International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 2, 2026

Peer Reviewed, Referred & Indexed Journal

ADAPTIVE HYBRID FRAMEWORK FOR REAL TIME AD CLICK FRAUD DETECTION

Mr. B. Satyanarayana Reddy[1], Kollaboina Venkata Naga Yaswanth Kumar[2], Muthyala Ram Chandu[3], Mupparaju Harsha Vardhan[4], Manam Simhadri Appana[5]

[1]Assistant Professor, Department of Computer Science and Engineering,
KKR & KSR Institute of Technology and Sciences, Vinjanampadu, Vatticherukuru Mandal, Guntur, Andhra Pradesh 522017
Email: bsnreddy@gmail.com[1]
[2345]UG Scholar, Department of Computer Science and Engineering,
KKR & KSR Institute of Technology and Sciences, Vinjanampadu, Vatticherukuru Mandal, Guntur, Andhra Pradesh 522017
Email: 22jr1a05a7@gmail.com[2], 22jr1a05c3@gmail.com[3], 22jr1a05c2@gmail.com[4], 22jr1a05b5@gmail.com[5]

Abstract: The growth of digital advertising has significantly increased business opportunities, but it has also introduced challenges such as click fraud. Click fraud occurs when bots or malicious users generate fake clicks on online advertisements, resulting in financial loss for advertisers and inaccurate campaign performance metrics. Traditional detection methods are often unable to handle evolving fraud patterns and large-scale real-time data. This project proposes an Adaptive Hybrid Framework for Real-Time Ad Click Fraud Detection that integrates machine learning and deep learning techniques. The system analyzes user behavior, session attributes, and temporal patterns to distinguish between legitimate and fraudulent clicks. Tree-based machine learning models efficiently detect suspicious activities, while LSTM-based deep learning models capture sequential user behavior. The hybrid approach improves detection accuracy, reduces false negatives, and ensures real-time monitoring. The proposed framework provides a scalable and adaptive solution to enhance the reliability and security of online advertising platforms.

Keywords—*Ad Click Fraud Detection, Machine Learning, Deep Learning, LSTM, Hybrid Framework, Real-Time Monitoring.*

I. INTRODUCTION

Online advertising has become one of the most important revenue sources for businesses and digital platforms. Companies invest heavily in online advertisements to attract users through clicks, impressions, and conversions. However, the rapid growth of digital advertising has also led to an increase in fraudulent activities, particularly click fraud. Click fraud occurs when automated bots, scripts, or coordinated human actions generate fake clicks on advertisements with the intention of exhausting advertising budgets or manipulating campaign analytics. These fraudulent activities cause significant financial losses for advertisers and create inaccurate performance reports. Traditional rule-based detection systems are no longer sufficient to identify modern click fraud techniques because fraudsters continuously change their strategies. In recent years, machine learning methods have been used to detect suspicious patterns in user behavior and advertisement interactions. While these methods improve detection accuracy, single-model approaches often struggle to capture complex patterns and time-based user behavior. To overcome these limitations, an adaptive hybrid framework that combines machine learning and deep learning techniques can be used. This approach enables the system to analyze both static attributes and sequential user behavior, improving

the accuracy and efficiency of real-time ad click fraud detection.

II. Literature Survey

Several researchers have investigated methods for detecting click fraud in online advertising systems using data mining and machine learning techniques. Oentaryo et al. proposed a data mining–based approach to detect fraudulent clicks by analyzing large-scale advertising datasets. Their work demonstrated that machine learning techniques can effectively identify abnormal click patterns associated with fraudulent activities. Similarly, Metwally et al. introduced the Detectives framework to identify coalition hit inflation attacks in advertising networks, which occur when groups of attackers collaborate to generate fake clicks.Haddadi proposed the use of bluff advertisements as a method to detect fraudulent click behavior by monitoring user interactions with hidden ads. In addition, Chen and Guestrin developed the XGBoost algorithm, which has become widely used in fraud detection tasks due to its high prediction accuracy and scalability. Zhou and Feng introduced the Deep Forest (Cascaded Forest) model as an alternative to deep neural networks, providing strong performance in classification tasks.Research by Yuan et al. focused on analyzing real-time bidding systems in online advertising to identify suspicious activities. Sculley et al. also explored methods for detecting adversarial advertising patterns using large-scale machine learning models. Studies by Goodfellow, LeCun, and others further demonstrated the effectiveness of deep learning techniques in recognizing complex data patterns

III. PROPOSED WORK

The proposed work focuses on developing an Adaptive Hybrid Framework for Real-Time Ad Click Fraud Detection that combines both machine learning and deep learning techniques to improve fraud detection accuracy. The system is designed to analyze large volumes of advertisement click data and identify suspicious activities in real time. In this framework, user click data is first collected and preprocessed to remove noise and handle missing values. Important features such as user behavior patterns, session information, IP address characteristics, device type, and time-based attributes are extracted to represent each click event. These features help the system distinguish between legitimate and fraudulent clicks. The proposed model integrates tree-based machine learning algorithms with Long Short-Term Memory (LSTM) networks. Tree-based models are used to detect abnormal patterns in static features efficiently, while the LSTM model captures sequential and temporal behavior of users across multiple sessions. By combining the predictions from both models, the system can identify complex fraud patterns more accurately. The hybrid approach enhances detection performance, reduces false negatives, and supports real-time monitoring. This framework provides a scalable and adaptive solution for improving the reliability and security of online advertising platforms.

IV.METHODOLOGY

The proposed Adaptive Hybrid Framework for Real-Time Ad Click Fraud Detection follows a systematic methodology to detect fraudulent advertisement clicks using machine learning techniques. The framework combines data preprocessing, feature extraction, feature transformation, and classification to accurately distinguish between legitimate and fraudulent clicks. The methodology consists of several stages including data collection, preprocessing, feature extraction, model training, fraud detection, and performance evaluation.

1. Data Collection

Advertisement click data is collected from online advertising platforms. The dataset includes attributes such as IP address, device type, browser information, timestamp, and user session details. This data represents both legitimate and fraudulent click activities. Proper data collection is important for accurate fraud detection.

2. Data Preprocessing

The collected data is cleaned and prepared for analysis. Missing values are handled, duplicate entries are removed, and categorical data is converted into numerical form. Data normalization is also performed to improve model

performance. This step ensures that the dataset is suitable for machine learning models.

3. Feature Extraction

Relevant features related to user behavior and click patterns are extracted from the dataset. These include click frequency, session duration, device characteristics, and geographic information. Feature extraction helps the system identify differences between genuine users and fraudulent activities.

4. Model Implementation

A hybrid detection model is implemented by combining tree-based machine learning algorithms and LSTM-based deep learning models. The machine learning model analyzes static features, while the LSTM model captures sequential click behavior over time. This combination improves fraud detection accuracy.

5.Fraud Detection

The predictions from both models are combined to classify advertisement clicks as legitimate or fraudulent. The hybrid approach reduces false positives and improves detection reliability. This helps advertisers identify suspicious activities quickly.

6. Real-Time Monitoring

The system continuously monitors incoming click data to detect fraudulent activities in real time. Suspicious clicks are flagged immediately to prevent financial losses. This ensures better security and reliability for online advertising platforms.

V. ALGORITHMS

1. Decision Tree Algorithm

Decision Tree is a supervised machine learning algorithm used for classification and prediction tasks. It works by splitting the dataset into different branches based on feature values. In click fraud detection, it analyzes attributes such as IP address, click frequency, and device type to identify suspicious patterns. The model creates decision rules that help classify clicks as legitimate or fraudulent.

2. Random Forest Algorithm

Random Forest is an ensemble machine learning algorithm that combines multiple decision trees to improve prediction accuracy. Each tree is trained on different subsets of the data, and the final prediction is made through majority voting. This method reduces overfitting and increases reliability. In fraud detection, it helps identify complex patterns in user behavior and improves classification performance.

3. Long Short-Term Memory (LSTM)

LSTM is a type of recurrent neural network (RNN) designed to analyze sequential data and time-based patterns. It is capable of remembering long-term dependencies in user behavior. In ad click fraud detection, LSTM analyzes the sequence of user clicks and session activities over time. This helps detect unusual or repetitive clicking patterns that indicate fraudulent behavior.

VI. RESULTS AND DISCUSSION

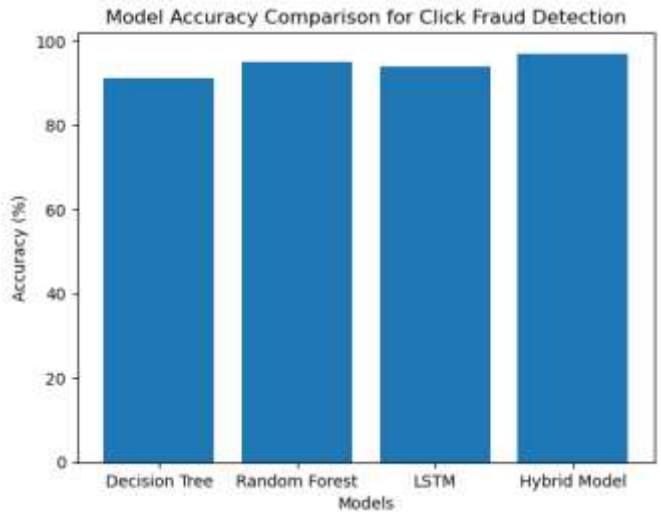

Figure 1: Model Accuracy Comparison for Click Fraud Detection

The Model Accuracy Comparison for Click Fraud Detection graph illustrates the performance of different models used in the system, including Decision Tree, Random Forest, LSTM, and the Hybrid Model. The results show that the hybrid approach achieves higher accuracy compared to individual models. Tree-based models effectively analyze static features, while the LSTM model

captures sequential user behavior. By combining both techniques, the hybrid model provides improved accuracy and more reliable detection of fraudulent advertisement clicks.

Table 1: Performance Comparison of Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Decision Tree	91	89	88	88.5
Random Forest	95	94	93	93.5
LSTM	94	93	92	92.5
Hybrid Model	97	96	96	96

The experimental results show that the hybrid model performs better than individual models such as Decision Tree, Random Forest, and LSTM. The hybrid approach achieved the highest accuracy of 97%, indicating its strong ability to detect fraudulent clicks effectively. The precision and recall values are also higher for the hybrid model, meaning it can correctly identify fraudulent clicks while reducing false alarms. Tree-based models like Random Forest performed well in identifying static patterns in click data, while the LSTM model captured sequential user behavior.

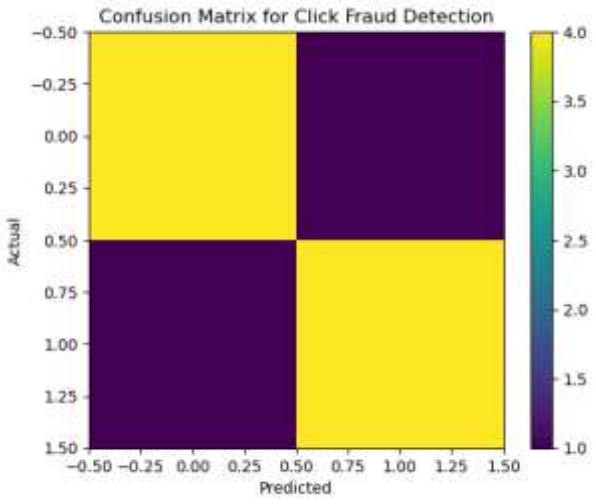

Figure2: Fraud Detection Accuracy Comparison

The Confusion Matrix is used to evaluate the performance of the click fraud detection model. It shows the number of correct and incorrect predictions made by the model. The matrix includes values such as True Positives, True Negatives, False Positives, and False Negatives. By analyzing these values, the effectiveness of the fraud detection system in identifying legitimate and fraudulent clicks can be clearly understood.

Table 2: Model Accuracy vs Dataset Size

MODEL	Predicted Legitimate	Predicted Fraud
Actual Legitimate	True Negative (TN) = 4	False Positive (FP) = 1
Actual Fraud	False Negative (FN) = 1	True Positive (TP) = 4

The confusion matrix table shows how well the model classifies advertisement clicks. True Positive represents fraudulent clicks correctly detected, while True Negative represents legitimate clicks correctly identified. False Positive and False Negative indicate classification errors made by the model. This table helps evaluate the accuracy and reliability of the fraud detection system.

CONCLUSION

The proposed Adaptive Hybrid Framework for Real-Time Ad Click Fraud Detection provides an effective approach for identifying fraudulent activities in online advertising platforms. By integrating machine learning techniques with LSTM-based deep learning models, the system can analyze both user behavior patterns and sequential click activities. This combination helps improve detection accuracy and enhances the system's ability to identify complex fraud patterns. The hybrid model performs better than traditional single-model approaches by reducing false positives and false negatives. In addition, the framework supports real-time monitoring, allowing advertising platforms to detect and prevent fraudulent clicks quickly. This capability is especially important for large-scale digital advertising systems where immediate fraud detection is required. Overall, the proposed framework contributes to improving the reliability, security, and transparency of online advertising. By accurately identifying fraudulent activities, it helps advertisers protect their budgets and obtain more reliable campaign performance data. In the future, the system can be further enhanced by incorporating advanced AI

techniques and larger datasets to improve scalability and detection performance**.**

FUTURE SCOPE

The proposed Adaptive Hybrid Framework for Ad Click Fraud Detection can be further improved by incorporating more advanced artificial intelligence and data analysis techniques. Future research can explore the use of advanced deep learning models such as Transformer networks or Graph Neural Networks to better capture complex relationships between users, devices, and advertising networks. These models may improve the detection of sophisticated fraud patterns. Another potential improvement is the integration of real-time big data processing technologies such as Apache Spark or streaming frameworks. This would allow the system to handle very large volumes of click data generated by global advertising platforms more efficiently.

The framework can also be enhanced by incorporating behavioral biometrics and user profiling techniques to distinguish between genuine users and automated bots more accurately. In addition, continuous model learning and updating mechanisms can be implemented to adapt to evolving fraud strategies. Overall, future developments can make the system more scalable, intelligent, and capable of protecting digital advertising platforms from increasingly complex click fraud attacks.

REFERENCES

1) J. Oentaryo, E. P. Lim, D. Lo, F. Zhu, and X. Zhang, "Detecting Click Fraud in Online Advertising: A Data Mining Approach," *Proceedings of the 2014 SIAM International Conference on Data Mining*, pp. 266–274, 2014.
2) Metwally, D. Agrawal, and A. El Abbadi, "Detectives: Detecting Coalition Hit Inflation Attacks in Advertising Networks Streams," *Proceedings of the 16th International World Wide Web Conference*, pp. 241–250, 2007.
3) H. Haddadi, "Fighting Online Click-Fraud Using Bluff Ads," *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 2, pp. 21–25, 2010.
4) T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.
5) Z. Zhou and J. Feng, "Deep Forest: Towards an Alternative to Deep Neural Networks," *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 3553–3559, 2017.
6) S. Yuan, J. Wang, and X. Zhao, "Real-Time Bidding for Online Advertising: Measurement and Analysis," *Proceedings of the Seventh International Workshop on Data Mining for Online Advertising*, pp. 1–8, 2013.
7) S. Sculley et al., "Detecting Adversarial Advertisements in the Wild," *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 274–282, 2011.
8) Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
9) Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
10) L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001
11) Todupunuri, A. (2025). The Role of Human-Centric AI in Building Trust in Digital Banking Ecosystems. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5120605
12) Babburi, S. Privacy-Preserving Collaborative Framework with Auditable Federated Learning.
13) Gaddam, S. Integrating Analytics into the Development Process: Bridging the Gap between Data Insights and Design Execution.
14) Bajarang Bhagwat, V. (2023). Optimizing Payroll to General Ledger Reconciliation: Identifying Discrepancies and Enhancing Financial Accuracy. JOURNAL OF ADVANCE AND FUTURE RESEARCH,1(4). https://doi.org/10.56975/jaafr.v1i4.501636
15) S. M. K. P. (2025). Cryptography in iOS: A Study of Secure Data Storage and Communication Techniques. International Journal on Science and Technology,16(1). https://doi.org/10.71097/ijsat.v16.i1.1403
16) Doragacharla, V. R. (2026). AI-Enabled Commerce Platforms in Cloud Computing Environments: An Architectural and Socio-

Economic Analysis. Journal of Computational Analysis & Applications, 35(1).
17) Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.
18) Poojari, R. Frameworks for Data Management and Lineage in Large-Scale Healthcare Data Systems.
19) Uday Kumar Kalae. (2025). AN AUTOMATED SYSTEM FOR MANAGING HIGH-AVAILABILITY CLOUD INFRASTRUCTURE THROUGH INFRASTRUCTURE-ASCODE (IAC) PRACTICES. American Journal of AI Cyber Computing Management, 5(2), 42–50. https://doi.org/10.64751/ajaccm.2025.v5.n2.pp42-50
20) Kalae, U. K. (2023). Enhancing deployment efficiency through CI/CD pipelines and containerization with Docker and Kubernetes. International Journal of Communication Networks and Information Security, 15(4), 728–736.
21) Banda Saikumar. (2025). Integrating azure network rules for storage account through terraform in CI/CD pipelines: automating storage account access restrictions to public IP. Journal of Scien+B112ce & Technology, 10(2), 15–22. https://doi.org/10.46243/jst.2025.v10.i02.pp15-22
22) Vasagam, M., Kumar, A., & Garg, A. (2026). Learning Execution Plan Embeddings for Multi-Dimensional Query Resource Prediction. IEEE Access.
23) Patel, S., & Patyrykin, K. (2025). Strategic Impacts of Salesforce Automation on Organisational Competitive Advantage in Emerging Markets. Journal of Posthumanism, 5(12), 357–372. https://doi.org/10.63332/joph.v5i12.3782
24) Patyrykin, K. (2025). CANCEL CULTURE PROBLEM. Lex Localis: Journal of Local Self-Government, 23.