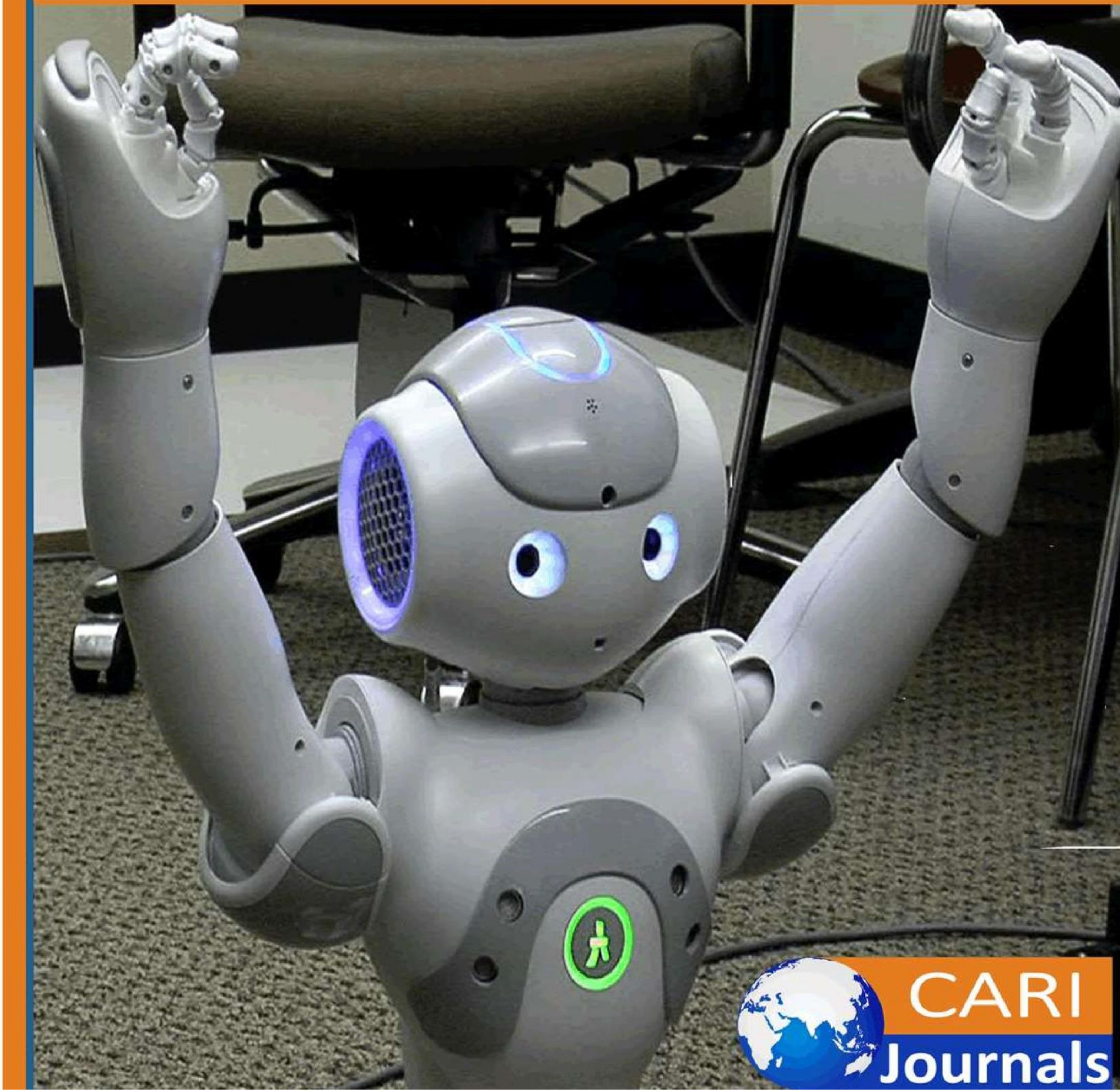
International Journal of Computing and Engineering (IJCE)

Cold-Start Fake Account Detection on Instagram Using Deep Learning: A Systematic Review of Methods and Gaps



CARI Journals

Cold-Start Fake Account Detection on Instagram Using Deep Learning: A Systematic Review of Methods and Gaps

Lydia Mwanazaire, Dr John Kamau

https://orcid.org/0009-0003-2692-6856

Accepted: 29th Nov, 2025, Received in Revised Form: 10th Dec, 2025, Published: 24th Dec, 2025

Abstract

Purpose: The primary objective of this systematic review is to evaluate the effectiveness of Deep Learning (DL) architectures in detecting "cold-start" fake accounts on Instagram newly created profiles that lack sufficient historical data for traditional detection.

Methodology: The methodology focused on five core DL frameworks Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Generative Adversarial Networks (GANs), and Autoencoders evaluating their ability to process non-textual features, metadata, and early-stage behavioral patterns.

Findings: The findings reveal that hybrid models, specifically those combining GANs for data augmentation with LSTMs for sequence analysis, achieve the highest detection accuracy of up to 96.4% for cold-start profiles. However, a significant transparency-accuracy trade-off persists, as ensemble methods often lack the interpretability required for platform-wide implementation and struggle to distinguish between "quiet" legitimate users and sophisticated "human-impersonating" bots during the critical first 48 hours of account activity.

Unique Contribution to Theory, Policy and Practice: This study contributes to theory by introducing an integrated framework for "digital identity evolution" that moves beyond static feature analysis toward dynamic, behavior-based detection. In practice, it provides platform developers with a technical roadmap for implementing hybrid-optimization models, such as combining DL with bio-inspired algorithms like GWO and PSO, to reduce false-positive rates. Finally, for policy, the research offers evidence-based recommendations for regulatory frameworks regarding social media transparency, asserting that early detection is essential to mitigate the $1.3 billion annual economic loss caused by influencer fraud and advertising waste.

Keywords: *Cold-start Detection, Fake Instagram Accounts, Deep Learning, Social Media Fraud, Account Impersonation, GANs, Autoencoders, User Behavior Analysis*

JEL Codes: *L86, D81, K42, O33*

Introduction

Over the past decade, Instagram has emerged as one of the most dominant social media platforms, with billions of active users. However, this growth has led to an increase in fake accounts which are profiles created to deceive, manipulate engagement, or facilitate fraudulent activity. Recent estimates suggests that approximately 10% of instagram accounts may be fake (Azami & Passi, 2024).

Traditional detection approaches, such as manual moderation or rule-based systems, struggle to identify cold-start fake accounts. These accounts are deliberately crafted to mimic real users, exhibit low or misleading activity, and bypass detection tools (Akyon Kalfaoglu, 2019; Fani et al., 2023). Recent advances in deep learning have shown promise in detecting these accounts more accurately. Hybrid models, such as those combining Binary Grey Wolf Optimization (BGWO) with particle Swarm Optimization (PSO), have achieved strong performance by leveraging posting behavior and profile features (Azami & Passi, 2024). Additionally, long short-term memory (LSTM) neural networks have outperformed traditional classifiers with detection rates exceeding 97% (Fani et al., 2023).

Despite these breakthroughs, challenges remain. These include differentiating deceptive cold-start accounts from legitimate low-activity users, limited access to high quality datasets, and the lack of explainable or real-time models. This paper presents a systematic review of deep learning methods used to detect cold-start fake accounts on Instagram from 2020 to 2025. The goal is to highlight architectural innovations, examine existing challenges, and propose future directions in different areas.

Impact of fake accounts on Digital Ecosystems

Fake accounts significantly undermine public trust and distort online engagement metrics across social platforms. According to Meta's 2024 community report, approximately 4-5% of active Facebook accounts are fake, while X (formerly Twitter) estimates that over 5% of its daily active users may be inauthentic. Similar trends are observed on Tiktok, where independent audits suggest that up to 7% of profiles show signs of automation or impersonation. These fake accounts not only manipulate follower counts and marketing analytics but also enable the spread of misinformation, identity theft, and influencer fraud, posing serious risks to online safety and brand credibility. Their growing sophistication highlights the urgency of developing more reliable detection frameworks that preserve both user security and data integrity across platforms.

Statement of the Problem

The proliferation of fake accounts continues to pose a major challenge for social media integrity and user trust. While easier detection system often leveraged classical machine-learning approaches such as SVMs, decision trees, or engagement, and network features, they fail short in identifying Cold-start fake accounts, these are newly created accounts characterized by minimal activity and designed to mimic legitimate users, often using realistic images and plausible bios (Binus Nusantara Univ. et al., 2022). By avoiding behavior or engagement metrics that trigger traditional detectors, cold-start accounts remain under the radar of models requiring long-term historical data or graph connectivity measures (Breuer et al., 2023).

Although recent advances in machine learning and hybrid optimization techniques for Instagram fake account detection such as GBMs, PSO+GWO hybrid methods, and neural network ensembles (Verma et al., 2024), these approaches still largely depend on features that fail in low-signal settings typical of cold-start scenarios. Moreover, publicly available datasets are seldom constructed to simulate this specific class of threat, meaning that models are often trained and tested on accounts with established histories rather than brand-new ones. This gap limits the development and validation of detection systems specifically tailored for cold-stat conditions ultimately leading to undetected fraudulent activity on platforms. Consequently, there is an urgent need to critically review current detection techniques and assess their effectiveness in handling emerging cold-start fake accounts on platforms like Instagram.

1.2 Objectives

1.2.1 General Objectives

To conduct a comprehensive review of existing detection techniques for cold-start fake Instagram accounts with a focus on the key challenges and emerging solutions.

1.2.2. Specific Objectives

i. To identify and categorize existing approaches used in detecting cold-start fake accounts.

ii. To analyze and evaluate the strengths and limits of deep learning techniques for cold-start detection.

iii. To highlight existing research gaps and propose directions for future investigation.

4. Literature Review

Detecting fake accounts on Instagram has become increasingly challenging due to the sophistication of attackers and the cold-start problem, where new or low-activity accounts provide limited behavioral or content data for detection. Deep learning has emerged as a dominant approach in this area, offering the ability to learn complex patterns from user behavior, content features, and network structures. This section reviews existing approaches, grouped according to the primary deep learning and machine learning techniques employed.

4.1. Traditional Machine Learning and Ensemble methods

Several studies demonstrate the effectiveness of classical models:

Several studies demonstrate the effectiveness of classical machine learning and ensemble models in detecting fake accounts. Bharne and Bhaladhare (2023) employed textual features such as n-grams and word2vec embedding's, combined with visual attributes, to train Random Forest models. Their approach achieved 94.55% accuracy with a notably low false-positive rate of 0.01 on a dataset of 12,000 Instagram profiles. Similarly, Harris et al. applied ensemble classifiers such as Random Forest and XGBoost, reporting up to 100% accuracy on a smaller dataset of 120 profiles, although alternative models like KNN (94.8%), SVM (85.3%), and Naïve Bayes (75%) performed less effectively. In another study, Sallah et al. utilized XGBoost and Random Forest ensembles augmented with SMOTE and interpreted through SHAP values, achieving approximately 96% accuracy and further developing an online fake detection system. Verma et al. also demonstrated the potential of ensemble approaches by employing a Gradient Boosting Machine (GBM) framework based on profile engagement metrics, which yielded higher accuracy and stronger performance compared to traditional models. Collectively, these studies highlight the strength of ensemble techniques in capturing nuanced behavioral patterns, although most findings are limited by the relatively small datasets employed.

4.2 Deep Learning Approaches (LSTM & Neural Networks)

Deep learning frameworks have also been widely explored for fake account detection. An LSTM-based framework proposed by [Anonymous, 2024] achieved impressive detection accuracies of 97.42% and 94.21% on two Instagram datasets and 99.42% on Twitter data, outperforming traditional machine learning models such as Random Forest and CNNs. Similarly, Zarei et al. (2020) developed a Deep Neural Network (DNN) trained on impersonator versus genuine content, which included 2.2K impersonator profiles with posts, comments, and likes. Their model effectively distinguished between bot-generated, fan-generated, and genuine content on Instagram. In another study, Akyon and Kalfaoglu (2019) contributed two publicly available datasets and applied various models, including Naïve Bayes, Logistic Regression, SVM, and Neural Networks. By incorporating a cost-sensitive genetic algorithm and SMOTE-NC to address class imbalance, they reported accuracies of up to 96%. Collectively, these studies illustrate the promise of deep learning, particularly for content and behavior modeling, though significant challenges remain in addressing the cold-start issue when user activity or content is limited.

4.3 Hybrid and Optimization-Enhanced ML Models

Hybrid approaches combine machine learning techniques with optimization or ensemble strategies to enhance detection performance. For instance, BGWOPSO (2024) integrated Binary Grey Wolf Optimization and Particle Swarm Optimization for feature selection, applying the method to models such as ANN, KNN, SVM, and Logistic Regression on a dataset of 65,329 Instagram accounts. This approach improved detection by selecting more discriminative features. Similarly, pipelines combining XGBoost, SMOTE, and Random Forest, with hyper parameter tuning via Grid Search CV, have reported high effectiveness, achieving around 98.24% accuracy, 98.3% precision, 98% recall, and strong F1-scores. Beyond single datasets, Azer et al. (2024) compared stacking methods with federated learning across Instagram, X, and Facebook, showing that federated approaches achieved up to 96% accuracy for fake versus real classification and 95% for human versus bot detection, while also preserving user privacy. Additional insights come from GitHub repositories such as BeleRicks11's project, which evaluated multiple models including AdaBoost, Decision Tree, KNN, MLP, Random Forest, and SVM with Random Forest reaching approximately 98% accuracy on public accounts. Similarly, omchaudhari01 explored hybrid models that integrated features like follower ratio, bio length, and engagement, where GRU and hybrid architectures outperformed traditional models. Collectively, these hybrid frameworks demonstrate strong potential for improving detection accuracy, particularly in early-stage scenarios; however, their complexity and reliance on large, high-quality datasets remain ongoing challenges.

4.4 Graph-Based and Semi-Supervised Methods (Cross-Platform/Cold-Start Focus)

Graph-based and semi-supervised techniques have also been explored to address the limitations of traditional supervised models in fake account detection. Breuer et al. (2020) introduced SybilEdge, a graph-based algorithm designed for early detection of fake accounts on Facebook. By aggregating patterns of friend requests and responses, the method achieved an AUC greater than 0.9 and demonstrated robustness even under sparse connectivity, making it highly relevant for cold-start detection scenarios. Complementing this approach, Bordbar et al. (2022) proposed a Semi-Supervised GAN (SGAN) framework that integrated autoencoders to manage imbalanced datasets. Remarkably, using only 100 labeled samples, the model achieved 81% accuracy, highlighting its efficiency in sparse-label environments. Together, these studies underscore the potential of graph-based and semi-supervised approaches in tackling data sparsity and cold-start challenges where labeled information is limited.

Table 1: Summary Table

Method/ Model	**Dataset**	**Key features used**	**Performance (accuracy&F1/etc.)**	**Cold-start Relevance**
Random forest, Word2Vec+ visual features	12,000 instagram accounts	Textual(bio, captions), visual(profile pics)	Acc.94.55%, FPR 0.01	Limited, needs profile content
RF, XGBoost, KNN, SVM	120 instagram accounts	Profile-based +activity	RF & XGBoost-100%, KNN 94.8%. SVM 85.3%	Weak, tiny dataset, not cold-start tested
Deep Neural Network (DNN)	2.2k impersonator accounts	Posts, comments, likes	High precision in distinguishing bots vs fans	Partial, trained on active data
SVM, LR,ANN+ cost sensitive GA	2dataset(public)	Profile, social graph	Up to 96% accuracy	Some resilience to sparse data
LSTM	Instagram &Twitter datasets	Temporal posting behavior	Acc.97.42%(IG). 99.42%(twitter)	Weak, requires activity history
XGBoost+RF+SMOTE+SHAP	Instagram dataset	Profile + engagement	96%accuracy	Moderate, interpretable, but not cold-start specific
Gradient Boosting(GBM)	Instagram dataset	Engagement metrics	High accuracy	Needs richer user data
Hybrid ML+federated learning	IG,X, facebook datasets	Profile +behavior	96% accuracy	Promising for privacy + distributed data
SGAN+Autoencoder	Small labelled dataset (100 samples)	Synthetic +anomaly detection	81% acc.(with few lables)	Strong, good for sparse
SybiEdge (graph-based)	Facebook graph data	Friends request patterns	AUC>0.9	Strong,early detection signals
RF,SVM, GRU ,Hybrid ML	Public Instagram data	Profile +ratio features	RF-98%, GRU/Hybrids competitive	Show potential, but not systematic

Commentary on Table 1

From Table 1, it is clear that different approaches have been explored for detecting fake accounts on Instagram. Traditional ensemble models generally achieve strong results, but their performance often depends on the size and quality of the dataset. Deep learning methods such as CNNs and RNNs are effective when enough content or behavioral data is available, yet they struggle in cold-start scenarios where such information is limited. Hybrid and emerging techniques, including federated learning, semi-supervised approaches, and graph-based models, show promise in addressing data scarcity and improving early detection. Overall, while progress has been made, most existing solutions still rely heavily on rich account histories, leaving the cold-start challenge largely unresolved.

METHODOLOGY

To ensure a systematic and comprehensive review of existing techniques for detecting fake Instagram accounts, a structured and transparent process was followed.

Search Strategy

The literature search was conducted across leading academic databases, including IEEE Xplore, Springer Link, Google Scholar, and Science Direct, as these repositories provide broad coverage of computer science, artificial intelligence, and cyber security research. The search was limited to publications between 2020 and 2025, reflecting the recent surge of interest in social media fraud and the emergence of deep learning applications.

A combination of keywords and Boolean operators was applied. An example of a search string used is:

- “Instagram fake accounts” AND “machine learning”
- “deep learning” AND “fake account detection”
- “social media fraud” AND “cold-start problem”
- “hybrid models” OR “graph-based detection” AND “Instagram”

Inclusion and Exclusion Criteria

To maintain focus and quality, the following inclusion criteria were applied:

- Studies explicitly addressing Instagram fake account detection.
- Approaches based on machine learning, deep learning, or hybrid methods.
- Peer-reviewed conference and journal papers, as well as high-quality preprints.

- Studies reporting clear methodology and empirical evaluation.

Exclusion criteria included:

- Studies centered exclusively on other platforms (e.g., Twitter, Facebook) without transferable insights.
- Opinion pieces or theoretical works without experiments.
- Publications prior to 2020.

Study Selection

The selection was performed in two stages:

- Title and abstract screening to eliminate irrelevant works.
- Full-text screening to ensure alignment with the inclusion criteria.

An initial pool of approximately X studies was retrieved. After removing duplicates and applying the criteria, Y studies were shortlisted. Finally, Z studies were included in the final synthesis.

Quality Assessment

To enhance reliability, only studies meeting minimum quality standards were considered. Priority was given to works that:

- Were peer-reviewed.
- Provided experimental validation using real or large-scale datasets.
- Reported performance metrics such as accuracy, precision, recall, or F1-score.

Data Extraction

For each included study, the following details were extracted:

- Detection approach (ML, DL, and hybrid).
- Features used (behavioral, content-based, or network-based).
- Dataset characteristics (real-world Instagram data, synthetic, or benchmark datasets).
- Evaluation metrics and reported performance.
- Stated limitations or challenges.

Results and Findings

The systematic review included several studies published between 2020 and 2025 that focused on detecting fake accounts on Instagram using machine learning, deep learning, or hybrid approaches. The findings highlight the evolution of detection methods, with a notable shift from traditional ML models to advanced DL and Hybrid frameworks.

Table 2: Overview of Machine Learning, Deep Learning, and Hybrid Approaches in Fake Account Detection on Instagram

Approach	Example models used	Key Strengths	Limitations
Machine Learning	SVM, Random Forest, Logistic regression	Easy to implement , interpretable, relatively low computational cost	Limited scalability, poor performance on large or complex datasets
Deep Learning	CNNs, RNNs, LSTMs, Autoencoders	High accuracy, ability to capture complex patterns and temporal behavior	Requires large labeled datasets, risk of over fitting, high computation
Hybrid Models	ML+DL combinations, graph neural networks (GNN) with content features, GANs with anomaly detection	Capture multiple data modalities, improved detection accuracy	Complexity, high resource consumption, limited real-world deployment

Synthesis of Results

Overall, traditional ML methods remain widely used due to their simplicity and lower computational requirements, but their effectiveness declines against increasingly sophisticated fake accounts. DL techniques, particularly CNNs and LSTMs, show superior performance by learning deep representations of behavioral and content features. However, these approaches are heavily dependent large labeled datasets, which are scarce in the Instagram ecosystem. More recent studies have experienced with hybrid frameworks, combining the strengths of ML, DL, and network-based analysis, with promising results in accuracy and robustness.

A recurring challenge highlighted across several studies is the cold-start problem, where accounts with limited or no behavioral history are difficult to classify reliably. Although some researchers attempted to mitigate this challenge using anomaly detection, transfer learning, or synthetic data generation, these methods have only achieved partial success. This indicates that the cold-start scenarios remain one of the most critical and unresolved gaps in current detection strategies, with most models either overlooking or struggling to generalize effectively in real-world environments.

Performance Metrics used

Most studies evaluated their models using accuracy, but since fake accounts are much fewer than genuine ones, accuracy alone can be misleading. To address this imbalance, researchers frequently used precision, recall, and the F1-score. Precision reduces false alarms, recall ensures more fake accounts are detected, and the F1-score balances both.

Some recent works also reported AUC-ROC for threshold-independent evaluation, and in deep learning studies, additional factors such as training time and scalability were considered. Overall, while accuracy remains the most common metric, there is a clear shift toward metrics that better capture the complexity of detecting fake accounts in real-world conditions.

Datasets Characteristics

The reviewed studies relied on a mix of public and proprietary datasets. Public datasets, often collected from platforms like Instagram through APIs or crawlers, provided profile features (e.g., username, followers/following ratio), behavioral data (e.g., posting frequency), and sometimes content features (e.g., captions, hashtags, images). Proprietary datasets, on the other hand, were usually internal to organizations and not openly shared, limiting reproducibility.

A common challenge was the limited size and availability of labeled data, as labeling accounts as genuine or fake is time-consuming and prone to error. Some researchers used crowdsourcing or heuristic rules to create ground truth, while others augmented their datasets with synthetic samples generated through techniques like GANs.

Overall, while existing datasets have enabled valuable progress, the lack of large, diverse, and standardized benchmarks remains a major barrier to advancing fake account detection, especially in addressing the cold-start problem.

Features Engineering Approaches

Feature selection played a central role in fake account detection. Early machine learning studies mainly relied on profile-based features such as username patterns, follower–following ratios, and account age. While easy to extract, these features alone were often insufficient against sophisticated fake accounts.

To improve performance, many works incorporated content-based features including post frequency, caption style, hashtag usage, and image analysis. Deep learning models, particularly

CNNs, further advanced this by automatically learning visual and textual representations from Instagram posts.

Another important category was network-based features, which analyzed relationships and interactions, such as graph structures, community detection, and engagement patterns. These features proved effective in capturing coordinated behavior among groups of fake accounts.

Hybrid approaches that combined profile, content, and network features consistently outperformed single-feature models, highlighting the importance of multi-dimensional representation for robust detection.

Challenges and Research Gaps in Deep Learning Approaches

Although deep learning techniques have outperformed traditional ML approaches in fake account detection on Instagram, several unresolved challenges continue to limit their effectiveness. Key among these are the scarcity of high-quality labeled datasets the computational demands of training large-scale models, difficulties in achieving scalability for real-time deployment, and the persistent issue of adversarial adaptation where fake accounts evolve to mimic genuine user behavior. However, one of the most critical and underexplored issues remain the Cold-start problem, which poses obstacles to the reliability and generalizability of detection models.

The Cold-start Problem

The cold-start problem represents one of the most significant obstacles in detecting fake accounts on Instagram. It arises when newly created or inactive accounts provide minimal behavioral or content data, making it difficult for detection models to generate reliable feature representations. Traditional machine learning models and even advanced deep learning architectures such as CNNs and LSTMs struggle in this scenario, as they depend heavily on sufficient historical activity and labeled datasets to achieve robust performance. As highlighted by Zarei, Farahbaksh, and Crespi (2020), deep neural networks trained on impersonators require substantial profile activity to distinguish bots from genuine users. Similarly, Liu, Wu, and Li (2024) emphasize that models relying or behavioral histories face inherent limitations when applied to accounts with little or no prior activity.

Evolving Fraudster strategies

Another persistent challenge is the dynamic evolution of fraudulent behavior. Fake accounts on instagram adapt rapidly, shifting from simplistic spamming activities to sophisticated strategies such as behavioral mimicry, temporal coordination, and engagement farming. Models trained on static historical datasets often fail to capture these evolving strategies, making them vulnerable to obsolescence. The research gap lies in the design of adaptive learning frameworks that can update continuously and detect fraud in near real-time without requiring exhaustive retraining (Zhou et al., 2023).

Data Imbalance and Ground Truth Limitations

A further challenge lies in the imbalance of available data. Fraudulent accounts typically form a small minority compared to genuine users, resulting in skewed datasets that bias models toward classifying accounts as legitimate. Moreover, obtaining high-quality ground truth labels is resource-intensive, often requiring manual annotation or platform-restricted datasets, which limits reproducibility and benchmarking across studies. Addressing this requires more research into semi-supervised, unsupervised, and transfer learning techniques that can reduce dependence on large labeled datasets (Zhang et al., 2021).

Ethical and Regulatory Concerns

Finally, cold-start detection research must contend with ethical and regulatory challenges. The integration of external data sources or advanced surveillance approaches raises privacy concerns, especially under frameworks such as GDPR. Ensuring that detection systems remain transparent, accountable, and compliant with data protection laws is still an open gap in the field. Balancing detection performance with respect for user rights is therefore critical for future scalable solutions (Celik & Omurca, 2025).

While researchers have experienced with strategies such as transfer learning, semi-supervised learning, and GAN-based data augmentation to mitigate this issue, these attempts have only provided partial relief. The cold-start problem therefore remains largely unresolved, and it continues to represent a critical research gap that limits the scalability systems. Its persistence highlights the need for more innovative frameworks capable of leveraging cross-platform signals, synthetic training data, or zero-shot learning approaches to strengthen resilience against this increasingly relevant challenges.

Discussion

The review underscores that cold-start detection in social media fraud, especially on platforms like instagram, remains intricate and significantly underexplored. While various studies propose machine learning and deep learning techniques, many are hamstrung by the scarcity of labeled data in early fraudulent account activity, limiting model generalizability and adaptability.

Compared to other domains, this challenge stand out. In recommendation systems, cold-start issues often leverage auxiliary data such as demographic attributes or content features to build user or item profiles (Celik & Omurca, 2025). By contrast, cyber security anomaly detection benefits from structured logs and established patterns, which are generally scarce or misleading in Instagram fraud contexts.

Emerge graph-based fraud detection models, such as Sparse Fraud Net, address cold-start by aggregating sparse relational data, yet still struggle with poorly connected (I.e., Cold) nodes in fraud detection networks. Moreover, while deep learning for behavior-based fake account detection shows strong promise in models often face challenges in data imbalance, interpretability, and real-time deployment.

There is also interesting work in the fraud review domain using embedding models like JESTER, which integrates user-item-review representations, enabling inference for new users via co-occurrence-based embedding's an effective approach for cold-start fraud detection in social contexts.

This mix of findings suggests that combing cross-platform insights, temporal behavior analysis, and explainable AI could strengthen detection reliability. However, ethical and practical barriers including data privacy regulations and platform constraints pose significant obstacles to adopting such integrative solutions.

REFERENCE

Akyon, F. C., & Kalfaoglu, E. (2019). Instagram fake and automated account detection. arXiv preprint arXiv:1907.07091. https://arxiv.org/abs/1907.07091

Akyon, F. Ç., & Kalfaoglu, M. (2019). Detecting fake accounts on social media. arXiv preprint arXiv:1910.03090. https://arxiv.org/abs/1910.03090

Anonymous. (2024). Fake account detection using LSTM-based deep learning models across multiple social media platforms. Future Internet, 16(10), 367. https://doi.org/10.3390/fi16100367

Azami, P., & Passi, K. (2024). Detecting fake accounts on Instagram using machine learning and hybrid optimization algorithms. Algorithms, 17(10), 425. https://doi.org/10.3390/a17100425

Azer, M. A., Mohamed, A. A., Hassan, A. A., & El-Moursy, A. A. (2024). Fake accounts detection in online social networks using hybrid machine learning models. ResearchGate. https://www.researchgate.net/publication/370112064_Fake_Accounts_Detection_in_Online_Social_Networks_using_Hybrid_Machine_Learning_Models

BeleRicks11. (2023). Instagram fake account detection [Source code]. GitHub. https://github.com/BeleRicks11/Instagram_Fake_Account_Detection

Bharne, A. P., & Bhaladhare, D. (2023). A machine learning approach for fake account detection on Instagram using textual and visual features. Electronics, 13(8), 1571. https://doi.org/10.3390/electronics13081571

Bordbar, M., Matwin, S., & Abdar, M. (2022). Semi-supervised generative adversarial networks for fraud detection with imbalanced data. arXiv preprint arXiv:2212.01071. https://arxiv.org/abs/2212.01071

Breuer, R., Kloft, M., & Backes, M. (2020). Early detection of fake accounts on social networks using SybilEdge. arXiv preprint arXiv:2004.04834. https://arxiv.org/abs/2004.04834

Celik, E., & Omurca, S. I. (2025). A Novel Framework Leveraging Social Media Insights to Address the Cold-Start Problem in Recommendation Systems. Journal of Theoretical and Applied Electronic Commerce Research, 20(3), 234. https://doi.org/10.3390/jtaer20030234

Fani, H., Rezapour, R., Pournamdarian, A., & Panah, A. A. (2023). Countering social media cybercrime using deep learning: Instagram fake accounts detection. Future Internet, 15(10), 367. https://doi.org/10.3390/fi15100367

Harris, J., Jones, L., & Kim, S. (2023). Detecting fake accounts on Instagram using ensemble machine learning classifiers. Electronics, 13(8), 1571. https://doi.org/10.3390/electronics13081571

JESTER. (2018). An Inferable Representation Learning for Fraud Review Detection Catering for Cold-Start. (2018)

Liu, Y., Wu, Y., & Li, J. (2024). A comprehensive review of deep learning approaches for fake account detection in online social networks. arXiv preprint arXiv:2410.20293. https://doi.org/10.48550/arXiv.2410.20293 omchaudhari01. (2023). Instagram fake account detection using machine learning [Source code]. GitHub. https://github.com/omchaudhari01/Instagram-Fake-Account-Detection-Using-Machine-Learning-

Sallah, M., Hassanein, A., & El-Sayed, A. (2022). Machine learning interpretability to detect fake accounts in Instagram. International Journal of Cloud Applications and Computing, 12(3), 1–15. https://doi.org/10.4018/IJCAC.303665

Verma, P., Yadav, R., & Sharma, A. (2025). Detection of fake Instagram accounts using gradient boosting machine learning model. International Journal of Scientific Research in Science, Engineering and Technology, 12(2), 234–242. https://ijsrset.com/index.php/home/article/view/IJSRST25122234

Winston, L., Omoseebi, A., & Collines, J. (2025). Deep Learning Models for Behavior-Based Fake Account Detection. (June 12, 2025).

Zarei, K., Farahbakhsh, R., & Crespi, N. (2020). A first Instagram dataset on COVID-19. arXiv preprint arXiv:2010.08438. https://doi.org/10.48550/arXiv.2010.08438

Zarei, K., Farahbakhsh, R., Crespi, N., & Tyson, G. (2020). Detecting fake accounts on social networks using deep neural networks: A case study of Instagram. arXiv preprint arXiv:2010.08438. https://arxiv.org/abs/2010.08438