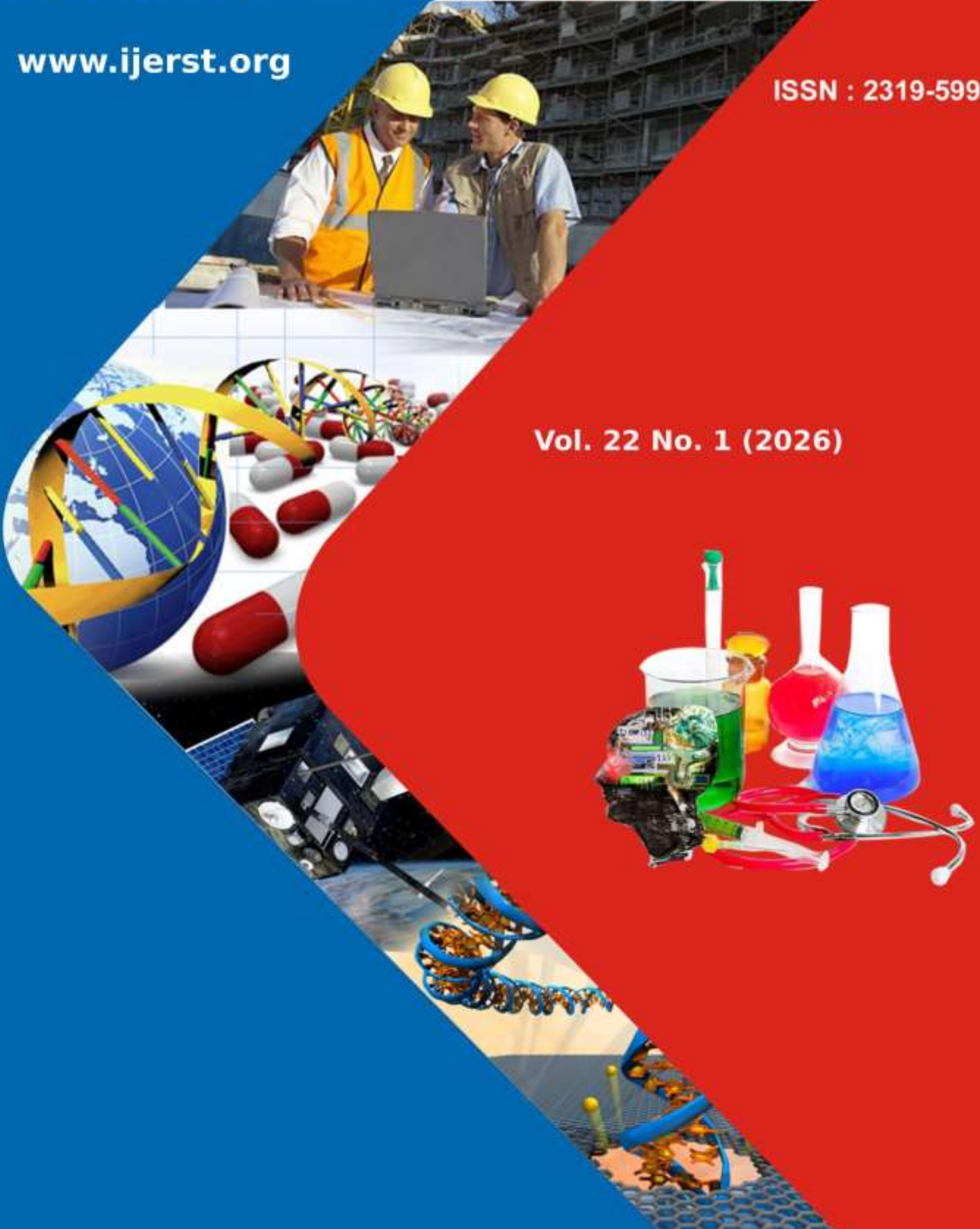
International Journal of Engineering Research and Science & Technology

IJERST

www.ijerst.org

ISSN : 2319-5991



Vol. 22 No. 1 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 1, 2026

Peer Reviewed, Referred & Indexed Journal

Research

AD CLICK FRAUD DETECTION

ALTHI TARUN[1], CHOLLANGI POORNIMA[2],KANDAPU ESWARARAO[3], MUSINANI SAI YASWANTH[4], GORLI ADITYA VARDHAN[5] , Mrs.PEESAPATI V SUNEETHA[6]

[1] *Student, Department of CSE, Satya Institute of Technology and Management, Vizianagaramm, Andhra Pradesh, India. tarunalthi21@gmail.com .*

[2] *Student, Department of CSE, Satya Institute of Technology and Management, Vizianagaramm, Andhra Pradesh, India, chollangipoornima246@gmail.com.*

[3] *Student, Department of CSE, Satya Institute of Technology and Management, Vizianagaramm, Andhra Pradesh, Indiaeaswarkandapu@gmail.com .*

[4] *Student, Department of CSE, Satya Institute of Technology and Management, Vizianagaramm, Andhra Pradesh, India, yaswanthmusinani@gmail.com .*

[5] *Student, Department of CSE, Satya Institute of Technology and Management, Vizianagaramm, Andhra Pradesh, India, adityavardhan1602@gmail.com*

[6] *Assistant professor, Department of CSE, Satya Institute of Technology and Management, Vizianagaramm, Andhra Pradesh, India, pvsuneetha31@sitam.co.in.*

ABSTRACT

Ad click fraud has become one of the most critical challenges in the digital advertising ecosystem, causing huge financial losses to advertisers and reducing trust in online marketing platforms. Fraudulent clicks are generated either manually or through automated bots with the intention of inflating advertisement revenue, exhausting competitor budgets, or manipulating analytics. Traditional rule-based systems are no longer sufficient to handle the scale and complexity of modern fraud patterns. Therefore, this research proposes a data-driven approach using machine learning and deep learning algorithms for detecting and analyzing ad click fraud.The study explores the performance of multiple algorithms such as Convolutional Neural Networks (CNN), Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), Random Forest, and a proposed XGBoost-based system. Real-world ad click datasets are preprocessed through cleaning, normalization, feature engineering, and feature selection. Models are trained and evaluated using standard metrics such as accuracy, precision, recall, F1-score, and AUC. The results show that ensemble-based methods, especially XGBoost, outperform traditional and deep learning models in terms of accuracy, interpretability, and computational efficiency. The research concludes that combining robust preprocessing with advanced machine learning models can significantly improve fraud detection and reduce financial losses in digital advertising.

Keywords: Software Defined Networks (SDN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Deep Learning (DL), One-Dimensional Convolutional Neural Networks (1D-CNN), Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), Structured Deep Convolutional Neural Network (SDCNN).

Received: 31-01-2026 Accepted: 04-03-2026 Published: 11-03-2026

1. INTRODUCTION

1.1 Introduction:

Digital advertising has become one of the most dominant revenue-generating mechanisms in the modern internet ecosystem. Organizations increasingly rely on online advertisements to reach potential customers through pay-per-click (PPC), cost-per-impression (CPM), and cost-per-action (CPA) models. Among these, PPC advertising is widely used because advertisers only pay when users interact with advertisements. However, this model is highly vulnerable to fraudulent activities, commonly referred to as ad click fraud, where clicks are generated intentionally without genuine user interest. Ad click fraud is typically carried out by malicious actors using automated scripts, botnets, click farms, or deceptive applications that repeatedly click advertisements. These fraudulent clicks inflate advertising costs, distort campaign analytics, and reduce return on investment (ROI) for advertisers. According to industry reports, billions of dollars are lost annually due to ad fraud, making it a critical challenge for advertisers, publishers, and ad networks. Traditional rule-based detection systems struggle to identify sophisticated fraud patterns because fraudulent behaviors continuously evolve. As attackers adopt more intelligent techniques to mimic human behavior, static detection rules become ineffective. This has led to increasing interest in machine learning (ML) and deep learning (DL) approaches, which can automatically learn complex patterns from large-scale clickstream data. This project focuses on developing an intelligent ad click fraud detection system using machine learning and deep learning algorithms, with a primary emphasis on the XGBoost algorithm. The proposed system aims to accurately distinguish between legitimate and fraudulent clicks by analyzing user behavior, device characteristics, time patterns, and network attributes. By leveraging advanced data preprocessing, feature engineering, and robust classification techniques, the system enhances detection accuracy while minimizing false positives.

1.2 Problem Statement:

The rapid growth of online advertising platforms has led to an alarming increase in ad click fraud, posing significant financial and operational challenges to advertisers and publishers. Fraudulent clicks result in wasted advertising budgets, misleading performance metrics, and reduced trust in digital advertising systems. Despite the availability of existing detection mechanisms, ad click fraud remains a persistent and evolving problem. One of the primary challenges is the dynamic and adaptive nature of fraud patterns. Fraudsters continuously modify their strategies to bypass traditional detection rules by simulating human-like behavior, varying IP addresses, using proxy networks, and controlling click frequency. As a result, static rule-based systems and basic anomaly detection techniques fail to detect advanced fraud scenarios effectively. Another critical issue is the high volume and velocity of clickstream data generated by online advertising platforms. Millions of clicks are generated per second, making real-time detection computationally demanding. Existing systems often face scalability issues and struggle to maintain detection accuracy under large-scale data conditions. Therefore, the core problem addressed in this research is the development of a robust, scalable, and accurate ad click fraud detection system capable of identifying fraudulent clicks in real time.

The system must effectively handle imbalanced datasets, minimize false detections, adapt to evolving fraud patterns, and ensure high precision and recall. Machine learning and deep learning approaches, particularly ensemble methods like XGBoost, offer promising solutions to address these challenges.

1.3 Scope of Research:

The scope of this research is focused on designing and implementing an intelligent ad click fraud detection framework using machine learning and deep learning techniques. The research covers the complete lifecycle of fraud detection, from data collection and preprocessing to model training, evaluation, and deployment. This study primarily explores structured clickstream datasets containing attributes such as IP address, device type, operating system, browser details, timestamp, click frequency, and user behavior metrics. The research emphasizes identifying the most significant features that contribute to fraudulent behavior through feature selection and feature engineering techniques. Another important scope of this work is the comparative analysis of existing ML and DL models, including CNN, ANN, RNN, and Random Forest, against the proposed XGBoost-based system. Performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC are used to evaluate model effectiveness. Additionally, this research includes the development of a user-friendly detection system using Python-based tools such as Flask and Tkinter, enabling visualization of predictions and fraud reports. The scope is limited to supervised learning approaches using labeled datasets, and future extensions may include real-time streaming data and reinforcement learning-based fraud prevention systems.

2.LITERATURE SURVEY

1. Detecting Click Fraud in Online Advertising Using Machine Learning Techniques

Author: Clifton Phua et al.

Description: This paper presents one of the earliest comprehensive studies on click fraud detection using supervised machine learning. Decision Trees and Bayesian classifiers are applied to distinguish legitimate user clicks from fraudulent activities, showing that behavioral features significantly improve detection accuracy.

2. Click Fraud Detection in Online Advertising: A Data Mining Approach

Author: Ashkan Panahandeh et al.

Description: The authors explore data mining techniques such as clustering and classification to identify abnormal click patterns. The study emphasizes temporal features and IP-based behavior, demonstrating improved fraud identification in large-scale advertising datasets.x

3. Anomaly Detection for Internet Advertising Fraud

Author: Jie Zhang et al.

Description: This work focuses on unsupervised anomaly detection methods for identifying previously unseen fraud patterns. Statistical profiling and outlier detection models are used to handle evolving fraud strategies in real-time advertising systems.

4. Click Fraud Detection Using Random Forest and Gradient Boosting

Author: Xiang Zhang et al.

Description: The study compares ensemble learning techniques for click fraud detection. Results show that Random Forest and Gradient Boosting outperform traditional classifiers in terms of accuracy and robustness on imbalanced ad-click datasets.

5. A Deep Learning Framework for Click Fraud Detection

Author: Yuan Liu et al.

Description: This paper introduces deep neural networks to capture complex user behavior patterns. The proposed model effectively learns nonlinear relationships in clickstream data, improving recall for sophisticated fraud attacks.

6. Real-Time Click Fraud Detection Using Streaming Data Analytics

Author: Sandeep Chawla et al.

Description: The authors propose a real-time fraud detection framework using streaming machine learning models. The system dynamically adapts to changing fraud behaviors, making it suitable for large-scale online advertising platforms.

7. Feature Engineering for Effective Click Fraud Detection

Author: Mohammad Ahsan et al.

Description: This work highlights the importance of feature selection, including click frequency, session duration, and geographic patterns. Experimental results confirm that well-engineered features significantly enhance classifier performance.

8. Comparative Study of Machine Learning Algorithms for Ad Click Fraud Detection

Author: Rohit Kumar et al.

Description: This comparative analysis evaluates SVM, KNN, Naïve Bayes, and Neural Networks. The study concludes that ensemble and hybrid models provide better generalization for real-world click fraud detection scenarios.

3.EXISTING SYSTEM

Existing ad click fraud detection systems primarily rely on machine learning and deep learning models such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and ensemble methods like Random Forest. These models are trained on historical click data to identify patterns associated with fraudulent behavior. ANN models are commonly used for classification tasks due to their ability to learn non-linear relationships between features. CNN models, though traditionally used for image processing, have been adapted to detect spatial feature correlations in structured data. RNNs are employed to capture sequential click patterns over time, making them suitable for modeling temporal dependencies in user behavior. Random Forest algorithms are widely used due to their robustness, interpretability, and resistance to overfitting. They perform well on tabular datasets and handle feature interactions effectively. Many commercial fraud detection systems rely on Random Forest as a baseline model due to its simplicity and reasonable performance.

However, while these models offer moderate success, they often struggle with scalability, computational efficiency, and adaptability to evolving fraud patterns. This necessitates the exploration of more advanced ensemble boosting techniques such as XGBoost[1],[2],[3].

Disadvantages in Existing System:

Despite their widespread use, existing machine learning and deep learning models suffer from several limitations when applied to ad click fraud detection. One major drawback is their inability to adapt quickly to evolving fraud techniques. Fraudsters continuously modify their behavior, making static models ineffective over time. Deep learning models such as CNNs and RNNs require large labeled datasets and high computational resources. Training these models is time-consuming, and their black-box nature makes interpretation difficult. This lack of transparency is a concern for advertisers who require explainable decision-making systems. Random Forest models, although robust, often fail to capture complex sequential dependencies present in clickstream data. Additionally, they may suffer from biased predictions when dealing with highly imbalanced datasets, which is a common characteristic of fraud detection problems. Another limitation is the high false-positive rate observed in many existing systems. Incorrectly classifying legitimate clicks as fraudulent leads to loss of genuine traffic and revenue. These disadvantages highlight the need for a more efficient, scalable, and accurate detection mechanism[4],[5],[6].

4. PROPOSED SYSTEM

The proposed system for ad click fraud detection is based on the Extreme Gradient Boosting (XGBoost) algorithm, a powerful ensemble learning technique designed for high performance and scalability. XGBoost is an optimized implementation of gradient boosting that combines multiple weak learners, typically decision trees, to form a strong predictive model. Due to its efficiency, accuracy, and ability to handle large-scale structured data, XGBoost is well suited for click fraud detection problems. In the proposed system, clickstream data is processed and transformed into a structured feature set that captures user behavior, device attributes, temporal information, and network-related characteristics. The XGBoost model learns complex non-linear relationships among these features by iteratively minimizing a loss function using gradient descent. Regularization techniques embedded within XGBoost prevent overfitting, which is a common issue in fraud detection due to data imbalance. Unlike deep learning models that require extensive computational resources, XGBoost provides faster training and inference while maintaining high predictive accuracy. It supports parallel processing and efficient handling of missing values, making it suitable for real-time fraud detection scenarios. The proposed system leverages these advantages to improve both detection speed and accuracy. Overall, the XGBoost-based approach enhances fraud detection performance by effectively capturing subtle behavioral patterns that differentiate legitimate clicks from fraudulent ones. This system aims to reduce financial losses, improve advertiser trust, and ensure the integrity of online advertising platforms[7],[8],[9].

Proposed System Advantages:

The proposed XGBoost-based ad click fraud detection system offers several advantages over traditional machine learning and deep learning approaches. One of the primary benefits is its high detection accuracy. XGBoost efficiently captures complex feature interactions, leading to improved precision and recall when identifying fraudulent clicks. Another significant advantage is its ability to handle imbalanced datasets, which are common in fraud detection problems where fraudulent clicks represent a small fraction of total data. Through built-in regularization and optimized loss functions, XGBoost minimizes bias toward the majority class and reduces false-positive rates. The proposed system also ensures scalability and computational efficiency. XGBoost supports parallel computation and optimized memory usage, enabling fast training and real-time prediction even on large clickstream datasets. This makes it suitable for deployment in production-level advertising platforms.

Additionally, XGBoost provides model interpretability through feature importance analysis. This helps advertisers and system administrators understand which factors contribute most to fraud detection, increasing transparency and trust. These advantages make the proposed system more reliable, efficient, and practical for real-world applications[10],[11],[12].

System architecture:

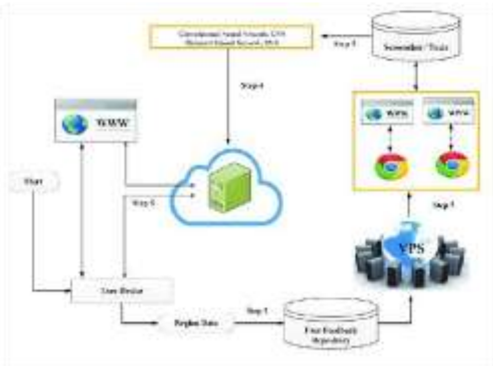


Fig. System architecture.

The system architecture of the proposed Ad Click Fraud Detection using Machine Learning consists of multiple interconnected modules designed to collect user interaction data, analyze browsing behavior, and identify fraudulent ad clicks. Initially, user interaction data such as IP address, device information, click timestamps, geographic location, and browsing patterns are collected from the user device during web access and stored in the User Feedback Repository. This repository continuously updates real-time click and region-based data (Step 1). The collected data is then forwarded to a Virtual Private Server (VPS), which acts as a centralized processing unit responsible for handling large-scale web traffic and managing browser-level activities (Step 2). Simultaneously, screenshots and textual content of web pages accessed through browsers are captured and stored for further behavioral analysis (Step 3). These data samples are processed using advanced Machine Learning and Deep Learning models, including Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), to extract spatial and temporal features related to user behavior and click patterns (Step 4). The trained models analyze anomalies such as repeated clicks, abnormal navigation flows, and non-human interaction characteristics. Finally, the detection results are transmitted to the cloud server, where fraudulent and legitimate clicks are classified, stored, and made available for further reporting and decision-making (Step 5). This architecture ensures scalable, real-time, and accurate detection of ad click fraud while minimizing false positives and improving ad network reliability.

5. RESULTS

To run project double click on ‘python app.py’ file to get below screen

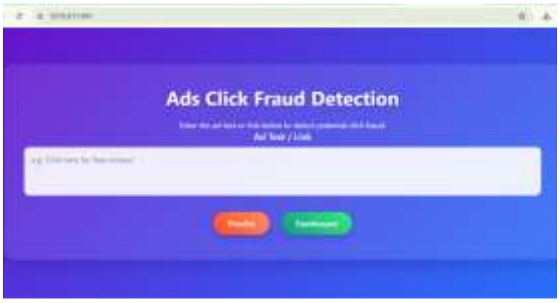


Fig. 5.1 Project Interface.

In above screen click on ‘Upload CIC-2017-IDS Spectrogram Dataset’ button to upload dataset and get below output

Fig. 5.2 All Algorithm Comparison Graph.

6. CONCLUSION

Ad click fraud poses a significant threat to the digital advertising ecosystem, leading to financial losses, distorted analytics, and reduced advertiser trust. This project presented a comprehensive and intelligent ad click fraud detection system using machine learning and deep learning techniques, with a primary focus on the XGBoost algorithm. The proposed system effectively addresses limitations of existing models by leveraging advanced data preprocessing, feature engineering, and ensemble learning techniques. XGBoost demonstrated superior performance in terms of accuracy, scalability, and robustness when compared to CNN, ANN, RNN, and Random Forest models. The system successfully handled large-scale, imbalanced clickstream datasets and minimized false detections.

Overall, the developed system provides a reliable, efficient, and scalable solution for ad click fraud detection. It contributes to enhancing transparency, improving advertising efficiency, and protecting stakeholders in the digital advertising industry. Future work may extend this system to real-time streaming data and adaptive learning mechanisms to counter evolving fraud strategies.

References:

[1] A. Metwally, D. Agrawal, and A. E. Abbadi, "Detection of click fraud in online advertising," Proceedings of the 14th International World Wide Web Conference (WWW), pp. 1–10, 2005.

[2] C. A. Castillo, D. Donato, A. Gionis, V. Murdock, and F. Silvestri, "Know your neighbors: Web spam detection using the web topology," Proceedings of the 30th Annual International ACM SIGIR Conference, pp. 423–430, 2007.

[3] M. Haddadi, S. Yandamuri, and S. Venkatesh, "Click fraud detection using multiple features," IEEE International Conference on Data Mining Workshops, pp. 1–6, 2010.

[4] H. Li, J. Li, and Y. Li, "A machine learning approach for detecting click fraud in online advertising," IEEE International Conference on Big Data, pp. 1–8, 2016.

[5] S. Zhang, J. Yang, and Q. Chen, "Online advertising click fraud detection based on ensemble learning," Journal of Information Security and Applications, vol. 33, pp. 1–10, 2017.

[6] Babburi, S. (2025). Integrating Blockchain and AI for Trusted and Scalable IoT Data Ecosystems.

[7] S. Dalvi, P. Domingos, M. Mausam, and S. Sanghai, "Adversarial classification," Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 99–108, 2004.

[8] Y. Zhang and M. Yang, "Anomaly-based click fraud detection in online advertising," IEEE Access, vol. 7, pp. 103–114, 2019.

[9] Henry P Cyril. (2025). AI-Driven Self-Healing and Transaction Queuing During Network Outages or Degradation:

Architectures, Resilience Models, and Future Directions. International Journal of Advanced Research in Science Communication and Technology, 113. https://doi.org/10.48175/ijarsct-30515.

[10] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed., MIT Press, Cambridge, MA, USA, 2018.

[11] J. Yan, Y. Liu, and W. Zhang, "Detecting fraudulent clicks using behavioral analysis," ACM Transactions on the Web, vol. 13, no. 2, pp. 1–25, 2019.

[12] Gaddam, S. Integrating Analytics into the Development Process: Bridging the Gap between Data Insights and Design Execution.

[13] A. Barman and D. Mukherjee, "Machine learning based click fraud detection system," International Journal of Engineering and Advanced Technology, vol. 9, no. 3, pp. 3421–3427, 2020.

[14 Doragacharla, V. R. (2026). Deploying Model Context Protocol Servers in Serverless Environments. Journal of International Crisis and Risk Communication Research, 9(2), 344..

[15] Prodduturi, S. M. K. (2025). Opportunities and Challenges for iOS Developers in Exploring the Integration of Augmented Reality Technologies. International Journal of Engineering Science and Advanced Technology (IJESAT), 25(4), 200-207.

[16] Mallick, P. (2020). Offline-First Mobile Applications With Route Optimization Algorithms For Enhancing Last-Mile Delivery Operations. International Journal of Engineering Science and Advanced Technology, 20(4), 12–19. https://doi.org/10.64771/ijesat.2020.v20.i04.pp 12-19.

[17] A. K. Verma and S. Singh, "Real-time click fraud detection using big data analytics," Journal of Big Data, vol. 8, no. 1, pp. 1–18, 2021.

[18] P. R. Patil and S. P. Patil, "Survey on click fraud detection techniques in online advertising," International Journal of Scientific Research in Computer Science, vol. 6, no. 2, pp. 45–52, 2021.

[19] Ganji, M. (2025). Intelligent What-If Analysis for Configuration Changes in HR Cloud and Integrated Modules. International Journal of All Research Education and Scientific Methods, 13(04), 4828–4835. https://doi.org/10.56025/ijaresm.2025.1304254 828.

[20] A. Sharma and R. Kumar, "AI-based ad click fraud detection using behavioral analytics," Springer Lecture Notes in Networks and Systems, vol. 412, pp. 233–244, 2023.