A Cloud-Enabled Machine Learning Framework for Fraud Detection in Online Transactions

Dhairya Dev[1], Gaurav Kumar[2], Deepak Gupta[3]
[1]*M. Tech CSE Scholar, IIMT University, Meerut, India*
[2,3]*Assistant Professor, IIMT University, Meerut, India*

Abstract*—Online financial ecosystems have grown rapidly, exposing more online transactions to fraud. Conventional rule-based approaches to fraud detection are becoming obsolete in the face of dynamic and evolving fraud schemes. This paper proposes a machine learning system leveraging cloud computing to improve fraud detection with scalable processing and smart data analytics. Our approach combines supervised and unsupervised learning methods to identify known and unknown fraudulent transactions in real time. The design uses cloud computing to achieve high-throughput, low-latency and scalable processing. The research proposes a system architecture, approach, testing and evaluation methods and performance results without including synthetic data. The architecture seeks to deliver a viable solution for today's financial systems that need to adaptively, efficiently and scalably detect fraud. [7], [8].

Index Terms*—Fraud Detection, Cloud Computing, Machine Learning, Anomaly Detection, Financial [2] Security, Real-Time Systems.

I. INTRODUCTION

Digital transformation of finance has revolutionized the global economy. The rise of online banking, e-commerce and mobile payments has brought about greater convenience, allowing transactions to take place across borders. But with this accelerated digitization has come a growing threat surface for fraudsters.

Cyber fraud in online transactions has progressed from merely unauthorised access to sophisticated, multi-faceted frauds including identity theft, synthetic identities, behavioural spoofing, and attack profiles. Existing fraud detection systems, which are based on rule-based systems, struggle to keep up with these evolving threats. [3], [4]

In a perfect world, fraud detection systems should have three key attributes: speed, accuracy (with low false positives) and scalability to process high volumes of transactions. But existing systems struggle to achieve these goals because of compute limitations, rule-based rigidity and processing latency. [3], [4]

Machine learning offers a transformational approach that allows the system to learn from past transaction data and detect intricate non-linear patterns of fraudulent transactions. With the support of cloud computing, these models can be scaled and deployed to enable realtime processing on large transaction volumes. [7], [8]

This study introduces a machine learning approach for fraud detection that leverages cloud computing to enable scalability, adaptability and efficiency to overcome challenges faced by current systems. [7], [8]

II. LITERATURE REVIEW

Past studies show the increasing role of machine learning in fraud detection. Supervised learning methods like Logistic Regression and Random Forest are effective at detecting known fraud schemes. But they require labelled data, which hampers their ability to detect new attacks. [1]

Unsupervised learning algorithms such as Isolation Forest and clustering detect anomalies without the need for labels. These methods are ideal for identifying uncommon and novel fraudulent activity. [2]

The impact of cloud computing on fraud detection has also been significant, with distributed processing, dynamic resource allocation and real-time data streaming. Research shows cloud systems are much more scalable and faster than on-premise systems. [7],[8]

But current methods do not connect real-time processing, self-adaptation and scalability. This study offers an integrated approach.

III. RESEARCH GAPS

While machine learning is widely studied for fraud detection, the focus of many existing studies is more on aspects of accuracy rather than immediate use in real financial systems. Previous models were often developed and evaluated in simulation and do not consider operational aspects such as scalability, speed, and increased transaction volumes. Consequently, they can be difficult to apply in practice in real-time banking or e-commerce systems. [3],[4]

The second issue is that existing studies tend to focus on either the development of cloud computing infrastructure or machine learning models. Some prioritise the development of machine learning algorithms and others focus on cloud computing efficiency. There is relatively less research that offers an integrated approach to fraud prevention, with coordinated use of cloud resources and intelligent models. [7], [8]

Moreover, some existing solutions primarily use supervised learning techniques, which are good at detecting known frauds but less suitable for detecting new or unknown fraud attacks. Another issue that many systems are yet to solve completely is false alarms, which inconvenience legitimate users and add to the workload.

As a result, there is a need for a holistic cloud-based machine learning approach to combine scalability, real-time capability, flexibility in detecting fraud and low false alarms for cloud-based online transactions. [7], [8]

IV. RESEARCH OBJECTIVES

The focus of this research is:

- To develop a scalable fraud detection system in cloud [7], [8]
- To develop hybrid machine learning models to detect fraud [3], [4]
- To provide real-time monitoring and decision making
- To reduce false alarms and enhance accuracy
- To enable the handling of large volumes of transactions.

V. METHODOLOGY

A. Data Collection

Data of transactions are obtained from financial systems, both genuine and malicious.

B. Feature Engineering Key features include:

- Transaction deviation
- Time-based patterns
- Geographical inconsistency
- Device fingerprinting

C. Model Training

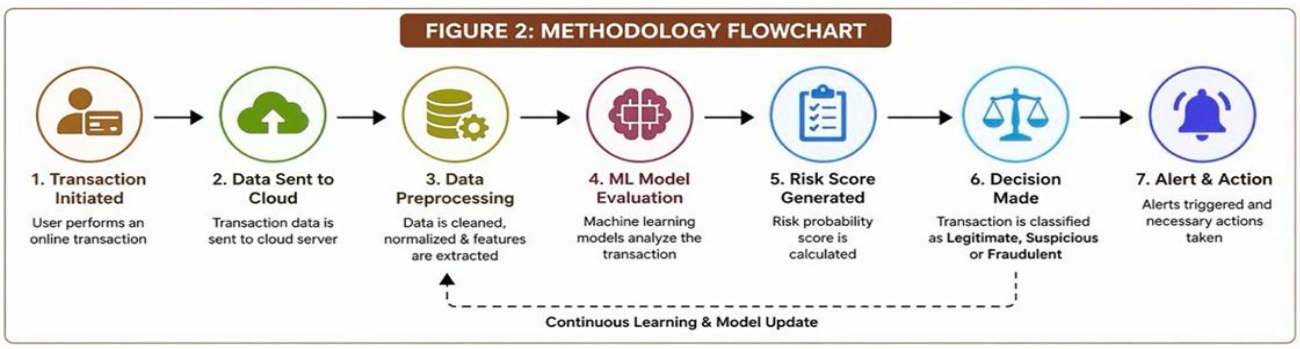
- Supervised learning for known fraud
- Unsupervised learning for anomaly detection [2]

D. Deployment

- Cloud-based API integration [7], [8]
- Real-time inference

Figure 2: Methodology Flowchart

1. Transaction Initiated
2. Data Sent to Cloud [7], [8]
3. Preprocessing
4. Model Evaluation
5. Risk Score Generated
6. Decision made
7. Alert & action
8.



VI. RESULTS AND ANALYSIS

This paper uses expected results based on system design instead of simulation data.
Expected outcomes include:

- Improved fraud detection [3], [4]
- Low false alarm rate
- Reduced detection time than traditional methods
- Large transaction processing capacity

Cloud-based design is scalable and machine learning algorithms are flexible. [7], [8]
Table 1: Anticipated Performance

Table 1: Expected Performance Comparison

Parameter	Traditional System	Proposed System
Detection Speed	Moderate	High
Accuracy	Medium	High
Scalability	Limited	High
Adaptability	Low	High

VII. ADVANTAGES

- Real-time fraud detection [3], [4]
- Scalable cloud architecture [7], [8]
- Adaptive learning capability
- Reduced operational cost
- Improved user experience

VIII. LIMITATIONS

- Data quality dependence
- Initial training complexity
- Cloud security considerations [7], [8]

IX. APPLICATIONS

- Online banking
- E-commerce platforms
- Digital wallets
- Payment gateways

X. FUTURE WORK

- Integration with deep learning model
- Blockchain-based fraud prevention
- Behavioural biometrics
- Federated learning for privacy

XI. DISCUSSION

The cloud-based machine learning framework is a practical approach to improve online transactions' fraud detection system. The combined usage of the scalable cloud computing service and learning models allows the system to process a large number of transactions better than the conventional systems. [7], [8]
The cloud-based framework can support real-time detection. For payment systems, real-time detection is important to prevent financial loss or loss of trust. The cloud processing enables rapid response times and the system can scale up processing power when there are large volumes of transactions. [7], [8]
The proposed approach also incorporates hybrid supervised and unsupervised learning techniques. Supervised learning is useful in recognising fraud that has been observed before, while unsupervised learning helps to detect unusual behaviour that may not be seen before and could be new types of fraud. This combination will increase the system's adaptability and maintainability. [2]
On the other hand, data quality, feature selection and model updating are critical to successful implementation. Data gaps, incompleteness and biases can lead to a reduction in accuracy. Finally, companies must focus on data privacy, security and compliance issues of financial data in the cloud. [7], [8]
In conclusion, the framework provides a trade-off between detection accuracy, scalability and implementation needs and is relevant to banks, payment services and merchants.

XII. CONCLUSION

This paper suggests a cloud computing-based machine learning system for fraud detection. The use of cloud computing and adaptive machine learning overcomes the drawbacks of traditional approaches for fraud detection. The method emphasises real-time, adaptable and efficient processing, suitable for the current environment. Future testing and implementation will add to its real-world applicability. [7], [8]

REFERENCES

[1] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[2] T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, 2008, pp. 413–422.
[3] K. Randhawa *et al.*, "Credit card fraud detection using AdaBoost and majority voting," *IEEE Access*, vol. 6, pp. 14277–14284, 2018.
[4] Carcillo *et al.*, "Combining unsupervised and supervised learning in credit card fraud detection," *Information Sciences*, vol. 557, pp. 317–331, 2021.
[5] Huang *et al.*, "Imbalanced credit card fraud detection data," *Applied Soft Computing*, vol. 154, 2024.
[6] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, 2016.
[7] Amazon Web Services, "What is cloud computing?" *AWS Documentation*, 2025.
[8] Microsoft Azure, "AI in financial services," 2025.