyperparameter
Convergence threshold ϵ
Output: Optimized hyperparameters
procedure JayaOptimization
    Initialize population P with random hyperparameter sets
    Initialize the best solution x_{best} with a random solution from P
    while Not converged **do**
        for Each solution x_i in P **do**
            Random number generation r distributed uniformly in the interval [0, 1].
            Update the solution:
                $x_i = x_i + r \cdot (x_{\text{best}} - x_i)$
            Ensure solutions stay within the exploration range:
                $x_i = \min(U, \max(L, x_i))$
        end for
        Select the solution with the best objective function value as x_{best}
    end while
end procedure

IV. SIMULATION RESULTS

In this section, within the Google Colab environment, we employed TensorFlow, taking advantage of its robust GPU capabilities to bolster the efficiency of our financial transaction fraud detection system. We have applied our proposed model to three financial transaction datasets. The results are discussed below.

A. RESULTS AND DISCUSSION ON THE IEEECIS DATASET

The dataset is first analyzed to check the frequency of benign and malignant transactions. Figure 3 shows that the fraudulent transactions are much less than non-fraudulent transactions. SMOTE algorithm is applied for data balancing. It balances the frequency of benign and malignant transactions by creating synthetic data, as shown in Figure 4.

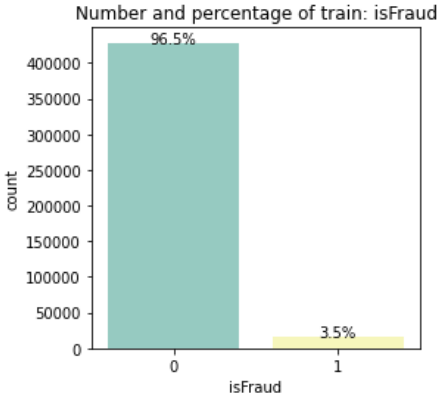

FIGURE 3. Target variable distribution in IEEE-CIS dataset.

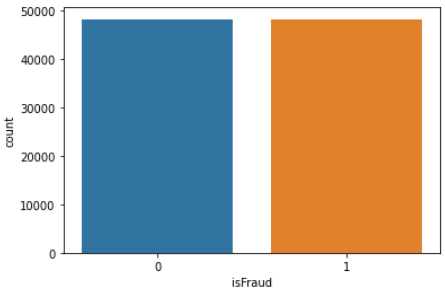

FIGURE 4. Target variable after applying SMOTE algorithm.

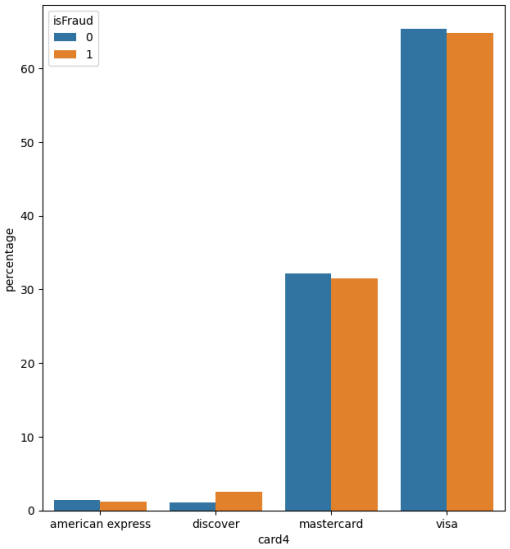

FIGURE 5. Card-4 variable data analysis.

After the data balancing, we performed some EDA analysis. As seen in Figure 5, Card-4 is a categorical characteristic; it is the card company through which the transaction was made. Visa and MasterCard are widely used, although American Express and Discover are far less common. This feature has significantly less separating power in this instance.

Another distinct characteristic is Card-6, which is the card type used for transactions. As anticipated, there are more

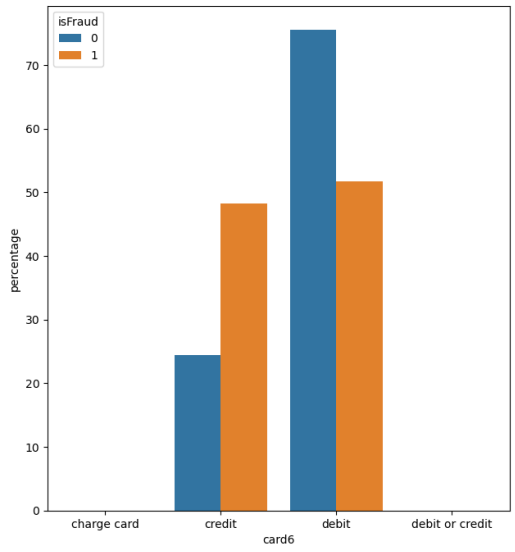

FIGURE 6. Card-6 variable data analysis.

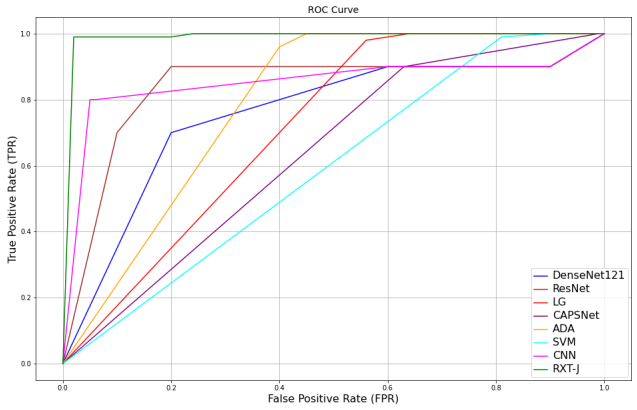

FIGURE 7. ROC curve of the proposed method and existing methods on the IEEECIS dataset.

credit card users than any other type, and a significant portion of users utilize either credit or debit charge cards.

Then, we moved on to the categorization stage. Our methodology began with the use of the JA to adjust weights and hyperparameters in the ResNeXt-GRU ensemble. The ensembler was used to run the optimal weights, and the model was trained and assessed for every combination of weights. The use of optimization techniques was essential for managing large amounts of malware data. The best set of parameters was found by JA via extensive testing, yielding 0.6742 and 0.4121 for WR and WG. These weights indicate which option best enhances our malware detection abilities.

We used the optimization technique JA to send the hyperparameters through ResNeXt-GA and evaluate the model's performance. Furthermore, we assessed our model by calculating its True Positive as well as True Negative values, which are displayed for the suggested and current approaches in Figure 7.

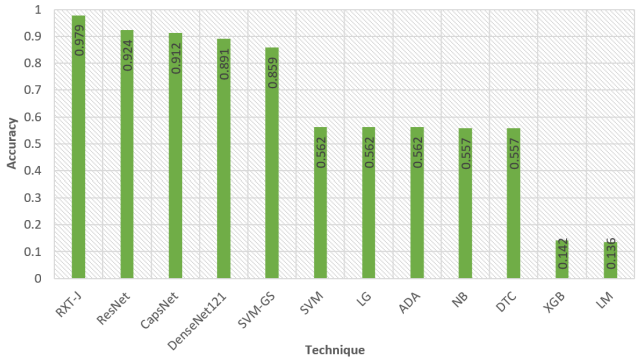

FIGURE 8. Accuracy of proposed and existing (IEEECIS dataset).

TABLE 4. Performance comparison of proposed RXT-J, BERT (transformer), and existing models.

Techniques	F1-Score	Accuracy	Precision	Recall	MCC	Log Loss	AUC	ROC
LM [13]	0.562	0.136	0.562	0.562	0.312	1.609	0.632	0.573
XGB [19]	0.660	0.142	0.506	0.930	0.305	1.042	0.769	0.697
DTC [19]	0.660	0.557	0.506	0.930	0.305	1.042	0.769	0.697
NB [19]	0.660	0.557	0.506	0.929	0.305	1.042	0.769	0.697
SVM [16]	0.663	0.562	0.509	0.928	0.312	0.981	0.785	0.709
LG [16]	0.663	0.562	0.509	0.929	0.313	0.977	0.788	0.712
ADA [27]	0.663	0.562	0.509	0.929	0.313	0.978	0.787	0.711
SVM-GS	0.867	0.859	0.817	0.863	0.794	0.214	0.943	0.907
DenseNet121 [12]	0.898	0.891	0.849	0.894	0.825	0.256	0.962	0.896
CapsNet [13]	0.920	0.912	0.870	0.915	0.847	0.200	0.973	0.920
ResNet	0.942	0.924	0.912	0.923	0.939	0.123	0.980	0.941
RXT-J (Proposed)	0.987	0.979	0.977	0.993	0.984	0.046	0.997	0.991
BERT (Transformer)	0.916	0.912	0.870	0.918	0.846	0.199	0.972	0.918

Figure 8 presents the accuracy values of various techniques applied to financial fraud detection using the IEEECIS dataset. Accuracy measures how well each method correctly identifies instances of fraud, with higher values indicating better performance.

Table 4 provides a compelling performance evaluation of the Proposed RXT-J model compared to a range of Established Models using the IEEECIS financial transaction dataset. RXT-J's standout performance is noteworthy, achieving an impressive 98% accuracy rate. This exceptional accuracy underscores RXT-J's remarkable ability to effectively identify fraudulent transactions, firmly establishing it as a highly proficient solution for detecting financial fraud within the dataset. In contrast to the existing models assessed, RXT-J's superior performance represents a significant leap forward in financial transaction security, offering a reliable and robust tool for fortifying fraud prevention systems and elevating the overall precision and dependability of financial transaction monitoring.

B. RESULTS AND DISCUSSION ON PAYSIM SYNTHETIC DATASET

Figure 9 illustrates the dispersion of fraudulent and normal transactions over time, revealing distinct patterns.

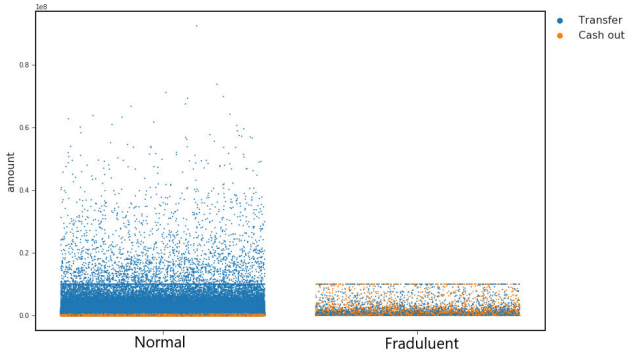

FIGURE 9. Dispersion of fraudulent and normal transactions over time.

FIGURE 10. Fraudulent and genuine transactions by analyzing their correlations using Heatmap.

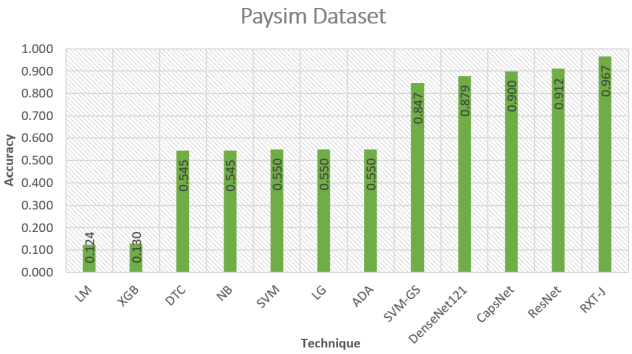

FIGURE 11. Performance of existing methods alongside the newly proposed RXT-J model.

TABLE 5. Performance evaluation results on the paysim dataset.

Techniques	F1-Score	Accuracy	Precision	Recall	MCC	Log Loss	AUC	ROC
LM [13]	0.550	0.124	0.550	0.550	0.311	1.621	0.620	0.561
XGB [14]	0.648	0.130	0.494	0.918	0.304	1.054	0.757	0.685
DTC [19]	0.648	0.545	0.494	0.918	0.304	1.054	0.757	0.685
NB [19]	0.648	0.545	0.494	0.917	0.304	1.054	0.757	0.685
SVM [16]	0.651	0.550	0.497	0.916	0.311	0.993	0.773	0.697
LG [16]	0.651	0.550	0.497	0.917	0.312	0.989	0.776	0.700
ADA [27]	0.651	0.550	0.497	0.917	0.312	0.990	0.775	0.699
SVM-GS	0.855	0.847	0.805	0.851	0.793	0.226	0.931	0.895
DenseNet121 [12]	0.886	0.879	0.837	0.882	0.824	0.268	0.950	0.884
CapsNet [5]	0.908	0.900	0.858	0.903	0.846	0.212	0.961	0.908
ResNet	0.930	0.912	0.900	0.911	0.938	0.135	0.968	0.929
RXT-J (Proposed)	0.975	0.967	0.965	0.981	0.983	0.058	0.985	0.979
BERT (Transformer)	0.904	0.896	0.896	0.912	0.919	0.079	0.92295	0.971

Fraudulent transactions exhibit a more uniform distribution across time compared to genuine transactions. A notable difference is observed in the transaction types: normal transactions predominantly involve CASH-OUTs, whereas fraudulent transactions display a more balanced distribution between CASH-OUTs and TRANSFERs. It's important to mention that the width of each 'fingerprint' is determined by the 'jitter' parameter in the plotStrip function, aimed at visually separating and plotting transactions that occur simultaneously but with different time points.

Figure 10 provides an insightful comparison between fraudulent and genuine transactions by analyzing their correlations using heatmaps. These heatmaps visually represent how these two types of transactions differ.

Figure 11 provides a comprehensive overview of the performance of existing methods alongside the newly proposed RXT-J model in accurately distinguishing between normal and fraudulent transactions. Notably, the RXT-J model emerges as the standout performer, exhibiting exceptional accuracy in classifying both types of transactions. It signifies a substantial advancement in fraud detection, as the model excels in precisely identifying fraudulent activities while minimizing false positives, reaffirming its effectiveness and reliability in combating fraudulent transactions. Table 5 presents a comprehensive performance evaluation of our model on the paysim dataset, highlighting its exceptional capabilities in accurately classifying normal and fraudulent transactions. Our model surpasses existing literature models in critical metrics such as AUC, ROC, loss, and overall evaluation scores. These results underscore the model's remarkable effectiveness and potential to enhance financial fraud detection efforts within the financial sector significantly. Our model's superior performance reaffirms its status as a robust and cutting-edge solution, poised to substantially impact combating fraudulent activities and improving the security of financial transactions.

Figure 12 provides a detailed insight into classifying normal and fraudulent financial transactions using the ROC Curve, as represented through a confusion matrix. This matrix helps us distinguish true positive instances, where the model correctly identifies fraudulent transactions, from false positive instances, where it incorrectly labels a normal transaction as fraudulent. The RXT-J model stands out by achieving notably superior accuracy in distinguishing between legitimate and fraudulent financial transactions when compared to other cutting-edge approaches. These

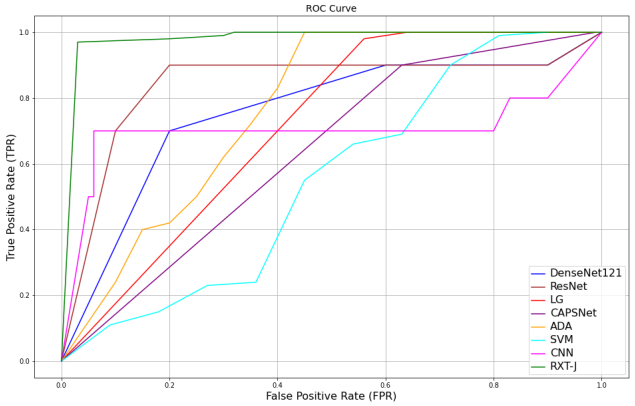

FIGURE 12. ROC curve of proposed RXT-J model and existing methods.

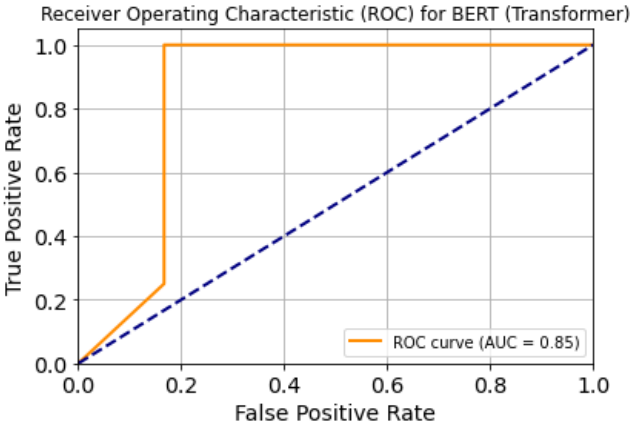

FIGURE 13. ROC curve of BERT (transformer).

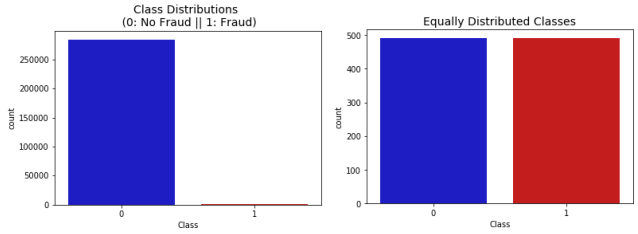

FIGURE 14. Before SMOTE balancing and after applying data balancing on UCI dataset.

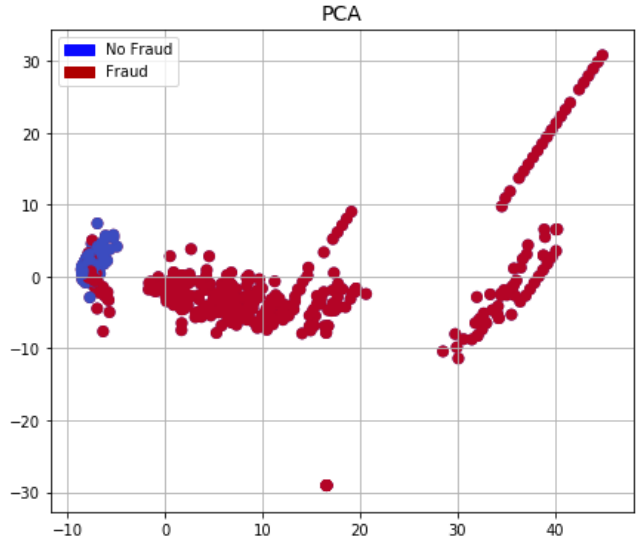

FIGURE 15. Essential data transformation steps i.e., PCA and clustering.

findings underscore the robustness and efficacy of the RXT-J model, which excels in accurately identifying fraudulent transactions while minimizing false alarms, making it a noteworthy advancement in financial fraud detection. Figure 13 illustrates the ROC Curve for the BERT (Transformer) model. This curve showcases the model's effectiveness in distinguishing between genuine and fraudulent transactions as the classification threshold varies. The high AUC of approximately 0.88 indicates that the BERT (Transformer) model excels at correctly identifying fraudulent transactions while minimizing false alarms, making it a reliable and effective tool for detecting online payment fraud.

C. RESULTS DISCUSSION ON UCI CREDIT CARD TRANSACTIONS DATASET

Figure 14, especially in the left graph, vividly illustrates the fundamental disparity within the dataset, showcasing a significant discrepancy between the number of legitimate transactions and the considerably smaller number of fraudulent transactions. This initial dataset distribution presents a challenge for our learning models, as they may become biased towards the majority class (normal transactions) and struggle to detect the minority class (fraudulent transactions). However, a significant transformation occurs after applying the SMOTE, as showcased in the right graph of Figure 13. SMOTE addresses the problem of imbalanced data by creating artificial examples of the minority class, ensuring a more even distribution between regular and fraudulent transactions. This balanced dataset is essential for developing reliable and equitable machine learning models that can effectively detect fraudulent activities accurately and precisely in practical financial fraud detection situations.

The UCI dataset necessitates additional dimensionality reduction through PCA for optimal analysis. Figure 15 comprehensively illustrates these essential data transformation steps, encapsulating class clustering and dimensionality reduction within a single framework. This integrated approach harmoniously combines class clustering, which reveals underlying data patterns and relationships, and dimensionality reduction through PCA, streamlining the dataset and enhancing manageability. This preparation is crucial for efficient analysis, particularly with high-dimensional financial data, as it simplifies the dataset, prevents overfitting, and empowers data analysts with a more refined and informative dataset.

TABLE 6. Performance evaluation results on the european dataset.

Techniques	F1-Score	Accuracy	Precision	Recall	MCC	Log Loss	AUC	ROC
LM [13]	0.564	0.138	0.564	0.564	0.325	1.601	0.622	0.563
XGB [19]	0.652	0.144	0.508	0.932	0.318	1.034	0.759	0.687
DTC [19]	0.692	0.559	0.508	0.932	0.318	1.034	0.759	0.687
NB [19]	0.762	0.559	0.508	0.931	0.318	1.034	0.759	0.687
SVM [16]	0.765	0.564	0.511	0.930	0.325	0.973	0.775	0.699
LG [16]	0.625	0.564	0.511	0.931	0.326	0.969	0.778	0.702
ADA [27]	0.795	0.564	0.511	0.931	0.326	0.970	0.777	0.701
SVM-GS	0.869	0.861	0.819	0.865	0.807	0.206	0.933	0.897
DenseNet121 [12]	0.900	0.893	0.851	0.896	0.838	0.248	0.952	0.886
CapsNet [5]	0.922	0.914	0.872	0.917	0.860	0.192	0.963	0.910
ResNet	0.944	0.926	0.914	0.925	0.952	0.115	0.970	0.931
RXT-J (Proposed)	0.989	0.981	0.979	0.995	0.997	0.038	0.987	0.981
BERT (Transformer)	0.9098	0.8857	0.8716	0.9190	0.9468	0.0519	0.9724	0.9687

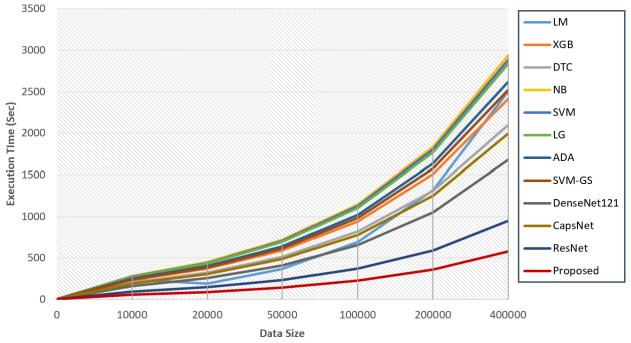

FIGURE 16. Computational complexity of proposed and existing models.

Table 6 illustrates the results of our performance assessment for European card transactions, and it is evident that our approach outperforms others on this dataset.

Additionally, we conducted simulations to assess the computational complexity of our model across all datasets, and the results are presented in Figure 16, which displays the average execution times. Our proposed model, the RXT-J, mainly exhibits the shortest execution time, demonstrating its efficiency in swiftly handling both high-dimensional and low-dimensional data and transactions. This enhanced processing speed optimizes the model's performance and significantly improves its ability to detect fraudulent users and corresponding transactions promptly, bolstering its effectiveness in real-time fraud detection scenarios.

Impact of SMOTE on Feature Selection and Data Balance: In Table 7, we conducted hypothesis testing to compare different feature selection methods using both balanced data (SMOTE) and unbalanced data. The results reveal a significant disparity between the two scenarios, highlighting the substantial impact of data balancing. Before applying SMOTE, the number of relevant features selected and the overall data balance differed significantly from the post-SMOTE situation. This underscores the crucial role of SMOTE in not only addressing data imbalance but also influencing the effectiveness of feature selection methods.

TABLE 7. Performance evaluation with different feature selections using PCA. (SF: feature sample).

Feature Selection using PCA		SF 1	SF	SF 3	SF 4	SF 5
	Methods	19	21	35	38	41
Accuracy with SMOTE	Proposed	98	97.1	94.223	88	83
	ResNet	91.2	87.2	82.2	82.21	82.1
	DenseNet121	87.9	83.9	78.9	75.2	75
	BERT	90.6	88.5	88.5	86.4	86.5
	SVM-GS	80.5	76.5	71.5	71.22	70.46
	ADA	55	51	54	53	52
Accuracy without SMOTE	Proposed	83	82.1	79.223	73	68
	ResNet	76.2	72.2	67.2	67.21	67.1
	DenseNet121	72.9	68.9	63.9	60.2	60
	BERT	75.6	73.5	73.5	71.4	71.5
	SVM-GS	65.5	61.5	56.5	56.22	55.46
	ADA	40	36	39	38	37

V. LIMITATIONS

It is important to identify some major constraints that impact the interpretation and generalization of our findings while seeking significant insights:

1) Data Imbalance: Despite our precise attempts to address data imbalance using approaches such as SMOTE, the risk of overfitting to the minority class remains, possibly adding unpredictability into the reliability of our conclusions.
2) Feature Selection: While Principal Component Analysis (PCA) is a good choice for feature selection, its linear nature may result in the unintended elimination of important characteristics. This linear feature is a restriction that might have an impact on the model's overall performance.
3) Generalizability: The extent to which our findings may be extended to various datasets or contexts is relatively limited. Our study' bases are founded on the particular properties of the IEEE CIS dataset, restricting its broader applicability.
4) Model Overfitting: Despite our strict procedures to prevent overfitting, there is still the possibility of model overfitting, particularly in real-world applications.

This alternate phrase keeps the substance of the constraints while presenting them in a slightly different way.

VI. CONCLUSION

Our study has taken a significant step forward in financial fraud detection, an ever-evolving challenge with profound outcomes for the financial sector. Despite the continuous advancement of technology, the complexities of financial fraud persist. In response to this pressing issue, we introduce an innovative financial fraud detection model based on RXT-J, meticulously designed for real-time transaction data analysis. Notably, our model consistently shows exceptional capabilities in handling the complexities of modern financial fraud, even when confronted with

extensive datasets. A key highlight is the model's ability to outperform existing solutions, substantially improving detection accuracy and swiftly identifying complex, previously unrecognized fraudulent patterns. Furthermore, our model effectively addresses the inherent inefficiencies of traditional approaches. We have conducted a comprehensive performance evaluation, comparing our model with conventional machine learning methods and other deep learning techniques using real-time financial transaction fraud data. While our model has exhibited remarkable potential, future research endeavors may further enhance its capabilities by incorporating additional features, such as fraud location and timing analysis, as relevant data becomes available. This research represents a substantial leap forward in the ongoing battle against financial fraud, promising heightened security and efficiency in financial transactions. In the broader context of wireless communications defense, where innovative algorithms bolster security, data availability, and resilience against interference, our work plays a pivotal role in securing financial transactions against the threat of fraud.

CONFLICT OF INTEREST

The authors declares no conflict of interest.

REFERENCES

[1] M. H. U. Sharif and M. A. Mohammed, "A literature review of financial losses statistics for cyber security and future trend," *IEEE Access*, vol. 15, pp. 138–156, 2022.
[2] Y. Bao, G. Hilary, and B. Ke, "Artificial intelligence and fraud detection," in *Innovative Technology at the Interface of Finance and Operations* (Springer Series in Supply Chain Management, Forthcoming), vol. 1. Springer, 2022, pp. 223–247. [Online]. Available: https://ssrn.com/abstract=3738618
[3] K. G. Al-Hashedi and P. Magalingam, "Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019," *IEEE Access*, vol. 40, 2021, Art. no. 100402.
[4] F. Y. Osisanwo, J. E. T. Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi, and J. Akinjobi, "Supervised machine learning algorithms: Classification and comparison," *Int. J. Comput. Trends Technol. (IJCTT)*, vol. 48, no. 3, pp. 128–138, 2017.
[5] P. R. Vlachas, J. Pathak, B. R. Hunt, T. P. Sapsis, M. Girvan,E. Ott, and P. Koumoutsakos, "Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics," *Neural Netw.*, vol. 126, pp. 191–217, Jun. 2020.
[6] S. Thudumu, P. Branch, J. Jin, and J. Singh, "A comprehensive survey of anomaly detection techniques for high dimensional big data," *J. Big Data*, vol. 7, no. 1, pp. 1–30, Dec. 2020.
[7] A. Cherif, A. Badhib, H. Ammar, S. Alshehri, M. Kalkatawi, and A. Imine, "Credit card fraud detection in the era of disruptive technologies: A systematic review," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 35, no. 1, pp. 145–174, Jan. 2023.
[8] J. Forough and S. Momtazi, "Ensemble of deep sequential models for credit card fraud detection," *Appl. Soft Comput.*, vol. 99, Feb. 2021, Art. no. 106883.
[9] Y. Lucas, P.-E. Portier, L. Laporte, L. He-Guelton, O. Caelen, M. Granitzer, and S. Calabretto, "Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs," *Future Gener. Comput. Syst.*, vol. 102, pp. 393–402, Jan. 2020.
[10] G. Douzas and F. Bacao, "Effective data generation for imbalanced learning using conditional generative adversarial networks," *Expert Syst. Appl.*, vol. 91, pp. 464–471, Jan. 2018.
[11] S. Bagga, A. Goyal, N. Gupta, and A. Goyal, "Credit card fraud detection using pipeling and ensemble learning," *Proc. Comput. Sci.*, vol. 173, pp. 104–112, Jan. 2020.
[12] H. Fanai and H. Abbasimehr, "A novel combined approach based on deep autoencoder and deep classifiers for credit card fraud detection," *Expert Syst. Appl.*, vol. 217, May 2023, Art. no. 119562.
[13] I. Benchaji, S. Douzi, B. El Ouahidi, and J. Jaafari, "Enhanced credit card fraud detection based on attention mechanism and LSTM deep model," *J. Big Data*, vol. 8, no. 1, pp. 1–21, Dec. 2021.
[14] P. C. Y. Cheah, Y. Yang, and B. G. Lee, "Enhancing financial fraud detection through addressing class imbalance using hybrid SMOTE-GAN techniques," *Int. J. Financial Stud.*, vol. 11, no. 3, p. 110, Sep. 2023.
[15] S. Nami and M. Shajari, "Cost-sensitive payment card fraud detection based on dynamic random forest and k-nearest neighbors," *Expert Syst. Appl.*, vol. 110, pp. 381–392, Nov. 2018.
[16] J. Chung and K. Lee, "Credit card fraud detection: An improved strategy for high recall using KNN, LDA, and linear regression," *Sensors*, vol. 23, no. 18, p. 7788, Sep. 2023.
[17] R. Saia and S. Carta, "Evaluating the benefits of using proactive transformed-domain-based techniques in fraud detection tasks," *Future Gener. Comput. Syst.*, vol. 93, pp. 18–32, Apr. 2019.
[18] R. Laurens, J. Jusak, and C. C. Zou, "Invariant diversity as a proactive fraud detection mechanism for online merchants," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2017, pp. 1–6.
[19] L. Zheng, G. Liu, C. Yan, C. Jiang, M. Zhou, and M. Li, "Improved TrAdaBoost and its application to transaction fraud detection," *IEEE Trans. Computat. Social Syst.*, vol. 7, no. 5, pp. 1304–1316, Oct. 2020.
[20] M. E. Islam, T. Tasnim, M. Y. Arafat, and A. Z. S. B. Habib, "Credit card fraud detection techniques: A comparative analysis," in *Proc. 25th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2022, pp. 716–721.
[21] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba, and G. Obaido, "A neural network ensemble with feature engineering for improved credit card fraud detection," *IEEE Access*, vol. 10, pp. 16400–16407, 2022.
[22] Y. Chen, N. Ashizawa, C. K. Yeo, N. Yanai, and S. Yean, "Multi-scale self-organizing map assisted deep autoencoding Gaussian mixture model for unsupervised intrusion detection," *Knowl.-Based Syst.*, vol. 224, Jul. 2021, Art. no. 107086.
[23] R. Jain, B. Gour, and S. Dubey, "A hybrid approach for credit card fraud detection using rough set and decision tree technique," *Int. J. Comput. Appl.*, vol. 139, no. 10, pp. 1–6, Apr. 2016.
[24] A. Singh and A. Jain, "Cost-sensitive metaheuristic technique for credit card fraud detection," *J. Inf. Optim. Sci.*, vol. 41, no. 6, pp. 1319–1331, Aug. 2020.
[25] A. Ali, B. A. S. Al-rimy, A. A. Almazroi, F. S. Alsubaei, A. A. Almazroi, and F. Saeed, "Securing secrets in cyber-physical systems: A cutting-edge privacy approach with consortium blockchain," *Sensors*, vol. 23, no. 16, p. 7162, Aug. 2023.
[26] Y. Xie, G. Liu, C. Yan, C. Jiang, M. Zhou, and M. Li, "Learning transactional behavioral representations for credit card fraud detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 1, no. 1, pp. 1–14, Oct. 2022.
[27] S. V. Suryanarayana, G. N. Balaji, and G. V. Rao, "Machine learning approaches for credit card fraud detection," *Int. J. Eng. Technol.*, vol. 7, no. 2, p. 917, Jun. 2018.
[28] A. A. Almazroi, F. Mohammed, M. A. Qureshi, A. A. Shah, I. A. T. Hashim, N. H. Al-Kumaim, and A. Zakari, "A hybrid algorithm for pattern matching: An integration of Berry–Ravindran and Raita algorithms," in *Proc. Int. Conf. Reliable Inf. Commun. Technol.* Cham, Switzerland: Springer, Dec. 2021, pp. 160–172.
[29] T. T. Nguyen, H. Tahir, M. Abdelrazek, and A. Babar, "Deep learning methods for credit card fraud detection," 2020, *arXiv:2012.03754*.
[30] V. Chang, L. M. T. Doan, A. Di Stefano, Z. Sun, and G. Fortino, "Digital payment fraud detection methods in digital ages and Industry 4.0," *Comput. Electr. Eng.*, vol. 100, May 2022, Art. no. 107734.
[31] Y. Alghofaili, A. Albattah, and M. A. Rassam, "A financial fraud detection model based on LSTM deep learning technique," *J. Appl. Secur. Res.*, vol. 15, no. 4, pp. 498–516, Oct. 2020.
[32] M. A. Gill, M. Quresh, A. Rasool, and M. M. Hassan, "Detection of credit card fraud through machine learning in banking industry," *J. Comput. Biomed. Informat.*, vol. 5, no. 1, pp. 273–282, 2023.

[33] A. A. Almazroi, "An empirical investigation of factors influencing the adoption of Internet of Things services by end-users," *Arabian J. Sci. Eng.*, vol. 48, no. 2, pp. 1641–1659, Feb. 2023.
[34] S. A. Ebiaredoh-Mienye, E. Esenogho, and T. G. Swart, "Artificial neural network technique for improving prediction of credit card default: A stacked sparse autoencoder approach," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 11, no. 5, p. 4392, Oct. 2021.
[35] B. Stojanović and J. Božić, "Robust financial fraud alerting system based in the cloud environment," *Sensors*, vol. 22, no. 23, p. 9461, Dec. 2022.
[36] K. Sengupta and P. K. Das, "Detection of financial fraud: Comparisons of some tree-based machine learning approaches," *J. Data, Inf. Manage.*, vol. 5, nos. 1–2, pp. 23–37, Jun. 2023.
[37] *IEEE-CIS Dataset*. Accessed: Sep. 22, 2023. [Online]. Available: www.kaggle.com/competitions/ieee-fraud-detection/data
[38] *European Transaction Dataset*. Accessed: Sep. 21, 2023. [Online]. Available: www.kaggle.com/datasets/mlg-ulb/creditcardfraud
[39] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez, and F. Herrera, "Big data preprocessing: Methods and prospects," *Big Data Anal.*, vol. 1, no. 1, pp. 1–22, Dec. 2016.
[40] M. Mukherjee and M. Khushi, "SMOTE-ENC: A novel SMOTE-based method to generate synthetic data for nominal and continuous features," *Appl. Syst. Innov.*, vol. 4, no. 1, p. 18, Mar. 2021.
[41] K. R. Shahapure and C. Nicholas, "Cluster quality analysis using silhouette score," in *Proc. IEEE 7th Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2020, pp. 747–748.
[42] W. Wu and Y. Chen, "Application of isolation forest to extract multivariate anomalies from geochemical exploration data," *Global Geol.*, vol. 21, no. 1, pp. 36–47, 2018.
[43] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2017, pp. 1597–1600.
[44] R. V. Rao, "Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems," *Int. J. Ind. Eng. Comput.*, vol. 7, no. 1, pp. 19–34, 2016.

ABDULWAHAB ALI ALMAZROI received the M.Sc. degree in computer science from the University of Science, Malaysia, and the Ph.D. degree in computer science from Flinders University, Australia. He is currently an Associate Professor with the Department of Information Technology, College of Computing and Information Technology at Khulais, University of Jeddah, Saudi Arabia. His research interests include parallel computing, cloud computing, wireless communication, and data mining.

NASIR AYUB (Student Member, IEEE) received the M.S. degree in smart grid and data science from COMSATS University Islamabad, Pakistan. He is currently pursuing the Ph.D. degree in computer science with the School of Electrical Engineering and Computer Science (SEECS), National University of Science and Technology (NUST), Islamabad, Pakistan. He is also a Lecturer with the Department of Creative Technologies, Air University, Islamabad. Previously, he was a Lecturer and a Senior Lecturer with FUUAST Islamabad and CUST University, Islamabad, respectively. His research interests include smart grids, machine learning, deep learning, natural language processing, and blockchain. He actively contributes to the academic community as a Reviewer of IEEE Access, Elsevier, MDPI, and peer journals.