Journal of Theoretical and Applied Information Technology
15th April 2026. Vol.104. No.7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

STTN-CP: A SPATIAL-TEMPORAL TRANSFORMER WITH CONTRASTIVE PRETRAINING MODEL FOR CREDIT CARD FRAUD DETECTION

KATHIRESAN JAYABALAN[1] , SETHURAMAN RADHAKRISHNAN[2]

[1]Research Scholar, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, 600119, Tamil Nadu, India.

[2]Associate Professor, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, 600119, Tamil Nadu, India.

E-mail: [1]kathiresan.jayabalan@gmail.com, [2]sethuraman.cse@sathyabama.ac.in

ABSTRACT

Credit card fraud detection remains a critical challenge due to highly imbalanced datasets, evolving fraud patterns, and complex transaction behaviors, leading to significant financial losses and reduced user trust. A novel Spatial-Temporal Transformer Network with Contrastive Pretraining (STTN-CP) is proposed to effectively detect fraudulent transactions. The STTN-CP model is designed to capture temporal dependencies throughout the entire transaction sequence and to identify spatial correlations among the transaction features. This study introduces a unified framework that integrates spatial–temporal transformer modeling with contrastive representation learning to enhance feature discrimination and detection accuracy. A two-step methodology is employed to tackle the class imbalance issue commonly found in credit card datasets. Initially, Min-Max normalization is conducted, followed by the application of the Synthetic Minority Oversampling Technique (SMOTE) in the data preprocessing phase. Subsequently, the spatial-temporal transformer blocks collaboratively generate hierarchical embeddings, while the contrastive pretraining module enhances feature discrimination by clustering analogous transactions and separating dissimilar ones inside the feature space. The proposed framework is evaluated using the Credit Card Fraud Detection dataset that is publicly available from Kaggle, and achieves an accuracy of 99.12%, precision 99.00%, recall 98.86%, F1-score 98.92%, and specificity 97.96%. The ablation studies show the importance of spatio-temporal modeling, contrastive pretraining, and data preprocessing with respect to high detection performance. The proposed model provides a scalable and robust solution for real-time fraud detection, improving financial security and decision-making reliability. Future work will focus on multi-class fraud detection, lightweight architectures for real-time deployment, and explainable AI integration.

Keywords: *Credit Card Fraud Detection; Deep Learning; Spatial–Temporal Transformer; Contrastive Pretraining; SMOTE*

1. INTRODUCTION

The rapid transition to a cashless society has rendered credit cards a main contributor of digital commerce, which has been valued for its easy use, efficiency, and general acceptance. However, on the other hand, the extensive and wide use of credit cards has brought about a big rise in fraud cases. Fraudulent activities are the main reason for huge financial losses and the destruction of trust in e-payment systems. According to reports, the card fraud losses of UK banks in 2020 were more than €574.2 million. However, the continuous application of countermeasures by banks and payment networks has not completely wiped-out credit card fraud, which still poses a huge risk of financial operational and reputational harm for banks, merchants, and consumers. This constant problem has led the development of automated fraud prevention systems backed by significant research. The latest systems are expected to work at a very large scale, usually in almost real-time, and at the same time, significantly reduce false alarms to avoid inconveniences during legitimate transactions [2].

Credit card fraud detection (CCFD) keeps running into major obstacles despite the advancements made by the current detection techniques. One of the most significant problems is the extreme imbalance in the transaction datasets, that highly consist mostly of

non-fraudulent transactions. This makes the challenge of detecting fraud accurately harder. Another issue is the continuous change of fraudulent activities, which makes the models very adaptive to the new and advanced patterns. Additionally, the large amount and rapid rate of transaction data contribute as complicating factors since systems have to ensure real-time processing of all transactions without causing any interruptions to the legitimate ones. Also, the current techniques have not been successful in properly capturing the temporal and spatial dependencies that are naturally present in sequential transaction data limiting their capability to detect coordinated or complex fraud schemes [3,4].

Credit card fraud manifests in many ways like card-not-present (CNP) fraud, skimming, account takeover, phishing, and transaction laundering. The attackers take advantage of vulnerabilities both online and offline to either steal card details or trick users into revealing their sensitive information. The attacks mentioned are not static; instead, the fraudsters are changing their methods to get through the security systems undetected. The variety and constant changes in illegal activities make it tough for the automated detection systems, which is the reason behind the need for a strong Deep Learning (DL) model that can recognize established and new fraudulent attempts effectively by dealing with the complex temporal and spatial patterns [5].

The CCFD method usually adhering to an approved procedure starts with the collection of transactional data such as the amount of the transaction, classification of merchant, date, Location, and specifications of the device. The unprocessed data undergoes then the preprocessing which entails the cleaning, standardizing, and modifying of data thereby bringing it up to the model level. Feature engineering is applied to reveal the important patterns that can separate the legal transactions from the fraudulent ones, while, techniques like dimensionality reduction and representation learning work together to reduce complexity and enhance system performance at the same time [6]. The data after processing is then used to train the detection models which are assessed by several criteria, considering the fraud rate, and thus guaranteeing that fraudulent transactions are spotted correctly and quickly without disrupting legitimate ones. Although this workflow offers a systematic procedure, the selection of detection techniques has a large bearing on the overall efficiency of the system, especially in adverse situations like class imbalance and shifting fraud patterns [7].

To detect credit card fraud, several methods have been developed, and these methods have a main focus on the issues of changing fraud patterns, class imbalance, and high-dimensional features. Supervised methods such as logistic regression, random forests, decision trees, and gradient boosting can recognize the patterns from the labeled data, but they suffer a lot when the fraud is either rare or new. Conversely, unsupervised, and semi-supervised techniques, like clustering, autoencoders, and isolation forests, find out the anomalous or unseen transactions. Fusion and hybrid models are built by combining different classifiers or connecting DL with anomaly detection, and thus, the stability is increased. The use of DL techniques like RNNs, LSTMs, CNNs, and attention-based models is capable of capturing very intricate temporal and sequential patterns. Data-level methods like oversampling, synthetic data creation, and incremental learning are further developing the problem of imbalance, concept drift, and feature complexity. Although these techniques have produced quite good results, there are still some problems remaining in the practical implementation, especially those related to DL-based models [9,10].

Many algorithms struggle to adapt to transaction data that is very imbalanced and changes rapidly. These data processing methods are likely to miss either the unusual or the emergence of fraud trends. One of the reasons for the limited use of deep learning models in the investigation context is the lack of interpretability that sometimes hinders the clarification of complex temporal and sequential relationships. The problems relating to privacy issues and the lack of labeled fraud data make it even harder for practitioners in different organizations to cope with the situation. These limitations point to the need for a powerful, flexible, and understandable deep learning model that can detect credit card fraud under actual high-velocity transaction conditions.

1.1. Problem Statement & Scope of the Research

Despite significant advancements in credit card fraud detection, existing approaches still face critical limitations, as identified in recent literature. Hybrid deep learning models such as CNN–LSTM and transformer-based architectures have demonstrated strong capability in capturing spatial and temporal transaction patterns; however, many of these models focus on either sequential dependency modeling or feature representation, lacking a unified framework that effectively integrates both. Similarly, contrastive learning and attention-based methods have improved feature discrimination and interpretability, but often introduce increased

computational complexity, limiting their scalability in real-time applications. Furthermore, federated learning frameworks address data privacy concerns but suffer from challenges related to heterogeneous data distribution, communication overhead, and complex multi-party training. Feature selection-based optimization methods improve classification performance; however, they heavily rely on handcrafted or optimized features and lack deep semantic representation capabilities. In addition, many existing studies primarily depend on conventional evaluation metrics such as accuracy and ROC-AUC, with limited emphasis on imbalance-aware metrics and time-aware validation strategies, which are crucial for real-world fraud detection scenarios.

Therefore, there exists a clear need for a robust, scalable, and unified deep learning framework that can simultaneously capture spatial–temporal dependencies, enhance feature representation through advanced learning mechanisms, and effectively handle highly imbalanced transaction data. To address these challenges, this study proposes a novel Spatial–Temporal Transformer Network with Contrastive Pretraining (STTN-CP), which integrates transformer-based modeling with contrastive representation learning to improve detection accuracy, generalization, and robustness in real-world fraud detection environments.

1.2. Research Objectives

The research contributes to CCFD by proposing a novel DL-based framework using the Kaggle Credit Card Fraud dataset. The model, leveraging Spatial-Temporal Transformer architecture with Contrastive Pretraining, captures complex temporal and spatial dependencies in transaction sequences to accurately detect and classify fraudulent activities. The aims of the research are as follows:

- To accurately detect and classify fraudulent transactions from the Kaggle Credit Card Fraud dataset using the Spatial-Temporal Transformer model.
- To preprocess transactional data with Min-Max normalization for standardized scaling and SMOTE for mitigating class imbalance.
- To enhance the representation learning capability of the model through spatial-temporal transformer + Contrastive Pretraining for improved fraud detection.
- To evaluate the effectiveness of the proposed model using metrics such as accuracy, precision, recall, and F1-score, and specificity.
- To compare the proposed framework with existing state-of-the-art approaches to demonstrate its improved performance and practical applicability.

1.3. Research Contribution and Novelty

Unlike conventional credit card fraud detection approaches that rely on either spatial–temporal modeling or feature-level optimization independently, the proposed STTN-CP framework introduces a unified architecture that integrates spatial–temporal transformer learning with contrastive pretraining for enhanced feature representation. This integration enables the model to effectively capture complex transaction dependencies while improving discrimination between fraudulent and legitimate patterns. In contrast to existing studies that primarily focus on improving accuracy through ensemble or resampling techniques, this work emphasizes robust representation learning and imbalance-aware evaluation using metrics such as recall, F1-score, and specificity, which are more suitable for fraud detection tasks. Furthermore, the proposed framework follows best practices in handling imbalanced data through SMOTE and standardized preprocessing, while also ensuring generalization and stability as demonstrated through statistical analysis and ablation studies. Therefore, this study contributes beyond incremental improvements by providing a scalable and unified deep learning framework that addresses key limitations in existing literature and advances the state-of-the-art in credit card fraud detection.

The work is presented in the following structure: Section 2 reviews related works on CCFD, focusing on machine learning and DL-based classification models. Section 3 comprehensively describes the methodology that was proposed; it includes the process of data gathering, implementation of Min-Max normalization and SMOTE for data sampling, and the use of the Spatial-Temporal Transformer with Contrastive Pretraining for classification purposes. The performance of the proposed model is measured in Section 4 by means of precision, accuracy, recall, specificity, and F1-score, and it is compared with the present best practices in the field. The research is wrapped up in Section 5, more so to the advancements in interpretability, scalability, and the ability to adapt to changing fraud trends, but also to the proliferation of new areas for study.

2. LITERATURE REVIEW

A sophisticated technique called Balanced Variational Autoencoder with Attention (Bal-VAE-Attention) was presented to deal with the problems

of CCFD caused by extremely imbalanced datasets in the paper [11]. Along with its new oversampling method, the model effectively dealt with the disparity between legitimate and fraudulent transactions. An automatic feature selection process was performed, which resulted in the extraction of the most informative features, while a deep multi-head attention mechanism was used to reveal the complex relationships in the data. These changes improved the accuracy, precision, and recall of detection and reduced noise and learning complex data patterns.

In [12], a DL stacking ensemble-based approach was proposed that would boost the performance of CCFD systems. GRU and LSTM networks were used as the main learners, and the MLP was the meta-learner in this mixed model. The experiments were performed on the CCFD dataset (Kaggle). To solve the issue of class imbalance between legitimate and fraudulent transactions, the hybrid SMOTE-ENN method for data resampling was chosen. This combination not only strengthened the model but also improved its classification accuracy to the level of surpassing traditional classifiers and standalone deep learning models.

The study [13] introduced an interpretable ensemble-based model dubbed FraudX AI that was specifically meant for the CCFD in the scenario of high imbalanced data sets. It was a fusion of XGBoost and Random Forest, and the outcome of these two algorithms was combined through probability averaging and threshold optimization for a performance boost in the detection task. The model underwent testing on the European Credit Card Fraud Detection dataset, which revealed its natural imbalance for a very authentic validation. The framework substituted the conventional technique SMOTE for oversampling with the following strategies: class-weighted learning, MLP validation, and threshold tuning. Also, SHAP was applied to explain model predictions and identify the most important features that led to the predictions. The results of the experiments proved that FraudX AI was remarkably superior to other models in the aspect of detecting rare fraudulent transactions.

The hybrid DL ensemble model was proposed to deal with the increasing issue of credit card fraud in online transactions. The model smartly combines the advantages of LSTM, CNN, and Transformer networks as main learners, with XGBoost operating as a meta-learner. The evaluation of the datasets from Europe and Taiwan shows that the model not only exceeds the performance of each separate model but also that of the traditional methods, with higher sensitivity, specificity, and AUC-ROC metrics being achieved. The ensemble model indeed raises the bar for newcomers in terms of the robustness, adaptability, and huge potential for real-world application. Still, the model is computationally intensive and has been tested on a limited number of datasets only.

The FinGraphFL, which was a new system presented in [15], combines graph-based learning with federated learning technologies so that the CCFD can be improved while preserving the data private. The system employs Graph Attention Networks (GATs) and a unique differential privacy method, which allows for several banks and financial institutions to work together on fraud detection without any compromise on data security. The findings of the public dataset experimentations indicated that the new method is more accurate than the classic ones. One of the main reasons making FinGraphFL excellent assistance for small and medium-sized finance companies is its capability to capture complex transaction patterns and to adapt to changing fraud strategies by providing flexibility. Moreover, the whole process addressed the common problems in real deployment, such as having varying infrastructures, being in line with the regulations, and making moral decisions.

In [16], a hybrid framework for CCFD was suggested that combined CatBoost and Deep Neural Networks (DNN) to enhance the accuracy and trustworthiness in classification. The model applied a user-based segmentation strategy, classifying users into two groups, old or new, in order to uncover different behavioral patterns. This method was validated on the IEEE-CIS Fraud Detection dataset, which included diverse real-life transaction records. To ensure the best possible performance of the model, advanced preprocessing strategies like feature engineering, transformation, and imbalance management were applied. An architecture for real-time fraud detection using Apache Flink was created, which allowed for the quick processing of a large number of transactions.

In [17], a CCFD method that combined an advanced Variational Autoencoder Generative Adversarial Network (VAEGAN) with the XGBoost classification algorithm was presented as a solution to the problem of data imbalance. The augmented VAEGAN model was successfully applied for generating numerous synthetic samples of the minority (fraudulent) class, which were very much like the real ones thus enlarging the training data set. The open-source credit card transactions dataset was used for the assessment of the model which revealed

that it was superior to the routine oversampling techniques such as VAE, GAN, and SMOTE regarding precision and F1-score. The results highlighted that the classifiers using ensemble techniques, notably XGBoost, were more effective than the traditional ones.

In [18], a new hybrid CCFD approach was adopted that included CNN, LSTM networks, and an attention mechanism with the goal of detection accuracy and robustness enhancement. CNN used to get the spatial features while LSTM is responsible of retrieving the temporal dependencies and the attention mechanism focused on the most important transaction features for better understanding of the model's decisions and thus, its interpretation. The model was built and evaluated on the European credit cards data set, where the issue of data imbalance was handled by the application of the SMOTE. Extensive preprocessing was carried out to ensure that the data was in the input format that the model accepted. The experimental results indicated that the hybrid CNN-LSTM-Attention model outperformed the traditional algorithms.

In [19], a model that combined LSTM and attention mechanism was introduced as a CCFD model that capable to detect sequential patterns in transaction data. The model was made robust through feature selection with swarm intelligence, dimensionality reduction via UMAP, and data balancing with SMOTE. While LSTM focused on the long-term fluctuations of the transaction sequence, the attention mechanism selected the features that were most pertinent for classification. The approach was applied to two credit card datasets and reached high sensitivity and good separation between the two classes of transactions-fraudulent and legitimate. The experimental findings indicated that the model had surpassed the performance of recent comparable techniques.

A federated learning framework for CCFD was developed in [20] by using the transactional datasets sourced from different financial institutions as a foundation. The datasets were extremely imbalanced indeed, so hybrid resampling methods such as SMOTE and AdaSyn were applied to balance the classes and thus to support the model training. Preprocessing involved feature selection, normalization, and imputation of missing values to ensure the quality of the data. The framework incorporated a list of classifiers like Random Forest, K-Nearest Neighbors, Decision Tree, Logistic Regression, and Gaussian Naive Bayes. The model was developed on the PyTorch and TensorFlow federated platforms, where PyTorch slightly surpassed TensorFlow in accuracy but required more computation.

The research presented in [21] showcases the creation of two advanced LightGBM models, CB-CHL-LightGBM and OS-CHL-LightGBM, aimed at resolving the customer loss prediction problem in cases of severe data imbalance. The authors relied on three datasets consisting of credit card transactions that showed substantial class and cost imbalances for their study. Preprocessing included class balancing, oversampling, and applying a cost-harmonization loss function to decrease false negatives and thus enhance the detection of small fraud cases. The model predictions were further refined, and the feature selection was guided by determining the feature relevance using SHAP analysis. The performance of the models was measured against that of standard LightGBM using F2-score, recall, and cost savings, in which they performed better than the latter. The researchers pointed out that it was very important to deal with both class imbalance and cost in the fraud detection process for datasets with large imbalances.

In [22], a hybrid deep learning model termed CLST was introduced. The architecture of this model consisted of three main elements: Convolutional Neural Networks (CNNs) were responsible for spatial feature extraction, Long Short-Term Memory (LSTM) networks detected temporal patterns in transactions, and a fully connected layer was finally applied for the unambiguous classification of credit card frauds. To train the model, the researchers used real transaction records and applied SMOTE to address the issue of class imbalance. To enhance the model's prediction accuracy, preprocessing of data and tuning of hyperparameters were performed. The results of the experiments demonstrated that CLST outperformed both individual DL models and traditional fraud detection techniques. The method could give a fast and precise response to financial fraud detection in real-time, but its excessive computational requirement might restrict its use in particular situations.

A framework called OptDevNet which was aimed at fraudulent credit cards transaction detections within the Comprehensive CCFD scheme was introduced in [23]. The paper utilized the common CCFD dataset and performed extensive data cleaning and preprocessing to secure the quality and the most consistent data possible. The deep-learning-based CCFD framework is backed up by a thorough evaluation against traditional classifiers like SVM, random forests, logistic regression, and KNN. Moreover, the model was able to efficiently deal

with the class imbalance problem and it was observed that the model was better in predicting new unseen data with higher imbalance ratios than the training data.

The application of a hybrid Deep Learning architecture combining Recurrent Neural Networks (RNNs) and Generative Adversarial Networks (GANs) has [24] introduced a new method for credit card fraud detection. Researchers utilized two different credit card datasets—one from Europe and the other from Brazil—applied numerous preprocessing methods for better quality of data and class balancing. The GAN component of the system assisted in creating fake transactions mainly from the minority class thus elevating the smaller group of transactions. In addition, the RNN models—Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU)—were excellent due to their capability of detecting temporal dependencies in the transaction groups. The joint treatment was highly effective in class balancing and detection rate enhancement. It is very possible that the considerable computational power demanded by the method will limit its adoption in large, real-time systems.

In [25], a cutting-edge method grounded on CCFD was developed that integrates machine learning and a Genetic Algorithm (GA) for optimal feature selection. The dataset of European credit card transactions of cardholders was utilized by the researchers, and their primary concern was to discover the parameters which most influence fraud detection. Eventually, the feature optimization performed by the genetic algorithm led to the application of different machine learning classifiers for the classification process. The genetic algorithm worked alongside the fitness function of the Random Forest to both refine the feature selection and boost the model's accuracy. The models built with GA were superior to previous systems which proved the effectiveness of evolutionary optimization together with machine learning, however, their dependence on specific datasets made them less suitable for application in actual scenarios. The details of the existing models are compared in Table 1 where their methodologies, datasets, and performance metrics are emphasized.

More recently, studies in 2025 have focused on advanced hybrid architectures, contrastive learning, and federated frameworks to further enhance fraud detection performance. A study [26] proposed a hybrid deep learning model combining CNN and RNN with LSTM units for credit card fraud detection. The model effectively captured spatial features and temporal dependencies, achieving high performance on the Kaggle dataset with the support of SMOTE for class imbalance handling. However, the evaluation mainly relied on accuracy and ROC-AUC, with limited emphasis on imbalance-aware metrics and time-aware validation. These limitations indicated the need for more robust evaluation frameworks and enhanced feature representation methods. A hybrid contrastive learning framework integrating Siamese networks with attention-based neural networks for fraud detection in digital payment systems was introduced in [27]. The model effectively enhanced feature representation and discrimination, achieving high recall, precision, and ROC-AUC on real-world datasets. Additionally, the incorporation of attention mechanisms and SHAP analysis improved feature interpretability and model transparency. However, the approach involved increased computational complexity, indicating the need for more efficient and scalable architectures.

The study [28] proposed a federated transfer learning framework (FED-SPFD) to address credit card fraud detection under heterogeneous data conditions. The model effectively handled data privacy and distribution differences by separating shared and private features and aligning them through statistical mechanisms. Experimental results demonstrated improved performance in terms of recall, precision, and F1-score compared to existing federated approaches. However, the complexity of multi-party training and feature alignment highlighted challenges in scalability and real-time deployment. A hybrid feature selection approach combining Big Bang–Big Crunch and Cuckoo Search algorithms for credit card fraud detection was proposed in [29]. The method effectively balanced exploration and exploitation to select optimal features, improving classification performance when integrated with DCNN and EDCNN models. Experimental results showed enhanced accuracy compared to individual optimization techniques. However, the approach relied heavily on feature selection and lacked advanced representation learning, indicating scope for more robust and adaptive deep learning frameworks.

Recent studies have demonstrated significant advancements in credit card fraud detection through hybrid deep learning models, federated learning frameworks, and optimization-based feature selection techniques. These approaches effectively capture spatial–temporal transaction patterns and address class imbalance, while contrastive learning and attention mechanisms enhance feature representation and model interpretability. However, existing methods primarily focus on either spatial–

temporal modeling or advanced representation learning in isolation, lacking a unified framework that effectively integrates both aspects. Furthermore, challenges such as high computational complexity, limited scalability, and insufficient emphasis on imbalance-aware and time-aware evaluation metrics continue to hinder real-world deployment. Therefore, a critical research gap exists in the design of a unified, efficient, and scalable framework that simultaneously captures spatial–temporal dependencies and enhances feature representation while ensuring robust performance on highly imbalanced datasets. To address this gap, this study proposes a novel Spatial–Temporal Transformer Network with Contrastive Pretraining (STTN-CP), which integrates transformer-based modeling with contrastive representation learning to improve detection accuracy, generalization capability, and robustness.

Table 1: Comparative Analysis of Reviewed Current Models

Ref	Models	Application	Advantages	Disadvantages
[11]	Balanced Variational Autoencoder with Attention (Bal-VAE-Attention)	Credit card fraud detection on highly imbalanced datasets	Effectively handles class imbalance using novel oversampling and attention-based feature selection.	Computationally intensive and requires high-quality feature extraction
[12]	Deep learning-based stacking ensemble (LSTM + GRU + MLP)	Credit card fraud detection using the Kaggle dataset	Hybrid SMOTE-ENN resampling improves robustness and accuracy; the ensemble enhances generalization	Complex model architecture and higher training cost
[13]	FraudX AI (Random Forest + XGBoost ensemble)	Fraud detection on imbalanced datasets	Interpretable model using SHAP; realistic validation without oversampling	May underperform on highly non-linear feature relationships
[14]	Hybrid CNN-LSTM-Transformer + XGBoost ensemble	Online credit card fraud detection	Combines spatial, temporal, and sequential learning; achieves high sensitivity and AUC	High computational demand; limited dataset diversity
[15]	FinGraphFL (Graph Attention Networks + Federated Learning)	Federated credit card fraud detection with privacy preservation	Enables cross-institution learning without data sharing.	Deployment complexity and heterogeneous system challenges
[16]	CatBoost + DNN hybrid	Real-time fraud detection (IEEE-CIS dataset)	Captures behavioral patterns via a user-based strategy; supports streaming detection with Apache Flink	Real-time processing increases hardware and latency requirements.
[17]	Improved VAEGAN + XGBoost	Fraud detection on imbalanced datasets	Generates realistic minority samples.	Training instability and computational cost in GAN optimization
[18]	Hybrid CNN+LSTM with Attention	CCFD (European dataset)	Integrates spatial, temporal, and attention-based learning.	Overfitting risk due to high model complexity
[19]	LSTM + Attention + Swarm Intelligence feature selection	Sequential transaction-based fraud detection	Captures long-term dependencies; feature selection improves efficiency	Requires extensive tuning of hyperparameters
[20]	Federated Learning with multiple classifiers	Cross-institution credit card fraud detection	Preserves data privacy; balances class distribution via hybrid resampling	Computationally heavy and complex to synchronize across clients
[21]	CB-CHL-LightGBM & OS-CHL-LightGBM	Extremely imbalanced fraud detection	Incorporates cost-sensitive learning; interpretable via SHAP	Less effective on dynamic transactional sequences
[22]	CLST (CNN + LSTM + Fully Connected)	Real-time fraud detection	Strong spatial-temporal modeling; high classification performance	High computational complexity; may struggle with real-time scalability

[23]	OptDevNet (Optimized Deep Event-based Network)	Credit card fraud detection	Handles high imbalance and improves generalization; performs well on unseen data	Requires large computational resources
[24]	GAN + RNN (GRU/LSTM) hybrid	Fraud detection (European & Brazilian datasets)	Generates synthetic fraud samples; captures temporal transaction patterns	Training complexity and scalability limitations
[25]	Genetic Algorithm-based Feature Selection + ML classifiers	Feature optimization for fraud detection	Efficient feature selection and improved model accuracy	Overfitting risk due to high model complexity.
[26]	Hybrid CNN–RNN (LSTM)	Credit card fraud detection (Kaggle dataset)	Effectively captures spatial and temporal features; high accuracy with SMOTE handling imbalance	Relies mainly on accuracy and ROC-AUC; lacks imbalance-aware and time-aware evaluation
[27]	Hybrid Contrastive Learning + Siamese Network + Attention	Fraud detection in digital payment systems	Enhances feature representation and discrimination; improves interpretability using SHAP	High computational complexity; scalability issues
[28]	FED-SPFD (Federated Transfer Learning Framework)	Credit card fraud detection under heterogeneous data	Preserves data privacy; handles heterogeneous data; improves recall and F1-score	Complex multi-party training; challenges in scalability and real-time deployment
[29]	HB3C2S (BB-BC + Cuckoo Search) + DCNN/EDCNN	Credit card fraud detection with feature selection	Effective feature selection; improves classification accuracy; balances exploration and exploitation	Lacks deep representation learning; depends heavily on feature selection

3. PROPOSED RESEARCH METHODOLOGY

The CCFD framework that is presented is structured upon a methodical series of procedures, which initiates by collecting data and concludes with the application of the Spatial–Temporal Transformer with Contrastive Pretraining for proficient classification. The complete workflow of the suggested research model is illustrated in Figure 1, which displays the successive phases of data gathering, preprocessing, classification, evaluation, and result generation. The experimentation and assessment are performed using the Credit Card Fraud Detection dataset obtained from Kaggle. This dataset comprised a total of 284,807 transactions made with European cards during September 2013 which include 492 fraudulent transactions and 284,315 transactions that are legitimate. There is a significant disparity in the dataset, with only 0.172% of it representing the fraudulent cases. The next step in the collection of data is to preprocess the dataset, which is essential for the successful training of the model. The first technique used is the Min–Max normalization which scales the features to a ranging scale, thus, removing any bias caused by the different magnitudes. The second technique is the Synthetic Minority Oversampling Technique (SMOTE), which produces new samples of the minority (fraud) class; hence, a dataset that is equal in size is the outcome. All the preprocessing steps ensure the model's input is standard and balanced, thus improving the training and classification effectiveness.

Journal of Theoretical and Applied Information Technology
15th April 2026. Vol.104. No.7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

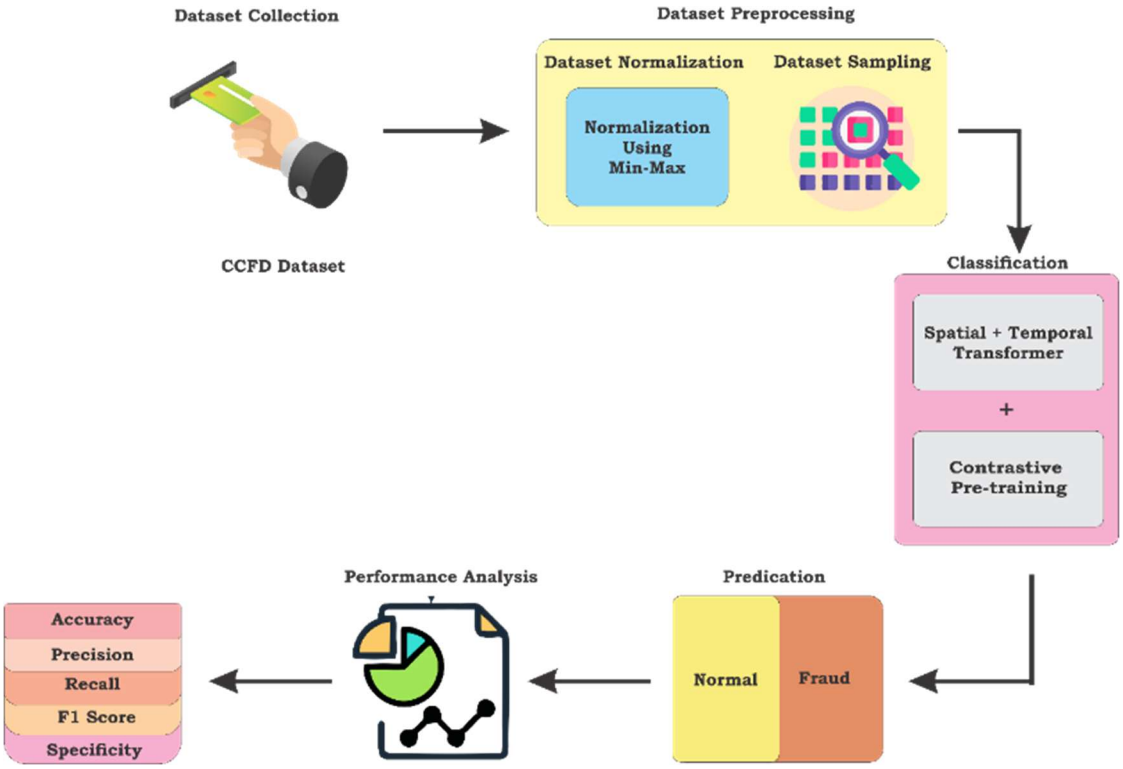

Figure 1: Workflow of the Research Model

The STTN together with Contrastive Pretraining applies to preprocessing the transaction sequences. The spatial-temporal transformer blocks are responsible of discovering the very complex relationships between transaction features and detecting the change of those features over time. Meanwhile, the discrimination of features is improved through contrastive pretraining, which unites the similar embeddings and separates the dissimilar ones. The enhanced embeddings are then classified using a fully connected layer with a sigmoid activation, which outputs the transaction's probability of being fraudulent. The STTN-CP model's performance was assessed through the application of common metrics: Accuracy, F1-Score, Recall, Precision, and Specificity. Accuracy estimates the general correctness of the predictions, while precision gives the ratio of properly predicted frauds to all predicted frauds. The use of all these metrics together provides a complete insight of the model's success in CCFD.

3.1 Dataset Description

The CCFD dataset from Kaggle had been selected as the data source for experimentation and evaluation. This particular dataset contains 284,807 credit card transactions from European cardholders, which were all made during September 2013. Out of these transactions, 492 were detected as fraudulent, and the rest 284,315 were identified as legitimate. Hence, it leads to a very imbalanced dataset, where only 0.172% of the transactions are fraud cases. The transactions are described through 30 different attributes, out of which, 28 are the anonymized numerical variables V1-V28 that are obtained from the Principal Component Analysis (PCA) approach that ensures confidentiality of data, plus Time, Amount, and the Class label (1 for fraud, 0 for legitimate). The dataset has become a standard benchmark for fraud detection research, due to the challenges of high dimensionality, class imbalance, and realistic distribution of genuine and fraudulent transactions. The dataset is the basis for training and testing the proposed Spatial-Temporal Transformer with Contrastive Pretraining model, which is capable of detecting fraudulent behavior [11]. The dataset consists of multiple features that capture important aspects of each transaction, and they are listed in Table 2.

Table 2: Summary of Dataset Features

Features	Description
Time	It shows the amount of time that has passed (in seconds) from the present transaction to the very first transaction stored in the dataset. This gives a time frame for evaluating the transaction patterns.

V1–V28	These are the principal components obtained using PCA applied to anonymized original features. PCA was employed to ensure user privacy while retaining critical information from the transaction data.
Amount	Denotes the monetary value of the transaction, which can be an indicator of abnormal spending behavior in fraudulent cases.
Class	Serves as the target label, where 0 indicates a legitimate transaction and 1 denotes a fraudulent transaction.

The dataset does not explicitly identify the different kinds of credit card fraud, but the fraudulent transactions can still be interpreted according to the common fraud behaviors that are usually seen in real-world cases. The different types of attack characterize the different techniques that the fraudsters use to take advantage of the weak points in the payment systems. In Table 3, the main categories of credit card fraud attacks that are significant in relation to the dataset are presented.

Table 3: Representative Types of Credit Card Fraud Attacks

Attack Type	Description
Card-Not-Present (CNP) Fraud	Fraudulent transactions are conducted online or over the phone without the physical card, using stolen card details.
Skimming	Duplication of card data from the magnetic stripe using illegal skimming devices at ATMs or POS terminals.
Account Takeover	Unauthorized access to a legitimate user's account to perform illicit transactions or change account credentials.
Phishing and Social Engineering	Deceptive techniques to trick users into revealing sensitive information, such as card numbers or OTPs.
Lost or Stolen Card Fraud	Fraudulent use of a misplaced or stolen credit card for unauthorized purchases.
Application Fraud	Use of stolen or fake identities to apply for and obtain new credit cards.

3.2 Data Preprocessing

Data preprocessing is the basis of the entire process which involves model training and detection of fraud, and it secures both the effectiveness and reliability of the process. To achieve the required data consistency and to keep its amount balanced, the two main preprocessing methods, Min-Max normalization and SMOTE, were employed in this research.

Normalization:

The research implemented Min-Max Normalization as a method to maintain feature uniformity throughout, which is a technique that adjusts all feature values within the limits of 0 to 1. This normalization prevents the training procedure from being dominated by features with high numerical ranges like Time and Amount. The Min-Max scaling is executed using the given formula for each feature:

$$Y_{scaled} = \frac{Y_i - \min(Y)}{\max(Y) - \min(Y)} \quad (1)$$

where Y_{scaled} denotes a normalized feature value, and $\min(Y)$ & $\max(Y)$ indicate the minimum and maximum values of a feature, respectively. After feature scaling, all the feature values will now be in the range of 0 to 1 [30].

Data Sampling:

Eventually, the significant class imbalance of the dataset was solved by the use of sampling strategies after preprocessing. The original dataset contained 275,190 legitimate transactions and 473 fraudulent transactions, which caused a tremendous 99.83% of the transactions to be classified as 'Class 0' (the majority class) and only 0.17% as 'Class 1' (the minority class). Such an imbalance may cause the deep learning models to be biased towards the majority class, and therefore, their ability to detect rare fraudulent transactions may be impaired. One of the tactics applied to counteract this problem is the adoption of SMOTE. To create synthetic samples for the minority class, SMOTE first picks out real fraudulent cases, then locates their closest neighbors and finally, carries out interpolation to yield the new instances. The count of synthetic samples produced can be varied, thus allowing for an adaptable dataset's distribution. A synthetic sample Z^\prime is calculated as follows theoretically:

$$Z' = Z + rand\ (0,1) * |c - d| \quad (2)$$

Z is a representation of a sample taken from the minority class, while c and d are neighboring

samples, and rand(0, 1) indicates a random number that falls within the range of 0 to 1. As a result, the representation of fraudulent transactions will be improved, and the model's capability of learning the fraud patterns will also be strengthened without the need for additional data.

By applying this preprocessing method, features will be scaled uniformly and the issue of class imbalance will be taken into consideration, thus allowing the model to detect trends in both normal and fraud transactions. Once the dataset has been standardized and balanced, the next step is to apply classification techniques, and during this phase, the proposed Spatial–Temporal Transformer with Contrastive Pretraining will be used to accurately detect fraudulent transactions [12, 30].

3.3 Classification using Spatial–Temporal Transformer with Contrastive Pretraining:

After preprocessing, the prepared dataset is utilized to train and evaluate the proposed DL framework. In this study, a Spatial–Temporal Transformer with Contrastive Pretraining is employed for the classification of credit card transactions into legitimate and fraudulent classes. This model is designed for capturing both the temporal dependencies in transaction sequence (e.g., transaction order and time gaps) and spatial correlations among features (e.g., relationships between transaction amount, merchant category, and other variables).

As illustrated in Figure 2, the proposed STTN with Contrastive Pretraining consists of stacked spatial–temporal blocks followed by a contrastive learning module and a classification layer. Each spatial–temporal block integrates a Spatial Transformer (S) and a Temporal Transformer (T) to jointly model spatial and temporal dependencies inherent in transaction sequences. The contrastive pretraining enhances feature discrimination between fraudulent and legitimate patterns before supervised fine-tuning.

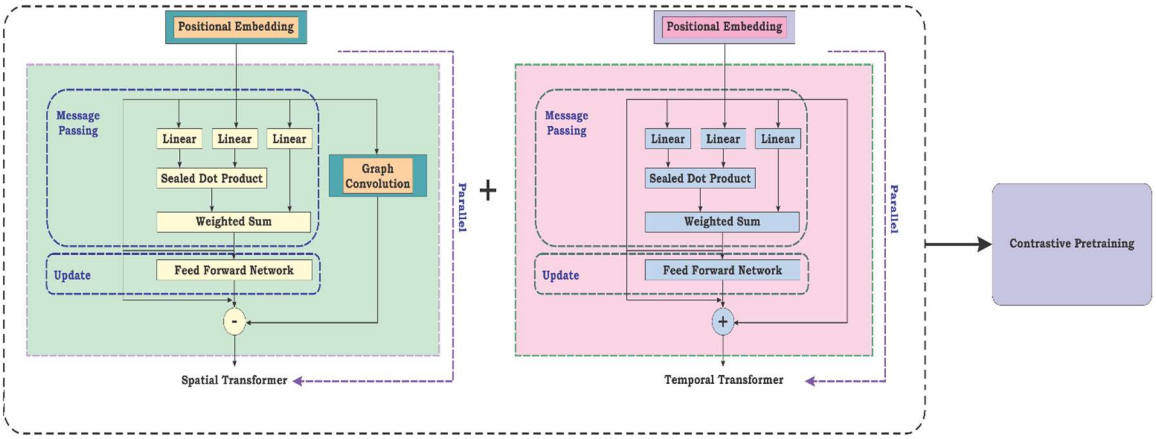

Figure 2: ST-Transformer with contrastive pretraining architecture.

The overall model consists of stacked spatial–temporal transformer (ST) blocks in a classification layer that distinguishes between fraudulent and legitimate transactions. By stacking multiple ST blocks, the model can learn deep hierarchical representations that jointly encode spatial–temporal interactions. Finally, the extracted high-level embeddings are aggregated and passed through a fully connected classification head activated by the sigmoid function for binary decision making [31]. The overall process can be formulated as:

The input to the lth ST block is a 3-D tensor.

$$M_l^{sp} \in R^{A \times T \times d_f} \quad (3)$$

which represents A batches of transaction sequences, each consisting of T time steps and d_f-dimensional transaction features (such as amount, time gap, merchant type, PCA components, etc.) extracted by the (l-1)th block.

Within each ST block, the Spatial Transformer (S) and Temporal Transformer (T) are applied in sequence to jointly extract high-level

spatial–temporal features that encode both feature interactions and temporal behavioral patterns.

Residual connections are incorporated around each transformer module to facilitate stable and efficient gradient flow during training. In the lth ST block, the Spatial Transformer sp captures inter-feature dependencies across all transaction attributes and generates spatial feature representations N_l^{sp} from the input tensor M_l^{sp}:

$$N_l^{sp} = sp(M_l^{sp}) \quad (4)$$

The output N_l^{sp} is then combined with the input M_l^{sp} via a residual connection to produce the input M_l^t for the subsequent Temporal Transformer, which models temporal dependencies across consecutive transactions:

$$N_l^T = T(M_l^T) \quad (5)$$

The output of the *lth* block is finally obtained by aggregating the temporal transformer output and its input through another residual connection:

$$M_{l+1}^{sp} = N_l^T + M_l^T \quad (6)$$

Without loss of generality, for simplicity in subsequent equations, the subscript l of M_l^{sp}, M_l^T, N_l^{sp} and N_l^T is omitted. Where,

sp (·) represents the spatial attention operation.

T (·) represents the temporal attention operation.

Contrastive Pretraining Module

After the stacked ST blocks extract high-level spatial–temporal embeddings, these representations are passed through a contrastive pretraining module before the supervised classification phase. The objective of this pretraining step is to improve the discriminative ability of the learned embeddings by encouraging transactions from the same class (legitimate or fraudulent) to have similar feature representations, while maximizing the distance between embeddings from different classes.

During the contrastive pretraining phase, the model learns to align similar representations (positive pairs) and separate dissimilar ones (negative pairs) using a contrastive loss function, typically based on cosine similarity. The pretraining process operates in a self-supervised manner, where the encoder (comprising stacked ST blocks) is trained to produce robust feature embeddings independent of explicit labels.

Given two augmented transaction samples x_i and x_j, their embeddings z_i and z_j are obtained as:

$$z_i = f(x_i),\ z_j = f(x_j) \quad (7)$$

where f(·) signifies the encoder network consisting of the spatial–temporal transformer layers. The contrastive loss used for representation learning is defined as:

$$L_{contrastive} = -log \frac{\exp(sim(z_i, z_j)/T)}{\sum_{k=1}^{N} \exp(sim(z_i, z_k)/T)} \quad (8)$$

where $sim(z_i, z_j)$ represents the cosine similarity between a pair of embeddings, T is the temperature coefficient, and N is the no. of samples in a batch. This formulation ensures that embeddings from similar transactions (e.g., repeated legitimate purchases) are clustered together, while those representing different behaviors (e.g., fraudulent vs. legitimate) are well separated in the latent feature space.

After the contrastive pretraining is finished, the weights of the encoder are moved to the next phase of the process, supervised fine-tuning, with the addition of a fully connected classification head having a sigmoid activation function. In this part of the process, the pretrained embeddings for binary classification are polished using binary cross-entropy loss function:

$$L_{classification} = -[y \log(\hat{y}) + (1 - y)\log(1 - \hat{y})] \quad (9)$$

Here, y is the real class label (which shows if a transaction is fraudulent or legitimate), and $\hat{y}$ denotes the expected possibility of fraud. The classification step with contrastive pretraining brings in a fantastic model to discern fine differences in transaction behaviors. The model can continuously augment its ability to represent intricate spatio-

temporal patterns hidden in transaction data, like unusual buying frequencies, unexpected merchant connections, or strange time gaps, by increasing the number of blocks stacked.

The final loss function combines both the objectives:

$$L_{total} = L_{classification} + \lambda L_{contrastive} \quad (10)$$

Here, λ is a trade-off parameter controlling the contribution of contrastive learning.

This hybrid strategy strengthens the STTN-CP model's robustness to slight transaction variations and effectively mitigates data imbalance.

Common Process in Transformer Architectures: In order to fully understand the ST Transformer's complexities, it is necessary to firstly show the principal transformer mechanisms which are shared by both the spatial and temporal aspects. Apart from that, the main part of the transformers was self-attention (SA) which makes it possible for the model to connect all the elements in a sequence regardless of whether the sequence is processed through recurrent or convolutional layers.

Input Embedding and Positional Encoding:

Initially, every transaction sequence is transformed into a continuous high-dimensional embedding space by means of a linear projection layer. As transformers do not have the ability to recognize the order of the sequence, positional encodings are added to the embeddings to keep the transaction's temporal sequence intact. In the spatial module, positional embeddings reveal the structural connections through the transaction features, while in the temporal module, they show the dependencies over time.

Multi-Head SA Mechanism:

The **SA** mechanism computes pairwise relationships between all tokens in a sequence using three learnable projections: Query (Q), Value (V) and Key (K) matrices. The attention weights are computed as:

$$Attention(Q,K,V) = Softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (11)$$

where d_k is the dimensionality of the key vectors.

To improve representational richness, the transformer uses **multi-head attention**, which performs multiple parallel SA operations, each focusing on various aspects of the relations between transactions. The outputs of all heads were linearly projected and concatenated to form the final representation.

Feed-Forward Network and Residual Connections:

After the self-attention layer, the token embedding of every input is passed to a feed-forward network (FFN) that is made up of two linear transformations, a non-linear activation function (usually GELU or ReLU), and a non-linear activation function. Residual connections and layer normalization are sublayers that are used to stabilize training and to allow continuous gradient flow. With this mechanism, each transaction representation can depend on the whole batch, thus, the contextual understanding of feature interactions is improved.

Stacking and Hierarchical Representation:

Deep hierarchical architecture consists of several transformer layers, with the lowest layer capturing detailed local dependencies (for example, transaction amount and time correlations) and the highest layer having the most abstract understanding of the behavior (for instance, repeated fraudulent spending, or gone beyond a customer's normal activity profile). The Spatial-Temporal Transformer architecture [32,33] is based on this representation as its backbone.

The proposed robust framework for CCFD through structured preprocessing, spatial-temporal feature extraction, and contrastive representation enhancement has the exception. Not only does the method effectively depict the basic correlations between the transaction attributes and the sequencing of the behavioral patterns, but it also rigorously fights the fraud data imbalance through normalization and synthetic oversampling. The encoder that was pretrained to produce very discriminative embeddings and subsequently fine-tuned for binary classification is incredibly efficient in locating anomalies that are typical of fraudulent activity. This part of the paper introduces the experimental study that defines the implementation setup, dataset configuration, training settings, and comparison of the STTN-CP model with the latest techniques in the field.

4. EXPERIMENTATION RESULTS & DISCUSSION

4.1 Experiment Setup

The Kaggle CCFD dataset, containing anonymized transactions classified as legitimate or fraudulent, was the foundation for the assessment of the STTN-CP model that was proposed. The dataset was subjected to several preprocessing operations in order to implement a training procedure of the highest quality such as Min-Max normalization application which scales the feature values and SMOTE which creates a balance between classes with reference to the minority (fraudulent) class. The dataset was divided into two sets, with the training set taking up 80% and the testing set having 20% of the overall data. The model's architecture is made up of a stack of spatial temporal transformer blocks a contrastive pretraining module and eventually binary classification. Each block is equipped with a residual connection and layer normalization to ensure a robust training process. The model was trained for 50 epochs with early stopping, a batch size of 128, a learning rate of 0.001, and using the Adam optimizer. The InfoNCE loss with a temperature coefficient of 0.07 was employed during contrastive pretraining to increase feature discrimination. Furthermore, the model performance was assessed not only by means of F1-score, accuracy, recall, precision, and specificity metrics but also by carrying out a general fraud detection effectiveness evaluation.

4.2 Evaluation Metrics

The STTN-CP model's effectiveness was assessed through the application of traditional classification measures, which consist of Accuracy, F1-Score, Recall, Precision, and Specificity. These metrics furnish an all-encompassing understanding of the model's competence to correctly discriminate between fraudulent and legitimate transactions. The metrics are explained as follows:

Accuracy (Acc): Accuracy detects the fact if the model is correct at overall level.

$$Accuracy = \frac{TP+T}{TP+TN+FP+F} \quad (12)$$

Precision (P): Measures the proportion of correctly predicted fraudulent transactions among all transactions predicted as fraudulent.

$$Precision = \frac{TP}{TP+FP} \quad (13)$$

Recall (R): Assesses the ratio of accurately identified fraudulent transactions by the STNN-CP model.

$$Recall = \frac{TP}{TP+F} \quad (14)$$

F1-Score: Harmonic mean of Recall and Precision to overcome the FP and FN.

$$F1Score = \frac{2\times precision\times recall}{precision+reca} \quad (15)$$

Specificity (SP): Measures the proportion of actual legitimate transactions correctly identified.

$$Specificity = \frac{TN}{TN+FP} \quad (16)$$

Where TP = True Positives, FP = False Positives, TN = True Negatives, and FN = False Negatives. These measures provide a thorough assessment of the strength and dependability of the fraud detection model.

4.3 Performance Analysis

The STTN-CP model, which stands for Spatial-Temporal Transformer with Contrastive Pretraining, was trained and subsequently evaluated using the CCFD dataset that was acquired from Kaggle. The standard evaluation metrics including accuracy, F1-score, precision, recall, and specificity are carefully calibrated so as to fully evaluate the performance of the model; thus, the detection of fraudulent and legitimate transactions is carried out in a balanced and trustworthy process. The testing and training sets are made from the CCFD dataset in such a way that the model's ability to generalize can be checked effectively on new data. The proposed model shows outstanding performance in all the assessment criteria applied during training and testing, thus achieving high detection accuracy and robustness.

Table 4: Results of The Proposed Model

Parameters	Training performance	Testing performance
Accuracy	99.23	99.12
Precision	99.16	99.00
Recall	98.98	98.86
F1-Score	99.00	98.92
Specificity	98.52	97.96

Table 4 presents the evaluation of STTN-CP model, comprising performance metrics of Spatial-Temporal Transformer for both training and testing phases. The training resulted in a positive outcome for the performance of the model, and its accuracy of 99.23% proved its ability to easily learn and perfect spatial-temporal representations for CCFD. The precision score of 99.16% shows that the model made minimal errors in detecting fraudulent transactions, which resulted in a very low false positive rate. The recall of 98.98% implies that the model almost detected all fraud transactions and therefore its performance in minority class instances was excellent. The F1-score of 99.00% reflects the balance between precision and recall, thus, revealing the model ability to produce similar results in both criteria. A specificity of 98.52% means that the model was very effective at consistently identifying legitimate transactions thereby reducing the rates of misclassification significantly.

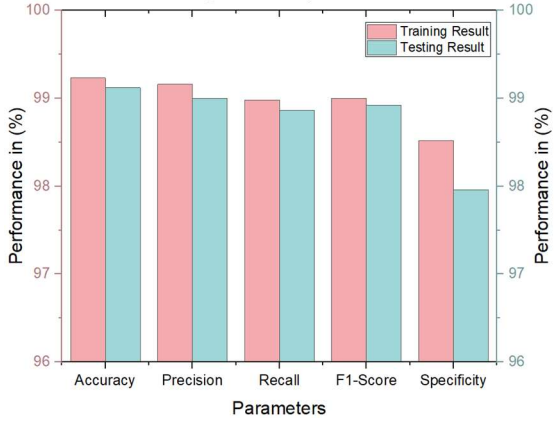

Figure 3: Graphical Illustration of Research Model's Results.

The proposed model has shown outstanding generalization during the evaluation phase, with an accuracy of 99.12% and a precision of 99.00%, which are very close to the training values. The closeness of the training and testing results is a sign that overfitting is not an issue and it also confirms the STTN-CP model's capability to transfer to new data. The recall and F1-score values of 98.86% and 98.92%, respectively, point to the fact that the model was very sensitive in detection and it also kept consistency in classification even though there was a lot of variability in the transactions of the real-world. The specificity of 97.96% is an indication of the STTN-CP model's ability to accurately and reliably identify legitimate transactions while at maintaining the false positive rate very low. To recap, the STTN-CP architecture not only copes with spatial-temporal interdependence but also isolates the unique properties, thus making it a very accurate and robust choice for CCFD.

Table 5: Statistical Performance Analysis of the Proposed Model (Mean ± STD).

Parameters	Training performance	Testing performance	Statistical Performance (Mean ±STD)
Accuracy	99.23	99.12	99.18 ± 0.08
Precision	99.16	99.00	99.08 ± 0.11
Recall	98.98	98.86	98.92 ± 0.08
F1-Score	99.00	98.92	98.96 ± 0.06
Specificity	98.52	97.96	98.24 ± 0.40

Table 5 shows the statistical performance analysis of the STTN-CP model, providing mean and standard deviation (STD) figures for every metric. The performance statistics of the proposed model are presented as Mean ± Standard Deviation and it is shown that the model is uniform and dependable during the training and testing phases. The model's average accuracy is 99.18 ± 0.08% which implies that the model is extremely reliable in distinguishing between fraudulent and non-fraudulent transactions. The precision (99.08 ± 0.11%) is very high meaning that the model's performance in eliminating false positives is good and those transactions marked as suspicious. In a similar vein, the recall rate of 98.92 ± 0.08% represents the power of the system in correctly spotting fraud, while the F1-score of 98.96 ± 0.06% illustrates the strong connection between the accuracy and the reliability of the detection. The specificity (98.24 ± 0.40%) infers that the model is able to accurately identify the legitimate transactions thus leading to reduction in false positives. The very small standard deviations for all metrics indicate that the proposed Spatial–Temporal Transformer with Contrastive Pretraining (STTN-CP) model offers inconsistent, weak, and untransferable performance, rendering it unsuitable for real-world CCFD applications.

The ablation study presented in Table 6 substantiates the STTN-CP framework by demonstrating the effect of each component sequentially.

Table 6: Ablation Study of the Proposed STTN-CP Model.

Configuration Description	Accuracy	Precision	Recall	F1-score	Specificity
Spatial Transformer only	97.25	97.06	96.85	96.94	95.80
Temporal Transformer only	97.10	96.88	96.52	96.70	95.60
Spatial + Temporal Transformer (without Contrastive Pretraining)	98.46	98.21	97.98	98.09	96.90
Spatial + Temporal Transformer + Pretraining (without SMOTE)	98.03	97.84	97.60	97.72	96.50
Proposed model Spatial + Temporal + Contrastive Pretraining + SMOTE	99.12	99.00	98.86	98.92	97.96

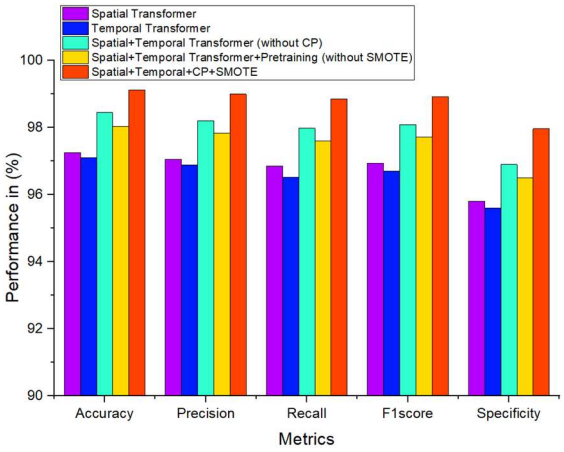

Figure 4: Graphical Illustration of Ablation Study Results Comparison.

The model receives a high score when either the Spatial Transformer or the Temporal Transformer is used independently, which means that the features that are obtained from each component are limited when they are working in isolation. The Spatial-Temporal Transformer improves the detection integration of both modules, showing that the spatial and temporal data unite, which is its strength. In addition, the contrastive pretraining acts as a filter that separates the strong from the weak feature representations produced by the model, making it easier to detect fraud in transactions that are compared to the genuine ones. The use of SMOTE has been credited with increasing both recall and specificity, thus highlighting the importance and advantages of dealing with class imbalance in fraud detection tasks. The entire STTN-CP setup, which contains all parts, has sealed the best accuracy (99.12%) and F1-score (98.92%), thus showing that the integration of each module works together to improve the detection of fraud.

Table 7: Comparison Analysis of Results

Models	Accuracy	Precision	Recall	F1score
Hybrid CNN-LSTM [14]	97.2	98.9	-	-
FinGraphFL [15]	98.39	-	-	-
Improved VAEGAN [17]	-	94.7	-	88.4
LSTM + Attention [19]	96.72	98.85	91.91	-
Adaboost+LGBM [21]	-	97	-	77
RF [37]	96	-	97	-
SVDD [35]	91	87	97	92
DCNN [36]	97.39	40.21	-	-
CNN-SVM [34]	91.08	90.50	90.34	90.41
Encoder–decoder GNN [38]	97.00	82.00	92.00	86.00
PROPOSED MODEL	99.12	99.00	98.86	98.92

In Table 7, a thorough comparison of the contemporary advanced methods, such as FinGraphFL, Hybrid CNN-LSTM, Improved VAEGAN, LSTM with Attention, Adaboost+LGBM, RF, SVDD, DCNN, CNN-SVM, and Encoder–Decoder GNN, along with the

proposed Spatial–Temporal Transformer with Contrastive Pretraining (STTN-CP) model is presented. The assessment is done through the major performance metrics of recall, accuracy, precision, and F1-score, which together evaluate the model's capability to classify fraudulent and legitimate transactions accurately.

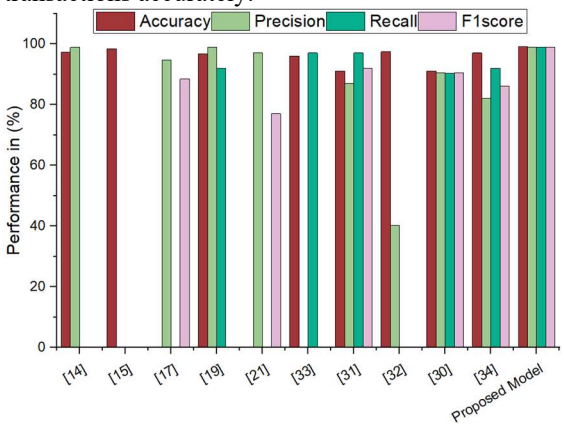

Figure 5: Graphical Illustration of Comparison of Results.

The proposed model, as illustrated in Figure 5, surpasses the existing techniques in terms of performance benchmarks all the time. The STTN-CP model recorded a 99.12% accuracy which is far higher than that of Hybrid CNN-LSTM (97.2%), FinGraphFL (98.39%), and Encoder-Decoder GNN (97.00%) models. This very increase proves beyond any shadow of doubt that the model is very effective in detecting the real and fraud classes accurately. The methods of contrastive pretraining and the combined spatial–temporal attention mechanism are the backbone of effective feature representation and accurate detection. The new model records a precision score of 99.00%, which signifies a great trustworthiness in eliminating false positives, or in-correctively classifying legal transactions as fraudulent. The proposed system has discrimination power over the previous techniques like Improved VAEGAN (94.7%) and CNN-SVM (90.50%) mainly due to proper embedding learning and noise reduction via contrastive pretraining that the model has accomplished.

The suggested strategy achieves a recall rate of 98.86%, surpassing the 91.91% recall rate of the LSTM + Attention model and the 97% of the SVDD model. A high recall rate indicates that the model is proficient in identifying genuine fraud situations, resulting in fewer undiscovered fraudulent transactions. The primary reason for this is the temporal transformer's capacity to comprehend the sequence of transactions and to detect even minor fluctuations in spending patterns over time. The proposed model attained an F1 score of 98.92%, which is the highest among all the other methods compared and it is the result of the balancing of precision and recall. For example, Improved VAEGAN and Encoder–Decoder GNN got 88.4% and 86.0% respectively. The higher F1 score indicates that the STTN-CP model is able to find the equilibrium between sensitivity and reliability, thus, it continues to be the best in fraud detection even though the data is not evenly distributed.

The comparison results show that the STTN-CP model always surpasses the baseline models of reference in all the evaluation metrics. The combination of the spatial-temporal attention mechanisms, contrastive pretraining, and SMOTE for balanced preprocessing contributes to the overall enhancing of the model's ability to separate legitimate from fraudulent credit card transactions.

Discussion and Justification of Findings:

The obtained results clearly demonstrate the effectiveness of the proposed STTN-CP model in addressing key challenges identified in the literature, including class imbalance, evolving fraud patterns, and complex transaction dependencies. Unlike many existing studies that primarily rely on accuracy and ROC-AUC, this work adopts a comprehensive evaluation framework incorporating precision, recall, F1-score, and specificity, which are more appropriate for highly imbalanced fraud detection tasks. In particular, recall and F1-score are critical, as failing to detect fraudulent transactions results in significant financial loss, while high precision minimizes false alarms and enhances user trust. Compared to prior approaches such as CNN–LSTM, VAEGAN, and federated learning models, the proposed method achieves superior and more balanced performance across all metrics, indicating improved robustness and generalization capability. The spatial–temporal transformer effectively captures both feature correlations and sequential transaction behavior, which are often only partially addressed in existing models, while contrastive pretraining enhances feature separability and representation quality. Furthermore, the integration of SMOTE with advanced deep representation learning enables better handling of minority class instances, leading to higher recall and F1-score. The low standard deviation values observed in the statistical analysis further confirm the consistency and stability of the model. Overall, these findings not

only align with but also extend previous research by demonstrating that a unified framework combining spatial–temporal modeling and contrastive learning can significantly improve fraud detection performance, thereby validating the proposed approach as a robust and scalable solution for real-world applications.

Advantages, Limitations, and Future Work:

The STTN-CP model that has been proposed is a notable beneficial to the field of CCFD. The applying of spatial-temporal transformers paired with contrastive pretraining has led to the very accurate pointing out of the inter-feature correlations and sequential behavioral patterns, which is shown by the very high accuracy, precision, recall, and F1-score metrics of the model. In addition to this, the utilization of SMOTE and Min-Max normalization in combination has resulted in the input data being balanced and standardized thus increasing the model's tolerance to class imbalance and scaling problems. The training process consumes a considerable amount of computer resources because of the complexity of the stacked transformer layers and contrastive learning, which could be a barrier to real-time deployment on devices with limited resources. The model's evaluation is based on one benchmark dataset alone; therefore, its applicability to other financial datasets and changing fraud patterns needs to be tested. One of the model's future potentials is its real-time adaptive learning capability, multi-class fraud detection, and cooperation with anomaly detection methods to discover new types of fraud. Moreover, an extensive study of lightweight transformer versions and explainable AI techniques could help to make the framework more user-friendly and adaptable, thus increasing its relevance in a financial sector that demands real-world applications.

5. CONCLUSION

This paper introduced the advanced CCFD system based on STTN-CP. The model was thoroughly examined in both the phases of training and evaluation with the use of the Credit Card Fraud Detection dataset that is publicly accessible in Kaggle. This dataset is characterized by a significant imbalance in the records of transactions, which means that a small fraction of the transactions are fraudulent ones. To mitigate this problem and to improve the performance of the model, Min-Max normalization and SMOTE were applied during preprocessing so as to produce standardized and balanced input features. The STTN-CP model processes a series of transformer blocks to elucidate spatial correlations among transaction attributes and temporal dependencies over transaction sequences, while the contrastive pretraining module boosts the discriminative power of the learned embeddings by clustering similar transactions and separating dissimilar ones. The findings of this study directly address the limitations identified in existing literature, particularly the inability of conventional models to jointly capture spatial–temporal dependencies and handle highly imbalanced datasets. The superior performance achieved across recall and F1-score demonstrates that the proposed model effectively detects minority class instances, which is critical in fraud detection scenarios. Additionally, the high precision and specificity confirm that the model minimizes false positives, thereby ensuring reliability in real-world financial applications. These results validate the effectiveness of integrating spatial–temporal transformers with contrastive pretraining as a unified framework for robust fraud detection. The experimental results demonstrate that the proposed model achieves a test set accuracy of 99.12%, precision of 99.00%, recall of 98.86%, F1 score of 98.92%, and specificity of 97.96%, thereby outperforming several state-of-the-art methods across multiple evaluation metrics. The ablation study further justifies the contribution of each component, demonstrating that the integration of spatial–temporal transformers, contrastive pretraining, and data preprocessing significantly enhances fraud detection performance compared to individual modules. These findings establish the proposed STTN-CP model as a scalable and effective solution for real-world credit card fraud detection, particularly in high-volume and dynamically evolving financial environments. The proposed model shows outstanding performance on the Kaggle dataset; however, the evaluation is limited to a single benchmark dataset, which may restrict the generalizability of the model across diverse financial environments. The stacked transformer layers' computational complexity might hinder the real-time implementation. Future work will focus on the development of techniques for transformer weight reduction, including multi-class fraud detection, real-time adaptive learning, and the use of explainable AI for model interpretability enhancement in practical banking and finance applications. Open Research Issues: Despite the strong performance of the proposed model, challenges such as real-time deployment, adaptability to evolving fraud patterns, model

interpretability, and validation across diverse real-world datasets remain open research issues.

AUTHOR CONTRIBUTIONS

Kathiresan Jayabalan: Writing – Original Draft, Methodology, Investigation, Formal Analysis, Data Curation, Conceptualization.

Sethuraman Radhakrishnan: Writing – Review & Editing, Supervision, Resources, Investigation, Conceptualization.

REFERENCES:

[1] T.A. Gaav, H.U. Adoga, T. Moses, "Recent advances in credit card fraud detection: An analytical review of frameworks, methodologies, datasets, and challenges," Journal of Future Artificial Intelligence Technology, Vol. 2, No. 3, 2025, pp. 343–369.

[2] N. Yousefi, M. Alaghband, I. Garibay, "A comprehensive survey on machine learning techniques and user authentication approaches for credit card fraud detection," arXiv preprint, 2019, pp. 1–29.

[3] A. Cherif, A. Badhib, H. Ammar, S. Alshehri, M. Kalkatawi, A. Imine, "Credit card fraud detection in the era of disruptive technologies: A systematic review," Journal of King Saud University – Computer and Information Sciences, Vol. 35, No. 1, 2023, pp. 145–174.

[4] F. Moradi, M. Tarrif, M. Homeai, "A systematic review of machine learnings in credit card fraud detections," MDPI Preprint, 2025, pp. 1–15.

[5] E. Oztemel, M. Issik, "A systematic review of intelligent system and analytical application in credit cards fraud detections," Applied Sciences, Vol. 15, No. 3, 2025.

[6] M. Alamri, M. Yikhlef, "Surveys of credit cards anomaly and fraud detections using sampling technique," Electronics, Vol. 11, No. 23, 2022, pp. 4003.

[7] I.D. Mienye, N. Jerre, "Deep learnings for credit cards fraud detections: A review of algorithm, challenge, and solution," IEEE Access, Vol. 12, 2024, pp. 1–20.

[8] E.A.L.M. Btoush, X. Zhau, R. Gururajam, K.C. Chen, R. Genrrich, P. Sankharan, "A systematic review of literatures on credit cards cyber frauds detections using machines and deep learnings," PeerJ Computer Science, Vol. 9, 2023, pp. 1–25.

[9] I.Y. Hafez, A.Y. Haffez, A. Salleh, A.A. Abd El-Maged, A.A. Abahany, "A systematic review of AI-enhanced techniques in credit cards fraud detections," Journal of Big Data, Vol. 12, 2025.

[10] R. Bin Sulaiman, V. Schettinin, P. Santt, "Reviews of machine learnings approach on credit cards fraud detections," Human-Centric Intelligent Systems, Vol. 2, 2022, pp. 55–68.

[11] S. Shi, W. Lou, G. Pua, "An attention-based balanced variational autoencoders method for credit cards fraud detections," Applied Soft Computing, 2025.

[12] I.D. Mienye, Y. Sun, "A deep learning ensembled with data resampling for credit cards frauds detections," IEEE Access, Vol. 11, 2023, pp. 30628–30638.

[13] N. Baisholan, J.E. Ditez, S. Gnatyak, M. Turdalyaly, E.T. Mattson, K. Baishalanova, "FraudXAI: An interpretable machine learnings framework for credit cards frauds detections on imbalanced dataset," Computers, Vol. 14, 2025, pp. 1–18.

[14] E. Ileberi, Y. Sun, "A hybrid deep learnings ensembled model for credit cards frauds detections," IEEE Access, Vol. 12, 2024, pp. 175829–175838.

[15] Z. Xia, S.C. Saaha, "FinGraphFL: Financial graphs-based federated learning for enhanced credit cards frauds detections," Mathematics, Vol. 13, No. 9, 2025.

[16] N. Nguyen, T. Duang, T. Chua, V.H. Nguayen, T. Trinah, D. Traan, T. Ho, "A proposed model for cards frauds detections based on CatBoost and deep neural networks," IEEE Access, Vol. 10, 2022, pp. 96852–96861.

[17] Y. Ding, W. Kag, J. Fang, B. Pang, A. Yeng, "Credit cards frauds detections based on improved Variational Autoencoders Generative Adversarial Networks," IEEE Access, Vol. 11, 2023, pp. 83680–83691.

[18] I. Akour, N. Mohammed, S. Saloum, "Hybrid CNN-LSTMs with Attention Mechanisms for Robust Credit Cards Frauds Detections," IEEE Access, Vol. 13, 2025, pp. 114056–114068.

[19] I. Benchaji, S. Doazi, B. El Ouahadi, J. Jafari, "Enhanced credit cards frauds detections based on attention mechanisms and LSTM deep models," Journal of Big Data, Vol. 8, No. 1, 2021.

[20] M. Abdul Sallam, K.M. Foaud, D.L. Ellbably, S.M. Elssayed, "Federated learning models for credit cards frauds detections with data balancing technique," Neural Computing and Applications, Vol. 36, No. 11, 2024, pp. 6231–6256.

[21] X. Zhao, Y. Lui, Q. Zhoa, "Improved LightGBMs for extremely imbalanced data and applications to credit cards fraud detections," IEEE Access, Vol. 12, 2024, pp. 159316–159335.

[22] M. Jabeen, S. Ramzzan, A. Razza, N.L. Fitrriyani, M. Siyafrudin, S.W. Le, "Enhanced Credit Cards Frauds Detections Using Deep Hybrid CLST Models," Mathematics, Vol. 13, No. 12, 2025.

[23] M. Adil, Z. Yinnjun, M.M. Jamjom, Z. Ulah, "OptDevNet: An Optimized Deep Events-based Networks Framework for Credit Cards Fraud Detections," IEEE Access, Vol. 12, 2024, pp. 132421–132433.

[24] I.D. Mienye, T.G. Siwart, "A hybrid deep learning approach with generative adversarial networks for credit cards fraud detections," Technologies, Vol. 12, No. 10, 2024.

[25] E. Ileberi, Y. Sun, Z. Weng, "A machine learning based credit cards fraud detections using the GA algorithms for feature selections," Journal of Big Data, Vol. 9, No. 1, 2022.

[26] Z.H. Mohammed, N.J. Ibrahim, A.K. Abbas, "Detecting Credit Card Fraud Using a Hybrid CNN-RNN Model," Journal of Information and Computer Technology Education, Vol. 9, No. 2, 2025, pp. 1–7.

[27] M.S.A. Mozumder, M.B.H. Sakil, M.R. Hasan, M.A. Hasan, K.N.R. Fuad, M.F. Mridha, M.R. Islam, Y. Watanobe, "Hybrid contrastive learning with attention-based neural networks for robust fraud detection in digital payment systems," IEEE Open Journal of the Computer Society, 2025, pp. 1–15.

[28] Y. Chen, K. Zhang, H. Zhu, Z. Qiu, "A novel federated transfer learning framework for credit card fraud detection under heterogeneous data conditions," Risks, Vol. 13, No. 11, 2025.

[29] M.S.A. Yajid, N. Bhosle, G. Sudhamsu, A. Khatibi, S. Sharma, R. Jeet, R. Sivaranjani, A. Bhowmik, A. Johnson Santhosh, "Hybrid Big Bang-Big Crunch with cuckoo search for feature selection in credit card fraud detection," Scientific Reports, Vol. 15, No. 1, 2025.

[30] M.V. Dinesh, "Comparative Analysis of Machine Learning Models for Credit Cards Fraud Detections," International Journal of Engineering Research and Technology, Vol. 13, No. 4, 2024, pp. 1–12.

[31] L. Wang, M. Coa, Y. Zheng, X. Yun, "Spatial-temporal transformers for videos snapshots compressive imaging," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 45, No. 7, 2022, pp. 9072–9089.

[32] M. Xu, W. Dai, C. Lu, X. Gou, W. Li, G.J. Qie, H. Xiang, "Spatial-temporal transformers network for traffic flows forecasting," arXiv preprint, 2020, pp. 1–15.

[33] H. Yao, T. Lu, R. Zau, S. Deng, Y. Xa, "A spatial-temporal transformers architecture using multi-channel signals for sleep stages classifications," IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 31, 2023, pp. 3353–3362.

[34] T. Berhane, T. Melesse, A. Waleligin, A. Mohamed, "A hybrid convolution neural networks and support vector machines-based credit cards fraud detections model," Mathematical Problems in Engineering, 2023.

[35] A. Mniai, M. Tarrik, K. Jebbari, "A novel framework for credit cards frauds detections," IEEE Access, Vol. 11, 2023, pp. 112776–112786.

[36] J. Kartika, A. Sentilselvi, "Smart credit cards frauds detection systems based on dilated convolution neural networks with sampling techniques," Multimedia Tools and Applications, Vol. 82, 2023, pp. 31691–31708.

[37] J.K. Afriyie, K. Tawah, W.A. Pells, S. Adai-Hene, H.A. Dwamenna, E.O. Owirredu, S.A. Ayah, J. Eshan, "A supervised machine learning algorithm for detecting and predicting frauds in credit cards transactions," Decision Analytics Journal, Vol. 6, 2023.

[38] A. Cherif, H. Amar, M. Kallkatawi, S. Alshehari, A. Immine, "Encoders–decoders graph neural networks for credit cards fraud detections," Journal of King Saud University – Computer and Information Sciences, Vol. 36, No. 3, 2024.