INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 6, June 2026 | ISSN: 2456-4184 | IJNRD.ORG

Online Advertisement Fraud Detection Using Hybrid Deep Learning and Attention Mechanism

GANDA STHUTHIRAJU
Master Of Computer Applications
Ideal College Of Arts & Sciences,
Autonomous, Affiliated To Adikavi Nannaya University - Rajamahendravaram
Kakinada

M. KAMESWARA RAO
Assistant.Prof, Master Of Computer Applications
Ideal College Of Arts & Sciences,
Autonomous, Affiliated To Adikavi Nannaya University - Rajamahendravaram
Kakinada

Dr. V.S.V DEEPAK
HOD, department of computer science
Ideal College Of Arts & Sciences,
Autonomous, Affiliated To Adikavi Nannaya University - Rajamahendravaram
Kakinada

***Abstract*— Online advertising platforms face major financial losses due to fraudulent ad clicks generated by automated bots and malicious users. Detecting these fake clicks using traditional rule-based methods is difficult because such approaches cannot effectively learn complex behavioral patterns from large-scale click data. This work proposes an enhanced ad click fraud detection system using machine learning and deep learning algorithms with an extension model based on CNN integrated with an Attention mechanism. The system performs preprocessing, feature selection using RFE, and classification of clicks as legitimate or fraudulent. Multiple algorithms were evaluated, including Random Forest, XGBoost, RNN, and CNN. Experimental results demonstrate that the proposed CNN with Attention model achieved 99.73% accuracy, outperforming existing models and improving the reliability of online advertisement security systems.**

***Keywords*— Deep Learning, Ad Click, Attention Mechanism, CNN**

I. INTRODUCTION

Digital advertising has become one of the most important marketing strategies for businesses, where advertisers promote products and services through online platforms and websites. Most advertising companies use pay-per-click models in which website owners receive revenue whenever users click on advertisements. Due to the rapid growth of online advertising, fraudulent activities such as fake ad clicks have also increased significantly. Fraudsters use automated bots, scripts, and malicious techniques to generate invalid clicks in order to gain unfair financial benefits and exhaust advertisers' budgets. These fraudulent activities reduce advertisement reliability and create major financial losses for advertising companies.

Traditional fraud detection techniques mainly depend on predefined rules and manual monitoring methods. However, these approaches are unable to effectively identify complex and continuously changing fraud patterns present in large-scale online traffic data. The increasing volume of user interactions, browser activities, device information, and network traffic has made ad click fraud detection a challenging task. As a result, intelligent data-driven approaches are becoming essential for improving detection accuracy and reducing false predictions.

Machine learning and deep learning techniques provide efficient solutions for analyzing large datasets and identifying hidden behavioral patterns associated with fraudulent activities. These methods can automatically learn relationships between multiple features and improve classification performance over time. By applying advanced analytical techniques, online advertisement platforms can improve security, protect advertiser investments, and ensure fair usage of digital advertising systems.

II. RELATED WORK

Stone-Gross et al. (2011) presented one of the earliest studies on fraudulent activities in online advertisement exchanges and explained how fake clicks generate financial losses for advertisers. Berrar (2012) introduced the Random Forest algorithm for detecting fraudulent clicks and demonstrated the effectiveness of machine learning approaches over traditional rule-based systems. Phua et al. (2012) emphasized the importance of feature engineering by analyzing user sessions, click frequency, and browsing behavior for fraud identification. Yan and Jiang (2013) explored classification techniques for automated fraud detection and proved that intelligent learning methods improve prediction capability. Perera et al. (2013) proposed an ensemble learning approach that combined multiple classifiers to improve detection accuracy on imbalanced datasets. Minastireanu and Mesnita (2019) investigated the LightGBM algorithm and showed that boosting methods provide efficient and accurate click fraud detection. Dash and Pal (2020) highlighted the importance of automated machine learning systems for reducing manual fraud monitoring efforts. Sisodia and Sisodia (2021) applied Gradient Boosting methods for detecting fraudulent publisher behavior and achieved improved classification performance. Aljabri and Mohammad (2023) compared multiple machine learning algorithms and emphasized the growing need for intelligent fraud prevention systems in online advertising. Recently, Kirkwood et al. (2024) evaluated advanced machine learning models for detecting advertisement fraud and confirmed their effectiveness in securing digital advertising platforms against malicious automated activities.

Table: Summary of Key Literature Contributions and Their Impact on Current Research:

Author	Contribution	Impact on Research
Stone-Gross et al. (2011)	Studied fake activities in online advertisements.	Helped researchers understand the problem of click fraud.
Berrar (2012)	Used Random Forest for fraud detection.	Showed that machine learning improves fraud prediction accuracy.
Phua et al. (2012)	Worked on feature	Improved the identification

	selection and user behavior analysis.	of fraud-related patterns.
Yan and Jiang (2013)	Applied classification methods for fraud detection.	Supported the use of intelligent automated detection systems.
Perera et al. (2013)	Proposed ensemble learning for fraud detection.	Increased detection accuracy on large datasets.
Minastireanu and Mesnita (2019)	Used LightGBM for click fraud detection.	Improved processing speed and prediction performance.
Dash and Pal (2020)	Developed automated fraud detection methods.	Reduced dependence on manual monitoring systems.
Sisodia and Sisodia (2021)	Applied Gradient Boosting for fraud prediction.	Improved classification performance in advertising systems.
Aljabri and Mohammad (2023)	Compared different machine learning algorithms.	Helped identify efficient models for fraud detection.
Kirkwood et al. (2024)	Evaluated machine learning methods for ad fraud detection.	Strengthened research on secure digital advertising systems.

III. PROPOSED APPROACH

TAdvertisement click data is initially collected from online advertising platforms containing information related to user behavior, browser details, device type, operating system, and click activities. Since raw datasets often contain inconsistent and non-numeric values, preprocessing techniques are applied to improve data quality and prepare the dataset for accurate fraud detection. Label Encoding is used to transform categorical attributes into numerical form, while normalization and missing value handling improve the consistency of the training data. This stage helps reduce noise and increases the reliability of the classification process.

After preprocessing, feature selection is performed using the Recursive Feature Elimination (RFE) algorithm to identify the most important attributes influencing fraudulent click behavior. Selecting relevant features reduces unnecessary complexity and improves overall prediction efficiency. The refined dataset is then divided into training and testing datasets, where the majority of records are used for learning and the remaining records are used for performance evaluation.

Several machine learning algorithms including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, XGBoost, LightGBM, Naïve Bayes, SVM, and KNN are trained and tested to analyze their fraud detection capability. Deep learning models such as CNN, DNN, and RNN are also implemented to learn hidden patterns from advertisement click behavior.

To improve prediction performance, the extension model combines a Convolutional Neural Network with an Attention mechanism. The Attention layer helps the model concentrate on important behavioral features associated with fraudulent clicks and improves contextual learning during classification. This enhancement enables better identification of complex fraud patterns generated by automated bots. Finally, all models are evaluated using performance metrics such as accuracy, precision, recall, and F1-score to determine the most effective fraud detection approach for secure online advertising systems.

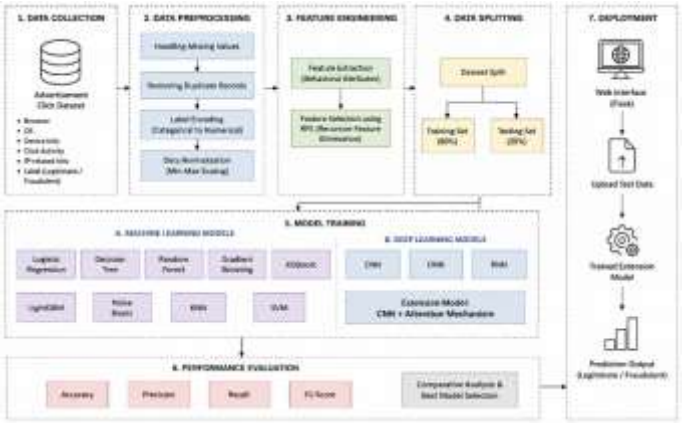

Figure 1: Click fraud detection workflow

IV. METHODOLOGIES

Algorithm: CNN with Attention for Ad Click Fraud Detection

Input:
Advertisement Click Dataset Z

Output:
Predicted Class (Legitimate / Fraudulent)

---

1. Load advertisement click dataset Z

2. Preprocess Dataset
a. Remove duplicate records
b. Handle missing values
c. Convert categorical values to numerical values using Label Encoder
d. Normalize dataset values

3. Extract input features X and target labels Y

4. Apply Recursive Feature Elimination (RFE)
a. Select important features
b. Remove irrelevant features

5. Split dataset
a. Training Data = 80%
b. Testing Data = 20%

6. Initialize CNN Model
a. Add Convolution Layer
b. Add ReLU Activation Layer
c. Add Max Pooling Layer
d. Repeat convolution and pooling operations

7. Apply Attention Mechanism
a. Compute feature importance weights
b. Assign higher weights to important fraud-related features
c. Generate attention output

8. Flatten extracted feature maps

9. Add Fully Connected Dense Layers

10. Apply Softmax/Sigmoid Classifier
a. Classify click as Legitimate or Fraudulent

11. Train Extension Model using Training Data

12. Test Model using Testing Data

13. Calculate Performance Metrics
a. Accuracy
b. Precision
c. Recall
d. F1-Score

14. Compare performance with existing ML and DL algorithms

15. Deploy trained model in Flask web application
a. Upload test data
b. Predict fraud result
c. Display output

End

Dataset Collection

The first step involves collecting the advertisement click dataset from online sources related to click fraud detection. The dataset contains details such as browser type, operating system, device information, click frequency, IP-related activity, and click labels representing legitimate or fraudulent behavior. The collected data serves as the foundation for training and evaluating the fraud detection models.

Data Inspection and Analysis

The collected dataset is analyzed to understand its structure, attribute types, and class distribution. Statistical analysis and visualization techniques are used to identify data imbalance, duplicate records, and inconsistencies. Graphical analysis helps in understanding user behavior patterns and fraudulent click trends across different browsers and operating systems.

Data Preprocessing

Preprocessing is performed to improve the quality of the dataset before training the models. Missing values are handled, duplicate records are removed, and noisy data is cleaned. Categorical attributes such as browser names and operating systems are converted into numerical form using Label Encoding techniques for efficient model processing.

Data Normalization

The numerical features in the dataset are normalized to maintain uniformity among different attribute values. Normalization prevents large-value attributes from dominating the prediction process and improves the convergence speed of machine learning and deep learning algorithms during training.

Feature Extraction

Important behavioral attributes related to fraudulent clicks are extracted from the processed dataset. Features such as click count, browser usage, device activity, and operating system patterns are analyzed to identify meaningful fraud-related information useful for classification.

Feature Selection Using RFE

Recursive Feature Elimination (RFE) is applied to select the most relevant features from the dataset. The algorithm removes less important attributes iteratively and retains only the significant features that contribute effectively to fraud prediction. This process reduces computational complexity and improves model efficiency.

Dataset Splitting

The selected dataset is divided into training and testing datasets. Around 80% of the data is used for model training, while the remaining 20% is used for testing and validation. This separation ensures fair performance evaluation and prevents overfitting issues during prediction.

Machine Learning Model Training

Different machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, XGBoost, LightGBM, Naïve Bayes, KNN, and SVM are trained using the processed dataset. Each algorithm learns the relationship between user behavior and fraud patterns to classify advertisement clicks accurately.

Deep Learning Model Implementation

Deep learning models including CNN, DNN, and RNN are implemented to capture complex hidden patterns from advertisement click data. These models automatically learn high-level representations from the dataset and improve fraud detection capability compared to traditional approaches.

Extension Model Development

The proposed extension model integrates an Attention mechanism with the CNN architecture. The Attention layer helps the model focus on the most important behavioral features related to fraudulent clicks. This enhancement improves contextual understanding and strengthens the ability of the system to detect complex bot-generated click activities.

Performance Evaluation

All machine learning and deep learning models are evaluated using performance metrics such as accuracy, precision, recall, and F1-score. Comparative analysis is performed to identify the best-performing model. Experimental results show that the CNN with Attention extension model achieves superior fraud detection accuracy compared to existing models.

Fraud Prediction

The final trained extension model is integrated into a web-based prediction system using Flask. Users can upload test datasets through the web interface, and the system predicts

whether the advertisement clicks are legitimate or fraudulent. This deployment stage demonstrates the practical applicability of the proposed fraud detection framework in real-world online advertising environments.

VI RESULTS & DISCUSSION

	Algorithm Name	Accuracy	Precision	Recall	FSCORE
0	Logistic Regression	97.267	96.573	96.343	96.457
1	Decision Tree	98.933	98.044	99.277	98.639
2	Random Forest	77.467	76.877	59.049	59.310
3	KNN	73.200	48.649	49.922	43.215
4	ANN	72.400	60.286	55.698	55.523
5	Gradient Boosting	98.000	96.454	98.645	97.475
6	LightGBM	98.533	97.349	99.006	98.137
7	XGBoost	98.667	97.579	99.097	98.304
8	Naive Bayes	89.333	85.492	92.527	87.606
9	SVM	73.800	36.900	50.000	42.463
10	CNN	95.667	93.212	96.326	94.599
11	DNN	73.800	36.900	50.000	42.463
12	RNN	98.733	97.825	98.978	98.382
13	Extension CNN with Attention	99.733	99.655	99.655	99.655

The experimental results demonstrate the effectiveness of the proposed advertisement click fraud detection system using multiple machine learning and deep learning algorithms. The performance of each model was evaluated using accuracy, precision, recall, and F1-score metrics. Among the traditional machine learning models, Logistic Regression achieved 97.267% accuracy with 96.573% precision and 96.343% recall. Decision Tree produced a high accuracy of 98.933% with an F1-score of 98.639, showing strong classification capability for fraudulent click detection. Gradient Boosting, LightGBM, and XGBoost also achieved excellent performance with accuracies of 98.000%, 98.533%, and 98.667% respectively. XGBoost obtained 97.579% precision and 99.097% recall, indicating effective identification of fraudulent clicks.

Some algorithms such as KNN, ANN, and SVM showed lower performance due to limitations in handling complex behavioral patterns present in the advertisement click dataset. KNN achieved only 73.200% accuracy, while ANN and DNN produced 72.400% and 73.800% accuracy respectively. Naïve Bayes achieved moderate performance with 89.333% accuracy.

Among deep learning models, CNN achieved 95.667% accuracy, while RNN performed significantly better with 98.733% accuracy and 98.382 F1-score. The proposed extension model, CNN integrated with an Attention mechanism, achieved the highest performance among all models with 99.733% accuracy, 99.655% precision, 99.655% recall, and 99.655% F1-score. These results confirm that the Attention mechanism improved feature learning and enhanced the capability of the CNN model to detect complex fraudulent advertisement click patterns effectively.

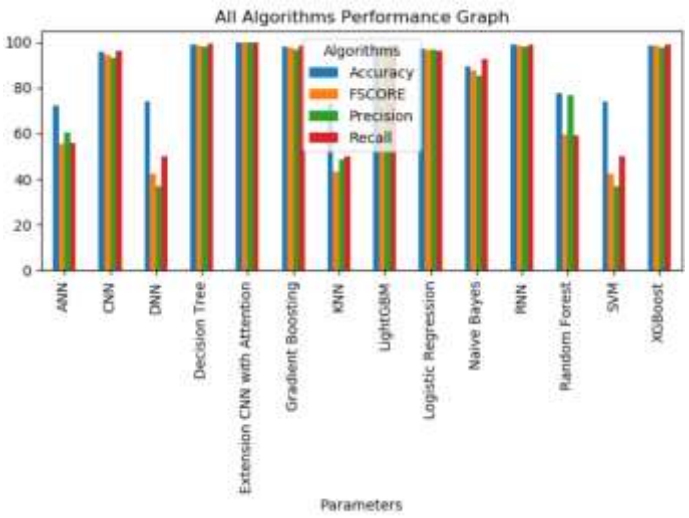

Figure 2: All Algorithms Performance Graph

The experimental analysis shows that machine learning and deep learning techniques can effectively detect fraudulent advertisement clicks from large-scale online traffic data. Traditional algorithms such as Decision Tree, Gradient Boosting, LightGBM, and XGBoost achieved strong prediction accuracy because they efficiently learned behavioral patterns from the dataset. Deep learning models also performed well, especially RNN, which captured sequential relationships in user click behavior. However, models such as KNN, ANN, and SVM produced lower accuracy due to their limitations in handling complex and highly imbalanced fraud data.

The proposed extension model, CNN integrated with an Attention mechanism, achieved the best overall performance with 99.733% accuracy. The Attention layer improved the model's ability to focus on important fraud-related features and enhanced contextual understanding during classification. This helped the system identify complex bot-generated click activities more accurately than existing methods. The obtained results confirm that combining deep learning with Attention mechanisms can significantly improve advertisement fraud detection performance and provide a reliable solution for secure digital advertising systems.

VII. CONCLUSION

The research successfully developed an intelligent advertisement click fraud detection system using machine learning and deep learning techniques. The study analyzed multiple classification algorithms to identify fraudulent and legitimate advertisement clicks based on user behavioral data. Experimental results proved that advanced learning models provide better fraud detection capability compared to traditional rule-based methods. Among all evaluated algorithms, the proposed extension model using CNN with an Attention mechanism achieved the highest accuracy of 99.733%, demonstrating superior performance in detecting complex fraudulent click patterns. The Attention layer

improved feature learning by focusing on important fraud-related information during classification. The developed system can help online advertising platforms reduce financial losses caused by fake clicks and improve advertisement security. The proposed framework also provides a scalable and reliable solution for future intelligent digital advertising systems.

REFERENCES

[1] Juniper Research, Hampshire, U.K. Quantifying the Cost of Ad Fraud: 2023–2028. Accessed: Jul. 12, 2024. [Online]. Available: https://fraudblocker.com/wp-content/uploads/2023/09/Ad-Fraud-Whitepaper_Juniper-Research.pdf

[2] X. Zhu, H. Tao, Z. Wu, J. Cao, K. Kalish, and J. Kayne, Fraud Prevention in Online Digital Advertising. Cham, Switzerland: Springer, 2017.

[3] A. K. Wood and A. M. Ravel, ''Fool me once: Regulating fake news and other online advertising,'' S. Cal. L. Rev., vol. 91, p. 1223, Jan. 2017.

[4] B. Stone-Gross, R. Stevens, A. Zarras, R. Kemmerer, C. Kruegel, and G. Vigna, ''Understanding fraudulent activities in online ad exchanges,'' in Proc. ACM SIGCOMM Conf. Internet Meas. Conf., Nov. 2011, pp. 279–294.

[5] (2024). Wasted Ad Spend Report 2024. [Online]. Available: https://lp.lunio.ai/wp-content/uploads/2023/09/Lunio_Wasted_Ad_Spend_Report_2024_V2.pdf

[6] D. Berrar, ''Random forests for the detection of click fraud in online mobile advertising,'' in Proc. Int. Work. Fraud Detect. Mob. Advert. (FDMA), Singapore, 2012, pp. 1–10. [Online]. Available: http://berrar.com/ resources/Berrar_FDMA2012.pdf

[7] J. H. Yan and W. R. Jiang, ''Research on information technology with detecting the fraudulent clicks using classification method,'' Adv. Mater. Res., vol. 859, pp. 586–590, Dec. 2013, doi: 10.4028/www.scientific.net/amr.859.586.

[8] K. S. Perera, B. Neupane, M. A. Faisal, Z. Aung, and W. L. Woon, ''A novel ensemble learning-based approach for click fraud detection in mobile advertising,'' in Mining Intelligence and Knowledge Exploration (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 8284. Berlin, Germany: Springer, 2013, pp. 370–382, doi: 10.1007/978-3-319-03844-5_38.

[9] C. Phua, E.-Y. Cheu, G.-E. Yap, K. Sim, and M.-N. Nguyen, ''Feature engineering for click fraud detection,'' in Proc. Work. Fraud Detect. Mob. Advert., 2012, pp. 1–10. [Online]. Available: http://palanteer.sis.smu.edu. sg/fdma2012/doc/FirstWinner-Starrystarrynight-Paper.pdf%5Cnpapers2: //publication/uuid/9290A6CF-A861-4058-99F4-D39706B0619A

[10] E.-A. Minastireanu and G. Mesnita, ''Light GBM machine learning algorithm to online click fraud detection,'' J. Inf. Assurance Cybersecur., vol. 2019, pp. 1–12, Apr. 2019, doi: 10.5171/2019.263928.

[11] D. Sisodia and D. S. Sisodia, ''Gradient boosting learning for fraudulent publisher detection in online advertising,'' Data Technol. Appl., vol. 55, no. 2, pp. 216–232, Apr. 2021, doi: 10.1108/dta-04-2020-0093.

[12] A. Dash and S. Pal, ''Auto-detection of click-frauds using machine learning Auto-detection of click-frauds using machine learning,'' Int. J. Eng. Sci. Comput., vol. 10, pp. 27227–27235, Sep. 2020.

[13] R. Mouawi, M. Awad, A. Chehab, I. H. E. Hajj, and A. Kayssi, ''Towards a machine learning approach for detecting click fraud in mobile advertizing,'' in Proc. Int. Conf. Innov. Inf. Technol. (IIT), Nov. 2018, pp. 88–92, doi: 10.1109/INNOVATIONS.2018.8605973.

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

[15] S. Shaik and V. Kakulapati, ''Fraud detection of AD clicks using machine learning techniques,'' J. Sci. Res. Rep., vol. 29, no. 7, pp. 84–89, Jun. 2023, doi: 10.9734/jsrr/2023/v29i71762.

[16] D. Sisodia and D. S. Sisodia, ''Stacked generalization architecture for predicting publisher behaviour from highly imbalanced user-click data set for click fraud detection,'' New Gener. Comput., vol. 41, no. 3, pp. 581–606, Sep. 2023, doi: 10.1007/s00354-023-00218-1.

[17] D. Sisodia, D. S. Sisodia, and D. Singh, ''Evaluating feature importance to investigate publishers conduct for detecting click fraud,'' in Machine Intelligence Techniques for Data Analysis and Signal Processing (Lecture Notes in Electrical Engineering), vol. 997. Berlin, Germany: Springer, 2023, pp. 515–524, doi: 10.1007/978-981-99-0085-5_42.

[18] R. Dekou, S. Savo, S. Kufeld, D. Francesca, and R. Kawase, ''Machine learning methods for detecting fraud in online marketplaces,'' in Proc. CEUR Workshop, vol. 3052, Jan. 2021, pp. 3–7.

[19] D. Sisodia and D. S. Sisodia, ''Quad division prototype selection-based Knearest neighbor classifier for click fraud detection from highly skewed user click dataset,'' Eng. Sci. Technol., Int. J., vol. 28, Apr. 2022, Art. no. 101011, doi: 10.1016/j.jestch.2021.05.015.

[20] B. Kirkwood, M. Vanamala, and N. Seliya, ''Click fraud detection of online advertising using machine learning algorithms,'' in Proc. IEEE Int. Conf. Electro Inf. Technol. (eIT), May 2024, pp. 586–590. [Online]. Available: https://api.semanticscholar.org/CorpusID

GANDA STHUTHIRAJU is currently pursuing MCA (Master of computer application) in ideal college of arts and science, Vidyuthnagar, Kakinada. His research interests include Cyber Security

M. Kameswara Rao is currently serving as the Additional Head of the Department of Computer Science at Ideal College of Arts & Sciences(A). He possesses more than 20 years of academic and administrative experience in the field of Computer Science.

Dr. V. S. V. Deepak is currently serving as the Head of the Department of Computer Science at Ideal College of Arts & Sciences (A). He possesses more than 18 years of academic and administrative experience in the field of Computer Science and Engineering. His areas of interest include Medical Image Processing, Cyber Security, Artificial Intelligence, Software Testing and Networking. He completed his Ph.D. research in Medical Image Processing from Swami Vivekananda University.
He has actively contributed to curriculum development, academic planning, and student mentoring. He has served as Chairman of the Board of Studies (BOS) for BCA, B.Sc. (Computer Science), B.Sc. (Artificial Intelligence), and MCA programs.