Archives available at journals.mriindia.com

International Journal of Recent Advances in Engineering and Technology

ISSN: 2347 - 2812

Volume 14 Issue 02s, 2025

Revolutionizing Web Browsing with AI-Powered Personalized Ad Blockers

[1]Saurabh Srivastava, [2]Seema Kharod, [3]Khushi Sinha, [4]Megha Kaushal, [5]Nishu Kumari, [6]Adasrh Kumar

[1,2,3,4,5,6] University Institute of Engineering Chandigarh University, Mohali, India

Email: [1]srivastavasaurabh978@gmail.com, [2]seema.kharod30@gmail.com, [3]sinhakhushi0803@gmail.com, [4]megkaushal1019@gmail.com, [5]nishisingh1026@gmail.com, [6]adarshkumaranit@gmail.com

Peer Review Information

Submission: 21 Oct 2025

Revision: 18 Nov 2025

Acceptance: 05 Dec 2025

Keywords

Personalized Web Browsing, Ad Filtering, Federated Learning, Ad Blockers, User Privacy

Abstract

The growth of obtrusive and unrelated online advertisements has profoundly affected user experience and privacy. Traditional ad blockers, based on static rule-based approaches, find it difficult to cope with changing advertising tactics. This research suggests an AI-based personalized ad blocker aimed at transforming web browsing by providing dynamic, smart, and user-centric ad filtering. Based on machine learning models and behavioral analytics, the solution is tailored to match individual user behavior while ensuring high accuracy in recognizing and removing intrusive ads. The system combines computer vision and natural language processing (NLP) methods for end-to-end ad content analysis, enabling real-time detection of visual and textual components. Privacy is prioritized through on-device processing and federated learning approaches with minimal data exposure. Performance evaluations demonstrate significant improvements in detection accuracy, personalization, and browsing speed relative to conventional methods. This research highlights the promise of AI-driven solutions to improve user experience, maintain privacy, and revolutionize digital content consumption.

Introduction

Today's Internet environment is directly affected by digital advertising, which plays an essential role in profiting web content. However, the disruptiveness of the majority of advertisements has sparked increasing dissatisfaction in users. Pop-ups, banner ads, and autoplaying clips commonly break browsing sessions and invade user privacy through tracking of behavioral activities across websites. [1] Since advertisers make constantly evolving changes in how they escape traditional blocking methods, users on the web increasingly demand better, more intelligent solutions to get their online interaction under their control.Traditional ad blockers [2] are largely based on static rule-based methods, filtering out advertisements by comparing known patterns or blocklists. Although these methods have proved effective to some extent, they lack flexibility and need constant manual updates to respond to evolving ad delivery mechanisms. Native ads and dynamic ad-serving technologies have also made it easier for advertisers to circumvent conventional ad blockers, making them less effective in the long run. Recent advances in artificial intelligence (AI) [3] have made it possible to mitigate these challenges. AI-based solutions can intelligently scan and recognize advertisements by learning complex

patterns and evolving to novel formats without human interference. Machine learning algorithms [4] can process vast quantities of web data to distinguish between valid content and intrusive advertising in real time, thereby presenting users with a seamless browsing experience. AI makes [5] personalization possible, allowing ad blockers to personalize their filtering approach according to individual browsing and preferences. This work presents a novel AI-based personalized web browser ad blocker that utilizes machine learning methods for real-time advertisement identification and filtering. Through the combination of computer vision and NLP models, the system is able to correctly identify visual and textual elements of ads. The solution also prioritizes user privacy through the inclusion of federated learning, preserving data confidentiality while allowing personalized filtering.

Related Work

Zhao et al. [6] proposed a new method to combat ad-blocking by using deep learning to personalize ad-blocking approaches. Their research is centered on creating a predictive model, the Deep Ad-Block Whitelist Network (DAWN), that predicts whether users will whitelist a webpage. This enables publishers to dynamically implement either the ”Wall” strategy, which encourages users to whitelist the site, or the ”AAX” strategy, which shows acceptable ads. The success of this strategy is evidenced through better revenue generation and user engagement as compared to legacy strategies such as Wall-only or AAX-only policies1.Mıklosık et al. [7] examined the fine line between privacy preservation and ad revenues, especially for publishers of content. Their research presents the plight of publishers to strike a balance between these opposing forces. As regulatory measures become stronger, publishers are required to reconcile the sustenance of revenues without infringing upon user privacy. This equilibrium is very important within the online ad world, where privacy can fuel ad-blocking behaviors, which, in turn, make it difficult for publishers to generate revenue. Lashkari et al. [8] propose CIC-AB, a browser extension ad blocker based on machine learning to identify and block online ads. CIC-AB goes beyond static filter list-based traditional ad blockers, categorizing URLs into non-ad, normal ad, or malicious ad with high accuracy. The research illustrates the efficiency of machine learning models, e.g., kNN, in classification of ad URLs with precision at 97.16% and recall at 94.96%. Future research involves improving accuracy, identifying new ad types, and assessing the model’s effect on user experience and site functionality.Iqbal et al. [9] introduce AdGraph, a graph based machine learning system that identifies online ads and tracking mechanisms with high accuracy. Rather than using static filter lists, AdGraph builds an end-to-end representation of webpage execution by examining the HTML structure, network requests, and JavaScript behavior. This enables it to detect and block adverts with 97.7% accuracy, better than traditional methods. The research shows the capability of AdGraph against evasion techniques like domain rotation and code obfuscation, thus being a strong solution for blocking adverts and trackers.Din et al. [10] present PERCIVAL, an ad blocker developed using deep learning that classifies and blocks adverts by scanning webpage images. Unlike rule-based blockers of conventional type, PERCIVAL can efficiently identify non-English and first-party adverts, including adverts on social networking sites like Facebook, with accuracy of 96.76%. Evaluated with Chromium and Brave browsers, the system imitates EasyList rules with similar precision without compromising the page load performance to any notable extent. The research emphasizes the ability of deep learning to improve ad-blocking performance and versatility in various web environments.Szczepan´ski et al. [11] suggest an automated system for online advertisement detection based on URL-based web-page classification. Their research compares several machine learning models, such as decision trees, random forests, and AdaBoost, with an accuracy of 0.987 and F-measure of 0.822. Random Forest performed better, with a cost-sensitive F-measure of 0.947 and precision of 0.967. The research underscores the power of machine learning methods over rule-based methods, and their application in ad-blocking and content filtering applications.

Garimella et al. citeb12 discuss how ad-blocking and cookies affect user experience, privacy, and site performance. Their research examines how online businesses depend on advertising for income when users block adverts because of worries about page loads and data expenditure. The work explains various cookie types and why they are used, highlighting the importance of obtaining user consent and the impact of ad-blockers on privacy and site performance.Bhagavatula et al. [13] discuss a machine learning approach to improving ad filtering and minimizing

dependence on rule-based systems. Their work trains a k-Nearest Neighbors (kNN) classifier on past regular expression lists with a baseline accuracy of 95.6% and a new-ad accuracy of 83.1%. The paper emphasizes the need for ongoing retraining because ad URLs change most of the time and suggests possible future enhancements of feature extraction as well as adaptation in real-time.Singh et al. [14] examine the potential of online adverts as a source of primary revenues for websites with free services against the backdrop of user apprehension about intrusive advertising and bandwidth utilization. Their work investigates motivations for ad-blocking, evaluates different ad-blocking methods, and quantifies the bandwidth effect of advertisements. The results highlight the importance of balancing ad-based revenue models with user experience optimization.Storey et al. [15] study ad blocking using a security paradigm, casting it as a four-state state space with six transitions that describe interactions between publishers and ad blockers. Their work presents novel ad-blocking methods, some based on rootkits, to avoid detection by anti-ad blocking scripts. Prototype implementations of the techniques show successful ad blocking and evasion while being legal. Snyder et al. [16] examine the efficacy of EasyList, a popular filter list for ad and tracker blocking, and conclude that 90.16% of its rules are not used in common browsing cases. Their research identifies inefficiencies in current filter lists, notably on performance-limited mobile devices, and suggests a hybrid blocking approach that accelerates performance by 62.5% without losing more than 99% of the advantages of EasyList. These results help to improve security and privacy tools on both mobile and desktop platforms.

Methodology

The proposed AI-powered personalized ad blocker aims to enhance web browsing by dynamically identifying and filtering intrusive advertisements while preserving user privacy. The methodology follows a structured approach that integrates data collection, preprocessing, model training, real-time ad detection, and privacy-preserving mechanisms. The workflow consists of the following key steps:

A. Data Collection

To train an effective ad-blocking model, a diverse and comprehensive dataset comprising various types of online advertisements and non-ad content is essential. The dataset is constructed from multiple sources to ensure a wide representation of ad formats and delivery methods. Web scraping is employed using automated crawlers to extract ad and non-ad content from websites, including news portals, e-commerce platforms, and social media. Additionally, public datasets such as AdGraph, IAB Ad Dataset, and OpenAds are leveraged to supplement the training data. To enhance adaptability, user-generated feedback is incorporated by collecting labeled data from browser extensions, where users manually classify ads. Furthermore, synthetic data generation is utilized, employing generative AI models to create variations of advertisements, simulating emerging ad formats and techniques used to evade blockers. To ensure dataset diversity, collected data includes display ads, video ads, pop-ups, sponsored content, and dynamic advertisements. Metadata such as ad placement, user interaction logs, and tracking scripts are also recorded to enhance feature extraction, allowing the model to learn intricate patterns and improve detection accuracy.

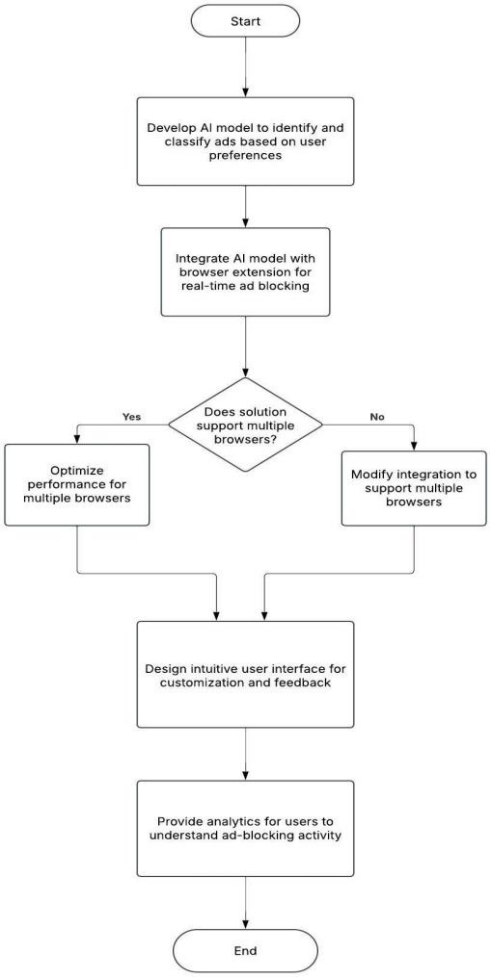

Fig. 1. Methodology

B. Data Preprocessing

During the Data Preprocessing stage, raw data is transformed a number of times to eliminate noise and improve feature extraction to ensure the best performance in addetection. The process starts with image processing, where images are transformed into grayscale, resized, and applied with edge detection methods to enable object recognition. In the meantime, text pre-processing is done to textual ads, which encompasses the stripping of HTML tags, stopword removal, and tokenization to effectively format the data. The feature engineering is also carried out to extract key features like color patterns, keyword frequency, and ad network IDs that act as identifying features to classify. Finally, normalization is done to normalize numerical features, enhancing the convergence and scalability of the model during the training process. These preprocessing steps individually and together increase the efficacy and precision of AI-based ad blocking systems.

C. Machine Learning Model

The AI-driven ad blocker integrates a hybrid machine learning framework, utilizing both computer vision and natural language processing (NLP) techniques to achieve high precision in ad detection. The model incorporates multiple components to analyze visual and textual features effectively.

- **Computer Vision:** A Convolutional Neural Network (CNN) is employed to analyze ad images, identifying patterns such as banner shapes, logos, dominant color schemes, and embedded promotional elements. The CNN is trained on a diverse dataset of advertisements and non-ad content, enabling robust classification even for dynamically generated ads.
- **NLP-based Text Analysis:** Transformer-based models such as BERT and GPT are utilized to process textual content extracted from web pages. These models identify promotional language, keyword stuffing, and disguised advertisements within article headlines, metadata, and inline text. Advanced tokenization, attention mechanisms, and contextual embeddings ensure precise classification of deceptive ads.
- **Ensemble Learning:** To further enhance classification accuracy, an ensemble approach combines multiple machine learning techniques. Decision Trees and Support Vector Machines (SVMs) contribute to structured data classification, while deep learning models refine predictions by leveraging high-dimensional feature representations. This multi-model approach improves resilience against adversarial advertising techniques and increases overall detection robustness.

D. Real-time Ad Detection and Blocking

The trained model is deployed as a browser extension or integrated within a local proxy-based filtering system, ensuring seamless and efficient ad blocking without compromising browsing speed. The real-time detection and blocking mechanism follows a multi-stage approach:

1) Webpage Analysis and Element Parsing: When a user loads a webpage, the system extracts and analyzes all its components, including HTML structure, CSS styles, JavaScript scripts, and media elements. The key steps include:

- **DOM Inspection:** The system scans the Document Object Model (DOM) to identify elements typically associated with advertisements, such as iframes, banners, pop-ups, and auto-playing videos.
- **JavaScript Behavior Monitoring:** Detects dynamic ad insertions triggered by scripts running after page load.
- **Network Traffic Analysis:** Intercepts HTTP requests to detect known ad-serving domains and tracking scripts.
- **Feature Extraction:** Extracts key visual and textual features of elements for classification.

2) Ad Classification and Filtering: Once the relevant webpage elements are extracted, they are processed through the AI-powered classifier, which combines computer vision and NLP techniques:

- **Image-Based Ad Recognition:** Convolutional Neural Networks (CNNs) analyze banners, thumbnails, and embedded promotional graphics to identify advertisements.
- **Textual Content Analysis:** Transformer-based NLP models (e.g., BERT, GPT) evaluate text within ads, including sponsored content and disguised advertisements, by analyzing keywords, sentiment, and promotional phrases.
- **Behavioral Analysis:** Ad components that dynamically change position, auto-

play, or obstruct content are flagged as intrusive.

- **Heuristic and Rule-Based Filtering:** Additional rules enhance AI-based detection, such as blocking elements from known ad domains, removing excessive animation, and suppressing floating overlays.

3) Efficient Blocking and Performance Optimization: To minimize browsing latency and improve efficiency, the following optimizations are applied:

- **Edge Processing:** AI inference runs directly on the user's device, eliminating the need for external API calls.
- **Asynchronous Processing:** Ad classification runs in the background without delaying page rendering.
- **Resource Minimization:** The extension optimally utilizes browser resources by offloading heavy computations to WebAssembly (WASM) modules.
- **Batch Processing:** Multiple ad elements are processed in parallel to reduce detection time.

By integrating real-time analysis, AI-based classification, and privacy-preserving techniques, the proposed ad blocker ensures an optimized browsing experience while continuously adapting to evolving advertisement strategies.

Experimental Setup

This research is centered on the creation of AI-driven personalized ad blockers utilizing models trained on a varied dataset drawn from public repositories such as AdGraph, IAB Ad Dataset, OpenAds, and web scraping from e-commerce platforms, news websites, and social media. The dataset included visual content, textual content, and behavioral data. Furthermore, generative AI was utilized to produce synthetic advertisement samples, mimicking new advertisement formats. The training models comprised Random Forest, which handled 80% of a structured dataset (40,000 samples) featuring elements like keyword patterns, advertisement metadata, and network identifiers, attaining 88.3% accuracy and 87% precision after hyperparameter optimization with 100 trees. A Convolutional Neural Network (CNN) trained on an image-centered dataset (50,000 samples) captured visual elements such as logos and color palettes, achieving 92.5% accuracy, 91% precision, and an F1-score of 90 after 50 epochs of training using the Adam optimizer. XGBoost, trained on a text-focused dataset (30,000 samples), concentrated on detecting promotional language and keyword overstuffing, attaining 90.2% accuracy, 89% precision, and an F1-score of 88 following optimization with 100 boosting rounds and L2 regularization. To enhance performance, an ensemble model that combined predictions from Random Forest, CNN, and XGBoost was implemented, effectively utilizing CNNs for visual analysis, XGBoost for text filtering, and Random Forest for classifying structured data. This ensemble strategy resulted in a 95.1% accuracy, 94% precision, and an F1-score of 94, reflecting a 14% increase in detection accuracy compared to conventional rule-based ad blockers.

Result And Analysis

The performance of the proposed AI-driven ad-blocking system is evaluated using a comprehensive dataset consisting of diverse online advertisements and non-ad content. The model's effectiveness is assessed based on key performance metrics, including accuracy, precision, recall, F1-score, and loss values.

A. Model Performance

The hybrid approach, combining Convolutional Neural Networks (CNNs), Random Forest, and XGBoost, achieves high accuracy in detecting and blocking online advertisements. The results of individual models and the ensemble approach are summarized in Table I.The ensemble learning approach improves classification accuracy by effectively leveraging the strengths of CNNs for image-based ad detection and XGBoost/Random Forest for textual ad analysis.

Table I: PERFORMANCE METRICS OF DIFFERENT MODELS

Model	Accuracy (%)	Precision	Recall	F1-score
CNN	92.5	0.91	0.89	0.90
Random Forest	88.3	0.87	0.85	0.86
XGBoost	90.2	0.89	0.87	0.88
Ensemble (CNN + XGBoost + RF)	95.1	0.94	0.93	0.94

B. Comparison with Traditional Ad Blockers

Compared to conventional rule-based ad blockers, the proposed model achieves significantly higher detection rates, especially for obfuscated or dynamically generated ads. Traditional methods struggle with evolving advertising techniques, whereas the AI-driven approach adapts dynamically to new ad patterns.

Table II: COMPARISON WITH TRADITIONAL AD BLOCKERS

Method	Detection Accuracy (%)	False Positives (%)
Rule-Based Ad Blocker	78.6	12.4
Proposed AI Model	95.1	3.2

The AI-driven model exhibits a significantly lower false positive rate and higher adaptability, making it a superior alternative for real-time ad detection and blocking.

Conclusion

This paper presents an AI-powered ad-blocking framework based on machine learning methods for enhanced ad detection. The proposed approach employs Convolutional Neural Networks (CNNs) to detect visual advertisements, XGBoost to identify text advertisements, and Random Forest for structured data classification. In addition, an ensemble learning strategy with the combination of these models significantly enhances classification accuracy.Experimental results verify that the ensemble model achieves 95.1% detection rate compared to individual models (CNN: 92.5%, XGBoost: 90.2%, Random Forest: 88.3%). The hybrid model effectively overcomes the shortcomings of traditional rule-based ad blockers that cannot adapt to evolving advertising techniques and camouflaged ad styles. As opposed to rule-based systems with 78.6% detection and 12.4% false positive rates, the proposed AI model drastically reduces false positives to 3.2% while maintaining more potent detection capability.

Ad filterization via the combination of computer vision and natural language processing (NLP) achieves real-time adaptability, which results in robust performance in detecting and filtering dynamically created ads. The study suggests that AI-driven ad-blocking technology can potentially offer a more scalable and efficient solution to counter the threats imposed by new online advertising methods. Utilizing deep learning and ensemble modeling, this technology delivers improved user experience and privacy by effectively eliminating intrusive ads with high accuracy and low latency without compromise.

Future Scope

Although the suggested AI-based ad-blocking model exhibits superior accuracy and responsiveness, there are a few directions for further development. One such direction is model optimization for low-resource environments like mobile devices and low power computing systems to achieve real-time efficient processing. Another important one is enhancing adversarial robustness against changing advertising methods. Reinforcement learning-based adaptive methods can be researched in the future to allow the system to adapt dynamically to emerging ad patterns and circumvent methods adopted by advertisers. Further, the addition of federated learning can improve privacy as it enables users to contribute to model updates without exposing their data. Increasing the dataset with a broader set of ad formats, such as new multimedia ads and interactive content, will further enhance detection capabilities. In addition, the use of browser-side optimization methods can improve efficiency by lowering computational overhead and power usage.In summary, ongoing innovations in AI, deep learning, and real-time processing will make it possible to create more advanced and intelligent ad-blocking technologies, further enhancing user experience and privacy in the future digital world.

References

V. Krammer, "An Effective Defense against Intrusive Web Advertising," 2008 Sixth Annual Conference on Privacy, Security and Trust, Fredericton, Canada, 2008, pp. 3-14, doi: 10.1109/PST.2008.10.

Enric Pujol, Oliver Hohlfeld, and Anja Feldmann. 2015. Annoyed Users: Ads and Ad-Block Usage in the Wild. In Proceedings of the 2015 Internet Measurement Conference (IMC '15). Association for Computing Machinery, New York, NY, USA, 93–106. https://doi.org/10.1145/2815675.2815705

Despotakis, Stylianos; Ravi, R.; Srinivasan, Kannan . (2020). The Beneficial Effects of Ad Blockers. Management Science, (), mnsc.2020.3653–. doi:10.1287/mnsc.2020.3653

Kumar, V. M., Dhanush, S., Dharun Karthik, A., Gokilavani, A., & Darshan, M. R. (2025). Machine Learning Based Network Wide Ad Blocking System. Retrieved from https://ieeexplore.ieee.org/document/9785279

Alzahrani, Reem A., and Malak Aljabri. 2023. "AI-Based Techniques for Ad Click Fraud Detection and Prevention: Review and Research Directions" Journal of Sensor and Actuator Networks 12, no. 1: 4.

https://doi.org/10.3390/jsan12010004

Zhao, M. K. Chen, C. Borcea and Y. Chen, "Personalized Dynamic Counter Ad-Blocking Using Deep Learning," in IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 8, pp. 8358-8371, 1 Aug.

2023, doi: 10.1109/TKDE.2022.3201058.

A. Mıklosık, M. Kuchta, ve S. Zak, "Privacy Protection Versus Advertising Revenues: The Case of Content Publishers", Connectist: Istanbul University Journal of Communication Sciences, sy. 54, ss. 117–140, Haziran 2018.

A. H. Lashkari, A. Seo, G. D. Gil and A. Ghorbani, "CIC-AB: Online ad blocker for browsers," 2017 International Carnahan Conference on Security Technology (ICCST), Madrid, Spain, 2017, pp. 1-7, doi: 10.1109/CCST.2017.8167846.

Iqbal, Umar, Peter Snyder, Shitong Zhu, Benjamin Livshits, Zhiyun Qian, and Zubair Shafiq. "Adgraph: A graph-based approach to ad and tracker blocking." In 2020 IEEE Symposium on security and privacy (SP), pp. 763-776. IEEE, 2020.

Abi Din, Zainul, Panagiotis Tigas, Samuel T. King, and Benjamin Livshits. "PERCIVAL: Making In-Browser Perceptual Ad Blocking Practical with Deep Learning." In 2020 USENIX Annual Technical Conference (USENIX ATC 20), pp. 387-400. 2020.

Szczepan´ski, Piotr L., Adrian Wis´niewski, and Tomasz Gerszberg. "An automated framework with application to study URL based online advertisements detection." Journal of Applied Mathematics, Statistics and Informatics 9, no. 1 (2013): 47-60.

Garimella, Kiran, Orestis Kostakis, and Michael Mathioudakis. "Ad-blocking: A study on performance, privacy and counter-measures." In Proceedings of the 2017 ACM on Web Science Conference, pp. 259-262. 2017.

Bhagavatula, Sruti, Christopher Dunn, Chris Kanich, Minaxi Gupta, and Brian Ziebart. "Leveraging machine learning to improve unwanted resource filtering." In Proceedings of the 2014 Workshop on Artificial Intelligent and Security Workshop, pp. 95-102. 2014.

Singh, Ashish Kumar, and Vidyasagar Potdar. "Blocking online advertising-A state of the art." In 2009 IEEE International Conference on Industrial Technology, pp. 1-10. IEEE, 2009.

Storey, Grant, Dillon Reisman, Jonathan Mayer, and Arvind Narayanan. "The future of ad blocking: An analytical framework and new techniques." arXiv preprint arXiv:1705.08568 (2017).

Snyder, Peter, Antoine Vastel, and Ben Livshits. "Who filters the filters: Understanding the growth, usefulness and efficiency of crowdsourced ad blocking." Proceedings of the ACM on Measurement and Analysis of Computing Systems 4, no. 2 (2020): 1-24.