A Review on Click Fraud Detection in Online Advertising Using Machine Learning Algorithms

Dr.Ganesh Gorakhnath Taware[1], Ms.Vaishali Balasaheb Pawar[2]
[1]*Department of Computer Engineering Dattakala Group of Institution , Faculty of Engineering Swami-Chincholi Bhigwan*
[2]*Department of Computer Engineering Dattakala Group of Institution , Faculty of Engineering Swami-Chincholi Bhigwan*

***Abstract*—Click fraud is one of the significant problems that keeps escalating in the digital advertising ecosystem. As a result, it causes a substantial loss of both revenue and trust from the advertisers' side. When someone performs click fraud, they make fake clicks on online advertisements to either artificially inflate the metric or exhaust a competitor's budget. Conventional rule-based methods of detection are not capable of keeping up with the complexity and the scale of today's advertisement data. Machine learning (ML) and deep learning (DL) algorithms have recently been considered promising tools for detecting click fraud, as they can learn to recognize behavioural patterns and distinguish between valid and fraudulent traffic. This review paper assesses machine-learning-based methods, which primarily include decision trees (DT), random forests (RF), as well as other ensemble methods, such as gradient-boosted decision trees (GBDT), XGBoost, and LightGBM. The paper summarizes the ML model architectures, their feature engineering methods, datasets, and the significant performance results extracted from the literature available in this field. Various experiments have demonstrated that tree-based ensemble models are more efficient than traditional classifiers in machine learning scenarios, as they can address the problems of data imbalances, temporal dependencies, and non-linear relationships that exist in clickstream data. Today's hybrid architectures, which utilize a combination of CNN, BiLSTM, and RF, achieve an extremely high level of accuracy (up to 99%) and are thus very suitable for practical applications. However, there are still issues of feature generalization, interpretability, adversarial robustness, and real-time scalability. In this paper, we identify the gaps in existing research and propose future research topics that consider Explainable AI (XAI), online learning, and privacy-preserving analytics as means to enhance the transparency and trustworthiness of advertising fraud detection systems. The present paper serves as a stepping stone towards future developments in intelligence, adaptation, and interpretability in machine learning models for identifying online advertising fraud, which in turn would provide robust protection mechanisms for the digital advertising ecosystem.**

***Index Terms*—Click Fraud Detection; Machine Learning; Decision Tree; Random Forest; Gradient Boosting; XGBoost; LightGBM; Ensemble Learning; Deep Learning; Online Advertising; Explainable AI; Fraud Analytics.**

I. INTRODUCTION

Online advertising has become the main driver of the global digital economy and is the primary source of income for a vast network of content producers, publishers, advertisers, and consumers. Worldwide spending on digital advertising exceeded $600 billion in 2024, underscoring the vital role digital ads play in the ecosystem that enables free online experiences and targeted advertising. Nevertheless, the rapid expansion of digital advertisement networks and automated bidding processes has led to the rise of illicit behaviours, such as click fraud, which, among other things, has become one of the most long-lived and harmful types of cyber deception.

Click fraud is a technique through which the reliability of advertising analytics is challenged by the emergence of fake or non-human clicks on pay-per-click (PPC) ads. Such clicks could be generated by bots, scripts, click farms, or fake affiliate advertisers, thereby fabricating engagement metrics, disbursing money without any productive returns, and consequently, causing advertisers to lose revenue. The type of click fraud thus depends on who the perpetrators are and the methods they use. Publisher fraud occurs when the

publisher, typically the owner of a website, generates fake clicks on ads on their site to inflate the appearance of revenue from the advertised impressions. Competitor fraud is a situation where an advertiser intentionally clicks on a competitor's ad to utilize the competitor's daily budget or to impact the competitor's performance negatively. In some sophisticated cases, attackers might assign botnets or automated click scripts that impersonate humans as they surf the web from the exact location but at different times, thereby avoiding detection.

The economic consequences of click fraud are frightening to the point that they keep the sleepless nights awake. The estimates made by the industries reveal that advertisers who pay for ads are the ones losing billions of dollars due to fraudulent clicks each year. There is also a report stating that non-human sources might be the reason for 20 percent of the total online advertisement traffic. However, these are merely the initial few dollars. The risk of miskicking has led to reduced advertiser trust, compromised campaign performance, and a degraded user experience, as well as inefficiencies in ad targeting and a gradual decline in market confidence.

Consequently, the deployment of exact and effective methods for identifying click fraud has been the top priority of research and industry. However, traditional rule-based and threshold-dependent detection systems have not been able to offer solutions to the highly dynamic, large-scale, and constantly changing fraudulent activities.

Initially, fraud detection systems leaned towards rule-based techniques, heuristic filters, and manually-set thresholds (click interval, IP repetition restrictions, device fingerprint). Although these techniques tend to be low-tech and computationally straightforward, they remain static in nature and reactive, identifying only existing patterns. Fraudsters continue to adapt their operations, whether by changing click patterns, rotating IP addresses, or deploying distributed botnets to circumvent these fixed-rule systems. Additionally, traditional methods may not be generalized across modern, large-scale, and heterogeneous datasets that advertising networks encounter, and are often not capable of managing imbalanced representations of data where the occurred fraud does not make up a significant subset of legitimate user interaction patterns. As a result, there tends to be a high false-positive rate (legitimate user activity flagged as fraud). In contrast, subtle or novel fraud patterns are often submit-rich and unnoticed, ultimately rendering the detection techniques effective in traditional advertising networks ineffective.

The increase in the volume, velocity, and variety of data from advertising interactions such as click timestamps, device metadata, geographic locations, or user session characteristics require data-driven, adaptive, and intelligent detection agents to address this need; for this reason, you notice Machine Learning (ML) and Deep Learning (DL) algorithms now widely used (adopted) in recent years.

Machine Learning provides a highly effective way to bypass the downsides of heuristic-based methods by allowing systems to learn patterns directly from the data. ML algorithms are capable of modelling intricate relationships between clickstream features, identifying subtle deviations from normal behaviour, and can take on new fraud patterns without explicit reprogramming. In particular, supervised learning methods are effective in click fraud detection, where labelled datasets (containing legitimate clicks and fraud clicks) are used to train classifiers that predict the likelihood of fraud. Decision Trees (DT), Random Forests (RF), and gradient-boosted decision trees (GBDT) have been shown to perform well in terms of interpretability, efficiency, and robustness to nonlinear relationships.

Decision tree (DT) algorithms create layered models by recursively splitting datasets using the values of features. These models are easy to interpret because we express the model's logic in terms of rules. However, in general, stand-alone DT models tend to over fit, especially with high-dimensional or noisy datasets. To combat this over fitting issue, ensemble learning strategies were developed. In particular, ensemble learning strategies can be described as combining multiple weak learners to create a more effective predictive model, and one of these strategies is random forests and boosting.

Random Forest (RF) is an ensemble algorithm based on bagging that constructs a multitude of decision trees using random subsets of features and data samples. The cumulative prediction was obtained through a majority vote across all trees, thereby enhancing generalizability and reducing variance. In relation to click fraud, an RF model has performed remarkably due to its ability to process heterogeneous features such as IP addresses, device types, and temporal intervals between the clicks. Additionally, RF models can indicate the importance

of each feature by showing how much each feature contributed to the detection assessment. We do, however, have more advanced ensemble modeling with gradient boosted decision trees (GBDT) and their optimizations, which are XGBoost, LightGBM, and CatBoost. Whereas bagging trees are grown in parallel, boosting trees are grown sequentially; thus, the next tree can correct the errors of the previous one. A model can therefore distinguish more complex patterns, such as nonlinearities, and very slight differences between genuine and fraudulent clicks. GBDT-based models are particularly effective for imbalanced datasets, where the number of fraudulent samples is very low and these samples are crucial to the dataset. The results of the experiments indicate that the use of the boosting methods improves the performance of fraud detection, which leads to the methods being preferred over traditional ML algorithms. In that case, the accuracy, recall, and F1-scores are higher with less computational cost.

Traditional ML-based methods have ensemble learning models as their mainstay. However, Deep Learning (DL) methods are increasingly being employed to decipher complex sequential and contextual dependencies in user click data. In particular, architectures such as Convolutional Neural Networks (CNNs) or Bidirectional Long Short-Term Memory (BiLSTM) are equipped to derive spatial and temporal relations directly from the ground-level click data, thereby eliminating the major pre-processing work.

On the other hand, Deep Learning models usually necessitate a large volume of labelled data and are inherently less interpretable than tree-based models. To mitigate these two issues, the research community has begun to employ hybrid architectures that leverage the advantages of both ML and DL. In fact, CNN or BiLSTM models are used for feature extraction, whereas tree-based classifiers are then utilized for the final classification stage (e.g., RF, LightGBM). These hybrid models have been demonstrated to achieve very high accuracies (e.g., up to 99%) and exhibit good performance across different datasets and varying sizes, compared to traditional models.

Another recent trend is the use of Generative Adversarial Networks (GANs) and auto encoders, which can enable semi-supervised learning for the identification of click fraud. Such models can infer from unlabelled data by determining the distributions of expected behaviours and recognizing deviations as possible fraud. Introducing deep-hybrid architectures represents a significant step toward more autonomous, adaptive, and explainable fraud detection systems.

First and foremost, the quality and diversity of input features have a significant impact on the performance of any machine learning or deep learning model. The identification of click fraud is usually achieved through a mixture of behavioural, contextual, and network-level features that may include

- Temporal features include click timestamps, session duration, click intervals, and burst patterns.
- Spatial features include IP addresses, geolocation, and country or region of origin.
- Device and browser attributes: operating system, device type, browser version, and user agent strings.
- Ad and campaign metadata: ad ID, publisher ID, click-through rates (CTR), conversion ratios, and impression history.
- Network-level indicators include packet transmission frequency, time-to-live (TTL) variance, and proxy usage patterns.

Modeling progress in literature has been measured against standards such as FDMA2012, Google Ads logs, and synthetic ad network datasets. Typically, the primary challenge in this field is the scarcity of large-scale, publicly available datasets that are well-labeled and annotated. The area has seen much algorithmic sophistication for the most part, but there are still many issues with the deployment of ML click-fraud detection systems in real-world situations.

1. Data Imbalance: Fraudulent clicks constitute a small fraction of the total advertisement traffic, causing classifiers to be biased toward legitimate clicks. Oversampling, undersampling, and synthetic data generation techniques (e.g., SMOTE) are often required to address these issues.
2. Evolving Fraud Strategies: Attackers continuously modify their techniques to evade detection, necessitating adaptive and online learning mechanisms.
3. Scalability: Real-time advertisement bidding systems process millions of clicks per second, requiring models with low latency and high throughput.

4. Interpretability: Complex ensemble and deep learning models act as "black boxes," making it difficult for analysts to explain the decisions or ensure fairness.
5. Privacy and Security: Integrating user-level behavioral data raises ethical and regulatory challenges, emphasizing the need for privacy-preserving analytics.

Dealingwith such problems demand a multidimensional strategy that involves algorithmic innovation, explainability, and real-time system design.

Considering the constraints of current warning systems and the increasingly complex fraudulent actors, this paper aims to provide a comprehensive overview of machine learning-based methods for detecting click fraud. The primary goals of this research were:

- To review and categorize Decision Tree, Random Forest, Gradient Boosting, and hybrid ML/DL models used in click fraud detection.
- The strengths, limitations, and comparative performance of various algorithms were analyzed.
- To identify research gaps in model generalization, interpretability, and adaptability.
- To propose future research directions that emphasize Explainable AI (XAI), adversarial robustness, and scalable online learning.

Through this review, we aim to provide a consolidated understanding of how ML algorithms have evolved to detect click fraud efficiently, robustly, and ethically.

Click fraud is a complicated issue that is constantly evolving and involves aspects of cybersecurity, data mining, and digital economics. The use of machine learning and deep learning has changed the fraud detection arena because they can create automated & scalable systems. As a result, fraud detection systems are now capable of adjusting to different scenarios of attack. Fraud detection systems based on machine learning algorithms incorporate analytical methods like Decision Trees, Random Forests, and Gradient Boosting as the main operations in the models. Hybrid deep models are becoming better both in terms of accuracy and adaptability. Providing explainable, privacy-preserving, and real-time detection is still mainly a research problem, in spite of these developments. This review paper serves as an intermediary between the present methods and the future directions, mainly for the establishment of a transparent and resilient digital advertising ecosystem.

II. LITERATURE SURVEY

Due to their interpretability for humans, efficiency in implementation, and capability to model complex decision boundaries, Decision Trees (DTs) and Random Forests (RFs) have been widely used in detecting click fraud. In MadTracer, Li et al. [1] introduced a browser-based detection system that surveyed ad infrastructure and behavior features from multiple ad networks. With DT-based detection rules, MadTracer successfully identified types of attacks, including drive-by downloads, scams, and unidentified click fraud variants, by leveraging knowledge of malicious ad paths.

Berrar [2] employed Random Forests (RFs) relying on skewed bootstrap sampling to classify publishers as either legitimate or fraudulent based on click profiles with IP-based temporal features. While feature engineering strengthened model interpretability, the overall accuracy was limited (49.99% validation, 42.01% testing), which limited the generalizability of the results.

According to Yen and Jiang [3], they employed multiple classifiers to model advertising logs employing MapReduce processing (e.g., RF, Bayesian networks, Naive Bayes). They consistently concluded that tree-based models outperform probabilistic models in terms of click distributions with imbalanced data. Perera et al. [4] created a new ensemble framework that extracted time-dependent statistical features (mean, variance, skewness) from raw click data. They found that among the six ensemble learners they used, bagging and boosting variants using J48 and REPTree achieved the best accuracy (59.39%). Oentaryo and Lim [5] extracted temporal and spatial features (e.g., click ratios and country-based distributions) and trained Logistic Regression (LR) and Extremely Randomized Trees, concluding that these features were essential for handling unbalanced datasets.

Perera [6] also highlighted the advantages of ensemble and sampling approaches. The author presented the detection rates of using SVM, RF, MLP, and DT models and verified that using a combined bagging model of C4.5 and cluster-based sampling improved detection rates, where the temporal and spatial click features were found to be significant predictors of

fraud. Oentaryo et al. [7] studied the top-performing models in the FDMA 2012 click fraud detection challenge. They found that the highest accuracy (52.3%) came from an ensemble-of-ensembles method (rotation forest with RF).

Xu et al. [8] present a behavioral verification model that differentiates bots from humans using JavaScript support and mouse movement tracking. Their real-time system, based on C4.5 and continuously updated over a ten-day period, achieved a 99.1% accuracy. He et al. [9] proposed a hybrid DT–LR framework that combines device type and CTR history as contextual and historical features, and suggested that continuous (daily) retraining improves accuracy. Ravi [10] studied the C4.5 model used to detect touch fraud in mobile gaming apps by using app-level metadata and ad constraint features to improve robustness when a constraint on ad visibility was enforced.

Beránek et al. [11] introduced the timeprint method by creating temporal feature sets (time of day, type of day) to represent user behavior. Timeprint-based preprocessing led to higher detection precision for various classifiers (NB, DT, and SVM). Berrar [12] also examined the FDMA2012 dataset and, using RF, pinpointed a recall of 36.2%, thereby emphasizing click-time features as the most revealing sources of fraud. Guo et al. [13] presented a traffic sampling approach based on IP addresses for CloudBot detection, while utilizing transport-layer features (e.g., TTL, packet variance) for privacy-preserving fraud detection. Lia and Jia [14] improved RF performance by adopting a hybrid sampling (over- and undersampling) strategy, thereby achieving an accuracy of 93% for non-fraudulent clicks and 91% for fraudulent clicks, respectively. This approach demonstrated that feature-rich ensemble tree models are an effective tool in combating class imbalance.

Wang et al. [15] invented a dual-layer hybrid system that applies a rule-based method at both user and traffic levels. By coupling the Gradient Boosting Decision Tree (GBDT) classifier with time-windowed hybrid features comprising IP and cookie recurrence, the detection capability was significantly enhanced. Jianyu et al. [16] addressed the problem of nominal feature modeling by introducing a novel encoding regime that preserves categorical feature information for large-scale advertising datasets. Their GBDT and XGBoost-powered models pinpointed fraud by focusing on feature frequency. Minastireanu and Mesnita [17] employed the LightGBM method to analyze the behavior of non-converting users from ad interactions, achieving 98% accuracy. Aside from that, it was more efficient in terms of time, computation, and memory than XGBoost and stochastic GB.

In a North Dakota study, Singh and Sisodia [18] demonstrated that Gradient Tree Boosting (GTB) exhibited excellent robustness on various datasets, specifying that it was capable of accommodating high-cardinality, imbalanced, and massive click data. Dash and Pal [19] developed adaptive and scalable feature sets using GBDT, achieving a reported accuracy of 97.2%. However, they did not provide details about user/behavioral features with geographical and temporal granularity.

Mouawi et al. [20] tested several classifiers, including ML and DL, for classification to detect fraudulent publishers with high click fraud in mobile advertising. They implemented methods such as SVM, KNN, and ANN models, incorporating click details and user information from the advertising network and advertisers, identifying click behavior from users with malicious intent. They generated synthetic ad traffic with 500,000 requests and 1,000 publishers each, then extracted features including the percentage of suspicious clicks, click duration, total number of clicks, the number of distinct IPs, obtained app downloads, and the distribution of click frequency. The K-nearest neighbors method achieved the highest predictor accuracy, 98 percent.

Likewise, other studies [21] also utilized FDMA 2012, another free and open-source dataset, for mobile advertising fraud detection using SVM, RF, Naïve Bayes, and Decision Tree (DT) algorithms. Oversampling of positive instances and undersampling of negative cases were implemented, yielding significant results that achieved 91% accuracy on the RF algorithm for both balanced and severely imbalanced datasets.

Espírito Santo [22] also proposed a machine-learning-based approach to detect click fraud in Google Ads, utilizing five models: support vector machines, random forest, K-nearest neighbors, gradient tree boosting (GTB), and XGBoost, in accordance with the CRISP-DM methodology. Their results revealed that tree-based models outperformed the others, most notably GTB and XGBoost. They also identified indicators of fraud, such as click frequency per IP and user ID, that can provide practical meaning for marketing agencies.

A strength is their relationship with an industry partner, which makes the research relevant to and grounded in solutions to real-world applications, aiming to combat click fraud in digital advertising.

Mahesh et al. [23] planned to build several machine learning models that would be capable of separating real users from bots. In this way, the researchers aimed to counter a practice known as click fraud. This fraudulent action intentionally increases the number of ad clicks, resulting in advertisers losing money and reputation. The paper authors run AI techniques and perform a performance comparison between different models. The results of various experiments suggest that machine learning methods are a powerful tool for addressing security issues in the online advertising domain.

To provide advertisers with tools to counteract fraudulent clicks, Thejas et al. [24] designed a supervised learning model, "CFXGB," which represents an integration of Cascaded Forest with XGBoost. Their method employed feature transformation in conjunction with classification algorithms, demonstrating superior performance over existing techniques on datasets of varying sizes.

Alzahrani et al. [25] involved highly advanced feature engineering techniques in the development of a strong click-fraud detection system. They made a comparative study of the performance of nine ML and DL models. Tree-based algorithms (Decision Trees, Random Forests, Gradient Boosting, LightGBM, and XGBoost) achieved an accuracy of over 98.9% after Recursive Feature Elimination, and the deep learning RNN model also demonstrated its effectiveness. The authors attest to the effectiveness of traditional and DL methods in the detection of fraudulent clicks at a very high level of confidence. At the same time, they foresee a potential lead for the dissemination of anti-fraud practices in the digital advertising sector.

Aljabri and Mohammad [26] contributed by suggesting a machine learning method that enables the identification of click fraud by distinguishing between human users and bots. In their work, they evaluated the performance of numerous machine learning (ML) models on real browsing data from users, which included descriptive features such as session time, page views, and user activities. The authors found that the Random Forest algorithm was the most efficient, yielding the highest and best-performing results among all metrics, underscoring the algorithm's significant capability in detecting fraudulent activities under the pay-per-click model.

Batool and Byun [27] introduced a novel hybrid ensemble model that combines CNN, BiLSTM, and Random Forest to artificially commit click fraud in online advertising. The deep learning algorithm is capable of automatically discovering patterns based on the various latent features of click data, whereas the RF model is employed for the classification task. The proposed model comprises numerous components, including a module for preprocessing categorical variables and addressing the data imbalance problem. The research results demonstrated an impressive performance; the proposed model achieved an accuracy, precision, and F1-measure of over 99%. The proposed model outperformed the standalone model and other ensemble models.

Batool et al. [28] developed an ensemble model that integrates CNN, BiLSTM, and RF models to enhance the detection of click fraud. The deep learning units can automatically extract the spatial and temporal features from the data, which the RF model then classifies. The combined model demonstrated significant enhancements in performance, reducing the total manual work required for feature engineering while simultaneously improving the classification performance of traditional ML models. Thus, the model achieved an accuracy of 99.19% along with high precision, F1-Measure, and recall. Minastireanu and Mesnita [29] proposed a method based on LightGBM-based fraud detection to manage the increasing risk of click fraud in online advertising. This study used a dataset of 200 million clicks from four days to identify suspicious IP addresses. This included high click volumes without generating app installations. The LightGBM algorithm, a gradient-boosting decision tree model, correctly identified fraud 98% of the time. The study highlighted the contribution of machine learning to better traffic filtering and illustrated the real-life application of sophisticated algorithms in modern advertising.

Thejas et al. [30] proposed a deep learning approach to mitigate the increase in click fraud in mobile in-app advertising. Their hybrid model consisted of a combination of Artificial Neural Networks (ANNs), autoencoders, and a semi-supervised Generative Adversarial Network (GAN) to detect fraudulent clicks in an adversarial environment where an attacker intentionally attempts to mislead the fraud detection

system. Their study addressed the shortcomings present in the existing literature. It proposed a hybrid deep learning approach that demonstrated increased accuracy compared to other models in dealing with innovative and evolving patterns of attack, surpassing the current state-of-the-art techniques.

The increasing complexity of fraudulent behavior in the online advertisement ecosystem has led to the need for research into the use of machine learning (ML) algorithms to detect click fraud. In Table 1, we provide a summary of essential studies that utilized different ML and ensemble-based models, such as Decision Trees (DT), Random Forests (RF), gradient-boosted decision trees (GBDT), XGBoost, LightGBM, and hybrid deep learning models, to detect fraudulent clicks in digital ad networks. Each study is summarized in terms of the algorithm utilized, dataset contents, characteristic features, and essential outcomes. The earliest studies prioritized decision tree-based models due to their interpretability and transparency, while more recent studies have implemented ensemble and boosting models to improve accuracy and scalability. Even more recently, researchers have incorporated deep neural architectures, including Convolutional Neural Networks (CNNs) and Bidirectional Long Short-Term Memory (BiLSTMs) networks, with traditional tree-based models, achieving near-perfect classification in both desktop and mobile advertisements. The literature indicates a gradual shift from simple classification models to more evolved hybrid models that utilize feature engineering to manage high-dimensional, imbalanced, and temporally dependent datasets. However, existing research to date remains limited in terms of real-time adaptability, explainability, and generalizability across multiple platforms.

Table Type Styles

Sr. No.	Author(s) & Year	Algorithm / Model Used	Features / Dataset	Major Findings	Limitations / Remarks
[1]	Li et al. (2011)	Decision Tree (DT) in MadTracer System	Browser-based ad infrastructure and behavioral features	Detected multiple attack types (drive-by, scam, and unknown fraud); DT rules improved interpretability	Limited scalability; static detection rules
[2]	Berrar (2012)	Random Forest (RF) with skewed bootstrap	Click profiles, IP-based temporal features	Classified publishers as legitimate/fraudulent; moderate accuracy (49.99%)	Low generalization; imbalance sensitivity
[3]	Yan & Jiang (2013)	RF, Bayesian Network, Naïve Bayes	Advertising logs processed via MapReduce	Tree-based models outperformed probabilistic methods	Limited dataset diversity
[4]	Perera et al. (2014)	J48, REPTree, Ensemble Models	Time-dependent statistical features (mean, variance, skewness)	Bagging and boosting improved detection accuracy (59.39%)	Moderate precision; feature bias
[5]	Oentaryo & Lim (2014)	LR, Extremely Randomized Trees	Temporal & spatial click ratios, country-based features	Temporal features crucial for unbalanced datasets	Requires high-quality data preprocessing
[6]	Perera (2015)	DT, RF, SVM, MLP with Cluster Sampling	Spatial & temporal click features	Bagging with C4.5 improved detection accuracy	Limited validation on real-world data
[7]	Oentaryo et al. (2015)	Rotation Forest + RF (Ensemble-of-Ensembles)	FDMA2012 Challenge Dataset	Achieved 52.3% accuracy, outperforming single models	Relatively low precision
[8]	Xu et al. (2015)	C4.5 Decision Tree	JavaScript behavior, mouse movement	Real-time system achieved 99.1% accuracy	Dataset limited to 10-day campaign
[9]	He et al. (2016)	Hybrid DT–LR Model	Contextual & historical features (CTR history, device type)	Frequent retraining improved accuracy	Computationally intensive

[10]	Ravi (2016)	C4.5 Decision Tree	Mobile gaming ad metadata & visibility constraints	Improved classifier robustness with visibility constraints	Limited to the mobile app context
[11]	Beránek et al. (2016)	Timeprint-based DT, SVM, NB	Temporal user behavior (time of day, day type)	Enhanced detection precision using time-based preprocessing	Sensitive to missing timestamps
[12]	Berrar (2017)	Random Forest	FDMA2012 Dataset (Click-time features)	36.2% precision; temporal features informative	Moderate detection accuracy
[13]	Guo et al. (2017)	RF with IP-based Traffic Sampling	Transport layer (TTL, packet variance)	Achieved privacy-preserving fraud detection	Needs real ad network validation
[14]	Lia & Jia (2018)	RF with Hybrid Sampling	Balanced dataset via over-/undersampling	93% accuracy for legitimate and 91% for fraud clicks	High cost in data preprocessing
[15]	Wang et al. (2019)	Hybrid GBDT + Rule-based System	Time-windowed hybrid features (IP, cookie recurrence)	The dual-layer system improved detection performance	High training complexity
[16]	Jianyu et al. (2019)	GBDT & XGBoost	Encoded categorical features in large datasets	Effectively identified fraudulent activities	Encoding overhead for large data
[17]	Minastireanu & Mesnita (2019)	LightGBM	Non-conversion user behavior data	98% accuracy; high efficiency, low memory use	May overfit on smaller datasets
[18]	Singh & Sisodia (2020)	Gradient Tree Boosting (GTB)	Multiple benchmark datasets	Robust with high-cardinality, imbalanced data	Limited feature diversity
[19]	Dash & Pal (2020)	GBDT	Adaptive and scalable feature sets	97.2% accuracy achieved	Lacked temporal & geographical features
[20]	Mouawi et al. (2020)	SVM, KNN, ANN	Synthetic ad traffic (500K requests, 1K publishers)	KNN achieved 98% accuracy in detecting fraudulent publishers	Synthetic data; lacks real-world noise
[21]	Anonymous (2020)	RF, SVM, NB, DT	FDMA2012 Dataset with Sampling	RF achieved 91% accuracy on balanced data	Dataset imbalance challenge
[22]	Do Espírito Santo (2021)	SVM, RF, KNN, GTB, XGBoost	Google Ads clickstream data	GTB & XGBoost outperformed others; identified key fraud indicators	Lacks a deep learning comparison
[23]	Mahesh et al. (2021)	Comparative ML Models (SVM, RF, ANN)	User behavior (session duration, actions)	ML improved cybersecurity in ad traffic	Focused on bot detection only
[24]	Thejas et al. (2021)	Cascaded Forest + XGBoost (CFXGB)	Clickstream from varied datasets	Outperformed existing ML models; scalable and effective	Lacks interpretability
[25]	Alzahrani et al. (2022)	DT, RF, GBDT, LightGBM, XGBoost, RNN	Feature-engineered dataset after RFE	Ensemble models achieved >98.9% accuracy	High computational demand
[26]	Aljabri & Mohammad (2022)	RF, SVM, KNN	Real-world browsing session data	RF achieved the highest accuracy across all metrics	Limited generalization to mobile apps

[27]	Batool & Byun (2022)	CNN + BiLSTM + RF Hybrid	Clickstream (spatial-temporal data)	Achieved >99% accuracy, precision, and F1-score	Complex architecture; high resource use
[28]	Batool et al. (2023)	CNN + BiLSTM + RF Ensemble	Temporal-spatial ad click data	99.19% accuracy; reduced manual feature engineering	Needs real-time validation
[29]	Minastireanu & Mesnita (2023)	LightGBM (GBDT variant)	200M ad clicks dataset	98% accuracy; practical for industrial deployment	Focused only on IP-level fraud
[30]	Thejas et al. (2024)	ANN + Autoencoder + Semi-supervised GAN	Mobile in-app ad data	Robust to adversarial attacks; superior to existing models	High training cost and data labeling requirements

III. RESEARCH GAP

While there has been a notable advancement in the use of machine learning (ML) and deep learning (DL) methods for the detection of click fraud, there still exist a few gaps in the achievement of a fraud mitigation system that is robust, explainable, and real-time. The majority of research works rely on features that are manually designed and static, such as repetition of IP addresses, click frequency, and session duration, which yield good results on specific datasets but do not generalize well to different ad networks. Furthermore, the issues of data imbalance and label scarcity have been significant challenges for detecting fraudulent activities, as fraudulent clicks account for only a small fraction of the total traffic. In addition, ensemble methods such as Random Forest (RF), gradient-boosted decision trees (GBDT), and Light Gradient Boosted Model (LightGBM) have demonstrated remarkable accuracy in laboratory-like experiments; however, their application in large-scale, streaming, and adversarial situations remains a topic of debate. Furthermore, there is a scarcity of research that has focused on the interpretability of the model, which is why advertisers and analysts often lack transparency in understanding the rationale behind a click being labeled as fraudulent.

Another significant research gap is mainly about how existing models can be scaled up and made more flexible. The majority of structures are trained on fixed datasets and, therefore, are unable to recognize new, gradually evolving adversarial click fraud techniques. Due to the high computational requirements and latency restrictions, real-time detection in dynamic ad exchanges is a relatively new area. Additionally, very little has been done to integrate Explainable AI (XAI), privacy-preserving learning, and federated architectures in line with the most recent data protection regulations. The absence of large-scale, publicly available, and standardized datasets also hampers the reproducibility and unprejudiced benchmarking of these models. Consequently, future research should be directed towards creating advanced, interpretable, and privacy-compliant hybrid models capable of efficiently processing large volumes of clickstream data while maintaining transparency and being resistant to the ongoing evolution of fraudulent activities.

IV.DISCUSSION

Over time, a review of the literature reveals that the ways of detecting click fraud have substantially changed. More specifically, detection methods have gradually moved from conventional rule-based systems to intelligent models that employ machine learning (ML) and deep learning (DL) techniques. The earliest studies, which primarily involved Decision Trees (DT) and simple classification models that were easily understandable and provided transparent decision boundaries, paved the way for this development. However, with the rise in the intricacy and volume of the clickstream data, these techniques are less effective in nonlinear relationships and imbalanced datasets. A wide range of new solutions based on ensemble methods, such as Random Forests (RF), gradient-boosted decision trees (GBDT), and XGBoost, has been described as a milestone in breaking through the limitations of weak learners by employing several models to both increase the power of the model and lower the risk of overfitting. The identification of very low and previously unnoticeable fraudulent behaviors has been carried out with high

accuracy in different advertising settings, thanks to these ensemble methods.

The literature survey reveals a clear trend: classification models have evolved from static to adaptive and hybrid structures, enabling them to comprehend changing behavioral patterns over time. XGBoost and LightGBM have consistently outperformed other algorithms. They are superior in efficiency, scalability, and their ability to capture complex feature interactions. Moreover, the current hybrid models, which integrate deep neural networks (e.g., CNNs and BiLSTMs) with tree-based classifiers (e.g., RF and LightGBM), boast accuracy rates close to 100% and are often above 98–99%. These structures are proficient in the automatic extraction of features; thus, they are less dependent on the manual feature engineering process, while also improving the temporal and contextual aspects of understanding click behavior. However, such accuracy is generally achieved at the expense of interpretability and computational efficiency, thereby posing difficulties in the geographical deployment of the system to real-time scenarios of large-scale ad vertising.

Feature engineering, being at the core of the model's performance, is another central point upon which the literature converges. Researchers have repeatedly found that temporal, spatial, and behavioral features, such as click intervals, IP frequency, and device consistency, are the most significant indicators of fraudulent activity. However, there is still no agreement on standard feature sets or benchmarking datasets, which makes it difficult to compare different studies and reproduce their results. Additionally, the further development of fraud tactics, such as the use of distributed botnets and the imitation of adversarial clicks, requires that models be capable of adjusting to new attack methods. While ensemble and hybrid models have acknowledged the problem and taken some steps towards the solution, the majority of them are still based on static datasets and do not provide for online or incremental learning.

The difficulty in understanding the models' decisions also figures among the main challenges raised by this research. Although ensemble and deep learning models exhibit good performance in terms of detection accuracy, it is often difficult to understand their reasoning. The opacity of these "black box" models hampers their implementation in the commercial systems that require transparency, accountability, and adherence to regulations. Only a handful of works have sought to harness the potential of Explainable AI (XAI) methods, such as SHAP or LIME, to make model predictions and the features taken into account more understandable. Future frameworks must strike a balance between prediction precision and explanation capability, thereby gaining the trust not only of advertisers but also of stakeholders.

Moreover, the problems related to scalability and real-time performance have not yet been fully solved. A good number of machine learning models are effective in a controlled, offline environment, but are far from being optimized for the high-throughput and low-latency requirements of real-world advertising exchanges. Solutions like LightGBM and distributed XGBoost partially resolve these issues by enabling parallel computation and efficient memory utilization, respectively. Nevertheless, enabling real-time fraud detection with high accuracy calls for the use of streaming processing frameworks, cloud computing scalability, and incremental learning.

On a higher level, the incorporation of privacy-preserving techniques and federated learning structures is the next big thing in research. Due to the imposition of data privacy regulations such as GDPR and CCPA, future systems for detecting click fraud should ensure that user-level behavioral data is handled securely and anonymously. Federated learning is a viable solution because it allows collaborative model training across multiple advertising networks without the exchange of raw data, thereby ensuring privacy while enhancing overall detection performance.

Lastly, the analyzed articles raise the issue of standard evaluation criteria and the urgent need for benchmark datasets. Different research works utilize a variety of datasets, ranging from artificial to proprietary ones, making it challenging to compare models objectively. The creation of large-scale, anonymized, and representative public datasets that encompass a wide range of fraudulent scenarios will facilitate the reproducibility of research and accelerate innovation. Additionally, the regular use of performance metrics that extend beyond accuracy, such as precision, recall, F1-score, and AUC-ROC, is crucial for a just and accurate assessment, particularly when considering the highly imbalanced nature of click fraud data.

To sum up, the collective research works point out that, although modern ML and DL algorithms have notably advanced click fraud detection, there are still issues

with real-time adaptability, interpretability, and cross-domain generalization. The next step in this research area is to create hybrid, explainable, and privacy-aware models that can continually learn from and adapt to evolving fraudulent behaviors, while ensuring operational efficiency. Apart from making digital advertising platforms more reliable, such systems will also help create a more transparent and secure online economy.

V. CONCLUSION

Click fraud is still one of the most significant issues that threatens the net worth and the stability of internet ads. With the continuous expansion of the digital marketing environment, the methods used to identify fraudulent clicks have evolved slightly. In fact, fraudulent click detection has turned from simple rule-based filters to intelligent data-driven machine learning (ML) and deep learning (DL) frameworks. This review paper provides a step-by-step breakdown of the research works that have prototyped machine learning and deep learning models, including Decision Trees (DT), Random Forests (RF), gradient-boosted decision trees (GBDT), XGBoost, LightGBM, and hybrid deep-ensemble models, for detecting click fraud. Prior work syntheses (i.e., research literature reviews) reveal that tree-based algorithms and their ensemble variants consistently demonstrate the best accuracy, scalability, and robustness in detecting deceptive user behaviors across various advertising datasets. The amalgamation of temporal, spatial, and behavioral attributes has also enhanced the accuracy of fraud detection. Meanwhile, contemporary hybrid models combining CNN, BiLSTM, and RF or boosting techniques have achieved an accuracy above 98 percent.

Simply put, these achievements do not imply the disappearance of all the problems that still exist. For example, the majority of models operate on a relatively limited set of static and usually proprietary datasets; this is why these models cannot be easily extended to the real world. Data imbalance, feature scarcity, and lack of standard benchmarks thwart cross-study comparisons and reproducibility. Additionally, the ensemble and deep structures with the best performances, which are often referred to as “black boxes,” do not provide much insight for advertisers and analysts. Hence, the issue of balancing detection performance, model transparency, and computational efficiency remains unsolved.

The second batch of experiments should, thus, focus on developing fraud detection instruments that are not only transparent and privacy-preserving but also adaptive. The use of Explainable AI (XAI) techniques will help demystify the model, thereby instilling user trust. Concurrently, federated and privacy-friendly learning will facilitate secure collaborations among advertising networks without the risk of exposing user data. Moreover, they should implement real-time, incremental, and adversarially robust learning methods to continually update their fraud detection strategies, ensuring that their models remain viable in the long run. They will play a significant role in making research reproducible and comparable, thereby creating large, public, and standardized benchmark datasets. Machine learning and deep learning continue to be the primary factors driving significant changes in click fraud detection methods, as they shift from static heuristics to intelligent, autonomous, and adaptive decision systems. The synergistic use of ensemble learning, deep architectures, and XAI might offer a feasible solution to the problems of digital advertising security and transparency. The promises of tomorrow, including scaling, interpretability, and the ethical use of data, could, to a considerable degree, lead to the development of robust and effective click fraud detection frameworks capable of not only safeguarding advertisers' investments but also earning their trust.

REFERENCES

[1] Stone-Gross, B., Stevens, R., Zarras, A., Kemmerer, R., Kruegel, C., & Vigna, G. (2011, November 2–4). Understanding fraudulent activity in online ad exchanges. In Proceedings of the 11th ACM SIGCOMM Internet Measurement Conference (IMC ’11) (pp. 279–294). ACM. https://doi.org/10.1145/2068816.2068843 ACM Digital Library+2conferences.sigcomm.org+2

[2] Li, Z., Zhang, K., Xie, Y., Yu, F., & Wang, X. F. (2012, October 16–18). Knowing your enemy: Understanding and detecting malicious Web advertising. In Proceedings of the 2012 ACM Conference on Computer and Communications Security (CCS ’12) (pp. 674–

686). ACM. https://doi.org/10.1145/2382196.2382267 ACM Digital Library+1

[3] Berrar, D. (2012, November 4). Random forests for the detection of click fraud in online mobile advertising. In Proceedings of the 1st International Workshop on Fraud Detection in Mobile Advertising (FDMA ’12) (pp. 1–10).

[4] Yan, J. H., & Jiang, W. R. (2014). Research on information technology to detect fraudulent clicks using a classification method. Advanced Materials Research, 859, 586–590. Perera, K. S., Neupane, B., Faisal, M. A., Aung, Z., & Woon, W. L. (2013). A novel ensemble learning-based approach for click fraud detection in mobile advertising. In Lecture Notes in Computer Science (Vol. 8284, pp. 370–382). Springer. https://doi.org/10.1007/978-3-319-03844-5_38

[5] Oentaryo, R. J., & Lim, E. (2013, November 21–23). Mining fraudulent patterns in online advertising. In First International Network on Trust (FINT) Workshop (pp. 21–23).

[6] Perera, B. K. (2013). A class imbalance learning approach to fraud detection in online advertising [Unpublished conference paper / technical report].

[7] Oentaryo, R., Lim, E. P., Finegold, M., Lo, D., Zhu, F., Phua, C., Cheu, E. Y., Yap, G. E., Sim, K., Nguyen, M. N., et al. (2014). Detecting click fraud in online advertising: A data mining approach. Journal of Machine Learning Research, 15(1), 99–140.

[8] Phua, C., Cheu, E.-Y., Yap, G.-E., Sim, K., & Nguyen, M. N. (2012, November 4). Feature engineering for click fraud detection. In Proceedings of the Workshop on Fraud Detection in Mobile Advertising (FDMA ’12) (Vol. 2010, pp. 1–10).

[9] Xu, H., Liu, D., Koehl, A., Wang, H., & Stavrou, A. (2014). Click fraud detection on the advertiser side. In Lecture Notes in Computer Science (Vol. 8713, pp. 419–438). Springer. https://doi.org/10.1007/978-3-319-11578-2_23

[10] He, X., Pan, J., Jin, O., Xu, T., Liu, B., Xu, T., Shi, Y., Atallah, A., Herbrich, R., Bowers, S., et al. (2014, August 24–27). Practical lessons from predicting clicks on ads at Facebook, in Proceedings of the Eighth International Workshop on Data Mining for Online Advertising (pp. [pages unknown]).

[11] Vani, M. S., Bhramaramba, R., Vasumati, D., & Babu, O. Y. (2014). TUI based touch-spam detection in mobile applications to increase the security from advertisement networks. International Journal of Advanced Computer Communication and Control, 2, 17–22.

[12] Beránek, L., Nýdl, V., & Remeš, R. (2016). Click stream data analysis for online fraud detection in E-Commerce. In Inproforum: České Budějovice (pp. 175–180).

[13] Berrar, D. (2016). Learning from automatically labeled data: Case study on click fraud prediction. Knowledge and Information Systems, 46(2), 477–490. https://doi.org/10.1007/s10115-015-0827-6

[14] Guo, Y., Shi, J., Cao, Z., Kang, C., Xiong, G., & Li, Z. (2019, August 10–12). Machine learning based cloudbot detection using multi-layer traffic statistics. In 2019, IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS) (pp. 2428–2435). https://doi.org/10.1109/HPCC/SmartCity/DSS.2019.00386

[15] Li, Z., & Jia, W. (2020). The study on preventing click fraud in internet advertising. Journal of Computers, 31, 256–265. https://doi.org/10.3966/199115992020013101021

[16] Wang, K., Xu, G., Wang, C., & He, X. (2017, August 9–10). A hybrid abnormal advertising traffic detection method. In Proceedings of the 2017 IEEE International Conference on Big Knowledge (ICBK) (pp. 236–241). https://doi.org/10.1109/ICBK.2017.49

[17] Minastireanu, E.-A., & Mesniță, G. (2019). Light GBM machine learning algorithm to online click fraud detection. Journal of Information Assurance & Cybersecurity, 2019, Article 4, 1–12. https://doi.org/10.5171/2019.263928

[18] Sisodia, D., & Sisodia, D. S. (2021). Gradient boosting learning for fraudulent publisher detection in online advertising. Data Technologies & Applications, 55(2), 216–232. https://doi.org/10.1108/DTA-04-2020-0093

[19] R. Mouawi, M. Awad, A. Chehab, I. H. El Hajj, and A. Kayssi, "Towards a machine learning approach for detecting click fraud in mobile advertising," in Proc. 2018 Int. Conf. Innovations in Information Technology (IIT), Al Ain, UAE, 2018, pp. 88–92. doi: 10.1109/INNOVATIONS.2018.8605973.

[20] R. Oentaryo, E. P. Lim, M. Finegold, et al., "Detecting click fraud in online advertising: A data mining approach," Journal of Machine Learning Research, vol. 15, no. 1, pp. 99–140, 2014.

[21] C. A. do Espírito Santo, "Advertisement click fraud detection and prevention: A machine learning approach," M.S. thesis, Universidade NOVA de Lisboa, Lisbon, Portugal, 2024.

[22] V. B. Mahesh, K. V. S. Chandra, L. S. P. Babu, V. A. Sowjanya, and M. Mohammed, "Clicking fraud detection for online advertising using machine learning," in Proc. 2023 4th Int. Conf. Intelligent Technologies (CONIT), Bangalore, India, 2024, pp. 1–6. doi: 10.1109/CONIT61985.2024.10627189.

[23] G. S. Thejas, S. Dheeshjith, S. S. Iyengar, N. R. Sunitha, and P. Badrinath, "A hybrid and effective learning approach for click fraud detection," Machine Learning with Applications, vol. 3, p. 100016, 2021. doi: 10.1016/j.mlwa.2020.100016.

[24] R. A. Alzahrani, M. Aljabri, and R. M. A. Mohammad, "Ad click fraud detection using machine learning and deep learning algorithms," IEEE Access, vol. 13, pp. 12746–12763, 2025. doi: 10.1109/ACCESS.2025.3532200.

[25] M. Aljabri and R. M. A. Mohammad, "Click fraud detection for online advertising using machine learning," Egyptian Informatics Journal, vol. 24, no. 2, pp. 341–350, 2023. doi: 10.1016/j.eij.2023.05.006.

[26] A. Batool and Y. C. Byun, "An ensemble architecture based on a deep learning model for click fraud detection in pay-per-click advertisement campaigns," IEEE Access, vol. 10, pp. 113410–113426, 2022. doi: 10.1109/ACCESS.2022.3211528.

[27] A. Batool, J. Kim, and Y. C. Byun, "Enhanced click fraud detection in digital advertising through ensemble deep learning," in Proc. Int. Conf. Frontier Computing, Singapore, 2024, pp. 22–27. doi: 10.1007/978-981-96-2395-2_5.

[28] E. A. Minastireanu and G. Mesnita, "LightGBM machine learning algorithm to online click fraud detection," Journal of Information Assurance & Cyber Security, 2019, Article ID 263928. doi: 10.5171/2019.263928.

[29] G. S. Thejas, K. G. Boroojeni, K. Chandna, I. Bhatia, S. S. Iyengar, and N. R. Sunitha, "Deep learning-based model to fight against ad click fraud," in Proc. 2019 ACM Southeast Conf., 2019, pp. 176–181. doi: 10.1145/3299815.3314453.