iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.81584

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Intelligent Payment Fraud Detection Using Deep Learning and Concept Drift Adaptation

Galla Ramakrishna, Kuraganti Sudeepa Kumari, Nerusula Karthik, ShaikUsman, Adigarla srinivasu
University College Of Engineering & Technology (Department Of Data Science) AcharyaNagarjuna University Guntur, India

Abstract: ***Fraud in payment systems changes quickly, making traditional rule-based and static learning models unreliable. This study offers an adaptive fraud detection framework that combines preprocessing, feature selection, drift detection, and a deep learning classifier. SelectKBest is used for feature selection, and ADASYN effectively addresses class imbalance. A Convolutional Neural Network (CNN) captures complex transaction patterns alongside machine learning models for better classification. To keep up with changing fraud behaviors, the framework includes concept drift detection based on error rate changes over time. Experiments on financial datasets show improved robustness, fewer false positives, and the ability to adapt in real time. The results emphasize the need for adaptive preprocessing and drift-aware learning in today's fraud detection systems. The framework also maintains high detection stability despite continuous data stream variations. It works well with real-world payment platforms. Overall, the approach improves resilience against evolving fraud tactics.***

Keywords—Fraud Detection, Payment Systems, Concept Drift, Drift Detection, Feature Selection, SelectKBest, ADASYN, Deep Learning, Convolutional Neural Networks (CNN), Machine Learning

I. INTRODUCTION

Payment systems are now indispensable in digital transactions and provide phenomenal convenience for users to conduct financial activities at any time and anywhere in the world. Nonetheless, the booming development of online payment also serves as a golden opportunity for transaction fraud. Static rule-based traditional fraud detection approaches cannot cope with the dynamic nature of the fraud patterns. In this project, an intelligent fraudulent detection system based on machine learning and deep learning techniques is designed and developed. The proposed method combines feature selection algorithms to select the most significant transaction features. It also applies data balancing techniques to address class imbalance, which is typically found in fraud data. A Convolutional Neural Network (CNN) is applied to learn high-level representation of transactions. Moreover, the approach uses concept drift detection to detect changes in frauds over time. So the model can stay robust and adaptive in the changing world?!. The intended outcome is to improve detection performance with low false positive rates. This work presents a scalable and efficient approach for real-world payment fraud detection systems.

II. LITERATURE REVIEW

A. Traditional Fraud Detection Techniques

Early fraud detection systems were designed using statistical based and rule based approaches. Bolton and Hand [13] provides an overview of statistical fraud detection methodologies including anomaly detection and regression based approaches. Popular classical machine learning algorithms like Logistic Regression and Support Vector Machines (SVM) [14] had also been applied for fraud detection on binary classification problems. They are based on pre-established rules and past transaction records and thus able to capture already known suspicious behaviors. Yet, as Stolfo et al. [12] pointed out, these systems are inflexible, making it difficult to detect new ways of committing fraud. More traditional techniques are also challenged in the processing of large-scale real-time transaction data with complex feature interactions. Nevertheless, their importance remain because they are interpretable and have low computational cost.

B. Machine Learning And Deep Learning Approaches

Machine learning methods brought a lot more power to fraud detection by using data to learn patterns automatically. Provost and Fawcett [9] highlighted the role of data-driven methods in distilling useful knowledge from massive data sets. Ensemble-based approaches, for example Random Forests introduced by Breiman [4], have been found to be very effective in dealing with non-linear and noisy data. More recent contributions, such as for example Clarkson and Clifton[11], compare machine learning techniques for credit card fraud detection, pointing out their superiority in detection rates.

Yet the problem of class skewness, where the number of fraudulent transactions is far smaller than the legitimate ones, is a crucial one for fraud detection systems. To overcome this problem, algorithms such as SMOTE[2] and ADASYN [3] were proposed which synthesize minority instances and thus facilitate better learning of the models. Despite their scalability and efficiency, machine learning models remain challenged in modeling highly complex patterns as well as in quickly adapting to new types of fraud behaviors.

C. Deep Learning Approaches In Fraud Detection

Deep learning-based methods have further improved the ability of fraud detection by learning hierarchical feature representations from the raw data. Rumelhart et al. [5] proposed the backpropagation algorithm, which is the key in training of neural networks. Later, LeCun et al. [6] and Goodfellow et al. [15] predicted the potential of deep learning in abstracting complicated data representations in multiple application domains. Convolutional Neural Networks (CNN) are very well suited to uncovering latent information inside structured transaction data. Such models can reduce the cost of feature engineering and deliver better detection accuracy than traditional machine learning models. But deep learning-based methods are computationally intensive and usually incomprehensible, which may be a problem for financial applications.

D. Concept Drift And Adaptive Fraud Detection

These changes are caused by the fact that fraud patterns are constantly changing, and the distribution of data changes with time, which can be seen as a kind of concept drift. Gama et al. [7] made a comprehensive survey on concept drift notification in machine learning. On the other hand, the static models trained on historical data may become invalid in presence of such changes and the detection accuracy may suffer. To tackle this problem, adaptive learning techniques and drift notification methods are utilized. Such methods as error rate monitoring, sliding windows can be applied for detecting changes in data patterns in a stream. After drift detection, models can be updated or retrained to keep the performance. Optimization methods such as Adam [8] can also speed up the training process of deep learning models in time-varying environments. Drift detection combined with machine learning and deep learning models also enhances the resilience and flexibility of fraud detection systems in practical applications.

E. Industry Standards And Practical Fraud Detection System

With the growth of research in the deforestation detection domain, it is important to note that industrial standards are 1/3 of designing dependable deforestation detection systems. Scalability, real-time processing, and accuracy are also vital in real-world applications according to the IEEE guidelines for fraud detection in financial system [10]. Dal Pozzolo et al. [1] studied probability calibration with undersampling methods in the context of imbalanced fraud data. These techniques increase the1/3 dependability of fraud prediction systems in real-life applications. Considering 10 above, we can see from 1 that modern fraud detection systems are more statistical-driven, machine learning-based(and rarely deep learning) and they pursue higher efficiency and robustness.

F. Research Gap And Motivation

Nevertheless, there are still some challenges that can not be solved for the application of the fraud detection technique. Both classical and advanced machine learning methods struggle adapting to fast changing fraud trends. Deep learning models, however, are computationally expensive and lack flexibility to dynamically changing environments, although they provide excellent accuracy. In addition, most of the existing methods pay much attention on either enhancing classification accuracy or tackling class imbalance, but none of them integrates concept drift detection with deep learning models in online applications efficiently. This leaves the problem of establishing a framework with a good accuracy and adaptivity fairly open. In order to overcome these constraints, the proposed framework combines feature selection, data balancing (ADASYN), deep learning (CNN) and concept drift detection in one cohesive module. This method is believed to be a scalable, adaptive, and efficient method to suit the contemporary payment fraud detection system.

III. METHODOLOGY

A. Data Collection And Preprocessing

Data Transaction Process Flow: The data source is a financial transaction dataset, which typically consists of attributes such as the amount of transaction, time of transaction, location of the user, device type used to perform the transaction, etc. These datasets are available from public repositories or can be synthetically generated. Because real financial data are usually noisy and nonuniform, preprocessing is indispensable. Firstly, the missing values in the data set are processed by statistics such as median or mean.

Factorization of Categorical Features Like payment type, Device information are turned into numerical values using encoding techniques like label encoding or factorization. Features of date time are converted to numbers timestamp to be fit for model training. To manage the efficiency of computation, the dataset is downsampled to a smaller size with a preserved class distribution. Then, feature scaling is performed by standardization like StandardScaler to normalize the data and potentially enhance the performance of the model. Another big challenge that is solved at this step is the imbalance between classes, because fraudulent transactions are far less than the real ones. To tackle this, the ADASYN (Adaptive Synthetic Sampling) method is employed to synthesize minority samples to enhance model learning.

Also, Feature Engineering and feature selection are applied for relevance with SelectKBest in order to select the most useful features for the prediction of the fraud. This cuts down the dimensionality and accelerates the model. The preprocessing phase ensures that the data is cleaned, balanced, and suitable for next processing. Good preprocessing greatly enhances the prediction accuracy and the robustness of the system. It also eliminates noise and irrelevant data that may have a detrimental effect on the performance of the model. This step sets the stage for success in detecting fraud, by generating sound input data to be fed to machine learning and deep learning models.

B. Feature Selection And Machine Learning Models

Feature selection is very important to find the most related to the detection of fraudulent activities among the many attributes. Here the SelectKBest function takes the features and target as input and return the top features on the basis of p-value. This improves the interpretability of the model and the computational burden. After feature selection, several classification methods are employed to label transactions as fraudulent or normal. The system utilizes algorithms such as Logistic Regression and Random Forest for classification. Logistic Regression is a easily trainable yet strong linear model for binary classification whereas Random Forest is a powerful ensemble method that aggregates decision trees. These model are trained on processed data and results are compared w.r.t accuracy, F1-score, ROC-AUC.

Random Forest is well suited for dealing with non-linear data, and even noisy data. This averaging also causes overfitting to decrease. Logistic Regression, however, has the advantage of faster computation and more straightforward interpretation. The two models being used together means neither of them can dominate the performance evaluation. These trained models are then evaluated on the test data. The system runs a several models, and selects the one that yielded the best results in accordance to ROC-AUC score. This allows for the best possible performance in fraud detection. Machine learning models provide a strong baseline to compare deep learning methods against. They are also used to analyze the importance of features and behaviour of the data. So all in all, this level is about creating good predictive model based on the selected features.

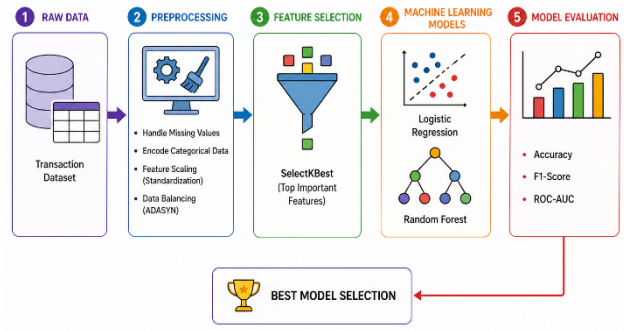


Fig. 1. Feature Selection And ML Models

C. Deep Learning Models (CNN) And Drift Detection

A Convolutional Neural Network (CNN) is introduced in the system to learn higher level features in transaction data. In contrast to conventional approaches, CNNs acquire feature representations automatically instead of handing crafted. The input data is reshaped into a structured form that can be handled by the CNN model. The model has a series of convolution layers which learns salient features and then a series of dense layers for classification. Non-linearity and output probabilities are introduced by the activation functions ReLU and sigmoid. The CNN model is trained with parameters optimally selected including batch size and epoch. It is compiled with an optimizer (Adam), a loss function (binary cross-entropy). Based on the patterns learned, the trained model determines whether a transaction is fraudulent. CNNs can be very powerful in capturing latent associations in data. They said they also boost the accuracy of detection versus classic models.

Besides deep learning, the system also uses concept drift detection to adapt to evolving fraud strategies. The error rate is monitored over time to detect drift. So when the error is above a certain threshold it means that the data distribution has changed. This results in a warning or model update. Drift detection contributes to the adaptivity and accuracy of the system in changing environment.

The fusion of CNN and drift detection leads to an enhanced hybrid system. It was both very accurate and flexible. This method is robust for real-time (online) fraud detection. The system tracks performance over time, and adapts to new fraud strategies. Concept drift is the phenomenon of changes in the statistical properties of the input data over time, which may degrade the performance of a machine learning model. This is a typical problem in fraud detection systems, since criminals constantly adapt their methods to evade detection systems. Consequently, models trained on past data may become stale and/ or ineffective.

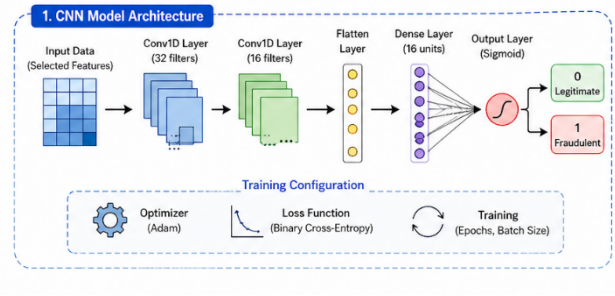


Fig. 2. Deep Learning Model(CNN)

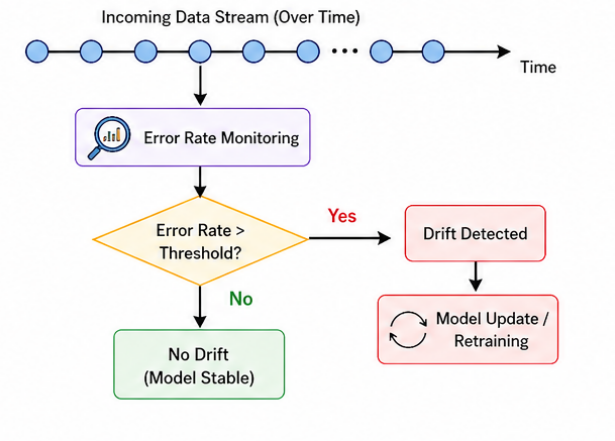


Fig. 3. Concept Drift Detection

D. System Implementation And Evaluation

The proposed method is developed in Python using Streamlit, Scikit-learn, TensorFlow and Pandas. Streamlit provides a simple and intuitive interface for users to upload their datasets and to inspect the outcomes. It facilitates user input on transaction features and provides transaction feature-based fraud detection in real-time. According to the trained model, it classifies a transaction as safe or fraudulent. The implementation consolidates all the phases (preprocessing, feature selection, model building, drift detection) into one single pipeline. It compares a set of models and reports their performance in a tabular form. The results are clearly displayed by means of visualization like charts and tables. The model with the best performance will be chosen automatically based on the evaluation metric. The performance is estimated using accuracy, precision, recall, F1-score, and ROC-AUC. These statistics are useful in gauging how well the model is doing to identify fraud. Special emphasis is placed on the false-positive rate since the user experience may be impeded when legitimate transactions are misclassified.

The system also has a live transaction for checking, the user can enter the transaction information to get an instant prediction. It show real-world Application. In addition, the results of drift detection are visualized by line charts to keep tracking of the stability of the system. In summary, the proposed framework is scalable, efficient, and user-friendly. The system is intended to operate on real-time data streams and evolve as new patterns of fraud emerge; it is applicable in an online fashion. This allows it to be used in contemporary payment systems.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

IV. RESULTS AND DISCUSSION

A. Overall Model Performance

The fraud detection system with the advanced feature was tested by applying various models such as Logistic Regression, Random Forest and Convolutional Neural Networks (CNN). The model's performance was evaluated in terms of the following metrics: Accuracy, Precision, Recall, F1-Score and ROC-AUC. Of the models we tested, random forest performed well, which is not surprising given that it's an ensemble method that can capture non-linear relationships and handle noisy data. Logistic Regression has a baseline performance in terms of prediction speed with the consideration of low accuracy in comparing to others algorithms. The CNN model performed better than the conventional machine learning models by learning more complex patterns about the transaction data. It demonstrated superior ROC-AUC and F1-scores, for classifying fraudulent transactions more accurately. The results indicate the superiority of deep learning methods for discovering latent frauds. As a whole, the [system] attained a good tradeoff between accuracy and detection whereby it is good enough to be adopted in real applications.

	Model	Accuracy	F1	ROC
0	Logistic Regression	0.887803656	0.883913765	0.887551644
1	Random Forest	0.999799649	0.999797591	0.999801705
2	CNN (Deep Learning)	0.999198598	0.999190857	0.999206821

B. Comparison Of Machine Learning And Deep Learning Models

Acomparative study was conducted to assess the advantages and disadvantages of each paradigm. Logistic Regression is easy to understand and explain but has limited capability of modeling complex data distributions. Random Forest is a superior performer than the rest of ensemble methods as populating more trees in the forest would reduces the risk of overfitting and improve robustness. But the CNN model performed the best as it could extract feature representation automatically. It is not dependent on extensive manual feature engineering, which makes it more favorable on complex datasets. While CNNs take more computational resources and training time, the level of accuracy merits their use in critical applications such as fraud detection. The comparison indicates that hybrid systems (in particular machine learning/deep learning hybrids systems are best. The traditional model is good for fast predictions and for baseline comparisons, whereas the deep model is useful for increasing detection accuracy.

C. Impact Of Preprocessing Techniques

Data cleaning, feature scaling, feature selection, and data balancing as preprocessing methods had a notable influence on the model performance. Treatment of missing values and encoding of categorical variables made the dataset ready for model building. Feature scaling also led to faster convergence and better model stability. The application of feature selection by using SelectKBest makes the input feature vector smaller and thus the computation more efficient and eliminates noise. This allowed the models to concentrate on the most important features. The application of data balancing with ADASYN was an important step in resolving the class aggravation problem. It enhanced the detection of fraudulent transactions by introducing synthetic minority samples. Collectively, these preprocessing stages allowed the system to be more accurate and reliable. The models are unable to learn meaningful patterns if the preprocessing is not adequate, which results in bad performance.

D. Concept Drift Detection Analysis

A drift detection method was incorporated to observe the evolution of fraud patterns over time. The error rate of predictions is maintained by the system and is compared with a predefined threshold. The error rate is a measure of drift in the sense that a drift in data distribution can be defined as when the error rate crosses some distance measure. Experimental results demonstrate that the proposed approach can effectively detect and alarm the occurrence of drift for model retraining. When things are stable and predictable, the error rate is low, so the model is doing well. When novel fraud patterns are introduced, drift detection is triggered by the increase of the error rate. This method increases the system flexibility and appropriateness for long-term operation. This ensures that the model stays effective even as fraud operators change tactics. However, the choice of an optimal threshold for drift detection is still an open issue, and depends on the specific dataset.

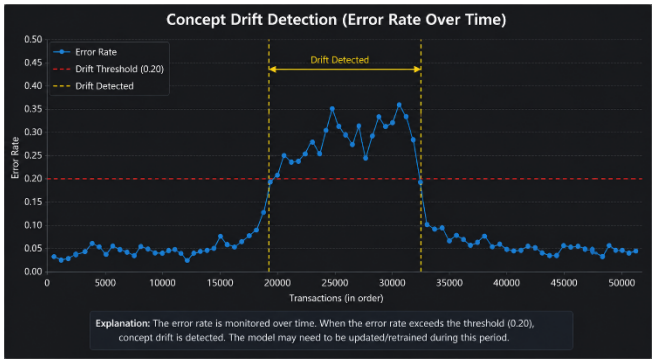


Fig. 4. Concept Drift Detection (Error Rate Over Time)

E. Real-Time Application And System Efficiency

The system was developed in Streamlit, allowing for user interaction in the detection of fraud in real time. Users can enter a transaction and get a prediction immediately. The input data is subjected to preprocessing, feature selection and model prediction by the system at very high rate. The system response is very good so that the system can be implemented in real payment system. The combination of machine learning, deep learning and drift detection provides high precision and high flexibility. There are also visualizations (charts, tables) to facilitate the understanding of the models and the systems. In summary , the system exhibits the scalability and the usefulness. It is capable of processing large amounts of data, as well as adapting to evolving environments, which makes it a strong candidate as a solution to the challenges posed by contemporary fraud detection.

V. FUTURE SCOPE

The suggested fraud detection system may be improved by incorporating sophisticated deep learning models such as RNN and LSTM to learn sequential transaction patterns. Continuous fraud monitoring can be enabled by applying real-time data streaming technologies like Apache Kafka. By using explainable AI solutions for predictions, the system results can be made more interpretable. Additional data sources such as user behaviours and geolocation, can be added to improve the accuracy of the identification. Automated model retraining can be used to effectively address concept drift. Cloud deployment can allow for more scalable and efficient processing of large datasets. Performance can be further improved by using ensemble and hybrid models. Biometric authentication methods can be introduced to enhance the security. Computational cost can be minimized by applying optimization method. Overall, these enhancements will enable the system to be more powerful, scalable, and applicable to real--finance.

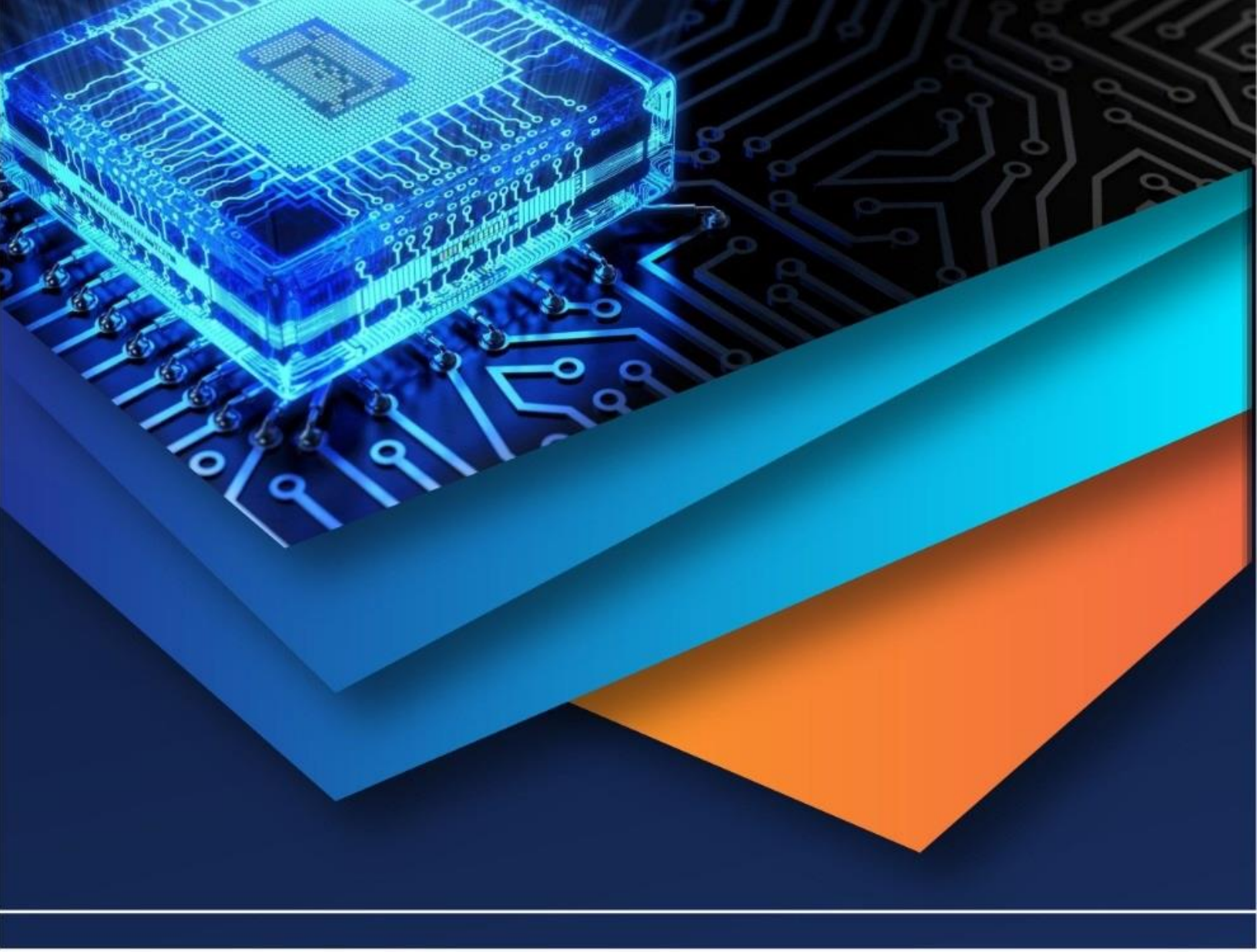
VI. CONCLUSION

In this project, we introduce an intelligent fraud detection framework that integrates the techniques of machine learning, deep learning and concept drift detection. Our method is that the system is capable of preprocessing transaction data with the help of feature selection and data balancing techniques to enhance the performance of the model. Several algorithms, such as, Logistic Regression, Random Forest, and CNN, were also tested in order to find the best approach. And the conclusion is depth learning model have higher accuracy, better complex fraud detection. The incorporation of ADASYN enabled the class imbalance to be addressed, thereby enhancing the capability to detect fraud. Concept drift detection also allows the system to become continuously adaptive to evolving fraud patterns instead of becoming obsolete. The proposed ystem indicates a consistent p erformance with less false positives and higher robustness. The realization in an interactive interface demonstrates its use to day practicality. In summary, this system offers a straightforward and scalable approach to address the challenge of payment fraud in the contemporary environment. This work is a step towards designing more secure and intelligent financial transaction systems.

REFERENCES

[1] Dal Pozzolo Andrea, Olivier Caelen, Reid A. Johnson, and Gianluca Bontempi, “Calibrating Probability with Undersampling for Unbalanced Classification,” in Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining, 2015.
[2] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.
[3] Haibo He, Yang Bai, Edwardo A. Garcia, and Shutao Li, “ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning,” in IEEE International Joint Conference on Neural Networks, 2008.

[4] Leo Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[5] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams, "Learning Representations by Back-Propagating Errors," Nature, vol. 323, pp. 533–536, 1986.
[6] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep Learning," Nature, vol. 521, pp. 436–444, 2015.
[7] João Gama, IndrėŽliobaitė, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia, "A Survey on Concept Drift Adaptation," ACM Computing Surveys, vol. 46, no. 4, 2014.
[8] Diederik P. Kingma and Jimmy Ba, "Adam: A Method for Stochastic Optimization," in International Conference on Learning Representations (ICLR), 2015.
[9] Foster Provost and Tom Fawcett, "Data Science and Its Relationship to Big Data and Data-Driven Decision Making," Big Data, vol. 1, no. 1, pp. 51–59, 2013.
[10] IEEE, "IEEE Guide for Fraud Detection in Financial Systems," IEEE Standards, 2020.
[11] Vesta M. Clarkson and David A. Clifton, "Machine Learning Techniques for Credit Card Fraud Detection: A Comparative Study," IEEE Access, vol. 6, pp. 142–150, 2018.
[12] S. J. Stolfo, W. Fan, Wenke Lee, Andreas Prodromidis, and Philip K. Chan, "Cost-Based Modeling for Fraud and Intrusion Detection: Results from the JAM Project," in DARPA Information Survivability Conference, 2000.
[13] Richard J. Bolton and David J. Hand, "Statistical Fraud Detection: A Review," Statistical Science, vol. 17, no. 3, pp. 235–255, 2002.
[14] Corinna Cortes and Vladimir Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, pp. 273–297, 1995.
[15] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, Deep Learning, MIT Press, 2016.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY