Combined Feature Set with Logistic Regression Model to Detect Credit Card Frauds in Real Time Applications

[1,2]Prabhakaran N and [3]Nedunchelian R
[1]Department of Computer Applications, Presidency College, Bangalore, Karnataka, India
[2]Research Scholar, Faculty of Information and Communication Engineering, Anna University, Chennai, India.
[3]Department of Computer Science and Engineering, Excel Engineering College, Namakkal, Tamil Nadu, India.
[1,2] prabhakaran.n@presidency.edu.in, [3]chelian1959@gmail.com

Correspondence should be addressed to Prabhakaran N : prabhakaran.n@presidency.edu.in

Article Info
Journal of Machine and Computing (http://anapub.co.ke/journals/jmc/jmc.html)
Doi: https://doi.org/10.53759/7669/jmc202404074
Received 25 December 2023; Revised from 27 April 2024; Accepted 28 June 2024
Available online 05 July 2024.

Abstract - Online payment methods are gaining popularity and are widely used, both in-store and online. Because to the Internet and smart mobile devices, conducting such transactions is quick, simple, and stress-free. However, online payment fraud is common due to the open nature of the internet, which allows criminals to use techniques such as eavesdropping, phishing, infiltration, denial-of-service, database theft, and man-in-the-middle assault. Online payment fraud is on the rise, and it is a big contributor to global economic losses. Financial services, healthcare, insurance, and other industries have long been plagued by fraud. Online fraud has developed in tandem with the use of digital payment systems such as credit/debit cards, PhonePe, Gpay, and Paytm. Furthermore, fraudsters and criminals are adept at evasion strategies, allowing them to steal more. Developing a secure system for client authentication and fraud protection is tough since there is always a workaround. This means that fraud detection systems play an important role in preventing financial crimes. Over time, victims of internet transaction fraud have incurred tremendous financial losses. The growth of cutting-edge technologies and global connection has led to a surge in online fraud. To reduce these expenses, it is critical to develop effective fraud detection systems. Machine learning and statistical tools make detecting dishonest money deals much easier. The scarcity of data, the sensitive nature of the data, and the uneven class distributions make it challenging to implement efficient fraud detection models. Given the delicate nature of the information, it is difficult to draw conclusions and construct more accurate models. This study offers a Linked Feature Set with Combined Feature Set with Logistic Regression (CFS-LoR) Model for accurate detection of online payment frauds. In comparison to extant models, the proposed model exhibits a highly accurate detection capability.

Keywords - Logistic Regression, Online Payment Fraud, Machine Learning, Feature Subset, Detection.

I. INTRODUCTION

When someone commits fraud, they do so to get some advantage, whether material or financial [1]. As a result, the two most significant approaches to avoid financial loss due to fraud are detection and prevention. The proactive approach to preventing fraudulent behaviour is known as fraud prevention [2], whereas the reactive method of discovering fraudulent transactions is known as fraud detection. Credit, charge, debit, internet payments, and prepaid cards, among others, are widely used nowadays. In some areas, they have replaced regular bank transfers as the favoured mode of trade [3]. As a result of the development of digital technologies, payment methods have undergone a significant transformation, shifting from a human approach to being executed electronically. Monetary policy as a whole has been turned upside down, as have the strategies and plans that businesses of all sizes use [4]. Internet payment scam is when someone uses your credit card or debit card to buy something without your permission [5]. These transactions can be executed both physically and digitally. Credit cards are typically used for in-person purchases. In contrast, digital transactions can be made over the phone or online [6]. It is customary for a cardholder to provide their card number, verification number, and expiration date over the phone or online [7]. After 2021, when internet shopping took off, the usage of online payment systems also surged.

Fraud occurs in every aspect of trade, including internet shopping, healthcare, banking, and finance. Every year, a scam brings adhering almost a trillion cents. While fraud is a significant risk to businesses, sophisticated methods such as rules engines and machine learning can help detect instances of it [8]. Online payment fraud has surged in tandem with the growth

of businesses that accept payments. This problem has been exacerbated by dealing with noisy and inconsistent data, as well as outliers. In this study, the use of machine learning to detect fraudulent activities is proposed. To detect and prevent online payment fraud, the proposed method uses augmented logistic regression to build a classifier [9]. An effective pre-processing phase is used to cope with contaminated data and ensure accurate detection. Two state-of-the-art techniques are employed during pre-processing to cleanse the data: the mean-based approach and the clustering-based method [10].

Historically, most methods for identifying fraud were based on statistical or multidimensional studies. The fundamental regulations that control transactional data are notoriously elusive [11] due to the verification tools' inherent characteristics. Transaction fraud may be effectively identified using big data technology and a machine learning algorithm [12]. Machine learning can identify key traits in a huge dataset that traditional statistical methods cannot. Using the appropriate machine learning approach, a model based on current transaction data can be developed to detect online transaction fraud, hence reducing fraud-related losses. [13].

From a pedagogical standpoint, this issue is especially difficult because of its complexity and the numerous contributing factors, including class imbalance. There are far more genuine transactions than fraudulent ones [14]. Furthermore, the statistical characteristics of the transaction patterns regularly change over time. However, these are not the only challenges that must be addressed to implement a fraud detection system in the actual world [15]. In practice, automated systems monitor a steady stream of payment requests and quickly choose which ones to authorize. Machine learning algorithms evaluate all legitimate transactions and report any irregularities [16]. Professionals investigate these accusations by contacting the cardholders to confirm the validity of the transaction. Over time, the automated system's fraud detection capabilities will improve as investigators' information is used to fine-tune the system's algorithm through training and upgrades [17]. As a result, the purpose of this study has been to create a model that employs machine learning to detect this type of online payment fraud [18].

Logistic regression is one approach for accomplishing this. To represent the independent variable, a logistic function is utilized. Because it is binary, the dependent variable has only two possible values. This study employs regression analysis, a relatively new technique in the field of data analytics for large online payment datasets [19]. For the most part, it focuses on large data challenges such as splitting data into two categories [20]. The literature clearly states that regression methods can only be employed with tiny datasets of no more than a few hundred elements. Applying regression to large datasets is thus a difficult task [21]. Regression studies are preferred because they rely simply on the interdependence of the variables under consideration [22]. Logistic regression in normal and anomalous transactions is demonstrated in **Fig 1.**

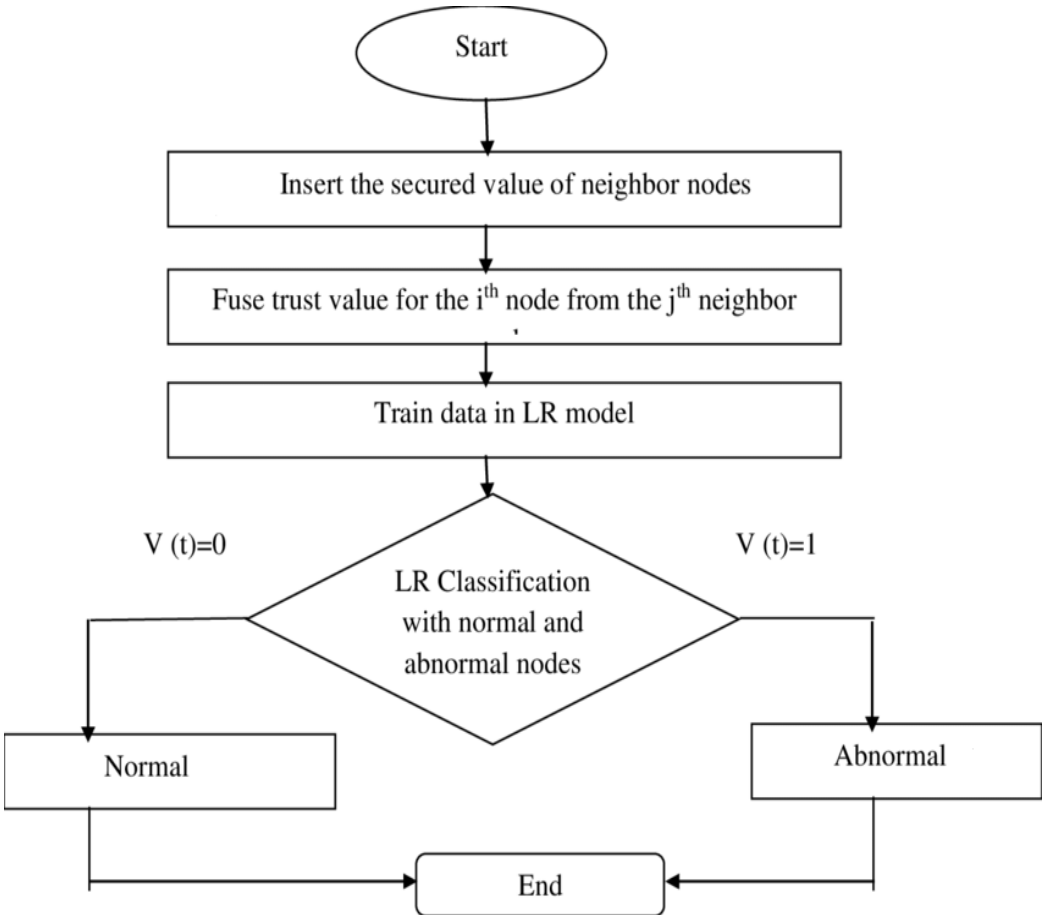

Fig 1. Logistic Regression Basic Model.

When the relationships between the qualities are considered, data detection becomes more logical [23]. That is why regression analysis was performed in this study. The focus of this study is on the two-class problem in logistic regression for big datasets [24]. Sampling is one solution to the challenge of gaining knowledge from a huge dataset, as it is difficult to obtain information by performing regression analysis on the complete dataset. During the initial phase, data is gathered as a sample for use in the subsequent regression analysis [25]. Finding a representative group to use in a regression analysis is what this article is all about. Step two entails using an autoregressive model to estimate the LR model's residual. In In the last phase, an ELR model is created by integrating the LR and RP models [26]. The ELR model's efficacy is empirically examined by analyzing a large dataset of people involved in online payment fraud. The presented model shows that the ELR model performs better than the ordinary logistic regression model in terms of accuracy.

II. LITERATURE SURVEY

Credit card fraud and nonpayment are two of the most worrying issues that might arise during a transaction. Data mining techniques are one of several strategies that researchers have been investigating as potential solutions to these issues. It can be challenging for researchers to interpret the credit card data that has been gathered. This is due to class samples overlapping and an unbalanced class distribution in the underlying data. Each of these factors increases the difficulty of detecting infrequent anomalies in data. To make matters worse, general learning algorithms tend to favour samples from the majority class, which complicates the task of identifying anomalies. [2] used two kinds of data for their study: credit card frauds (CCF) and credit card default payments (CCDP). The MCS's sequential decision combination technique ensures successful anomaly identification.

Credit card usage surged as e-commerce gained popularity and technology-enabled more individuals to obtain them. This increased the volume of financial transactions performed by banks. However, the high cost of financial transactions is required due to the significant increase in fraud. Research about the identification of fraudulent activities has garnered considerable attention. In this study, [3] considered altering hyperparameters for how much focus to place on various sorts of transactions by adjusting the weight allocated to each class. When optimizing the hyperparameters, the author considered real-world aspects such as unbalanced data and used an approach known as Bayesian optimization. To increase the performance of the Light GBM technique, the author proposed weight-tuning as a pre-process for imbalanced data, as well as Cat Boost and XG Boost, which take into account the voting mechanism. Finally, the author introduced a weight-tuning hyperparameter and used deep learning to fine-tune the other hyperparameters for improved performance. The author tested the recommended approaches using real-world data.

Using IBM Safer Payments, IBM Quantum Computers, and the Qiskit software stack, [4] created the first full-length implementation of a quantum support vector machine (QSVM) algorithm for a detection job in the financial payment industry. The author conducted a thorough study of cutting-edge computational machine-learning approaches in addition to the conventional strategy, using data from actual card payments. Through an analysis of the characteristics of the QSVM's feature map, an innovative method for identifying the most optimal features is explored. Studies employing human expertise, traditional machine learning approaches, and QSVM are compared in terms of fraud-specific key performance features. Furthermore, an ensemble model combining classical and quantum algorithms is examined as a method of improving fraud prevention decisions, resulting in a hybrid classical-quantum approach. The author discovered, as expected, that the results are very sensitive to the feature selections and methodologies used.

Credit card processing for online purchases is a simple and time-saving solution for consumers. Credit card fraud has become more common as card usage has increased. Credit card fraud causes significant costs for both people and financial organizations [7]. E-commerce and other FinTech applications have increased the frequency with which credit cards are used for online purchases. Credit card fraud, which affects businesses, banks, and card issuers, has significantly increased. As a result, it is vital to develop systems that ensure the honesty and integrity of credit card transactions [8]. Synthetic Minority over-sampling fixed the collection class misfit. SVM, LR, RF, EG, BT, DT, and ET were used to assess the framework. These machine learning algorithms were combined with the Adaptive Boosting (AdaBoost) technique to improve their data categorization accuracy. The Matthews Correlation Coefficient (MCC) and Area under the Curve (AUC) were used to assess the models' performance. Furthermore, the proposed methodology was tested using a highly skewed synthetic credit card fraud dataset to back up the study's findings.

The feature selection procedure has been recognized as an effective tool for dealing with unbalanced detection concerns. Finding a small feature subset with good detection accuracy can be approached as a multi-objective optimization problem (MOP). Usually, traditional MOP only looks at finding the best answer and doesn't consider other options. Because some characteristics are more difficult to obtain than others, offering users more freedom in selecting features would be beneficial. In this study, [9] used a multimodal MOP (MMOP) approach on feature selection, with the goals of locating a strong Pareto front in objective space and discovering as many similar Pareto optimal solutions as possible in feature space. Although several MMEAs have been provided, it is vital to remember that applying criteria reduces the number of possible solutions within a particular objective and feature space. To address this issue, a new competition-driven approach has been devised to assist existing multimodal MMEAs in identifying more comparable feature subsets as well as a desired Pareto front.

The detection of financial fraud is an important topic in the financial sector since it allows for the building of criminal profiles and the discovery of system flaws. Much of the research on fraud detection has shifted toward using complex, manually produced characteristics to comply with privacy standards and ensure interpretability. [10] Compared this method

to previous automated feature generation and fraud detection techniques established through feature engineering. The author also discovered that using random sampling to generate training/testing sets is not only inefficient but also results in erroneous assessments of machine learning model resilience. Several resampling algorithms for dealing with data imbalance in fraud detection have varying degrees of success when dealing with time-inhomogeneous events involving distinct modus operandi patterns.

III. PROPOSED MODEL

Credit card transactions in India increased from 390 million in 2012 to 652 million by 2021. As more people use online payments, the possibility of fraud increases. Despite the introduction of numerous verification techniques, online payment fraud remains at an alarming rate. The financial services industry, due to its constantly changing nature, as well as the significant risks involved, provides numerous opportunities for fraudulent activities.

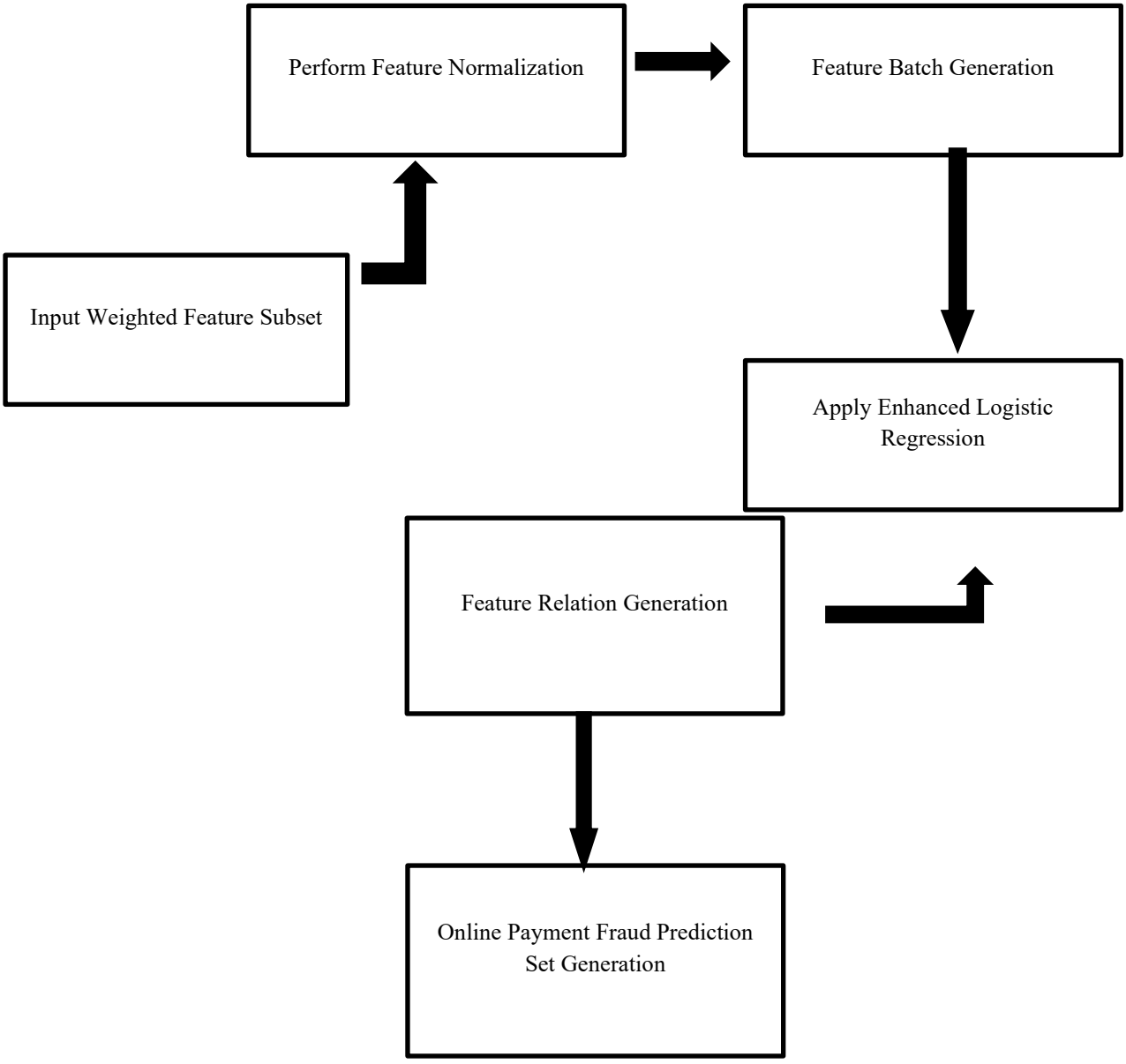

Fig 2. Proposed Model Framework.

Mischievous activities, like terrorist attacks, are often paid for with money gained through fake online payment transactions. Fraudsters frequent the internet because it allows them to conceal their genuine identities and locations. The increase in online payment fraud has a significant impact on the economy. Businesses suffer as a result, resulting in higher prices and fewer sales for customers. Thus, it is critical to reduce waste. To reduce fraud, a solid fraud detection system must be in place.

First and foremost, there is feature selection, which entails identifying the most essential variables in a dataset. Selecting the proper characteristics while removing the unneeded ones can assist reduce over fitting, improve accuracy, and speed up

the training process. The usage of visualization methods can help in this procedure. The feature selection approach produced a list of available features by picking the most relevant characteristics for training the model. If the distribution of categories in detection is not fairly even, machine learning algorithms will have difficulty learning. Because the data provided is so imbalanced, some balancing is required before the model can be trained efficiently. Common tactics for modifying the class distribution include under sampling the dominant class, oversampling the minority class, or a combination of both.

A reliable method of identifying online payment fraud that prevents consumers' money from being stolen is consequently required. Online money transactions are extremely confidential, hence this is important. So, the datasets used to teach machine learning models how to spot scams mustn't contain any information that could be used to find out who someone is. Furthermore, the form and patterns of fraudulent transactions are constantly changing, making fraudulent use of payment fraud detection a challenging task. Online payment fraud datasets are notoriously skewed, which is another problem that existing ML models for fraud detection haven't fixed as indicated by **Fig 2.**

Logistic regression makes an effort to forecast the logic of the fully dependent variable by utilizing a set of explanatory factors. If you have a categorical dependent variable and multiple independent variables, you may use logistic regression to study them. Logistic regression can help you understand enormous amounts of data. Logical approaches, such as time series modelling, are critical in the data-driven learning process. Logistic regression is a powerful analytic method for Big Data analysis. The first step is to select a representative sample from the larger dataset and run logistic regression on it. Because of the enormous expense or the fact that its members are always changing, it is not feasible to examine the population that comprises all potential parts of the topic under evaluation in logistic regression. However, big data makes it simple to examine a dataset that covers a significant portion of the population.

This study provides a Combined Feature Set with Logistic Regression (CFS-LoR) Model for accurately detecting online payment fraud. The revised logistic regression model performs regression analysis on dependent variables at several levels, resulting in a reduced feature set. The proposed LFS-ELR algorithm is thoroughly examined in this section.

Input: Feature Weighted Vector {F_{WV} set}
Output: Fraud Detection Set {F_D set}
Step-1: The dataset's weighted features are taken into account and examined for processing individual characteristics in the fraud detection procedure. The feature loading and processing is accomplished as follows:

$$FS(F_{WV}\,set[M]) = \sum_{f=1}^{M} \frac{\max\left(FS\left(Corr\left(F_g,F+1\right)\right)\right)-\min\left(fs\left(Corr(F,F+1)\right)\right)+\max\left(Wf(F)\right)}{\text{len(FS[m])}} \quad (1)$$

Here, FS denotes the weighted feature set evaluated for processing. The maximum feature correlation set is considered, while the others are eliminated because they have less weight.

Step-2: The weighted feature set is used as input, and the features are evaluated to achieve feature normalization. The goal of the normalization process is to alter the values of the features such that they all fall on the same scale. This improves the model's performance and stability throughout training. Normalization is applied as follows:

$$FN(fs[m]) = \prod_{f=1}^{M} m_g(fs(F)) - \sum_{f=1} \frac{std(FS(F,F+1)))}{\sqrt{\text{median(max (FS(F))}}} \quad (2)$$

Here, $m_g()$ is used to identify the mean values of the features so that normalization may be performed by balancing the feature values, and std is used to calculate the standard deviation among the relevant feature set.

Step-3: The identifying information provided by batch generators is known as batch feature generation, and batches can use this information to extract certain data features. They provide flexibility in the ways that can be utilized to collect data, such as transaction information and down sampling, as appropriate for the data source. Batch Feature Perspectives can be programmed to automatically materialize freshly added feature sets, resulting in accurate forecasts. The feature batch generation is carried out as

$$FB[\text{M}] = \left(\sum_{F=1}\sum_{corr=0} \frac{\text{getmin(FN(F))}}{\lambda} + \lim_{\text{F}\to WFVset}\left(\gamma + \frac{\text{max (FN(F+1))}}{\text{median(max (FS(F))}}\right)^2\right)\right) \quad (3)$$

λ → Batch size based on the normalized feature set.
γ → Handle batch vectors of maximum-normalized features.
Step-4: Enhanced Logistics Assuming the log odds of an event are linear combinations of one or more independent variables, regression modelling computes the event's probability. Logistic regression is a method for estimating variables in a logistic model as part of a regression study. For the purpose of logistic regression, a single dependent variable with two

possible values (zero and one) is utilized, whereas the explanatory variables might be either discrete binary or continuous. The ELR model analyzes individual attributes and connects them with both the batch feature set and the multi-level independent feature set, which is generated as

$$L_{reg} = \log\left\{\frac{\max(FB(F))}{\lambda}\right\} + S_{max}(corr(F, F+1)) \tag{4}$$

$$\mathrm{EL}_{reg} = \lambda_{\max}(L_{reg}) + \sqrt{\sum_{F=1}^{m} \min(\mathrm{FB}(\mathrm{F}+1)) + \sum_{F=1}^{m} \frac{C_{max}(F+1,M-F)}{m(L_{reg}(f))}} \tag{5}$$

Step-5: The correlation-based feature processing method can be used with any detection model because of its filter-like structure. To detect fraud, it evaluates feature subsets using solely the data's inherent qualities. A popular tool for measuring the level of connection between two feature set traits is the feature relation coefficient. If two characteristics are linearly dependent on one another, the relational coefficient between them is 1. A relational coefficient of 0 means that there is no relationship between the features. The process of creating feature relations is achieved as

$$F_{rel}[M] = \sum_{f=1}^{M} \frac{diff_{min}(F,F+1)}{std(EL_{reg}(F))} \tag{6}$$

$$\mathrm{UFrel[m]} = \frac{\sum_{\mathrm{F}=1}^{\mathrm{R}} \delta(F_{rel}(\mathrm{F},\mathrm{F}+1)) - \min\left(\mathrm{d}_f(F_{rel}\mathrm{l}(\mathrm{F},\mathrm{F}+1))\right)}{\sum_{\mathrm{F}=1}^{\mathrm{m}} \gamma * \mathrm{size(FB)}} \tag{7}$$

Step-6: The model will be trained with the ELR model, and the final online payment fraud set will be constructed based on the feature relations trained. The final prediction set is generated as

$$PSet(UFrel[M]) = \sum_{f=1}^{M} \mathrm{d}_f(UF_{rel}(\max(F_{rel}(F+1,F)) + \sum_{f=1}^{M} \frac{\lambda(UF_{rel}(F)) + \sum_{i=1}^{N} F_{rel}(\mathrm{F}+1)}{\mathrm{S}_z(\mathrm{UFrel})} \tag{8}$$

IV. Results

Criminals are more likely to resort to online payment fraud in an attempt to circumvent payment providers' security measures, as the popularity of such transactions has grown. With the ultimate goal of preventing fraud in an online payment system and devising countermeasures against attacks, there is a lot of pressure to investigate any security vulnerabilities that could be exploited. Detecting potentially fraudulent financial transactions as early as possible is an important aspect of this research. The development of online payment systems has led to an increase in demand for automated detection technologies that can detect and stop fraudulent transactions in real-time.

With the spread of smartphones, there is an increase in the usage of mobile payment methods, which piques the interest of scammers. Numerous fraud detection algorithms that employ supervised machine learning have been developed in response to the aforementioned body of literature. Nonetheless, suitable labelled data are scarce, and the considerable class imbalance in financial fraud data reduces detection performance. Given the monetary ramifications of fraud detection systems, this study seeks to propose an improved logistic regression framework for detecting fraudulent behaviour. The system was validated using a massive dataset of over 3 million Internet transactions. This study presents a Linked Feature Set with Enhanced Logistic Regression (LFS-ELR) Model for accurately detecting online payment fraud. Results from a comparison with the standard Fine-Grained Co-Occurrences for Behavior-Based Fraud Detection in Online Payment Services (FGCO-BFD-OPS) show that the proposed model delivers respectable results. The fraud prediction set is calculated using the formulas shown below.

Feature extraction is a form of dimensionality reduction that involves partitioning an initial set of raw data into more manageable subsets. One of the characteristics of these massive data sets is the large number of variables that must be processed, which requires a significant amount of computational power. The process of selecting and/or combining variables to create features is known as feature extraction. This effectively reduces the amount of data that must be handled while accurately and thoroughly characterizing the initial data set. **Table 1** shows the feature extraction time levels of the proposed and traditional models.

Table. 1. Feature Extraction Time Levels

Size of the Dataset	FGCO-BFD-OPS	Proposed CFS-LoR
5	12.0	6.0
10	14.0	7.8
15	15.9	9.3
20	17.7	11.2
25	21.4	13.5
30	21.4	14.6

The feature extraction accuracy levels of the suggested and current models are displayed in **Table 2**.

Table 2. Feature Extraction Accuracy Levels

Size of the Dataset	FGCO-BFD-OPS	Proposed CFS-LoR
5	78.2	82.7
10	80.9	84.5
15	82.1	86.4
20	84.5	88.2
25	86.7	90.2
30	88.3	92.6

Feature selection is the process of using only the data that is relevant to the model and removing noise from it to limit the number of variables that are fed into the model. It is the technique of automatically identifying appropriate features for a machine-learning model based on the type of problem being addressed. This can be accomplished by selectively including or removing significant features while leaving them unchanged. Thus, data noise and size can be reduced. **Table 3** shows the Feature Selection Accuracy Levels of the existing and proposed models.

Table 3. Feature Selection Accuracy Levels

Size of the Dataset	FGCO-BFD-OPS	Proposed CFS-LoR
5	82.1	88.1
10	82.6	90.5
15	84.7	92.1
20	87.4	94.8
25	89.4	96.2
30	90.7	97.8

The detection of modern payment fraud makes use of machine learning based ELR model and statistical analysis to continually monitor transactions and evaluate the level of risk that is connected with each transaction. This may require comparing lacks of different pieces of transactional data to various models of fraud that are already known to exist. Scammers take advantage of the payment request option that is available in apps that support the UPI in order to obtain the PIN or OTP that is required to authorize a transaction. They start the process of requesting payment and then contact the person in question to inquire about the OTP or PIN, claiming that this information is necessary on their end in order to finalize a transaction. The **Fig 3** represents the Online Payment Fraud Detection Accuracy Levels of the proposed and existing models.

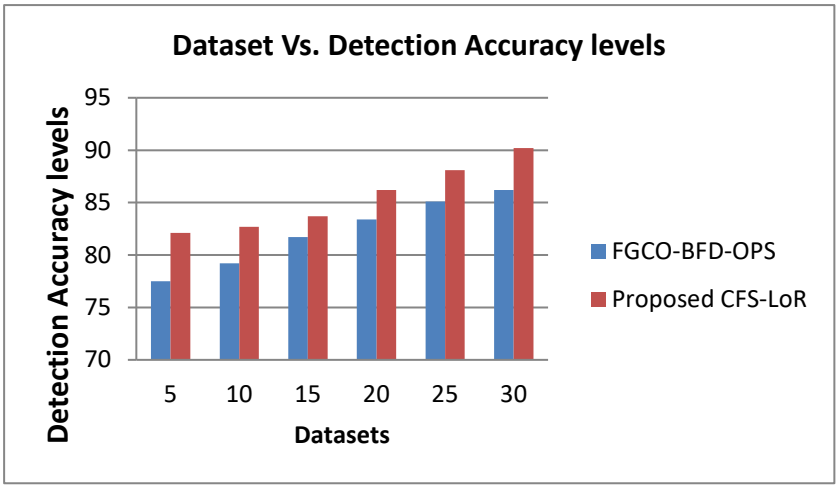

Fig 3. Dataset Vs. Detection Accuracy levels

V. CONCLUSION

Online Payments Fraud is one of the most common types of online payment fraud, however it can happen to anyone using any method. As a result, it is critical for online payment platforms to detect fraud in order to protect consumers from being charged for goods and services for which they did not authorize payment. It's no secret that online payment fraud and other forms of financial theft have increased in tandem with the rise of e-commerce and online payment systems. Financial services are highly complex due to their widespread use. As the number of people utilizing the internet to do business increases, so does the number of people seeking to perpetrate fraud via the internet. Because of this, you need to use a Fraud Detection

System that works automatically. Several approaches have been tried throughout the years in an effort to tackle this challenge. Due to the massive volume of everyday transactions, manual fraud detection tests are just not possible. As a result, such systems demand rapid and precise development. Machine learning is useful for detecting online payment fraud, but only when the correct features are used. This study presents a Combined Feature Set with Logistic Regression (CFS-LoR) Model for accurately detecting online payment fraud. The proposed model not only works in these environments, but it also provides more precision, resulting in less waste and lower costs. Integrated classifiers may be developed in the future to decrease the number of features and improve the rate of precision, and optimization may be incorporated into this model to achieve even higher levels of performance.

Data Availability
No data was used to support this study.

Conflicts of Interests
The author(s) declare(s) that they have no conflicts of interest.

Funding
No funding agency is associated with this research.

Competing Interests
There are no competing interests

References

[1]. S. K. Hashemi, S. L. Mirtaheri and S. Greco, "Fraud Detection in Banking Data by Machine Learning Techniques," in IEEE Access, vol. 11, pp. 3034-3043, 2023, doi: 10.1109/ACCESS.2022.3232287.
[2]. M. Grossi et al., "Mixed Quantum–Classical Method for Fraud Detection With Quantum Feature Selection," in IEEE Transactions on Quantum Engineering, vol. 3, pp. 1-12, 2022, Art no. 3102812, doi: 10.1109/TQE.2022.3213474.
[3]. H. Wang, W. Wang, Y. Liu and B. Alidaee, "Integrating Machine Learning Algorithms With Quantum Annealing Solvers for Online Fraud Detection," in IEEE Access, vol. 10, pp. 75908-75917, 2022, doi: 10.1109/ACCESS.2022.3190897.
[4]. E. Ileberi, Y. Sun and Z. Wang, "Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost," in IEEE Access, vol. 9, pp. 165286-165294, 2021, doi: 10.1109/ACCESS.2021.3134330.
[5]. S. A. Ebiaredoh-Mienye, E. Esenogho, and T. G. Swart, "Artificial neural network technique for improving prediction of credit card default: A stacked sparse autoencoder approach," International Journal of Electrical and Computer Engineering (IJECE), vol. 11, no. 5, p. 4392, Oct. 2021, doi: 10.11591/ijece. v11i5.pp4392-4402.
[6]. E. U. Savona and M. Riccardi, "Assessing the risk of money laundering: research challenges and implications for practitioners," European Journal on Criminal Policy and Research, vol. 25, no. 1, pp. 1–4, Mar. 2019, doi: 10.1007/s10610-019-09409-3.
[7]. 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.
[8]. T. Zhang, K. Zhu, and D. Niyato, "A Generative Adversarial Learning-Based Approach for Cell Outage Detection in Self-Organizing Cellular Networks," IEEE Wireless Communications Letters, vol. 9, no. 2, pp. 171–174, Feb. 2020, doi: 10.1109/lwc.2019.2947041.
[9]. P. Zhang, S. Shu, and M. Zhou, "An online fault detection model and strategies based on SVM-grid in clouds," IEEE/CAA Journal of Automatica Sinica, vol. 5, no. 2, pp. 445–456, Mar. 2018, doi: 10.1109/jas.2017.7510817.
[10]. H. Liu, M. Zhou, and Q. Liu, "An embedded feature selection method for imbalanced data classification," IEEE/CAA Journal of Automatica Sinica, vol. 6, no. 3, pp. 703–715, May 2019, doi: 10.1109/jas.2019.1911447.
[11]. Q. Kang, L. Shi, M. Zhou, X. Wang, Q. Wu, and Z. Wei, "A Distance-Based Weighted Undersampling Scheme for Support Vector Machines and its Application to Imbalanced Classification," IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 9, pp. 4152–4165, Sep. 2018, doi: 10.1109/tnnls.2017.2755595.
[12]. Arora, R. S. Leekha, K. Lee, and A. Kataria, "Facilitating User Authorization from Imbalanced Data Logs of Credit Cards Using Artificial Intelligence," Mobile Information Systems, vol. 2020, pp. 1–13, Oct. 2020, doi: 10.1155/2020/8885269.
[13]. J. Błaszczyński, A. T. de Almeida Filho, A. Matuszyk, M. Szeląg, and R. Słowiński, "Auto loan fraud detection using dominance-based rough set approach versus machine learning methods," Expert Systems with Applications, vol. 163, p. 113740, Jan. 2021, doi: 10.1016/j.eswa.2020.113740.
[14]. B. Branco, P. Abreu, A. S. Gomes, M. S. C. Almeida, J. T. Ascensão, and P. Bizarro, "Interleaved Sequence RNNs for Fraud Detection," Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Aug. 2020, doi: 10.1145/3394486.3403361.
[15]. F. Cartella, O. Anunciacao, Y. Funabiki, D. Yamaguchi, T. Akishita and O. Elshocht, "Adversarial attacks for tabular data: Application to fraud detection and imbalanced data",2021, arXiv:2101.08030.
[16]. Benchaji, S. Douzi, and B. E. Ouahidi, "Credit Card Fraud Detection Model Based on LSTM Recurrent Neural Networks," Journal of Advances in Information Technology, vol. 12, no. 2, pp. 113–118, 2021, doi: 10.12720/jait.12.2.113-118.
[17]. Y. Fang, Y. Zhang, and C. Huang, "Credit Card Fraud Detection Based on Machine Learning," Computers, Materials & Continua, vol. 61, no. 1, pp. 185–195, 2019, doi: 10.32604/cmc.2019.06144.
[18]. J. Forough and S. Momtazi, "Ensemble of deep sequential models for credit card fraud detection," Applied Soft Computing, vol. 99, p. 106883, Feb. 2021, doi: 10.1016/j.asoc.2020.106883.
[19]. B. Baesens, S. Höppner, and T. Verdonck, "Data engineering for fraud detection," Decision Support Systems, vol. 150, p. 113492, Nov. 2021, doi: 10.1016/j.dss.2021.113492.
[20]. 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," Information Sciences, vol. 557, pp. 302–316, May 2021, doi: 10.1016/j.ins.2019.05.023.
[21]. Y. Xie, G. Liu, R. Cao, Z. Li, C. Yan, and C. Jiang, "A Feature Extraction Method for Credit Card Fraud Detection," 2019 2nd International Conference on Intelligent Autonomous Systems (ICoIAS), Feb. 2019, doi: 10.1109/icoias.2019.00019.
[22]. Y. Y. Hsin, T. S. Dai, Y. W. Ti and M. C. Huang, "Interpretable electronic transfer fraud detection with expert feature constructions", Proc. CIKM Workshops, pp. 1-11, 2021.

[23]. D. Cheng, S. Xiang, C. Shang, Y. Zhang, F. Yang, and L. Zhang, "Spatio-Temporal Attention-Based Neural Network for Credit Card Fraud Detection," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 01, pp. 362–369, Apr. 2020, doi: 10.1609/aaai.v34i01.5371.

[24]. Y. Lucas et al., "Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs," Future Generation Computer Systems, vol. 102, pp. 393–402, Jan. 2020, doi: 10.1016/j.future.2019.08.029.

[25]. V. N. Dornadula and S. Geetha, "Credit Card Fraud Detection using Machine Learning Algorithms," Procedia Computer Science, vol. 165, pp. 631–641, 2019, doi: 10.1016/j.procs.2020.01.057.

[26]. K. Ashok, M. Ashraf, J. Thimmia Raja, M. Z. Hussain, D. K. Singh, and A. Haldorai, "Collaborative analysis of audio-visual speech synthesis with sensor measurements for regulating human–robot interaction," International Journal of System Assurance Engineering and Management, Aug. 2022, doi: 10.1007/s13198-022-01709-y.