# Credit Card Fraud Detection Using Machine Learning Models: A Study Review

**Al-Maha Hashem Al-Wadie** [1,*]
**Abdullah Hussein Al-Hashedi** [2,*]



[1] Department of Information Technology, University of Science and Technology, Sana'a, Yemen
[2] Department of Information Systems, University of Science and Technology, Sana'a, Yemen
* Corresponding authors: 202310101567@student.ust.edu.ye, a.alhashedi@ust.edu.ye

https://doi.org/10.59222/ustjet.4.1.5

# Credit Card Fraud Detection Using Machine Learning Models: A Study Review

## Abstract:

Advances in technology have led to the emergence of digital payments. As credit card transactions have become the most common payment method and increasingly sophisticated methods have been adopted by fraudsters, detecting fraudulent transactions has become crucial due to the financial losses incurred by both cardholders and financial institutions. This study aims to review existing research on Credit Card Fraud Detection (CCFD) to provide future researchers with insights into the latest Machine Learning (ML) models applied in this field. This study adopts a critical literature review approach to explore and analyze current studies. With the rapid advancements in CCFD using ML models, which have addressed numerous challenges in this domain, it has become increasingly difficult to identify the techniques that contribute most significantly to its development, as well as the research gaps that require further research. Therefore, this review identifies 20 research articles published between 2022 and 2025. The focus on studies from the past four years ensures coverage of up-to-date developments and the latest technologies. The findings of this review highlight the most effective ML models and CCFD datasets, identify key research gaps, and outline the key evaluation metrics utilized in the field of CCFD, thereby supporting future studies.

**Keywords:** Machine learning, credit card fraud detection, fraudsters methods, financial fraud.


https://doi.org/10.59222/ustjet.4.1.5

# الكشف عن الاحتيال في بطاقات الائتمان باستخدام نماذج التعلم الآلي: دراسة مراجعة

## الملخص:

**لقد أدى التقدم في التكنولوجيا إلى ظهور المدفوعات الرقمية. نظرا لأن معاملات بطاقات الائتمان أصبحت طريقة الدفع الأكثر شيوعا واعتمد المحتالون أساليب متطورة بشكل متزايد، فقد أصبح اكتشاف المعاملات الاحتيالية أمرا بالغ الأهمية بسبب الخسائر المالية التي يتكبدها كل من حاملي البطاقات والمؤسسات المالية. تهدف هذه الدراسة إلى مراجعة الأبحاث الحالية حول كشف الاحتيال في بطاقات الائتمان (CCFD) لتزويد الباحثين المستقبليين برؤى حول أحدث نماذج التعلم الآلي (ML) المطبقة في هذا المجال. تتبنى هذه الدراسة منهج مراجعة الأدبيات النقدية لاستكشاف وتحليل الدراسات الحالية. مع التقدم السريع في CCFD باستخدام نماذج ML، والتي عالجت العديد من التحديات في هذا المجال، أصبح من الصعب بشكل متزايد تحديد التقنيات التي تساهم بشكل كبير في تطويرها، فضلا عن الفجوات البحثية التي تتطلب مزيدا من البحث. ولذلك، تحدد هذه المراجعة 20 مقالة بحثية منشورة بين عامي 2022 و2025. ويضمن التركيز على الدراسات التي أجريت خلال السنوات الأربع الماضية تغطية التطورات الحديثة وأحدث التقنيات. تسلط نتائج هذه المراجعة الضوء على نماذج ML ومجموعات بيانات CCFD الأكثر فعالية، وتحدد الفجوات البحثية الرئيسية، وتحدد مقاييس التقييم الرئيسية المستخدمة في مجال CCFD، وبالتالي دعم الدراسات المستقبلية.**

**الكلمات المفتاحية : التعلم الآلي، كشف الاحتيال في بطاقات الائتمان، أساليب المحتالون، الاحتيال المالي.**

## 1. Introduction

Technology has flourished in recent years and digital payment services such as e-wallets have emerged, credit card transactions are the most common method of payment [1]. The extensive application of credit cards, coupled with the diverse transaction contexts that lack stringent verification and oversight, unavoidably results in financial losses; this is caused by credit card fraud [2]. Credit card fraud is characterized as the illicit appropriation of credit card data to facilitate transactions or acquire monetary resources, leading to considerable economic detriment for both individuals and organizations [3].

Reports indicated a loss of $24.2 billion in the world in 2018 due to credit card fraud. Considering that there are millions of people who use credit cards in the world, there are expectations that losses resulting from credit card fraud will reach $40 billion in 2027 [4]. Due to recent advances in the field of Machine Learning (ML), detection of online credit card fraud on transaction data becomes one of the hot topics [1].

The fraud in credit card is an imperative issue, and the widespread use of credit cards has led to increased fraudulent behavior [5]. The methods used by fraudsters to obtain financial account information in an unauthorized manner are considered sophisticated [6]. This study presents a review of the existing literature on Credit Card Fraud Detection (CCFD) using ML techniques. Furthermore, it discusses the most widely used algorithms and the key evaluation metrics applied in this field, while also identifying the limitations and research gaps in existing studies.

The remainder of this paper's organization begins with the types of credit card fraud and fraudulent methods in Section 2, the objective of the study in Section 3, background of CCFD in Section 4, the literature review of CCFD in Section 5, the evaluation metrics and datasets in Sections 6 and 7, the research gaps in Section 8, the methodology in Section 9, and the results with discussion in Section 10. The last, Section 11, is a conclusion of the study.

## 2. Types of Credit Card Fraud and Fraudulent Methods

There are two main types of credit card fraud, also there are other types that fall under these two types. Fraudsters use various methods to obtain cardholder information. The first and most common type is card not present (CNP), which involves using card data online. Its types include phishing


https://doi.org/10.59222/ustjet.4.1.5

where fraudsters trick the victim via calls or messages to obtain card-specific data such as the password, account takeover where fraudsters conduct transactions in the customer's name by obtaining the customer's login data and seizing his bank account, and application fraud when fraudsters using fake identity data or documents to obtain a new card. The second type is card present (CP), which includes using the actual card. Its types include skimming when fraudsters copy card data by swiping it through ATMs or points of sale, counterfeit cards when fraudsters create a fake printout of the original card and use it as if it were the original card, and lost or stolen card, these are the most common, but there are many methods that fraudsters are constantly developing [7].

## 3. Objective of the Study

This study aims to review existing studies on CCFD in order to provide future researchers with insights into the state-of-the-art ML techniques applied in this field.

## 4. Theoretical Background

This section provides a theoretical background on CCFD, machine learning and also introduces ensemble learning, its types, and deep learning.

### 4.1. Credit Card Fraud Detection (CCFD)

CCFD is crucial due to important economic losses for institutions and customers [8]. Is a complicated but vital field that combines cutting-edge technology and procedures to shield customers and monetary institutions from possible losses [9]. Owing to the substantial volume of credit card transactions that financial services are required to manage on a daily basis, it becomes imperative to automate the categorization of transactions as either fraudulent or legitimate. To develop such a system, algorithms must be trained on datasets comprising labeled transactions of fraud and non-fraud in order to identify patterns of fraudulent activity in forthcoming transactions, thereby eliminating the necessity for human intervention in the process [10].

### 4.2. Machine Learning (ML)

Machine Learning (ML) is a fundamental component of artificial intelligence, focusing on creating algorithms that use statistical inference and optimization to extract predictive patterns from data. It often outperforms rule-based systems in complex decision-making situations by identifying correlations and anomalies that traditional methods miss. ML models are trained on past

transaction data to distinguish between fraudulent and legitimate patterns in the context of CCFD [11].

### 4.3. Ensemble Learning (EL)

Ensemble methods are considered advanced paradigm in ML. The idea is multiple models are trained on the same task, and their outputs are then combined to obtain a final prediction that is often better than any single model. This means that a single model may be limited in either accuracy or stability, while ensemble enables the utilization of model and data differences to improve overall performance by capitalize on their distinct advantages, culminating in enhanced accuracy, robustness, and generalizability. Their approaches are stacking, boosting, and bagging [12]. Ensemble approaches have appeared as an effective strategy for detecting credit card fraud [13]. Figure 1 illustrates a general architecture for an EL model predicated on supervised classification algorithms [14].

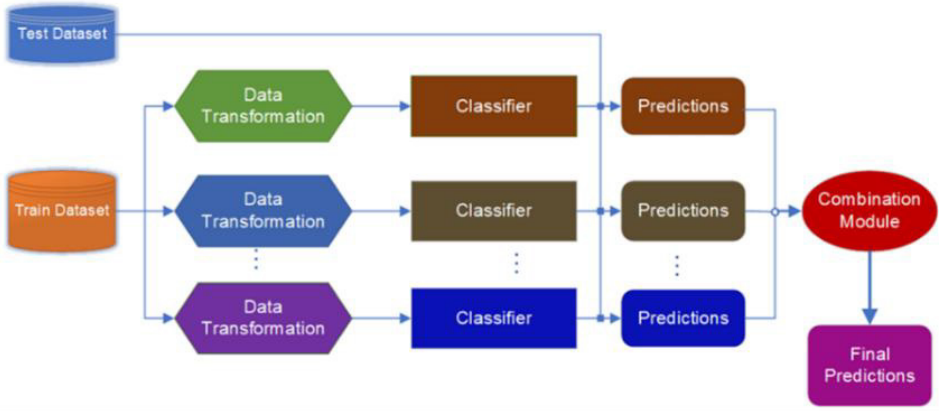


Figure 1: A general architecture for an EL model [14]

This subsection provides an overview about approaches of ensemble learning.

#### *4.3.1. Stacking*

The stacking ensemble approach is an effective machine learning methodology that combines multiple basic classifiers to increase prediction performance, used in classification and regression. Stacking includes independent training of several base learners and their predictions are integrated via a meta-learner. This technique generally utilizes a two-tiered framework whereby the outputs of base learners are input into a secondary model that generates the ultimate prediction [12].


https://doi.org/10.59222/ustjet.4.1.5

### *4.3.2. Boosting*

Boosting ensemble approaches are effective ways for improving prediction performance and improve weak learners, used in classification and regression. Gradient boosting frameworks like eXtreme Gradient Boosting (XGBoost), Light Gradient-Boosting Machine (LightGBM), and Categorical Boosting (CatBoost) are popular because to their effectiveness and precision in structured data workloads. These frameworks create models repeatedly, correcting prior faults to produce extremely accurate ensembles. This approach differs from other ensemble learning strategies in that the weak learner produces each new model with the intention of repairing the mistakes of the prior models [12]. Its ability to improve prediction accuracy, especially in complex datasets, has made it a strong choice in the field of fraud detection [15].

### *4.3.3. Bagging*

Bagging known as Bootstrap Aggregating, used in classification and regression. It works by generating multiple bootstrap samples from the original dataset, then training independent base models of the same model type on each sample, then lastly combining their predictions using voting in classification or averaging in regression. This approach minimizes variability and improves stability of predictions. Random Forest (RF) is an example of bagging approaches [12]. Figure 2 illustrates how bagging different from boosting.

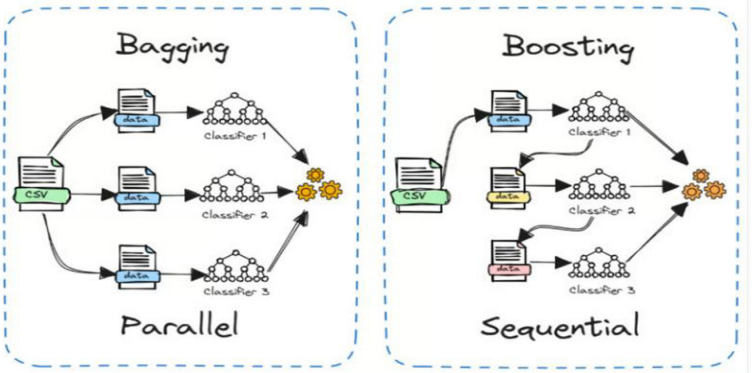


Figure 2: Bagging versus Boosting (figure source: [16])

### 4.4. Deep Learning

Deep Learning (DL) has gained importance in academic study [17]. DL is a branch of artificial intelligence and has emerged as an advanced paradigm in Machine Learning (ML) that utilizes deep neural architectures to extract complex patterns from data. Focuses on building and training deep neural networks comprising a large number of hidden layers [18]. DL methods automatically understand representations required for detection or classification unlike conventional ML methods that essential a lot of feature engineering. DL refers to ML approaches that employ neural networks with several layers, including unsupervised, semi-supervised, and supervised models [19].

DL architectures considerably outperform conventional ML methods for CCFD in huge datasets. DL models excel at investigating complex patterns in credit card transaction data where conventional methods struggle to handle complex patterns effectively [20].

## 5. Literature Review

In recent years, numerous studies have been conducted in the field of CCFD using ML techniques. This section provides a review of published studies and identifies existing efforts.

Albalawi & Dardouri [21] proposed several ML models, including Logistic Regression (LR), Decision Tree (DT), RF, and XGBoost evaluated on the European Credit Card Transactions Dataset in 2013 (ECCD 2013), and PaySim synthetic dataset. The study also incorporated focal loss into a deep learning model to further enhance CCFD performance. The study explicitly indicated that the RF model achieved the best overall performance, with an accuracy of 99.95%, F1 score of 0.8256, and Receiver Operating Characteristic - Area Under Curve (ROC-AUC) of 0.9759. The deep learning model provided the highest precision, demonstrating its potential in minimizing false positives. However, when reviewing the tables and actual results of all models implemented in the study, it was confirmed that the XGBoost model achieved the highest accuracy, precision, F1 score. There is a contradiction in the results mentioned in the study textually with those mentioned in the tables. The RF model actually achieved 99.69% accuracy, not 99.95% and F1 score of 74.42%. Anyway, the RF model achieved 100% in recall which indicates its efficiency in detecting fraudulent transactions.


https://doi.org/10.59222/ustjet.4.1.5

Similarly, Alrasheedi [22] introduced a comprehensive approach to enhancing credit card fraud detection by employing a diverse set of ML algorithms such as Support Vector Machine (SVM), DT, RF, XGBoost, Adaptive Boosting (AdaBoost), Multilayer Perceptron (MLP), and Artificial Neural Network (ANN). The study utilized three datasets and reported that XGBoost, MLP, DT, and RF achieved the highest accuracy of 99% on the European Credit Card Transactions Dataset in 2023 (ECCD 2023). Salunke et al. [23] presented ML models, including LR, DT, RF, and stacking ensemble. The models were evaluated on the ECCD 2013 dataset. The stacking model outperformed the other models, achieving a high accuracy of 99.95%. In another research, Siam et al. [24] proposed a hybrid framework for features selection to improve the performance of CCFD models, with applying EL algorithms across five datasets. Their results demonstrated that the Extra Trees (ET) model achieved high accuracy 99.99% on the ECCD 2023 dataset.

Furthermore, Tayebi & El Kafhali [25] developed an enhanced XGBoost model by tuning its hyperparameters using Bayesian optimization to detect fraudulent transactions on two datasets: the ECCD 2013, and the IEEE-CIS dataset. The model achieved accuracies of 99.96% on the ECCD 2013 dataset and 83.25% on the IEEE-CIS dataset. In a study by Wu et al. [26] the authors proposed a Continuous-Coupled Neural Network (CCNN) method, which achieved an accuracy of 99.98% on the ECCD 2013 dataset. In addition, Xia & Saha [27] introduces an innovative approach based on Graph Networks–driven federated learning (GraphFL). The study utilized both the ECCD 2023 and the ECCD 2013 datasets, achieving accuracies of 98.3% on the ECCD 2023 dataset and 97.8% on the ECCD 2013 dataset. Another study by Alamri & Ykhlef [28] presented an approach that combines feature aggregation with Exhaustive Feature Selection (EFS). The authors employed four models including RF, DT, LR, and Deep Neural Network (DNN). The reported accuracies were 99.97% for DT, 99.96% for RF, 99.94% for DNN, and 98.11% for LR on the PaySim dataset.

Furthermore, the study by Aslam & Hussain [29] applied various models including LR, RF, ET, LightGBM, XGBoost, and CatBoost. The reported accuracies were 99.99% for both RF and ET, 99.97% for XGBoost, 99.95% for CatBoost, 99.93% for LightGBM, and 96.50% for LR using ECCD 2023 dataset. In Ghrib et al. [13] an ensemble model that integrates Bidirectional GRU (BiGRU) and Bidirectional LSTM (BiLSTM) was presented. The model achieved an Area Under the Curve (AUC) of 91.23% on the ECCD 2013

dataset. The authors in Jemai et al. [30] proposed an ensemble learning methods to identifying fraudulent transactions in credit cards including XGBoost, RF, Naive Bayes (NB) classifier. The study utilized both ECCD 2013 and Sparkov simulated datasets. XGBoost achieved a high classification accuracy of approximately 96% on the ECCD 2013 dataset.

In another research, Khalid et al. [31] proposed a novel ensemble model that integrates SVM, RF, k-nearest neighbors (KNN), Bagging, and Boosting classifiers. The model achieved an accuracy of 99.95% on the ECCD 2013 dataset. Furthermore, Mienye & Swart [32] developed hybrid DL framework that integrates Generative Adversarial Network (GAN) with Recurrent Neural Networks (RNNs) including Simple RNN, Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU). Among the evaluated models, the GAN-GRU achieved the highest recall, recording 0.992 on the ECCD 2013 dataset and 0.920 on the Brazilian dataset. Sulaiman et al. [33] demonstrated that LSTM model outperformed Autoencoders (AE) and Convolutional Neural Network (CNN) models, achieving an accuracy of 99.2% on the ECCD 2013 dataset.

Afriyie et al. [34] proposed supervised machine learning models such as DT, LR, and RF for credit card fraud detection. Using a fraud detection dataset comprising simulated U.S. transactions conducted in 2020. The RF model outperformed the other classifiers, achieving an accuracy of 0.96, a recall of 0.97, a specificity of 0.96, and an AUC of 0.989. Mienye & Sun [35] introduced a stacking ensemble framework that employed LSTM, GRU, as base learners with MLP as the meta learner. The study utilized both the ECCD 2013 and Taiwan datasets. The framework achieved recall of 1.000 on the ECCD 2013 dataset and 0.930 on the Taiwan dataset. Similarly, Raval et al. [36] proposed an LSTM-based model incorporating an explainability method, denoted to as X-LSTM. The study utilized two datasets: the ECCD 2013 dataset and the Credit Approval dataset from the UCI Machine Learning Repository. The proposed model achieved an accuracy of 99.8% on the ECCD 2013 dataset, and 80% on the Credit Approval dataset.

Another study by Alfaiz & Fati [37] developed a hybrid All K-Nearest Neighbors- CatBoost (AllKNN-CatBoost) model, which achieved an AUC 97.94%, Recall 95.91%, F1-Score 87.40% on the ECCD 2013 dataset. The authors in Esenogho et al. [38] presented an ensemble model that integrates LSTM as base learner within the AdaBoost framework. The proposed model


https://doi.org/10.59222/ustjet.4.1.5

achieved a sensitivity of 0.996, a specificity of 0.998, and an AUC of 0.990 on the ECCD 2013 dataset. Malik et al. [39] developed a hybrid model that integrates AdaBoost and LightGBM, which achieved a precision of 97.00 on the IEEE-CIS dataset. Table 1 summarize the previous studies of CCFD.

Table 1: Summary of previous CCFD studies

| No. | Author(s) & Year | Publisher/ Journal | Method(s) | Dataset(s) | Result | Limitation |
|---|---|---|---|---|---|---|
| 1 | Albalawi & Dardouri (2025) | Frontiers Media/ Frontiers in Artificial Intelligence | LR, DT, RF, XGBoost, Neural Network with Focal Loss (NN) | ECCD 2013, PaySim synthetic | XGBoost achieved the highest accuracy (Acc), precision, and F1 score, while LR, RF, and NN achieved the best recall | There are no explanations for why a transaction was classified as fraudulent, which reduces confidence in the model. |
| 2 | Alrasheedi (2025) | Springer/ Computational Economics | SVM, DT, RF, XGBoost, AdaBoost, MLP, ANN | ECCD 2023, Customer credit card, and United states transactions | The highest Acc was 99% by each the XGBoost, MLP, DT, and RF models on the ECCD 2023 | Some of the datasets used are small in size. |
| 3 | Salunke et al. (2025) | Springer/ Cureus Journal of Computer Science | LR, DT, RF, Stacking Ensemble | ECCD 2013 | The stacking model outperformed the other models, achieving a high Acc of 99.95%. | Advanced techniques for addressing skewness in the dataset, such as SMOTE and ADASYN, were not used. |
| 4 | Siam et al. (2025) | Public Library of Science/ PLOS ONE | RF, ET, XGBoost, AdaBoost, CatBoost | ECCD 2013, German dataset, ECCD 2023, Australian dataset, and Abstract dataset | The ET model achieved high Acc 99.99% on the ECCD 2023 | Cost-sensitive aspects were not emphasized. |
| 5 | Tayebi & El Kafhali (2025) | Elsevier/Cyber Security and Applications | XGBoost with Bayesian optimization | ECCD 2013 and IEEE-CIS | The Acc achieved 99.96% on the ECCD 2013 and 83.25% on the IEEE-CIS | Despite the model's performance, there are no explanations to clarify the decisions. |

Table 1: Continued

| No. | Author(s) & Year | Publisher/ Journal | Method(s) | Dataset(s) | Result | Limitation |
|---|---|---|---|---|---|---|
| 6 | Wu et al. (2025) | MDPI / Mathematics | Continuous-Coupled Neural Network (CCNN) | ECCD 2013 | The Acc is 99.98% | Old dataset may not reflect current fraud patterns. |
| 7 | Xia & Saha (2025) | MDPI / Mathematics | Graph Networks based federated learning (GraphFL) | ECCD 2013, and ECCD 2023 | The Acc is 98.3% on the ECCD 2023 and 97.8% on the ECCD 2013. | The evaluation metrics are insufficient and there is no explanation for the model's decisions. |
| 8 | Alamri & Ykhlef (2024) | MDPI/ Electronics | Logistic regression, DT, RF, DNN | PaySim | The Acc of RF 99.96%, DT 99.97%, LR 98.11%, DNN 99.94% | Lack of exploration into feature engineering techniques and the use of the synthetic PaySim dataset. |
| 9 | Aslam & Hussain (2024) | Tech Science Press/Journal on Artificial Intelligence | LR, RF, ET, LightGBM, XGBoost, CatBoost | ECCD 2023 | The Acc of each RF and ET 99.99%, XGBoost 99.97%, CatBoost 99.95%, LightGBM 99.93%, LR 96.50% | Implementation details such as parameters were not mentioned, and there is no explanation for the model's decisions. |
| 10 | Ghrib et al. (2024) | Polish Association for Knowledge Promotion/ Applied Computer Science | Integrate BiGRU-BiLSTM Model | ECCD 2013 | AUC of 91.23% | The dataset used is outdated and doesn't reflect current fraud patterns, thus limiting the ability to generalize the results to other systems. |


https://doi.org/10.59222/ustjet.4.1.5

Table 1: Continued

| No. | Author(s) & Year | Publisher/ Journal | Method(s) | Dataset(s) | Result | Limitation |
|---|---|---|---|---|---|---|
| 11 | Jemai et al. (2024) | IEEE/ IEEE Access | Ensemble Learning Methods XGBoost, RF, NB classifier | ECCD 2013 and Sparkov simulated datasets | XGBoost achieved a high classification Acc of approximately 96% on the ECCD 2013 | Classifier performance was much poorer on synthetic datasets compared to real ones. |
| 12 | Khalid et al. (2024) | MDPI/Big Data and Cognitive Computing | Ensemble model integrates SVM, RF, KNN, Bagging, and Boosting classifiers | ECCD 2013 | Acc of 99.95% | Old dataset may not reflect current fraud patterns. |
| 13 | Mienye & Swart (2024) | MDPI/ Technologies | Hybrid DL framework GAN-RNN, GAN-LSTM, GAN- GRU | ECCD 2013 and Brazilian credit card | GAN-GRU achieved high result recall of 0.992 on the ECCD 2013, and 0.920 on the Brazilian dataset | There may be computational limitations when training the model and it takes a long training time. |
| 14 | Sulaiman et al. (2024) | Tech Science Press/ Computers, Materials & Continua | AE, CNN, LSTM | ECCD 2013 | The LSTM outperformed AE and CNN, achieving an Acc of 99.2% | Old dataset may not reflect current fraud patterns. |
| 15 | Afriyie et al. (2023) | Elsevier/ Decision Analytics | DT, LR, RF | Fraud Detection dataset (simulated U.S. transactions in 2020) | The Acc of RF 96%, DT and LR achieved same Acc of 92% | Advanced models were not used. |
| 16 | Mienye & Sun (2023) | IEEE/ IEEE Access | Stacking ensemble framework LSTM, GRU, as base learners with MLP as the meta learner | ECCD 2013 and Taiwan datasets | The model achieved recall of 1.000 on the ECCD 2013, and 0.930 on the Taiwan dataset | A lack of analysis on how evolution of transaction patterns may affect the model's capacity to adapt to shifting fraud practices. |

Table 1: Continued

| No. | Author(s) & Year | Publisher/ Journal | Method(s) | Dataset(s) | Result | Limitation |
|---|---|---|---|---|---|---|
| 17 | Raval et al. (2023) | MDPI/ Mathematics | LSTM model with Explainable method (X-LSTM) | ECCD 2013 and Credit Approval datasets | The model achieved acc of 99.8% on the ECCD 2013, and 80% on the Credit Approval | Despite the model's performance, it does not work efficiently with non-sequential data. |
| 18 | Alfaiz & Fati (2022) | MDPI/ Electronics | Hybrid model (AllKNN -CatBoost) | ECCD 2013 | AUC 97.94%, Recall 95.91%, F1-Score 87.40% | Insufficient validation and some resampling procedures have performed poorly, highlighting the need for more effective methods. |
| 19 | Esenogho et al. (2022) | IEEE/ IEEE Access | An ensemble model, LSTM as base learner within the AdaBoost | ECCD 2013 | Recall 0.996, Specificity 0.998, AUC 0.990 | The dataset is outdated and will not cover modern fraud patterns. |
| 20 | Malik et al. (2022) | MDPI/ Mathematics | Hybrid model AdaBoost and LightGBM | IEEE-CIS | The precision 97.00 | The study didn't focus on parameter adjustment for the algorithms utilized, and hybrid models can be complex. |

## 6. Evaluation Metrics

Many evaluation metrics were utilized to evaluate the efficiency of the models used in existing studies. The evaluation metrics include Accuracy, Recall, F1-Score, Precision, and AUC. The following subsection discusses the most significant evaluation metrics employed in this field.

### 6.1. Accuracy

Accuracy is the most often utilized metric for measuring the performance of a classification issue. Which represents the proportion of correctly predicted


https://doi.org/10.59222/ustjet.4.1.5

cases that includes both positive and negative out of the total number of cases [40]. The accuracy is expressed mathematically as in the Equation 1.

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

Where TP is True Positive, referring to the number of positive cases correctly predicted by the model; TN is True Negative, referring to the number of negative cases correctly predicted by the model. FP is False Positive referring to the number of cases that the model incorrectly predicted as positive when they were actually negative; FN is False Negative, referring to the number of cases that the model incorrectly predicted as negative when they were actually positive.

### 6.2. Precision

Precision is a performance metric that measures the proportion of correctly predicted positive cases (fraudulent transactions) out of all cases classified as positive by the model [11]. Precision measures the model's capability to avoid false alarms. In fraud detection, a low precision model often misclassifies normal transactions as fraudulent, causing customer frustration. Therefore, high precision is crucial for reducing operating costs and maintaining customer trust by minimizing false alarms. The precision is expressed mathematically as in the Equation 2.

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

### 6.3. Recall

Recall, also known as sensitivity or true positive rate (TPR), measures a model's ability to correctly classify all actual positive cases in a dataset [41]. In fraud detection, recall shows how effectively the model classifies fraudulent transactions among all actual fraudulent transactions. The recall is expressed mathematically as in the Equation 3.

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

### 6.4. F1-score

The F1-score is a performance metric that balances precision and recall [42]. F1-score is characterized by not being inclined towards precision or recall alone, but rather balancing them. The F1-score is expressed mathematically as in the Equation 4.

$$\text{F1-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

## 7. Datasets

In this section, we provide an overview of the datasets utilized in the studies covered in this review, Table 2 presents the datasets including their characteristics such as the number of transactions, the number of features, whether they are public or private, fraud ratio, and remarks about the datasets.

Table 2: The datasets of CCFD studies

| Utilized In | Dataset | Provider | No. of Transactions | No. of Features | Fraud Ratio | Type | Remarks |
|---|---|---|---|---|---|---|---|
| [13], [21], [23], [24], [25], [26], [27], [30], [31], [32], [33], [35], [36], [37], [38] | ECCD 2013 | ULB, Worldline | 284,807 | 30 | 0.172% | Public | The dataset is outdated and will not cover modern fraud patterns. |
| [22], [24], [27], [29] | ECCD 2023 | ULB (Elgiriyewithana – Kaggle) | 568,630 | 31 | 50% | Public | Updated dataset for the ECCD 2013 version. |
| [21], [28] | PaySim | UFU | 6.3 million | 11 | 0.129% | Public | The dataset is artificial. |
| [22] | Customer credit card | Data provider not disclosed | 689 records | 16 | 49.9% | Public | The dataset size is very small. |
| [22] | United states transactions | Data provider not disclosed | 14,446 | 15 | 12.77% | Public | The dataset size is very small. |
| [24] | German dataset | UCI ML Repository | 1,000 | 21 | 30% | Public | The dataset size is very small. |
| [24] | Australian dataset | UCI ML Repository | 690 | 15 | 55.51% | Public | The dataset size is very small. |
| [24] | Abstract dataset | Data provider not disclosed | 3,075 | 12 | 85.43% | Public | The dataset size is very small. |
| [25], [39] | IEEE-CIS | Vesta Corporation | 590,540 | 432 | 3.50% | Public | High dimensionality. |
| [30] | Sparkov simulator | Western Michigan University | 1,604,292 | 22 | 34.64% | Public | Transactions are generated by an emulator and not from a real banking system. |
| [32] | Brazilian credit card | Obtained from a large Brazilian bank | 374,823 | 17 | 3.74% | Private | Limiting the generalizability to other banks and countries. |


https://doi.org/10.59222/ustjet.4.1.5

Table 2: Continued

| Utilized In | Dataset | Provider | No. of Transactions | No. of Features | Fraud Ratio | Type | Remarks |
|---|---|---|---|---|---|---|---|
| [34] | Fraud Detection (simulated US transactions in 2020) | Data provider not disclosed | 555,719 | 23 | - | Public | Transactions are generated by an emulator and not from a real banking system. |
| [35] | Taiwan dataset | National Chung Cheng University, UCI ML Repository | 30,000 | 23 | - | Public | This data relates to default prediction, not fraud detection. |
| [36] | Credit Approval | UCI ML Repository | 690 | 15 | 22.1% | Public | The dataset size is very small. |

## 8. Research Gaps

This study highlights several important research gaps. Specifically, there is a lack of advanced interpretability methods in credit card fraud detection models. This limitation reduces confidence among cardholders and financial institutions, as they are unable to determine or justify which specific features contributed to a given decision. Additionally, current models exhibit limitations in feature engineering, particularly in the effective creation of temporal and behavioral features.

Furthermore, many previous studies have relied on outdated datasets that may not adequately capture or reflect modern fraudulent patterns. These gaps highlight the need for more advanced models capable of providing clear explanations for their decisions, for example, explain what features contributed to classifying a particular transaction as fraudulent. They also emphasize the importance of enhancing model performance through the creation of temporal and behavioral features, which can significantly improve the effectiveness of credit card fraud detection models.

## 9. Methodology

This study adopts a critical literature review approach to explore and analyze current research on ML techniques for credit card fraud detection. This approach makes it easier to conduct a comprehensive and flexible review of the literature, incorporating different study results, methodological perspectives, and technological developments in the field.

The review process involved identifying relevant scholarly articles published in peer-reviewed journals. These articles were obtained from reputable academic digital libraries and publishing platforms, including ScienceDirect, IEEE Xplore, SpringerLink, MDPI, PLOS, and Tech Science Press. The search strategy included sets of keywords such as «machine learning», «deep learning», «credit card fraud detection», and «fraudsters methods». Priority was given to recent publications to reflect the latest progresses in the credit card fraud detection field.

The collected studies were examined in terms of the algorithms employed, the datasets utilized, the evaluation metrics applied, the key findings, and the limitations. Following that, a critical analysis was conducted of the collected studies, to identify limitations in existing studies, performance comparisons, and research gaps.

## 10. Results and Discussion

According to the reviewed studies, ML models, particularly DL and EL models, have emerged as the most widely utilized. The ET and RF models achieved the highest performance among the EL models, demonstrating 99.99% accuracy and 100% recall. Similarly, the XGBoost, CatBoost, and LightGBM models exhibited strong predictive performance, with accuracy values ranging from 99.93% to 99.97%, while recall rates ranged between 99.91% and 100%. This is due to their ensemble structure and ability to model complex non-linear relationships inherent in CCFD datasets [29].

In deep learning models, the most commonly used models were DNN, LSTM, CCNN, MLP, and ANN. The CCNN model achieved the highest accuracy, reaching 99.98%, with a recall of 100% and a precision of 99.96%. This excellent performance due to the CCNN's ability to handle dynamically changing data more effectively and accurately than traditional neural networks [26]. Furthermore, the DNN, MLP, ANN and LSTM models exhibited strong predictive performance, with accuracy values ranging from 99.00% to 99.94%.

DL architectures considerably outperform conventional ML models for CCFD in huge datasets and investigating complex patterns in credit card transaction data where conventional models struggle to handle complex patterns effectively [20].


https://doi.org/10.59222/ustjet.4.1.5

However, traditional ML models such as LR, DT, SVM, and NB have also been widely applied for CCFD, often achieving high accuracy values ranging from 98.00% to 99.97%.

Advanced ML techniques such as EL and DL often outperform these traditional models, particularly with complex and large datasets. However, the traditional models themselves still demonstrate high effectiveness when trained on good features and with proper preprocessing [31]. Table 3 presents the evaluation metrics used in the studies. Several metrics were used, including Accuracy, Recall, Precision, F1-Score, AUC, and other metrics that are important in evaluating the performance of CCFD models. The symbol ✓ indicates that the metric was used in the study, and the symbol ✗ indicates that the metric was not used in the study.

Table 3: The evaluation metrics of CCFD studies

| Study Ref. | Evaluation Metrics | | | | | |
|---|---|---|---|---|---|---|
| | Accuracy | Recall | Precision | F1-Score | AUC | Other |
| [25] | ✓ | ✓ | ✓ | ✓ | ✓ | AUC-PR |
| [21], [23], [31], [37] | ✓ | ✓ | ✓ | ✓ | ✓ | |
| [35], [38] | ✗ | ✓ | ✗ | ✗ | ✓ | Specificity |
| [13] | ✗ | ✓ | ✓ | ✓ | ✓ | |
| [39] | ✗ | ✓ | ✓ | ✓ | ✓ | AUC-PR, Specificity, Misclassification Rate |
| [28] | ✓ | ✓ | ✓ | ✓ | ✗ | AUC-PR |
| [22], [26], [29] | ✓ | ✓ | ✓ | ✓ | ✗ | |
| [24] | ✓ | ✓ | ✓ | ✓ | ✓ | AUC-PR, MCC |
| [30] | ✓ | ✗ | ✓ | ✓ | ✗ | |
| [36] | ✓ | ✓ | ✓ | ✗ | ✓ | |
| [27] | ✓ | ✗ | ✗ | ✗ | ✓ | |
| [32] | ✗ | ✓ | ✓ | ✓ | ✓ | Specificity |
| [33] | ✓ | ✓ | ✗ | ✗ | ✓ | |
| [34] | ✓ | ✓ | ✓ | ✓ | ✓ | Specificity |

Table 4 presents a comparison of studies based on two main aspects: feature engineering and advanced interpretation techniques. With respect to feature

engineering, the focus is on the creation of temporal and behavioral features due to their contribution to improving model performance and understanding complex patterns in the data. Regarding advanced interpretation techniques, particular attention is given to Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) due to their ability to explain model decisions in classifying transactions as fraudulent or non-fraudulent, as well as their role in improving the reliability and dependability of the model for financial institutions. The symbol ✓ indicates that the aspect was used in the study, and the symbol ✗ indicates that the aspect was not used in the study.

Table 4: Comparison between studies

| Study Ref. | Feature Engineering | | Interpretation | |
|---|---|---|---|---|
| | Temporal features | Behavioral features | SHAP method | LIME method |
| [21], [25], [26], [27], [29], [13], [30], [31], [32], [33], [35], [37], [38], [39] | ✗ | ✗ | ✗ | ✗ |
| [28] | ✗ | ✓ | ✗ | ✗ |
| [24] | ✗ | ✗ | ✓ | ✗ |
| [36] | ✗ | ✗ | ✓ | ✓ |
| [34] | ✓ | ✓ | ✗ | ✗ |

We note that most studies have not focused on applying these aspects, despite their importance in the CCFD field.

Advanced interpretation methods such as SHAP, play a crucial role in enhancing the understanding of complex model decisions. This is particularly important in financial environments including the domain of CCFD, where transparency and explainability are essential [24].

## 11. Conclusion

In conclusion, this review highlights the significant progress made in CCFD using ML models, particularly ensemble learning and deep learning models, which have demonstrated excellent performance in detecting fraudulent transactions. Despite these advancements, several research gaps remain, such as the lack of use of advanced interpretation techniques, insufficient focus on creating temporal and behavioral features during feature engineering, and reliance on outdated datasets that may not reflect current fraud patterns.


https://doi.org/10.59222/ustjet.4.1.5

Future work should focus on developing more robust models capable of explaining their decisions in classifying transactions as fraudulent by employing advanced interpretation techniques such as SHAP to improve transparency and reliability. Moreover, feature engineering, by adding temporal and behavioral features, should also enhance the model's understanding of complex patterns within the data. Furthermore, utilize modern datasets to enhance generalizability.

## References

[1] 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, Dec. 2021, doi: 10.1186/s40537-021-00541-8.

[2] X. Zhang, Y. Han, W. Xu, and Q. Wang, "HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture," *Inf. Sci. (N Y).*, vol. 557, pp. 302–316, 2021, doi: 10.1016/j.ins.2019.05.023.

[3] M. Z. Ali, S. Masood, F. U. Rehman, R. Rasool, and Z. Sadiq, "An Overview of Credit Card Fraud Detection Techniques," *Bulletin of Business and Economics (BBE)*, vol. 13, no. 3, pp. 444–449, 2024, doi: 10.61506/01.00519.

[4] R. Nimashini, R. Rathnayake, and W. Wickramaarachchi, "A Machine Learning Approach for Detecting Credit Card Fraudulent Transaction," 2021.

[5] E. Oztemel and M. Isik, "A Systematic Review of Intelligent Systems and Analytic Applications in Credit Card Fraud Detection," Feb. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/app15031356.

[6] 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," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 1, pp. 145–174, Jan. 2023, doi: 10.1016/j.jksuci.2022.11.008.

[7] N. Ranjan and S. Kamble, "Implementation of Machine Learning Algorithm to Detect Credit Card Frauds," 2022, doi: 10.5120/ijca2022921959.

[8] Prof. Dipali Dube, Siddhesh Kharge, Abhay Nighot, Omkar Fulsundar, and Prasad Naykodi, "Credit Card Fraud Detection," *International Journal of Advanced Research in Science, Communication and Technology*, pp. 137–139, Apr. 2024, doi: 10.48175/IJARSCT-16923.

[9] I. Y. Hafez, A. Y. Hafez, A. Saleh, A. A. Abd El-Mageed, and A. A. Abohany, "A systematic review of AI-enhanced techniques in credit card fraud detection," *J. Big Data*, vol. 12, no. 1, p. 6, Jan. 2025, doi: 10.1186/s40537-024-01048-8.

[10] R. Vegter, "Optimizing the financial gain for credit card fraud detection systems using machine learning techniques," *Master's Thesis, University of Groningen*, 2022.

[11] X. Feng and S. K. Kim, "Novel Machine Learning Based Credit Card Fraud Detection Systems," *Mathematics*, vol. 12, no. 12, Jun. 2024, doi: 10.3390/math12121869.

[12] Z.-H. Zhou, *Ensemble methods: foundations and algorithms*. CRC press, 2025.

[13] T. Ghrib, Y. Khaldi, P. S. Pandey, and Y. A. Abusal, "Advanced fraud detection in Card-Based financial systems using a bidirectional Lstm-Gru ensemble model," *Applied Computer Science*, vol. 20, no. 3, 2024, doi: 10.35784/acs-2024-28.

[14] M. Amini and M. Rabiei, "Ensemble learning for fraud detection in E-commerce transactions: a comparative study," *Journal of Applied Intelligent Systems and Information Sciences*, vol. 3, no. 2, pp. 65–73, 2022, doi: 10.22034/JAISIS.2022.377265.1057.

[15] S. F. Pratama and A. M. Wahid, "Fraudulent transaction detection in online systems using random forest and gradient boosting," *Journal of Cyber Law*, vol. 1, no. 1, pp. 88–115, 2025, doi: 10.63913/jcl.v1i1.5.

[16] "DataCamp. (n.d.). What is bagging in machine learning? A guide with examples. Retrieved from https://www.datacamp.com/tutorial/what-bagging-in-machine-learning-a-guide-with-examples."

[17] A. Mathew, P. Amudha, and S. Sivakumari, "Deep Learning Techniques: An Overview," *Advanced Machine Learning Technologies and Applications*, p. 599, 2021, doi: 10.1007/978-981-15-3383-9_54.

[18] M. A. Wani, S. Ali, M. A. Sofi, and B. Sultan, *Advances in Deep Learning*, Volume 2, vol. 12. Springer, 2025, doi: 10.1007/978-981-15-6321-8.

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[20] X. Wang, Y. Zhao, and F. Pourpanah, "Recent advances in deep learning," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 4, pp. 747–750, Apr. 2020, doi: 10.1007/s13042-020-01096-5.


https://doi.org/10.59222/ustjet.4.1.5

[21] T. Albalawi and S. Dardouri, "Enhancing credit card fraud detection using traditional and deep learning models with class imbalance mitigation," *Front. Artif. Intell.*, vol. 8, 2025, doi: 10.3389/frai.2025.1643292.

[22] M. A. Alrasheedi, "Enhancing Fraud Detection in Credit Card Transactions: A Comparative Study of Machine Learning Models," *Comput. Econ.*, 2025, doi: 10.1007/s10614-025-11071-3.

[23] Y. Salunke, S. Phalke, M. Madavi, P. Kumre, and G. Bobhate, "Fraud Detection: A Hybrid Approach With Logistic Regression, Decision Tree, and Random Forest," *Cureus Journal of Computer Science*, Jan. 2025, doi: 10.7759/s44389-024-02350-5.

[24] A. M. Siam, P. Bhowmik, and M. P. Uddin, "Hybrid feature selection framework for enhanced credit card fraud detection using machine learning models," *PLoS One*, vol. 20, no. 7, p. e0326975, 2025, doi: 10.1371/journal.pone.0326975.

[25] M. Tayebi and S. El Kafhali, "A novel approach based on XGBoost classifier and Bayesian optimization for credit card fraud detection," *Cyber Security and Applications*, vol. 3, Dec. 2025, doi: 10.1016/j.csa.2025.100093.

[26] Y. Wu, L. Wang, H. Li, and J. Liu, "A Deep Learning Method of Credit Card Fraud Detection Based on Continuous-Coupled Neural Networks," *Mathematics*, vol. 13, no. 5, Mar. 2025, doi: 10.3390/math13050819.

[27] Z. Xia and S. C. Saha, "FinGraphFL: Financial Graph-Based Federated Learning for Enhanced Credit Card Fraud Detection," *Mathematics*, vol. 13, no. 9, May 2025, doi: 10.3390/math13091396.

[28] M. Alamri and M. Ykhlef, "Hybrid Feature Engineering Based on Customer Spending Behavior for Credit Card Anomaly and Fraud Detection," *Electronics (Switzerland)*, vol. 13, no. 20, Oct. 2024, doi: 10.3390/electronics13203978.

[29] A. Aslam and A. Hussain, "A Performance Analysis of Machine Learning Techniques for Credit Card Fraud Detection," *Journal on Artificial Intelligence*, vol. 6, no. 1, pp. 1–21, 2024, doi: 10.32604/jai.2024.047226.

[30] J. Jemai, A. Zarrad, and A. Daud, "Identifying Fraudulent Credit Card Transactions Using Ensemble Learning," *IEEE Access*, vol. 12, pp. 54893–54900, 2024, doi: 10.1109/ACCESS.2024.3380823.

[31] A. R. Khalid, N. Owoh, O. Uthmani, M. Ashawa, J. Osamor, and J. Adejoh, "Enhancing credit card fraud detection: an ensemble machine learning approach," *Big Data and Cognitive Computing*, vol. 8, no. 1, p. 6, 2024, doi: 10.3390/bdcc8010006.

[32] I. D. Mienye and T. G. Swart, "A Hybrid Deep Learning Approach with Generative Adversarial Network for Credit Card Fraud Detection," *Technologies (Basel).*, vol. 12, no. 10, Oct. 2024, doi: 10.3390/technologies12100186.

[33] S. S. Sulaiman, I. Nadher, and S. M. Hameed, "Credit Card Fraud Detection Using Improved Deep Learning Models," *Computers, Materials & Continua*, vol. 78, no. 1, pp. 1049–1069, 2024, doi: 10.32604/cmc.2023.046051.

[34] J. K. Afriyie et al., "A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions," *Decision Analytics Journal*, vol. 6, Mar. 2023, doi: 10.1016/j.dajour.2023.100163.

[35] I. D. Mienye and Y. Sun, "A Deep Learning Ensemble With Data Resampling for Credit Card Fraud Detection," *IEEE Access*, vol. 11, pp. 30628–30638, 2023, doi: 10.1109/ACCESS.2023.3262020.

[36] J. Raval et al., "RaKShA: A Trusted Explainable LSTM Model to Classify Fraud Patterns on Credit Card Transactions," *Mathematics*, vol. 11, no. 8, Apr. 2023, doi: 10.3390/math11081901.

[37] N. S. Alfaiz and S. M. Fati, "Enhanced Credit Card Fraud Detection Model Using Machine Learning," *Electronics (Switzerland)*, vol. 11, no. 4, Feb. 2022, doi: 10.3390/electronics11040662.

[38] 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, doi: 10.1109/ACCESS.2022.3148298.

[39] E. F. Malik, K. W. Khaw, B. Belaton, W. P. Wong, and X. Chew, "Credit Card Fraud Detection Using a New Hybrid Machine Learning Architecture," *Mathematics*, vol. 10, no. 9, May 2022, doi: 10.3390/math10091480.

[40] A. A. S. Alsuwailem, E. Salem, and A. K. J. Saudagar, "Performance of Different Machine Learning Algorithms in Detecting Financial Fraud," *Comput. Econ.*, vol. 62, no. 4, pp. 1631–1667, Dec. 2023, doi: 10.1007/s10614-022-10314-x.

[41] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," in *2017 International Conference on Computing Networking and Informatics (ICCNI)*, IEEE, Oct. 2017, pp. 1–9. doi: 10.1109/ICCNI.2017.8123782.

[42] H. Zhu, G. Liu, M. Zhou, Y. Xie, A. Abusorrah, and Q. Kang, "Optimizing Weighted Extreme Learning Machines for imbalanced classification and application to credit card fraud detection," *Neurocomputing*, vol. 407, pp. 50–62, Sep. 2020, doi: 10.1016/j.neucom.2020.04.078.


https://doi.org/10.59222/ustjet.4.1.5