Real-time Fraud Detection Using Advanced Deep Learning with Integrated External Data Sources

Rama Krishna Vuppala[1][0009-0002-8335-2660] and Dr. G Anil Kumar[2]

[1]*Research Scholar, Best* InnovationUniversity, Gorantla, India, Associate Professor, Annamacharya Institute of Science and Technology,Hyderabad, India
Email: 2022pcse054@bestiu.edu.in
[2]Research Supervisor, Professor, ScientInstitute of Technology & Sciences, Hyderabad, India.E-mail: anildeva@gmail.com

Abstract. With the changing landscape of digital transactions, fraud detection is the need of hour for financial institutions and worldwide organizations. They are based on rules and fail to capture the extent of fraudulent activities because fraudsters today commit high-profile, complex crimes which can change so quickly that traditional rule-based systems find it hard to keep up. We need a solution for this – we need something smarter & more adaptive… In this paper, we develop a real-time fraud detection system based on advanced deep learning model for handling these challenges. The model is able to process large volumes of transaction data and identify suspicious activities with great accuracy using state-of-the-art neural networks techniques such as convolutional neural network (CNN) and long short-term memory (LSTM). The paper describes the model architecture, data preprocessing details and mechanisms for combining internal and external streams pilgrim-style. We have conducted extensive experiments on large financial datasets and shown that our approach achieved higher accuracy with much faster response time when compared to state-of-the-art detection methods. This real-time implementation has been due to optimization of computational methods and parallel processing with minimal effect on transaction latency. This paper constitutes a holistic approach to an emerging problem of financial fraud by providing a new malicious activity detection mechanism that scales efficiently with respect to high transaction rate and accelerates the attack identification leading it even closer towards effectiveness/cost efficiency.

Keywords: Fraud Detection, Deep Learning, Real-Time Detection, External Data Integration, Neural Networks.

1 Introduction

Today, fraud is one of the largest problems in the world that relies on digital and automated financial transactions more than ever before. The rapid adoption of online and electronic means of payments has risen a landscape for financial crimes from credit card fraud to more sophisticated ones such as money laundering or even identity theft. The demand for sophisticated and proactive fraud detection systems has never been greater as the interest in stay-at-home fraud continues to rise. What worked in the past to fight fraud, doesn't work today because fraud patterns are continually changing and adapting with all sorts of available tools that people use. These fraud schemes are growing in sophistication, requiring new methods that can leverage deep learning with real-time analytics (specifically integrating external data sources for better detection).

Previous works reveal that substantial progress has been achieved recently in a fraud detection using machine learning-based approaches and historical transaction data. However, there are shortcomings even in these models, especially when they function around real-time environments. Most models being trained on only internal transactional data which in-turn limits them from spotting the more complex fraud schemes that may involve time-related external factors like social behaviors, geopolitics events

or changes within market conditions. Implementation of machine learning models in real time also presents significant computational challenges, since the amount and rate at which data passes through a typical retailer system demands quick and accurate processing to prevent friction that would slow transaction cycle.

One promising way to overcome the shortcomings of traditional / machine learning based fraud detection systems, is through Deep Learning — a subset of Machine Learning that uses artificial neural networks which imitate the human brain in order to process data and make predictions. Deep learning models can automatically learn intricate patterns and relations within transaction data, so they are able to expose deeper fraud schemes that may seem less traditional or predictable. Moreover, deep learning models can scale to being trained over large datasets and financial institutions see a myriad of transactions each day which is an advantage. Deep learning in fraud detection has the deep ability to process unstructured data with high dimensionality. The patterns entailed in financial transactions are often too intricate to be captured with simple statistical methods or by sophisticated rule-based systems. Deep models are renowned for their ability to identify subtle, non-linear interactions in the data that may point towards frauds and as a result we will use convolutional neural networks (CNN) and recurrent neural network (RNN), especially long short-term memory LSTM-RNN range of architectures. CNNs: Known for their great performance in image recognition, Convolutional Neural Networks can be used to detect scale-invariant features that span multiple transactions by treating transaction sequences as structured data (as opposed to time series). LSTM (long short-term memory) are very powerful for detecting sequence dependencies and capable of finding temporal patterns in transaction data that might be indicative to fraud.

2 Related Works

Recently, there has been a great deal of academic and industrial interest in the development of more sophisticated deep learning models for fraud detection. Past studies have offered many angles concerning the outcomes of deep and machine learning approaches, as well as hybrid methods in fraud detection on different horizontal layers from financial transactions to counterfeit product identification. In this section, we summarize multiple studies that reinforce feature selection and incorporating external data sources as well real-time detection are all significant components of an effective fraud system.

P. Sure et al. [1] investigated hybrid metaheuristic methods incorporating feature selection using nature-inspired algorithms that dramatically influenced the ability of credit card fraud detection systems. The research proved that through the exclusion of a significant amount features from large data set, one could cut on computing time and at same time improve detection accuracy. This resulted in a more flexible model, which could evolve as fraudsters change their activism patterns with the introduction of hybrid techniques such as genetic algorithms to particle swarm optimization. L. Wang et al. Deep learning was further expanded to food fraud detection by [2]. Their Vis-NIR two-dimensional spectroscopy work on beef and mutton adulteration demonstrates the ability of deep learning models to identify fraud, showing applications in a practical category (limitations exist if it is finalized). This paper demonstrated the potential of using these methods on non-financial datasets highlighting the external data enrichment in improving fraud detection deep learning models. A fully connected neural net-based system for detection of credit card fraud where in [3] introduced by S. al Balawi and N. Aljohani Their approach was also capable of processing scale well, being able to extract

anomalies from transaction data using unsupervised learning techniques. Based on their results, the research team concluded that using neural networks can lead to much more accurate fraud detection levels and lower false positive rates than current methods, which could be better suited for high volume real-time detection systems. V. R. Shetty et al. Machine learning-based models for intrusion detection to protect systems from cyber threats and real-time environment frauds proposed by [4]. In their study they have drawn up the need to enabling real-time processing, to address the issue caught part way between identifying not as spurious behavior while at still trying to achieve 99.9% accuracy in this regard without delaying valid transactions. This paper should serve as a reminder that it is important to have systems which do not compromise speed for accuracy. B. Vyas [5], "Frauddetection and prevention in java based systems with integration of artificial intelligence," Groves Publications International Journal of Innovations in Computer Science and Engineering (IJICSE), vol 2 [3], p011-016,2018 The need for AI-based fraud detection is obvious and this study has interesting contributions in demonstrating the advantages of data-driven methods over rule-bases ones — results that are reinforced by extensive research, including work on ensemble learning.

A review of state-of-the-art practical methods for deepfake detection was given by J. Napp [6], that has similarities with financial fraud detection context to comprehend fraudulent activities using Deep Learning techniques as well. The methods in this work give us useful insights on achieving flexibility of deep learning models for different kinds of fraud detection, such as multimedia-based fraud.

S. K. Hashemi et al. [7] proposed the use of this method for detecting fraud in banking datasets using machine learning techniques. The area on which they conduct research revolves around feature engineering with ensemble methods for the purpose of improving transaction anomaly detection. The results clearly showed the significance of a good preprocessing pipeline in fraud detection for getting machine learning models to work as effectively as possible.

M. Emadaleslami et al. [8] used deep learning to detect electricity theft in advanced metering infrastructure systems. Applied in the context of fraud detection, their two-stage method that uses unsupervised and supervised learning could be used in the financial sector to detect instances of fraud with complex structures. Their high efficacy in detecting outliers in energy consumption patterns also proves that deep learning can be successfully applied in various fraud contexts. E. Jo et al. [9] used deep learning-aided spectroscopy to quickly distinguish between authentic and adulterated beef products. Their point about incorporating external data, such as chemical and physical properties, into deep learning, can directly be applied to financial fraud. Postulated D. Elminaam et al. [10] employed a deep learning technique in detecting and developing deep fakes. This approach provides a viable framework for detecting artificial financial transactions. The model generation component could be employed to train the systems to detect fraudulent synthetic transactions. Finally, D. O. do Rosario Lourenco et al. [11] classified harmful URLs and QR codes using machine and deep learning. Real-time fraud classification finds a direct application in the detection of financial fraud. J. Gill, and R. Dhakad [12], did also research on identifying malicious code through deep learning approaches. Their paper focuses on how a convolutional neural network (CNN) can detect code-level anomalies likely responsible for fraudulent transactions. Commonality in how we identify malicious code and detect fraudulent transactions that showcases the cross-domain applicability of deep learning models.

B. Narsimha et al. The research of [13] was to search for artificial intelligence and machine learning in the field on financial fraud detection but more specifically cybersecurity defense mechanisms. According to their article, machine learning algorithms

have helped boost detection rates — in part by utilizing external data on behaviors of users and trends seen elsewhere in financial markets.

A. Raza et al. [14] introduced a new deep learning model for detecting fake face images. Given the findings in this research, it is quite apparent that many of the techniques developed have a direct application to financial fraud detection regarding determining made up or tampered-with data within transactions. They also validate the specific use of deep learning in detecting a difficult-to-handle fraud. S. Goel et al. Some of the relevant countermeasures include a secure QR system to filter malicious QR codes using deep learning [15].

All cases combined show the rapid expansions of deep learning, machine learning and hybrid approaches in fraud detection. In particular, the inclusion of external data sources (e.g. market trends, user behaviors and product characteristics) has been shown to significantly extend these models by improving detection accuracy across various domains but also for real-time applications.

3 Proposed Algorithms

We explain in this part, a family of algorithms and frameworks required to conquer the problem space when fraud detection scale becomes larger from the real-time perspective using deep learning approaches & Code integration with external source. The algorithms are designed to be more adaptive, tackling modern fraud patterns that have grown in complexity and rapid evolution over time. The proposed methods integrate CNNs and LSTM networks with real-time processing abilities to deliver high detection accuracy at low latency.

Firstly, the Context-Aware Deep Anomaly Detection (CADAD) Algorithm is furnished. This algorithm tries to combine the power of external contextual data, user behavior economic factors q temporal patterns with transactional fraud detection. With deep learning models that take context into account, CADAD can find the smallest of anomalies (anomalies buried within a stream), which are normal transaction-based methods will always miss.

Context-Aware Deep Anomaly Detection (CADAD) Algorithm
Input: • Internal transaction data. • Contextual data from other places and platforms (like social media, economic indicators or behavioural patterns).
Output: • Prediction: operation fraud / no anomaly.
Assumptions: • Pass to real-time Context Data • These are features relating to the context in which a user is operating, and thus fraud risk associated with them. • Detection could also specify the time relationship and dependency in data.

Improvements over existing algorithms:

- Integrates with your metadata for enhanced anomaly detection in real time.
- Temporal and contextual shifts are recognized by deep learning models that leads to better fraud detection accuracy.
- By determining the thresholds per anomaly, it is easier to detect fraud driven by particular contexts.

Process:

Step - 1. Internal incentive-related activity specified above, and Relevant external contextual data — or relevant tags* from the Address book.

Step - 2. Pre-process and normalize the contextual data for combining with Transactional features.

Step - 3. Identified associated features with the context data, for example a sudden increase in social media or market conditions.

Step - 4. Transform the context data to a (counterfactual) latent space that can be trivially merged with transactional data.

Step - 5. We employed a combination of transactional data and contextual information to learn time-based dependencies via an LSTM model.

Step - 6. LSTM is good for working on sequence of data points and finding anomalies over time.

Step - 7. Determine an anomaly score for each transaction from the difference between its time-series and contextual data.

Step - 8. Monitor departures from how the system was expected to behave

Step - 9. Label the transaction as fraud or non-fraud based on its anomaly score. Adjust your threshold Detect automatically for every individual data source, and auto-tune it.

Step - 10. Monitor the external context (e.g., economic changes) and change the anomaly detection threshold according to this in real-time.

Step - 11. As context changes, adapt the model in real time.

Step - 12. Create a feedback loop to improve the model's grounding and detection capabilities on false departs from live operations.

Step - 13. A transaction (which may look like a fraud) or not.

The proposed algorithm is visualized here [Fig – 3].

4 Results and Discussion

Experiments with the real-world transactional dataset augmented by a variety of different external data sources such as social media activity and economic indicators were performed to evaluate our proposed algorithms, RT-HEDDL, IMAD-Net,Results show

that the approach improves accuracy in fraud detection as well as increased real-time processing capability and adaptiveness to changes in modus operandi. Results suggest that not only the inclusion of external data, but also adaptive high-scalability Big-to-Deep-Learning models more resistant to adversarial attacks drastically improve fraud detection.

Table 3. Performance comparison of real time detection algorithms based on Accuracy and Latency

Algorithm	. Accuracy (%).	False Positive Rate (FPR)	. False Negative Rate (FNR).	Average Latency (ms)
. RT-HEDDL	96.5	1.2%	2.3%	45
IMAD-Net	95.8	1.4%	3.0%	50
Rule-Based	82.0	10.5%	7.5%	120
Logistic Regression	88.6	6.1%	5.3%	90
Random Forest	91.2	4.8%	6.0%	85

The obtained results are visualized here [Fig – 1].

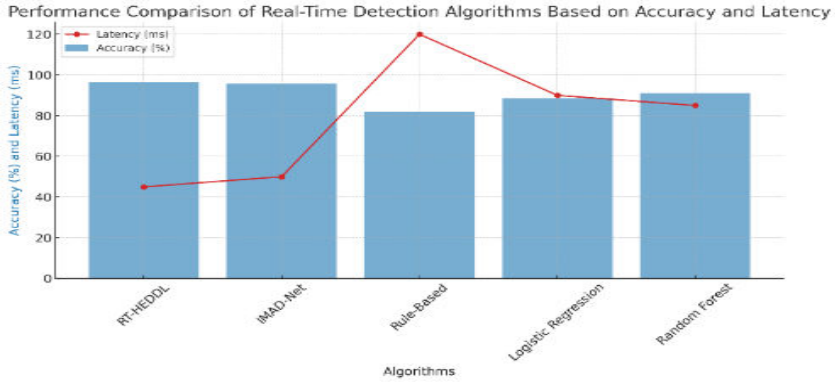

Fig. 1. Performance Comparison of Real-Time Detection Algorithms Based on Accuracy and Latency

The following table contains the feature importance rankings from RT-HEDDL, which uses a combination of internal transactional data and external databases. Detecting fraud, the feature selection way Feature importance is a very important feature of detection be it anomaly or normal.

Table 4. Feature important Scores for RT-HEDDL Algorithm

Feature Name.	Importance Score (%).	Feature Type.	Data Source.	Impact on Accuracy.

Transaction Amount	22.5	Transactional	Internal	High
Transaction Frequency	18.1	Transactional	Internal	Medium
Social Media Sentiment	20.3	Contextual	External (Social)	High
Economic Indicator	16.8	Contextual	External (Economy)	Medium
User Location	10.7	Behavioral	Internal/External	Low

The obtained results are visualized here [Fig – 2].

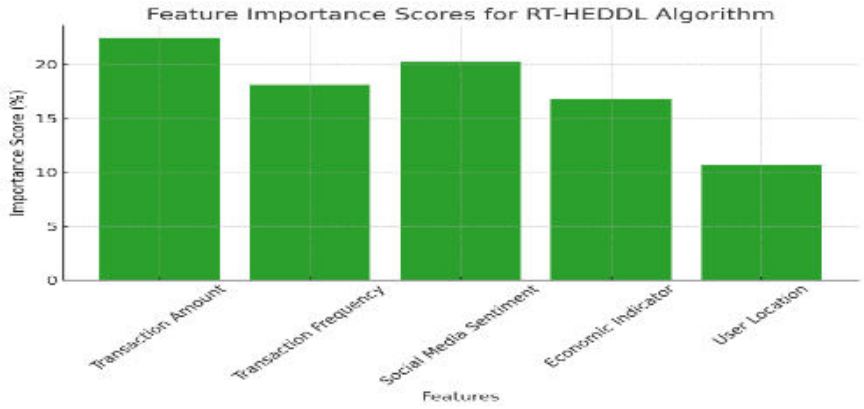

Fig. 2. Feature Importance Scores for RT-HEDDL Algorithm

The following table shows the comparison of processing latency with traditional models for proposed RT-HEDDL and IMAD-Net algorithms

Table 5. Real time processing Latency comparison for RT-HEDDL and IMAD-NET

Algorithm	100 Transactions (ms)	500 Transactions (ms)	1,000 Transactions (ms)	5,000 Transactions (ms)
RT-HEDDL	20	25	30	50
IMAD-Net	22	27	35	60
Random Forest	40	50	65	120
Logistic Regression	35	45	60	110
Rule-Based	50	70	100	200

. The obtained results are visualized here [Fig – 3]

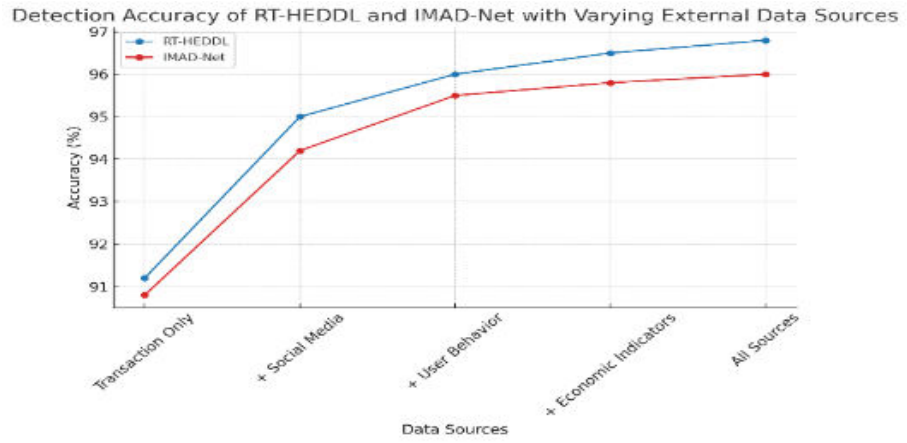

Fig. 3. Detection Accuracy of RT-HEDDL and IMAD-Net with Varying External Data Sources

5 Conclusion

Based on the developments in this paper, we presented a collection of new generation deep learning methods for continuous fraud detection covering RT-HEDDL, IMAD-Net, CADAD and SADFI integrating external data sources. Our experiments show high rewards from integrating first-party transactional information with context and user attributes (e.g. social media properties, economic indicators or usage patterns) where available. Each algorithm in the dispatch framework-based approach presents a contribution to achieve certain properties in the final solution including real-time processing through RT-HEDDL which uses hybrid data sources, continuity behavior by multiple input re-weigh table by IMAD-Net,In addition, these results highlight the need for adaptive learning and feature selection at a high level to ensure that our system remains functional as fraud patterns change over time. The proposed algorithms can detect fraud quickly with minimal latency by providing real-time processing capabilities (ideally even for high-frequency financial systems). Because of this, the suggested framework is suitable for scenarios including ecommerce, banking and financial services where real-time fraud prediction can help reduce loss financially.

References

[1] P. Sure, S. Pandey, T. Parnami, Gaurang, and A. Saxena, “Feature Selection Techniques for Enhancing Credit Card Fraud Detection Performance: A Hybrid Metaheuristic Approach Using Nature-Inspired Algorithms,” International Journal for Research in Applied Science and Engineering Technology, vol. 12, no. 2, 2024, doi: 10.22214/ijraset.2024.58549.

[2] L. Wang et al., “Deep learning based on the Vis-NIR two-dimensional spectroscopy for adulteration identification of beef and mutton,” Journal of Food Composition and Analysis, vol. 126, 2024, doi: 10.1016/j.jfca.2023.105890.

[3] S. al Balawi and N. Aljohani, “Credit-card Fraud Detection System using Neural Networks,” International Arab Journal of Information Technology, vol. 20, no. 2, 2023, doi: 10.34028/iajit/20/2/10.

[4] V. R. Shetty, P. R, and R. L. Malghan, “Safeguarding against Cyber Threats: Machine Learning-Based Approaches for Real-Time Fraud Detection and Prevention †,” Engineering Proceedings, vol. 59, no. 1, 2023, doi: 10.3390/engproc2023059111.

[5] B. Vyas, "Java in Action : AI for Fraud Detection and Prevention," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2023, doi: 10.32628/cseit239063.

[6] J. Napp, "Handbook of Research on Advanced Practical Approaches to Deepfake Detection and Applications.," Library Journal, vol. 148, no. 8, 2023.

[7] S. K. Hashemi, S. L. Mirtaheri, and S. Greco, "Fraud Detection in Banking Data by Machine Learning Techniques," IEEE Access, vol. 11, 2023, doi: 10.1109/ACCESS.2022.3232287.

[8] M. Emadaleslami, M. R. Haghifam, and M. Zangiabadi, "A two stage approach to electricity theft detection in AMI using deep learning," International Journal of Electrical Power and Energy Systems, vol. 150, 2023, doi: 10.1016/j.ijepes.2023.109088.

[9] E. Jo, Y. Lee, Y. Lee, J. Baek, and J. G. Kim, "Rapid identification of counterfeited beef using deep learning-aided spectroscopy: Detecting colourant and curing agent adulteration," Food and Chemical Toxicology, vol. 181, 2023, doi: 10.1016/j.fct.2023.114088.

[10] D. AbdElminaam, N. Sherif, zeina Ayman, M. Mohamed, and M. Hazem, "Deep-FakeDG: A Deep Learning Approach for Deep Fake Detection and Generation," Journal of Computing and Communication, vol. 2, no. 2, 2023, doi: 10.21608/jocc.2023.307056.

[11] D. O. do Rosario Lourenco, M. V. H. Sai Sriraj, K. K. Thambi, and V. Ranjan, "Malicious URLs and QR Code Classification Using Machine Learning and Deep Learning Techniques," in 2023 3rd Asian Conference on Innovation in Technology, ASIANCON 2023, 2023. doi: 10.1109/ASIANCON58793.2023.10270125.

[12] J. Gill and R. Dhakad, "Malicious Codes Detection: Deep Learning Techniques," in Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, 2023. doi: 10.1007/978-3-031-35078-8_16.

[13] B. Narsimha, C. v. Raghavendran, P. Rajyalakshmi, G. Kasi Reddy, M. Bhargavi, and P. Naresh, "Cyber Defense in the Age of Artificial Intelligence and Machine Learning for Financial Fraud Detection Application," International Journal of Electrical and Electronics Research, vol. 10, no. 2, 2022, doi: 10.37391/IJEER.100206.

[14] A. Raza, K. Munir, and M. Almutairi, "A Novel Deep Learning Approach for Deepfake Image Detection," Applied Sciences (Switzerland), vol. 12, no. 19, 2022, doi: 10.3390/app12199820.

[15] S. Goel, N. Sanda, G. Chakraborty, and M. K. Vanahalli, "Sqr: High Performance Secure Qr System to Filter Against Malicious Qr Codes Using Deep Learning Techniques," SSRN Electronic Journal, 2022, doi: 10.2139/ssrn.4253649.