Improving Accuracy and Efficiency of Online Payment Fraud Detection and Prevention with Machine Learning Models

Noman Abid
1120 Mac Arthur Dr Apt 1203 Carrollton Tx 75007

Abstract:- These days, many incidents of internet fraud are handled by cyber forensics. The likelihood of online fraud is being amplified by the widespread use of the internet. Automated fraud detection in online transactions is a challenging task, as fraudsters are constantly developing new and sophisticated methods. This study focuses on improving an accuracy and efficiency of online payment fraud detection and prevention by integrating advanced data preprocessing, feature extraction, and model optimisation techniques. A robust dataset preprocessing pipeline, including handling missing values, outlier removal, data standardisation, and balancing through undersampling, ensures high-quality input. Key features are extracted to enhance model interpretability and efficiency. Several ML models, including LR, SVM, KNN, and CNN, are employed to classify transactions. Models are assessed by calculating their F1-score, accuracy, precision, and recall. By obtaining an astounding 95% accuracy, 97.72% precision, 99.41% recall, and 98.56% F1 score, the CNN model surpasses conventional ML approaches to provide better outcomes. These results highlight CNN's superiority in capturing complex fraud patterns and maintaining high performance across all metrics. The proposed approach offers a robust solution for real-time fraud detection in online payment systems, ensuring accuracy, efficiency, and scalability.

Keywords:- *Online Fraud, Detection, Prevention, Credit Card Fraud, Machine Learning, Digital Payment Security, Risk.*

I. INTRODUCTION

New types of fraud have emerged with the rise of the digital age, most notably in the area of online transactions [1]. The convenience and widespread adoption of digital payment methods have simultaneously created opportunities for fraudsters to exploit vulnerabilities in online systems[2][3]. Fraud has long been a pervasive issue affecting various sectors[4], disrupting financial stability, and eroding trust in systems[5]. The convenience and widespread adoption of digital payment methods have simultaneously created opportunities for fraudsters to exploit vulnerabilities in online systems[6]. The rise of sophisticated forms of online fraud—including phishing, identity theft, and credit card scams—poses a serious danger to the safety and reliability of our digital financial systems [7].

The rise of online financial transactions—spanning e-commerce, internet banking[8], and virtual payment platforms—has not only transformed the global economy but also heightened the need for robust fraud detection and prevention mechanisms[9][10][11]. Traditional methods, while effective to some extent, often struggle to cope with the scale, complexity, and evolving nature of online fraud[12][13]. This necessitates the integration of advanced technologies to identify and mitigate fraudulent activities with greater precision[6][14].

Combating online fraud has been revolutionised by ML [15][16]. ML algorithms are able to identify fraud in real-time, adapt to new tactics, and find complicated patterns in massive amounts of transaction data [17][18]. These models utilise a variety of features, including transaction amount locations, user behaviour, timestamps, and device information, to identify anomalies and predict potential threats[19][20][21]. ML's inherent capacity for continual learning and evolution makes it a potent instrument for bolstering the precision and effectiveness of systems designed to identify fraud [22][23][24].

➢ *Aim and Contribution of Paper*

The motivation for this study stems from the growing threat of fraudulent activities in online payment systems, which undermine financial security and consumer trust. Traditional methods fail to address the complexity and scale of modern fraud schemes, necessitating the development of advanced ML models for accurate and efficient fraud detection and prevention. The main contributions of this study are:

- Implemented robust preprocessing steps, including missing value handling, outlier removal, and data standardisation, to enhance data quality and model reliability.
- LR, SVM, KNN, and CNN were among the ML models examined in order to determine the best algorithms for detecting fraud.
- When evaluating a model's performance, sophisticated evaluation measures like as F1-score, recall, accuracy, and precision may be used.
- Improved the efficacy and accuracy of systems that safeguard online payments by suggesting a scalable method for real-time fraud detection.

➢ *Structure of Paper*

The following is a synopsis of the paper's main body. Give some context for using machine learning models to detect and prevent online payment fraud in Section 2 and Section 1.

The approach is described in depth in Section 3. A comparison of the findings, analysis, and discussion is presented in Section 4. Section 5 presents the study's results as well as suggestions for further research.

II. LITERATURE REVIEW

In recent years, researchers have shown a growing interest in the development of Predictive Analytics in Online Payment Fraud Detection and Prevention with Machine Learning Models. The following are a few studies that provide background:

This research work, Sharma and Sharma (2024) is concerned with the comparative effectiveness of machine learning and DL models, crafting CNNs and RNNs as an illustration, in upgrading fraud detection power within digital finance infrastructures. Utilising a rich database of transactional data, both supervised and unsupervised learning tools are employed to establish the suspicious ones. This methodology consists of data cleaning using automatic approaches, machine learning algorithms such as CNNs and RNNs for modelling, and key metrics that involve accuracy, sensitivity, specificity, AUC, and ROC curves when evaluating the model. The analysis shows that the RNN architecture performs even better than the CNN model, observing an incredible accuracy of 95.8%, sensitivity93.7%, and specificity97. 5 % with an AUC of 0. 972. Besides, analysis showed that the models consistently performed well across various transaction amounts, indicating robustness and applicability in various situations. This underlines the fact that deep learning models are most effective when dealing with the occurrences of financial transactions that are fraudulent [25].

This study, Sharma and Babbar (2023) provides an in-depth analysis of several fraudulent activities prevalent in the ecosystem of virtual currency. Next, they had been looking at other ML methods that might detect suspicious patterns that could indicate fraud, such as AdaBoost, RF, and XGBoost. In order to determine how successful the proposed method is, experiments have been conducted using a crypto fraud detection dataset that includes instances of fraud as well as real-world cryptocurrency transactions. Model robustness and accuracy are evaluated using performance indicators such as F1-score, recall, and precision. Various methods have been tested for scalability and efficiency to determine which algorithms work best for real-time fraud detection. The findings demonstrate the potential of ML approaches to enhance the safety of bitcoin networks. In terms of accuracy, the XGBoost method ranks first with 98%, followed by AdaBoost with 67%, and RF with 90%. In terms of detecting fraudulent activity, the proposed models perform well, with notable achievements in detecting attack patterns that had not been seen before [26].

In, Garg and Gupta (2024) Effective fraud detection and prevention strategies are necessary to shield people and organisations from significant financial losses. The scarcity of publicly available datasets with fraud examples, however, is a major obstacle to this effort. In order to tackle these problems, our study uses state-of-the-art machine learning methods. Using DTs, which are effective at extracting insights from data, for fraud detection in real time is one solution that may be considered. On a similar note, DL and ANN may detect complex fraud patterns. Logistic Regression: The method that is used to predict the likelihood of fraud. Analyse the test set accuracy of LR (99.8%), DTs (99.9%), and ANNs (99.94%) when they are retrained to this model. Businesses who want to reduce the prevalence of fraud must pay close attention to these findings. Additionally, they pointed out a significant need in this area of ML for fraud detection by providing suggestions for how algorithms should be tailored to the particular financial context of actual fraud detection [27].

In, Charizanos, Demirhan and İçen (2024) to address the effects of non-stationary shifts in fraudulent transaction patterns, suggest a novel strategy. Given the massive number of datasets, it allows for efficient model training. In order to tackle the challenges caused by the special transaction features and the very low fraud rate in the dataset, they used a robust fuzzy LR model that accounts for class imbalance and separation. With sensitivity and specificity above 0.90 and Matthew's correlation over 0.80, the proposed framework reliably exhibits good performance, even when working with tiny samples. Furthermore, the suggested methodology's performance vs efficiency nexus study shows that it delivers extremely accurate findings, distinguishing between fraudulent and non-fraudulent transactions with an accuracy of more than 0.99. A suggested framework outperforms ML and other methods for detecting fraudulent transactions, while also detecting a larger percentage of legitimate transactions, according to benchmarking. Better classification performance means less financial losses and client satisfaction by more accurately identifying fraudulent transactions and preventing the misclassification of valid transfers [28].

In, Wahid et al. (2024) provide a model for detecting online fraud that makes use of a NFA to examine client calling habits in order to identify phoney calls. In order to simulate client call patterns, the model makes use of a memory module NFMs and an Autoencoder (AE). They evaluate our method using a massive dataset of actual call detail records and compare it to many state-of-the-art methods. Our strategy surpasses the baselines in terms of performance, as seen by our 91.06% AUC, 91.89% TPR, 14.76% FPR, and 95.45% F1-score. These findings imply that our method might be a useful tool for avoiding telecom fraud and show that it is successful in real-time fraud detection [29].

This study Afriyie et al. (2023) evaluate the efficacy of DTs, LR, and RF as ML models for credit card fraud detection, prediction, and classification. With an AUC of 98.9% and a maximum accuracy of 96%, RF outperformed the other methods tested for predicting and detecting fraudulent credit card transactions. Credit card fraud might be difficult to forecast and detect, but random forest is the best ML algorithm for the job. Most of these fraudulent transactions involved credit card users over the age of 60, and the peak hour for these crimes is between 22:00GMT and 4:00GMT [30].

Table 1 summarises current research on fraud detection, including studies' datasets, methods, results, and contributions.

Table 1 Summary of Literature Review for Online Payment Fraud Detection and Prevention using Machine Learning

Author	Data	Methodology	Findings	Limitations/Contributions/Future Study
Sharma and Sharma (2024)	Rich transactional data	CNNs and RNNs for fraud detection; supervised and unsupervised learning tools; metrics: accuracy, sensitivity, specificity, AUC, ROC curves	RNN outperformed CNN with 95.8% accuracy, 93.7% sensitivity, 97.5% specificity, and robust performance across transactions.	Demonstrates robustness of deep learning models for digital finance fraud detection; suggests further research on broader datasets.
Sharma and Babbar (2023)	Crypto fraud detection dataset	ML algorithms: XGBoost, AdaBoost, RF; metrics: precision, recall, F1-score	XGBoost achieved the highest accuracy of 98%; AdaBoost and RF achieved 67% and 90%, respectively.	Highlights ML techniques' scalability for real-time fraud detection; recommends further exploration of algorithm adaptability for large-scale applications.
Garg and Gupta (2024)	Limited fraud datasets	Decision Trees, ANNs, and Logistic Regression; test set evaluation with accuracy metrics	Logistic Regression: 99.8% accuracy, Decision Trees: 99.9%, ANNs: 99.94%.	Addresses dataset scarcity using ML techniques; suggests model customisation for specific financial settings.
Charizanos, Demirhan & İçen (2024)	Large-scale datasets with non-stationary fraud transaction patterns	Fuzzy Logistic Regression; handling class imbalance and separation problems; metrics: specificity, sensitivity, MCC, accuracy	Specificity and sensitivity > 0.90; MCC > 0.80; overall accuracy > 0.99; robust performance on imbalanced datasets.	Combines efficiency and accuracy for fraud detection; calls for further benchmarking with alternative ML techniques on larger datasets.
Wahid et al. (2024)	Real-world call detail records	Neural Factorization Autoencoder (NFA); Neural Factorization Machines (NFM); Autoencoder with memory module; metrics: AUC, TPR, FPR, F1-score	AUC: 91.06%, TPR: 91.89%, FPR: 14.76%, F1-score: 95.45%; effective for real-time fraud detection.	Demonstrates potential of NFA for fraud detection in telecommunications; proposes further optimisation for faster adaptation to changing patterns.
Afriyie et al. (2023)	Credit card transactions	Logistic Regression, Random Forest, Decision Trees; metrics: accuracy, AUC	RF achieved maximum accuracy (96%) and AUC (98.9%); highlighted demographic and temporal fraud patterns.	Recommends Random Forest for credit card fraud detection; suggests exploring advanced techniques for fraud occurring during specific time frames.

III. RESEARCH METHODOLOGY

To enhance an accuracy andefficiency of online payment fraud detection and prevention, this study employs a robust methodology integrating data preprocessing, feature extraction, and model optimisation techniques. The dataset is preprocessed through robust techniques like handling missing values, outlier removal, and data standardisation to ensure high-quality input. Data balancing is achieved through undersampling to address class imbalance, ensuring fair representation of fraudulent and non-fraudulent transactions.Key features are extracted to enhance model interpretability and computational efficiency while preserving critical fraud-related patterns. To evaluate the model, the data is divided into two sets: one for training and one for testing, with a ratio of 80:20. There are a number of sophisticated machine learning models used for transaction classification, including LR, SVM, KNN, and CNN. Confusion matrices provide the metrics needed to evaluate these models, such as recall, accuracy, precision, and F1-score. The proposed approach leverages these techniques to optimise fraud detection, ensuring the system's ability to identify fraudulent transactions accurately while maintaining computational efficiency for real-time implementation.

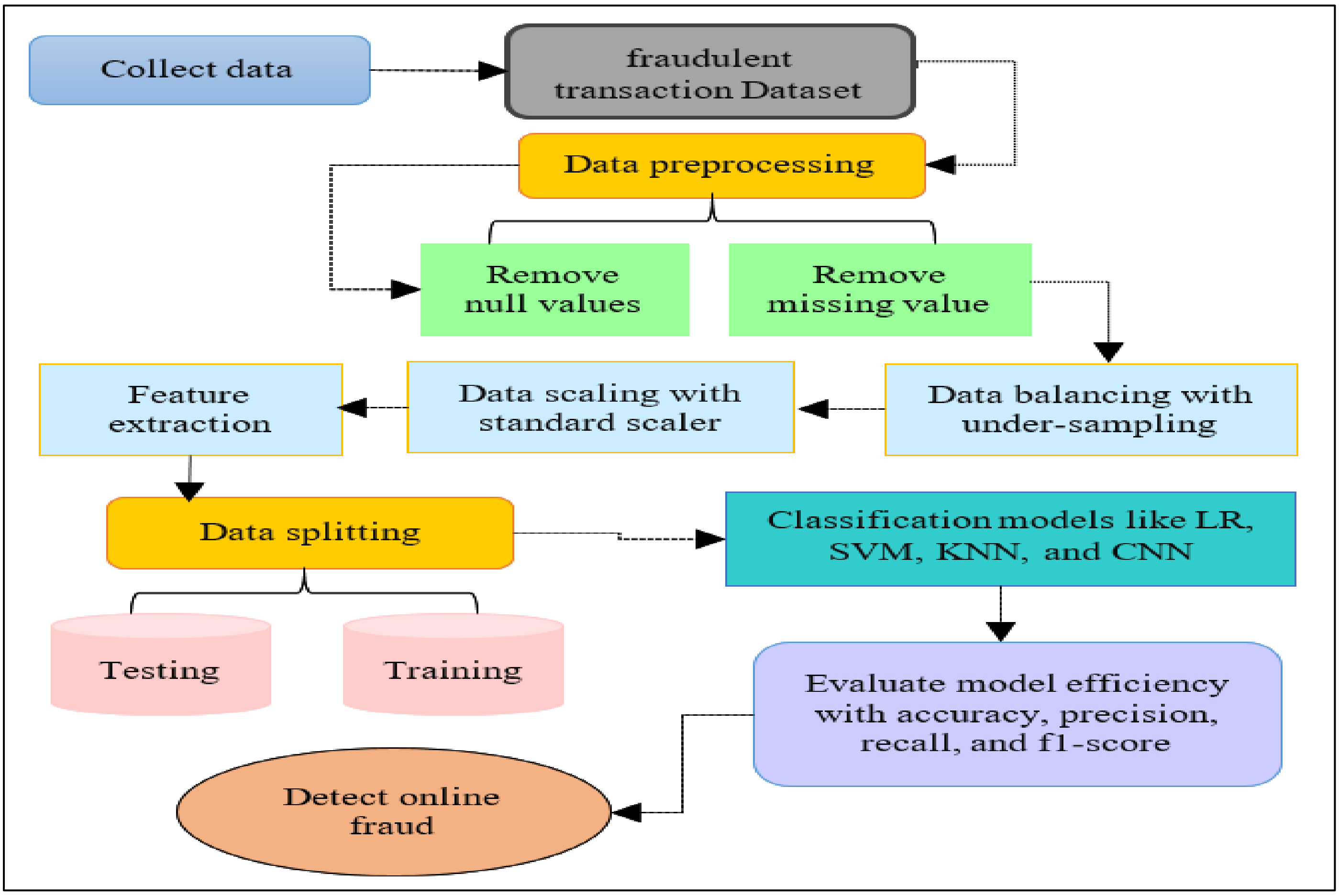
Collect data
fraudulent transaction Dataset
Data preprocessing
Remove null values
Remove missing value
Feature extraction
Data scaling with standard scaler
Data balancing with under-sampling
Data splitting
Classification models like LR, SVM, KNN, and CNN
Testing
Training
Evaluate model efficiency with accuracy, precision, recall, and f1-score
Detect online fraud

Fig 1 Flowchart for Online Fraud Detection and Prevention

Figure 1 provides a high-level overview of the diagram's steps.

- *Data Collection*

This research used the "fraudulent transaction" dataset, which was obtained from the Kaggle Depository, to identify fraudulent transactions. The collection has 6362620 records that have 10 attributes. There was one transaction with a value of 1991430 USD, while the average value of all transactions was 144972 USD. While a small percentage of transactions approach the limit, the great majority are of a far smaller magnitude. This striking difference is shown in the following graphic (see Figure 2)

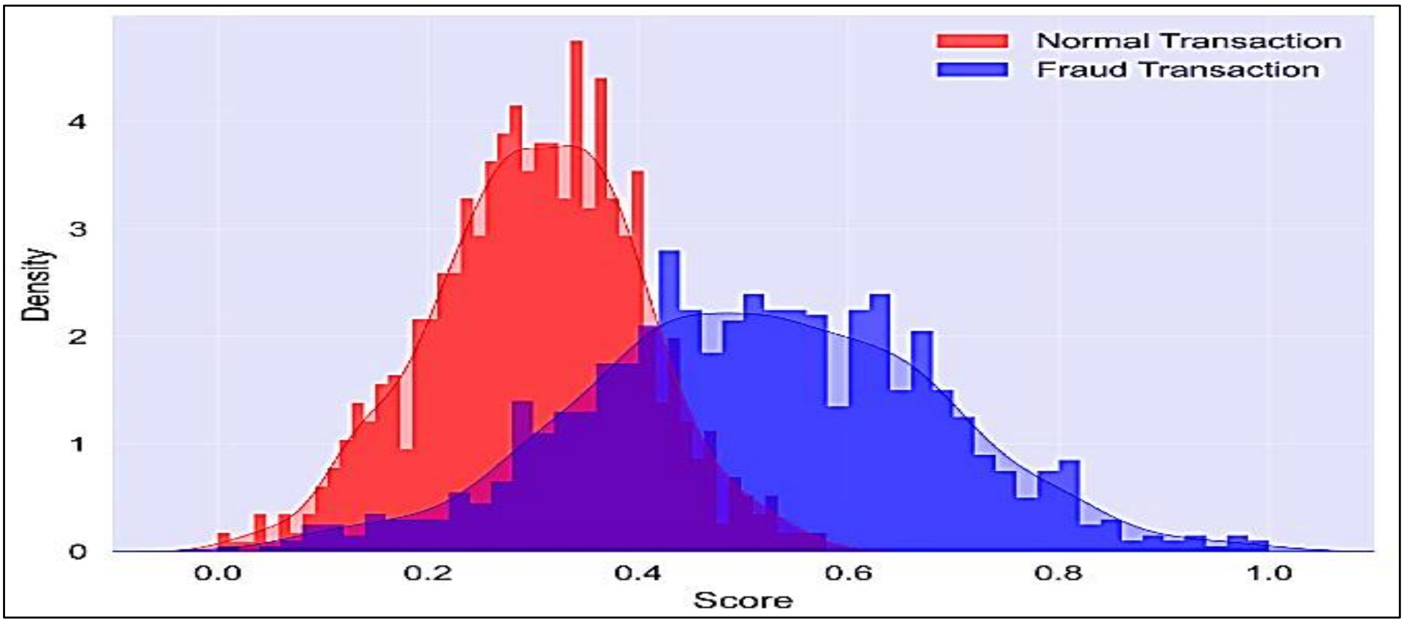

Fig 2 Kernel Density Curve of Fraud Score on Dataset

Figure 2 represents a Kernel Density Estimation (KDE) curve for fraud scores of Dataset. It illustrates the distribution of scores for normal transactions (in red) and fraudulent transactions (in blue). The overlapping region highlights cases where distinguishing between fraud and normal transactions is challenging. Fraudulent transactions tend to have higher scores, while normal transactions are concentrated at lower scores. This visualisation is crucial for understanding the separation between these classes, aiding in the design of effective fraud detection models.

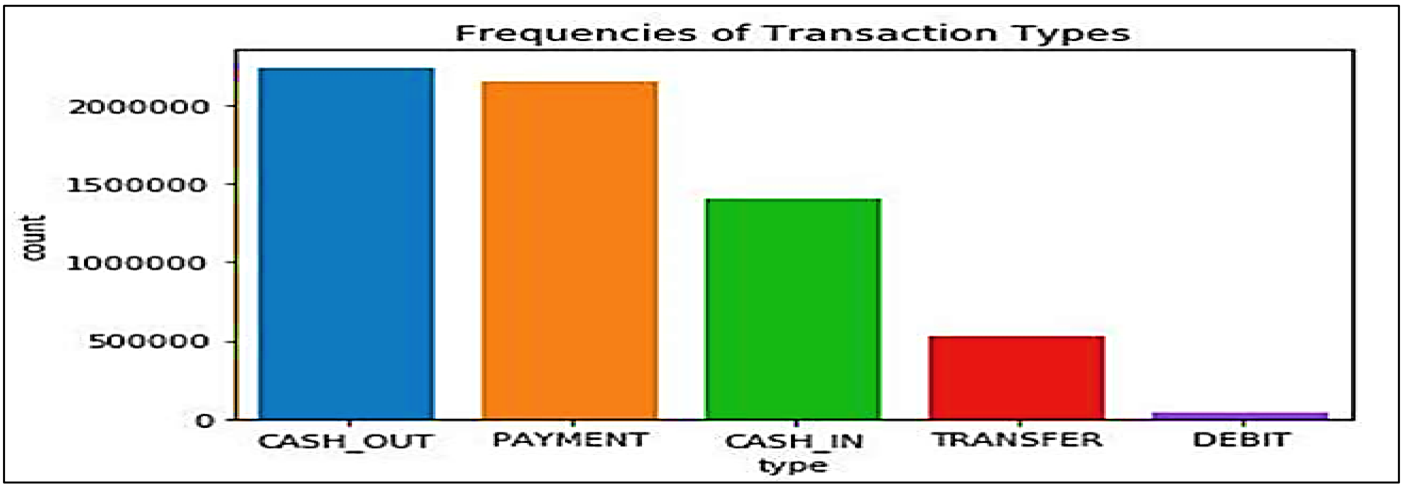

Fig 3 Frequencies of Transaction Types

The following Figure 3 shows the Transaction Types. The data shows that CASH_OUT and PAYMENT transactions are the most frequent, indicating active withdrawals and purchases by customers. CASH_IN transactions are less common, suggesting fewer deposits. The low frequency of TRANSFER and DEBIT transactions may reflect security concerns or transaction complexity.

➢ *Data Preprocessing*

An essential first step in discovering new information is data pre-processing. Among the several processes that are required are data reduction and transformation [31]. To make sure learning algorithms are effective and precise, it is essential to improve the raw data quality. Consequently, the collected data may be properly examined if the necessary data preparation methods are followed and suitable learning algorithms are used [32]. Here are some further processing important terms:

- **Remove missing & null values:** Use the isnull() method in Pandas to find datasets with null values. You may use this method to check whether any columns or fields are missing data [33][34].
- **Remove outliers:** Outliers in datasets may significantly impact the outcomes of statistical tests and models [35]. Reliable and resilient data analyses are fostered by robust data pretreatment approaches in ML, such as converting or cutting outliers, which guarantee that the effect of extreme values is minimised [36].

➢ *Data Balancing with Under-Sampling*

Data balancing is a ML approach that helps with datasets that are uneven, meaning that one class contains more entries than the other [37]. As a means of achieving statistical parity in a dataset, undersampling involves removing data points from the dominant class and replacing them with a fresh subset of the original data to train the models [38][39]. Undersampling is a method for achieving data parity by gradually removing records from the majority class.

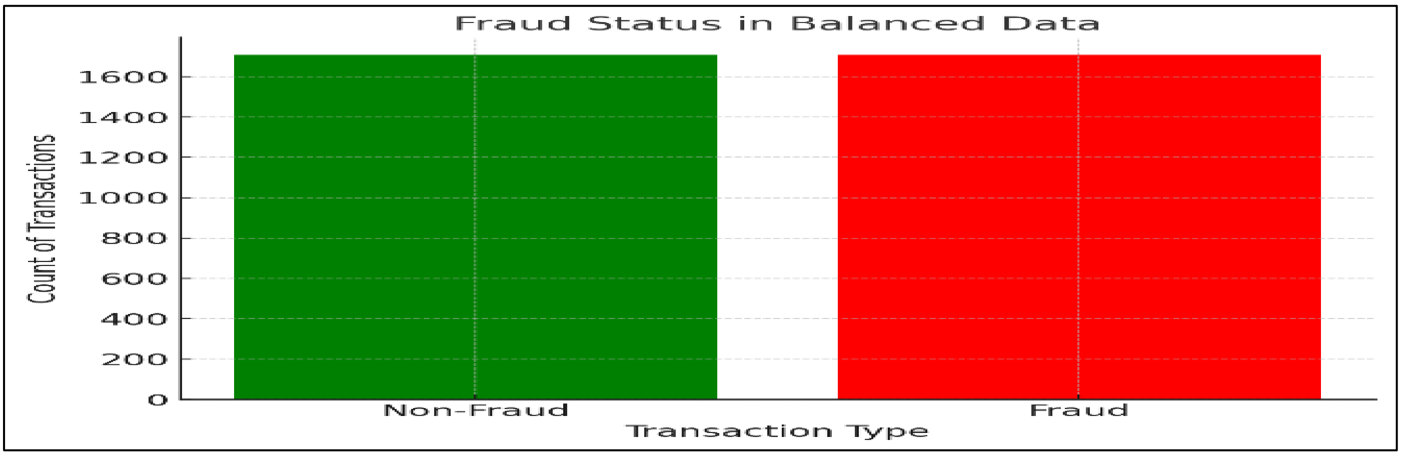

Fig 4 After Data Balancing

The dataset has been balanced with an equal number of fraudulent and non-fraudulent transactions, with 1707 instances of both categories, as shown in Figure 4. This balanced dataset ensures that the model is trained equally on both classes, helping to mitigate any bias towards the non-fraudulent transactions, which typically dominate in imbalanced datasets. The equal distribution allows for a more robust and fair evaluation of fraud detection models, enabling better generalisation and performance across both fraud and non-fraud cases.

➢ *Standardization the Data*

In ML, one of the most used feature scaling methods is the standard scaler, often known as standardisation [40]. Although this approach does not restrict the data to a particular interval or change its distribution, most of the data will fall quite near to 0[41]. This indicates that even after scaling, outliers remain in the data [42]. As seen in Equation 1, standard scaling is defined.

$$x_{scaled} = \frac{x-\bar{x}}{\sigma} \qquad (1)$$

Where:

xscaled = scaled sample point

- x = sample point
- x̄ = mean of the training samples
- σ = standard deviation of the training samples

➢ *Feature Extraction*

In DL, feature extraction refers to the transformation of raw data into useful features for training ML models [43]. The purpose of feature extraction is to simplify data while preserving important information. This can help machine learning algorithms perform better and more efficiently[44]. Feature Extraction and generally works well with large volume of data.

➢ *Data Splitting*

In order to train and test, Sklearn uses the train-test split function to randomly divide the dataset into two halves. Training will make use of 80% of the dataset, while testing will hold 20%.

➢ *Convolutional Neural Network Model (CNN)*

CNN is a DL method often used for picture processing and identification; it employs ANNs to identify patterns in pictures [45]. The layers that make up a CNN include fully connected, pooling, and convolutional layers [46]. Input parameters are reduced by the pooling layer, and images are converted to numerical values by the convolutional layer [47]. CNNs are a subset of DL algorithms that excel in processing and recognising images [48]. The many layers that make it up include fully connected, pooling, and convolutional layers [49]. CNNs are designed to mimic the human brain's visual processing, making them ideal for detecting hierarchical patterns and spatial correlations in pictures. They could find out how big a convolutional layer's output will be by using t. Here, the output is five characters long. The standard deviation of the output length is Eq. (2).

$$Output\ size = nx = 2P - nhS + 1 \qquad (2)$$

Output size = n x + 2 P − n h S + 1, where the input signal length is denoted by nx and the filter length is denoted by nh. [50].

Mathematical operations like the convolution operation (Conv_Op) find widespread use in computer vision, signal processing, and image processing [51]. It takes two signals or functions and uses their shapes to create a third signal that shows how one signal affected the other. The use of CNNs for feature extraction in computer vision is commonplace [52]. The convolution operation is defined mathematically as Eq. (3):

$$(f * g)[n] = \sum_{m=-\infty}^{\infty} f[m]g[n-m] \qquad (3)$$

Two functions, f and g, may be either discrete or continuous; n is the output signal's location or time index [53]. The function ∗ represents the convolution process. An alternative form of the Equation that accounts for discrete input signals is (4):

$$(f * g)[n] = \sum_{m=-\infty}^{\infty} f[m]g[n-m]\Delta m \qquad (4)$$

Two functions, f and g, may be either discrete or continuous; n is the output signal's location or time index [54]. The convolution operation symbol is ∗. The above Equation may be rewritten as (5) when dealing with discrete input signals:

$$(f * m)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau \qquad (5)$$

The output signal's time index is represented by t [55].

➢ *Evaluation Metrics*

A model's efficacy may be ascertained by use of assessment criteria. Criteria for assessment must be able to distinguish between various model outputs [56]. To measure the efficacy of the suggested method, this research used the following metrics: confusion matrix, accuracy, precision, recall, and f1 score [57]. An approach to demonstrating the efficacy of a classification system is the confusion matrix. After comparing the predicted and actual results, four columns are produced: true negative (TN), false positive (FP), true positive (TP), and false positive (FP). When an instance is not really diabetic, even when projections say otherwise, this is called a false positive. The confusion matrix is used to measure precision, accuracy, and recall [58].

- **Accuracy:** A model's accuracy may be described as the ratio of correct predictions to all predictions in the test dataset. It is given as (6)-

$$Accuracy = \frac{TP+TN}{TP+Fp+TN+FN} \qquad (6)$$

- **Precision:** The precision measures how many correct class activity predictions out of all the predictions in the testing dataset. It is expressed as (7)-

$$Precision = \frac{\text{TP}}{\text{TP+FP}} \qquad (7)$$

- **Recall:** The recall measures how many true positives for a given class there were relative to the total number of activities in the test dataset. In mathematical form it is given as (8)-

$$Recall = \frac{\text{TP}}{TP+FN} \tag{8}$$

- **F1 score:** The F1-Score, a derived effectiveness measure, is determined by taking the harmonic mean of recall and precision, as shown in Equation (9).

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{9}$$

These measures, when taken as a whole, show how well the model predicts the target variable.

IV. RESULTS AND DISCUSSION

This section uses the transaction dataset to evaluate several ML methods. The following results of models like LR[59], SVM[60], KNN[61], and CNN. In this section, firstly provide the CNN model performance for online fraud detection and prevention shown in Table 2. Table 3 then compares the model's results on that dataset.

Table 2 Performance of CNN model with Evaluation Matrix

Performance matrix	Convolutional Neural Network model (CNN)
Accuracy	95
Precision	97.72
Recall	99.41
F1-score	98.56

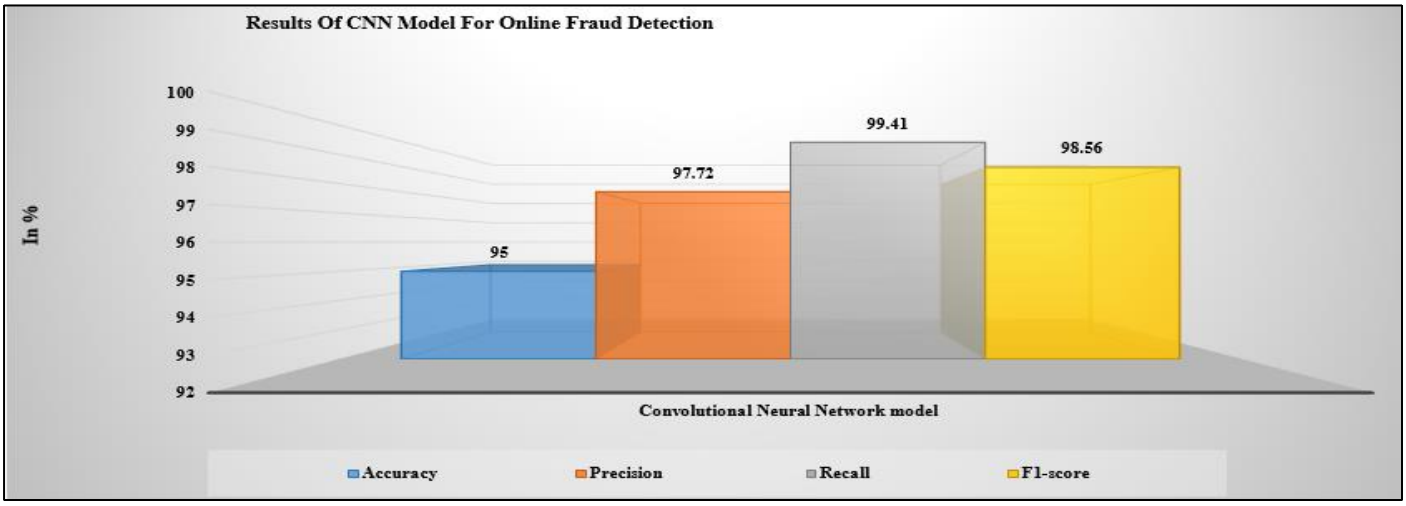

Fig 5 CNN model Performance on the Transaction Dataset

The CNN model's performance is shown in Figure 5, and Table 2 is located above. The CNN model achieved an impressive accuracy of 95%, indicating the model's overall correctness in predictions. The precision of 97.72% reflects its ability to accurately identify TP among all positive predictions, while the recall of 99.41% highlights its effectiveness in capturing nearly all actual positive instances. Furthermore, the model's F1-score of 98.56%, a harmonic mean of precision and recall, underscores its robust and balanced performance.

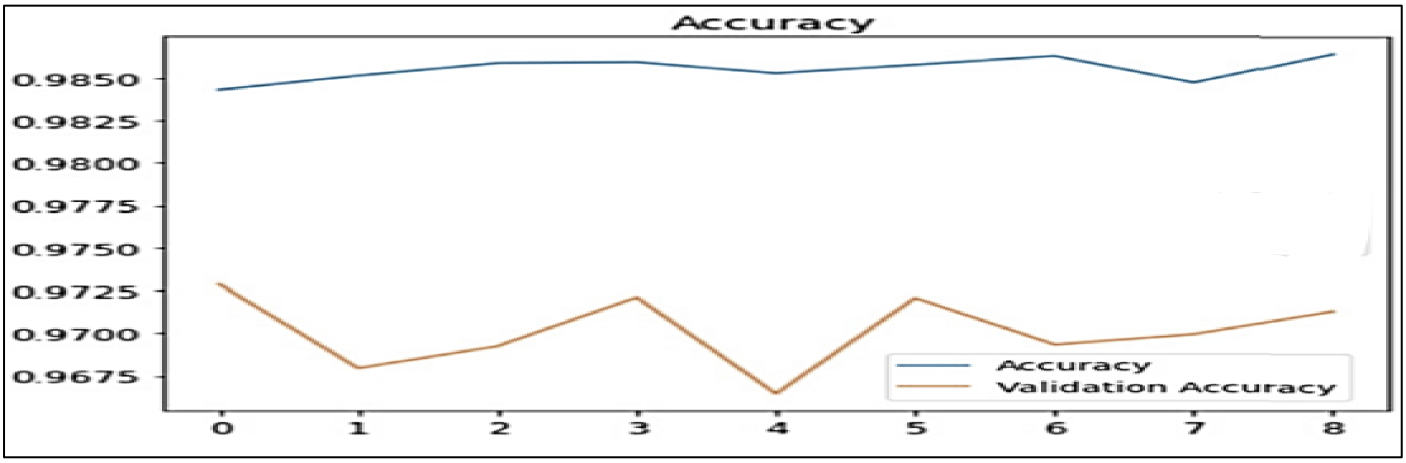

Fig 6 Accuracy Curve of CNN Model

Figure 6 shows the CNN model's accuracy curve as it goes through the training epochs. The training accuracy (blue curve) consistently remains high, fluctuating slightly around 98.5%, indicating that the model effectively learns patterns from the training data. Meanwhile, the validation accuracy (orange curve) displays a relatively lower yet stable trend, averaging around 97%, with minor oscillations across epochs. This performance suggests that the model is well-trained with minimal overfitting, maintaining good generalisation on unseen validation data.

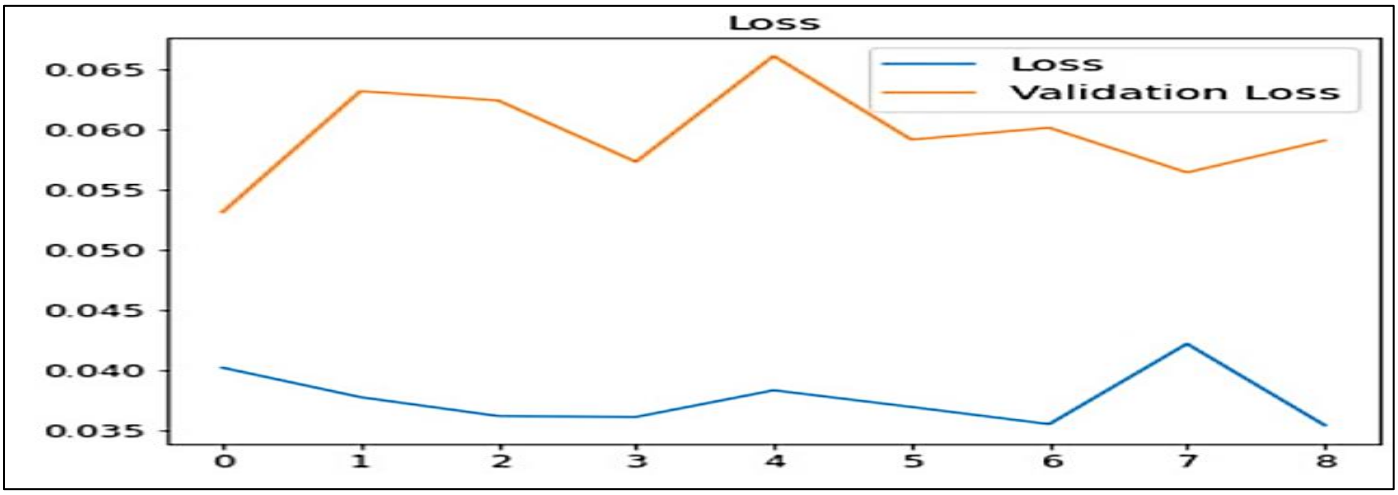

Fig 7 Loss Curve of CNN Model

Figure 7 illustrates the loss curve of the CNN model during training and validation. The training loss (blue curve) exhibits a steady decline, reaching approximately 0.035, which indicates effective optimisation and learning from the training data. Conversely, the validation loss (orange curve) fluctuates between 0.055 and 0.065, showing a less consistent trend compared to the training loss. While the gap between training and validation loss suggests some level of overfitting, the overall performance implies that the model maintains reasonable generalisation on unseen data.

Table 3 ML and DL Method Comparison on the Basis of Performance Measures

Matrix	**LR**	**SVM**	**KNN**	**CNN**
Accuracy	90.44	94.8	94.20	95
Precision	92.89	87.2	97.81	97.72
Recall	93.11	95.7	93.43	99.41
F1-score	92.11	88.4	96.51	98.56

Table 3 provides a comparison of different models. The comparative analysis of online fraud detection models on a transaction dataset reveals that the CNN model outperforms the traditional ML approaches, including LR, SVM, and Neighbors KNN, across all evaluated metrics. CNN achieves the highest accuracy, 95%, precision 97.72%, recall (99.41%), and F1-score 98.56%, showcasing its superior ability to handle complex patterns in the data. While SVM delivers a competitive accuracy of 94.8% and the highest recall among the traditional models at 95.7%, its precision of 87.2% and F1-score of 88.4% fall short. KNN demonstrates balanced performance with high precision 97.81% but slightly lower recall 93.43% and F1-score 96.51%. LR, while consistent, trails with lower metrics, achieving 90.44% accuracy, 92.89% precision, 93.11% recall, and 92.11% F1-score. This comparison highlights CNN's dominance in online fraud detection, offering robust and reliable performance.

V. CONCLUSION AND FUTURE STUDY

Online payment risk fraud has become a major issue because of the expansion of the Internet, which has led to a boom in e-commerce and online banking. The most pressing issues are inconsistencies in financial data and the reliability of methods used to identify online fraud. This study demonstrates the effectiveness of Convolutional Neural Networks (CNN) in online payment fraud detection, outperforming traditional ML models like LR, SVM, and KNN in all key performance metrics. This study shows the effectiveness of CNN in online payment fraud detection, achieving an accuracy of 95%, precision of 97.72%, recall of 99.41%, and an F1-score of 98.56%. These results significantly outperform traditional ML models such as LR with 90.44% accuracy, Support Vector Machine (SVM) with 94.8% accuracy, and K-Nearest Neighbors (KNN) with 94.20% accuracy. The integration of advanced data preprocessing and feature extraction techniques ensures high-quality input for model training, leading to accurate and efficient fraud detection. However, limitations include the potential lack of dataset diversity and challenges with model generalisation to new fraud patterns. Extending the dataset, investigating hybrid or ensemble models, improving the model's interpretability using explainable AI approaches to increase scalability and flexibility, and optimising real-time deployment are all potential areas for future development.

REFERENCES

[1]. B. Patel, V. K. Yarlagadda, N. Dhameliya, K. Mullangi, and S. C. R. Vennapusa, “Advancements in 5G Technology: Enhancing Connectivity and Performance in Communication Engineering,” Eng. Int., vol. 10, no. 2, pp. 117–130, 2022, doi: 10.18034/ei.v10i2.715.

[2]. S. Madan, S. Sofat, and D. Bansal, “Tools and Techniques for Collection and Analysis of Internet-of-Things malware: A systematic state-of-art review,” Journal of King Saud University - Computer and Information Sciences. 2022. doi: 10.1016/j.jksuci.2021.12.016.

[3]. S. Bauskar, “Enhancing System Observability with Machine Learning Techniques for Anomaly Detection,” Int. J. Manag. IT Eng., vol. 14, no. 10, pp. 64–70, 2024.

[4]. S. A. and A. Tewari, “AI-Driven Resilience: Enhancing Critical Infrastructure with Edge Computing,” Int. J. Curr. Eng. Technol., vol. 12, no. 02, pp. 151–157, 2022, doi: https://doi.org/10.14741/ijcet/v.12.2.9.

[5]. R. Arora, S. Gera, and M. Saxena, “Mitigating Security Risks on Privacy of Sensitive Data used in Cloud-based ERP Applications,” in 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom), 2021, pp. 458–463.

[6]. F. Carcillo, Y. A. Le Borgne, O. Caelen, and G. Bontempi, “Streaming active learning strategies for real-life credit card fraud detection: assessment and visualisation,” Int. J. Data Sci. Anal., 2018, doi: 10.1007/s41060-018-0116-z.

[7]. S. shrivastava Khare, Pranav, “Transforming KYC with AI: A Comprehensive Review of Artificial Intelligence-Based Identity Verification,” J. Emerg. Technol. Innov. Res., vol. 10, no. 12, pp. 525–531, 2023.

[8]. V. K. Y. Nicholas Richardson, Rajani Pydipalli, Sai Sirisha Maddula, Sunil Kumar Reddy Anumandla, “Role-Based Access Control in SAS Programming: Enhancing Security and Authorization,” Int. J. Reciprocal Symmetry Theor. Phys., vol. 6, no. 1, pp. 31–42, 2019.

[9]. T. Pencarelli, “The digital revolution in the travel and tourism industry,” Inf. Technol. Tour., 2020, doi: 10.1007/s40558-019-00160-3.

[10]. H. Sinha, “An examination of machine learning-based credit card fraud detection systems,” Int. J. Sci. Res. Arch., vol. 12, no. 01, pp. 2282–2294, 2024, doi: https://doi.org/10.30574/ijsra.2024.12.2.1456.

[11]. R. K. Arora, A. Tiwari, and Mohd.Muqeem, “Advanced Blockchain-Enabled Deep Quantum Computing Model for Secured Machine-to-Machine Communication.” Sep. 2024. doi: 10.21203/rs.3.rs-5165842/v1.

[12]. P. Khare and S. Srivastava, “Enhancing Security with Voice: A Comprehensive Review of AI-Based Biometric Authentication Systems.” 2023.

[13]. V. K. Yarlagadda and R. Pydipalli, “Secure Programming with SAS: Mitigating Risks and Protecting Data Integrity,” Eng. Int., vol. 6, no. 2, pp. 211–222, Dec. 2018, doi: 10.18034/ei.v6i2.709.

[14]. M. R. S. and P. K. Vishwakarma, “THE ASSESSMENTS OF FINANCIAL RISK BASED ON RENEWABLE ENERGY INDUSTRY,” Int. Res. J. Mod. Eng. Technol. Sci., vol. 06, no. 09, pp. 758–770, 2024.

[15]. S. Arora and P. Khare, “THE IMPACT OF MACHINE LEARNING AND AI ON ENHANCING RISK-BASED IDENTITY VERIFICATION PROCESSES,” Int. Res. J. Mod. Eng. Technol. Sci., vol. 06, no. 05, pp. 8246–8255, 2024.

[16]. R. Goyal, “EXPLORING THE PERFORMANCE OF MACHINE LEARNING MODELS FOR CLASSIFICATION AND IDENTIFICATION OF FRAUDULENT INSURANCE CLAIMS,” Int. J. Core Eng. Manag., vol. 7, no. 10, 2024.

[17]. B. Lebichot, G. M. Paldino, W. Siblini, L. He-Guelton, F. Oblé, and G. Bontempi, “Incremental learning strategies for credit cards fraud detection,” Int. J. Data Sci. Anal., 2021, doi: 10.1007/s41060-021-00258-0.

[18]. M. R. Kishore Mullangi, Vamsi Krishna Yarlagadda, Niravkumar Dhameliya, “Integrating AI and Reciprocal Symmetry in Financial Management: A Pathway to Enhanced Decision-Making,” Int. J. Reciprocal Symmetry Theor. Phys., vol. 5, no. 1, pp. 42–52, 2018.

[19]. E. Kurshan, H. Shen, and H. Yu, “Financial Crime Fraud Detection Using Graph Computing: Application Considerations Outlook,” in Proceedings - 2020 2nd International Conference on Transdisciplinary AI, TransAI 2020, 2020. doi: 10.1109/TransAI49837.2020.00029.

[20]. E.-A. MINASTIREANU and G. MESNITA, “An Analysis of the Most Used Machine Learning Algorithms for Online Fraud Detection,” Inform. Econ., 2019, doi: 10.12948/issn14531305/23.1.2019.01.

[21]. J. Thomas, H. Volikatla, V. V. R. Indugu, K. Gondi, and D. S. Gondi, “Machine Learning Approaches for Fraud Detection in E-commerce Supply Chains,” Innov. Comput. Sci. J., vol. 8, no. 1, 2022.

[22]. A. A. Almazroi and N. Ayub, “Online Payment Fraud Detection Model Using Machine Learning Techniques,” IEEE Access, 2023, doi: 10.1109/ACCESS.2023.3339226.

[23]. P. Khare and S. Srivastava, “AI-Powered Fraud Prevention: A Comprehensive Analysis of Machine Learning Applications in Online Transactions,” J. Emerg. Technol. Innov. Res., vol. 10, pp. f518–f525, 2023.

[24]. R. Bishukarma, “Adaptive AI-Based Anomaly Detection Framework for SaaS Platform Security,” Int. J. Curr. Eng. Technol., vol. 12, no. 07, pp. 541–548, 2022, doi: https://doi.org/10.14741/ijcet/v.12.6.8.

[25]. R. Sharma and A. Sharma, “Combatting Digital Financial Fraud through Strategic Deep Learning Approaches,” in 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS), 2024, pp. 824–828. doi: 10.1109/ICSCSS60660.2024.10625249.

[26]. A. Sharma and H. Babbar, “Machine Learning-Driven Detection and Prevention of Cryptocurrency Fraud,” in 2023 IEEE International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering, RMKMATE 2023, 2023. doi: 10.1109/RMKMATE59243.2023.10369055.

[27]. Y. Garg and N. Gupta, “Cyber Sentinel: Real-Time Fraud Detection in Online Transactions Using Advanced Machine Learning Techniques,” in 2024 International Conference on Futuristic Technologies in Control Systems & Renewable Energy (ICFCR), 2024, pp. 1–6. doi: 10.1109/ICFCR64128.2024.10762982.

[28]. G. Charizanos, H. Demirhan, and D. İçen, “An online fuzzy fraud detection framework for credit card transactions,” Expert Syst. Appl., vol. 252, p. 124127, 2024, doi: https://doi.org/10.1016/j.eswa.2024.124127.

[29]. A. Wahid, M. Msahli, A. Bifet, and G. Memmi, “NFA: A neural factorization autoencoder based online telephony fraud detection,” Digit. Commun. Networks, 2024, doi: 10.1016/j.dcan.2023.03.002.

[30]. J. K. Afriyie et al., “A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions,” Decis. Anal. J., 2023, doi: 10.1016/j.dajour.2023.100163.

[31]. Ramesh Bishukarma, “Privacy-preserving based encryption techniques for securing data in cloud computing environments,” Int. J. Sci. Res. Arch., vol. 9, no. 2, pp. 1014–1025, Aug. 2023, doi: 10.30574/ijsra.2023.9.2.0441.

[32]. Q. Li et al., “Using fine-tuned conditional probabilities for data transformation of nominal attributes,” Pattern Recognit. Lett., 2019, doi: 10.1016/j.patrec.2019.08.024.

[33]. N. Gameti and A. P. A. Singh, “Asset Master Data Management: Ensuring Accuracy and Consistency in Industrial Operations,” Int. J. Nov. Res. Dev., vol. 9, no. 9, pp. a861-c868, 2024.

[34]. J. Thomas, K. V. Vedi, and S. Gupta, “Enhancing Supply Chain Resilience Through Cloud-Based SCM and Advanced Machine Learning: A Case Study of Logistics,” J. Emerg. Technol. Innov. Res., vol. 8, no. 9, 2021.

[35]. H. Sinha, “ANALYZING MOVIE REVIEW SENTIMENTS ADVANCED MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING METHODS,” Int. Res. J. Mod. Eng. Technol. Sci. (, vol. 06, no. 08, pp. 1326–1337, 2024.

[36]. Sahil Arora and Apoorva Tewari, “Zero trust architecture in IAM with AI integration,” Int. J. Sci. Res. Arch., vol. 8, no. 2, pp. 737–745, Apr. 2023, doi: 10.30574/ijsra.2023.8.2.0163.

[37]. V. V. Kumar, A. Sahoo, and F. W. Liou, “Cyber-enabled product lifecycle management: A multi-agent framework,” in Procedia Manufacturing, 2019. doi: 10.1016/j.promfg.2020.01.247.

[38]. A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, Learning from Imbalanced Data Sets. Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-98074-4.

[39]. R. Goyal, “An Effective Machine Learning Based Regression Techniques For Prediction Of Health Insurance Cost,” Int. J. Core Eng. Manag., vol. 7, no. 11, pp. 49–60, 2024.

[40]. R. Goyal, “THE ROLE OF BUSINESS ANALYSTS IN INFORMATION MANAGEMENT PROJECTS,” Int. J. Core Eng. Manag., vol. 6, no. 9, pp. 76–86, 2020.

[41]. A. P. A. Singh and N. Gameti, “Leveraging Digital Twins for Predictive Maintenance: Techniques, Challenges, and Application,” IJSART, vol. 10, no. 09, pp. 118–128, 2024.

[42]. V. V. Kumar, A. Sahoo, S. K. Balasubramanian, and S. Gholston, “Mitigating healthcare supply chain challenges under disaster conditions: a holistic AI-based analysis of social media data,” Int. J. Prod. Res., 2024, doi: 10.1080/00207543.2024.2316884.

[43]. K. Ullah et al., “Ancillary services from wind and solar energy in modern power grids: A comprehensive review and simulation study,” J. Renew. Sustain. Energy, vol. 16, no. 3, 2024, doi: 10.1063/5.0206835.

[44]. R. Tandon, “The Machine Learning Based Regression Models Analysis For House Price Prediction,” Int. J. Res. Anal. Rev., vol. 11, no. 3, pp. 296–305, 2024.

[45]. M. S. Rajeev Arora, “Applications of Cloud Based ERP Application and how to address Security and Data Privacy Issues in Cloud application,” Himal. Univ., 2022.

[46]. Muthuvel Raj Suyambu and Pawan Kumar Vishwakarma, “Improving Efficiency of Electric Vehicles: An Energy Management Approach Utilizing Fuzzy Logic,” Int. J. Adv. Res. Sci. Commun. Technol., vol. 3, no. 2, pp. 737–748, Feb. 2023, doi: 10.48175/IJARSCT-9749V.

[47]. R. Bishukarma, “Optimising Cloud Security in Multi-Cloud Environments : A Study of Best Practices,” TIJER – Int. Res. J., vol. 11, no. 11, pp. 590–598, 2024.

[48]. H. Sinha, “Advanced Deep Learning Techniques for Image Classification of Plant Leaf Disease,” J. Emerg. Technol. Innov. Res. www.jetir.org, vol. 11, no. 9, pp. b107–b113, 2024.

[49]. V. Kumar and F. T. S. Chan, “A superiority search and optimisation algorithm to solve RFID and an environmental factor embedded closed loop logistics model,” Int. J. Prod. Res., vol. 49, no. 16, 2011, doi: 10.1080/00207543.2010.503201.

[50]. S. R. Bauskar and S. Clarita, “Evaluation of Deep Learning for the Diagnosis of Leukemia Blood Cancer,” Int. J. Adv. Res. Eng. Technol., vol. 11, no. 3, pp. 661–672, 2020, doi: https://iaeme.com/Home/issue/IJARET?Volume=11&Issue=3.

[51]. V. V Kumar, “An interactive product development model in remanufacturing environment : a chaos-based artificial bee colony approach,” 2014.

[52]. K. Patel, “Quality Assurance In The Age Of Data Analytics: Innovations And Challenges,” Int. J. Creat. Res. Thoughts, vol. 9, no. 12, pp. f573–f578, 2021.

[53]. R. Tandon, “Face mask detection model based on deep CNN techniques using AWS,” Int. J. Eng. Res. Appl., vol. 13, no. 5, pp. 12–19, 2023.

[54]. K. Ullah et al., “Short-Term Load Forecasting: A Comprehensive Review and Simulation Study with CNN-LSTM Hybrids Approach,” IEEE Access, vol. 12, no. July, pp. 111858–111881, 2024, doi: 10.1109/ACCESS.2024.3440631.

[55]. K. Patel, “A review on cloud computing-based quality assurance : Challenges , opportunities , and best practices,” Int. J. Sci. Res. Arch., vol. 13, no. 01, pp. 796–805, 2024.

[56]. H. Sinha, “Predicting Employee Performance in Business Environments Using Effective Machine Learning Models,” IJNRD - Int. J. Nov. Res. Dev., vol. 9, no. 9, pp. a875–a881, 2024.

[57]. M. R. S. Vishwakarma, Pawan Kumar, “A Study on Energy Management Systems (EMS) in Smart Grids Industry,” IJRAR, vol. 10, no. 2, pp. 558–563, 2023.

[58]. S. Narkhede, “Understanding Confusion Matrix.,” 2018.

[59]. Bharti Kudale, Swapnil Birajdar, Abhishek Hattekar and S. G. Sameer Kulkarni, “Credit Card Fraud Detection Using Machine Learning,” Proc. - Int. Conf. Dev. eSystems Eng. DeSE, no. 01, pp. 168–172, 2023, doi: 10.1109/DeSE60595.2023.10469583.

[60]. S. Komakula and M. Jagadeeshwar, “INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING An Exploration of Deep Learning Algorithm for Fraud Detection using Spark Platform,” 2024.

[61]. A. Taha, “A novel deep learning-based hybrid Harris hawks with sine cosine approach for credit card fraud detection,” AIMS Math., 2023, doi: 10.3934/math.20231180.