MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 50-55

A Hybrid Deep Learning Framework for Suspicious Profile Detection: Integrating Structured Metadata and Sequential Engagement Patterns

Ms Sumitra Menaria,
Research Scholar, Computer Engineering Department,
Gujarat Technological University,
Ahmedabad, Gujarat, India, sumitra.menaria@gmail.com,
Dr Viral H Borisagar
Associate Professor, Computer Engineering Department,
Vishwakarma Government Engineering College,
Chandkheda, Ahmedabad, Gujarat, India, viralborisagar@yahoo.com
* Corresponding author: Ms. Sumitra Menaria, sumitra.menaria@gmail.com,

Abstract
The swift rise in suspicious and fraudulent accounts on Instagram makes it necessary to use sophisticated detection strategies as opposed to traditional unimodal detection strategies. This paper introduces a sequential modeling platform that will be used to systematically estimate the value of various data modalities in the detection of suspicious profiles. The experimental design is planned in phases to determine the baseline performance and also to determine the limitations of each methodology.In Experiment 1, a metadata-only Dense Feedforward Neural Network (FNN) with a 78% accuracy rate but no contextual and behavioral depth was used. Experiment 2, with temporal dynamics run through Long Short-Term Memory (LSTM) networks, increases accuracy to 82 percent by taking into account sequential engagement patterns, such as interaction spikes. Experiment 3 uses a hybrid late fusion model to integrate structural metadata with behavioral features, which significantly increases the performance to 87% accuracy and an F1-score of 85%.These results indicate that temporal behavior models are more effective than the use of static metadata analysis, and that the multimodal fusion approach is more effective in terms of classification robustness and false negative minimization, which is a critical security requirement in a platform. In spite of these enhancements, the hybrid model does not have textual and visual semantic comprehension, which limits its overall performance.They give solid empirical reasons supporting the formulation of a multimodal framework with textual and visual elements, which is suggested as the next step of the current study.

Keywords: Multimodal Deep Learning, Instagram Fake Profile Detection, Adversarial AI, Class Imbalance, Sequential Engagement Modeling

Introduction

With the burst of Online Social Networks (OSNs), the digital environment has been turned into a rich ecosystem of communication, business, and social interaction [1]. Instagram, Facebook, and X (previously Twitter) are platforms that have developed as complex socio-technical systems, which facilitate economic and social interactions and the construction of social identities [1].In these systems, user credibility is increasingly measured by quantifiable metrics—followers, engagement rates, and verification status—turning digital reputation into a measurable economic and social asset [2,3].

But this fast development has brought in systematic weaknesses, that is, in the form of suspicious and fake user profiles [1]. These profiles abuse platform metrics to intervene in a system of trust, support malicious activities, and propagate misinformation [1]. The modern social media environments have ceased to be chronological feeds to being predictive, algorithmic content curation that aims at maximizing user retention, which is exploited by bad actors through coordinated inauthentic behavior [4,5]. In contrast to early spam-based bots, modern suspicious profiles are based on adversarial AI, synthetic media generation, and realistic behavioral models, rendering them much harder to detect [6,7].False profiles are divided into various types depending on their complexity, with examples including automated bots simulating rhythms of human interaction, cyborg accounts where automated scripts are overseen by humans, impersonation profiles, and organized inauthentic networks that amplify narratives to manipulate trending algorithms [8 - 10]. The effects of such accounts are far reaching as they lead to distortion in the economic marketplace, cyber security issues like phishing and erosion of social trust [11 - 13].One of the most important issues of the present day security research is that most of the existing detection systems are based on unimodal analysis that is, focusing on a single type of feature such as metadata or text, instead of cross-modal consistency [14]. Moreover, the mass-generation of synthetic identities with authoritative captions and profile images have made it easier to blur the distinction between authentic and fake accounts in the years 2022-2025 with the democratization of generative AI[15,16].

To address these challenges, there is an urgent need for scalable, imbalance-sensitive multimodal deep learning frameworks [17]. Such systems must move beyond simple classification to integrate heterogeneous data sources, including structured metadata (e.g., follower-following ratios), temporal engagement patterns (e.g., posting frequency), textual semantics (e.g., biography and captions), and visual artifacts [14, 17]. This research proposes a unified architecture designed to detect cross-modal inconsistencies and antagonistic patterns of adaptation, thereby enhancing digital trust and security in contemporary social networking environments [18, 19],.I have initiated the creation of an infographic that visualizes the Online Social Network ecosystem, the various types of fake profiles, and the core components of the proposed multimodal detection framework.

2. Literature Review

The rapid proliferation of Online Social Networks (OSNs) has necessitated advanced research into the automated identification of fake and suspicious profiles. Over the last decade, detection methodologies have transitioned from simple heuristic rule-based filtering to sophisticated Machine Learning (ML) classification and, most recently, to deep learning-based multimodal architectures [1,2]. Despite these advancements, detecting fraudulent digital identities remains a dynamic challenge as adversarial strategies evolve[1]. This section provides a systematic review of the five major paradigms in detection research: metadata-based, behavioral, text-based, image-based, and multimodal fusion frameworks.

2.1 Evolution of Detection Paradigms: Initial detection systems relied heavily on manually crafted thresholds for profile features such as follower counts and interaction ratios[2]. While computationally efficient, these systems were highly vulnerable to adversarial manipulation. The integration of ML brought about supervised classifiers—including Logistic Regression (LR), Random Forest (RF), and Support Vector Machines (SVM)—which offered better generalization but remained dependent on structured feature engineering[3-5].

The introduction of Deep Learning (DL) architectures significantly enhanced representational power. Long Short-Term Memory (LSTM) networks enabled the modeling of temporal engagement, while transformer-based Natural Language Processing (NLP) models identified semantic inconsistencies in profile descriptions[5-7]. Current research now emphasizes the rising sophistication of adversarial accounts that utilize AI-generated content and disjointed engagement strategies[8-10].

2.2 Metadata-Based Detection Approaches: Metadata-based techniques utilize structured profile attributes—such as follower-following ratios, account age, and posting frequency—to distinguish between authentic and suspicious behavior[11]. These models typically employ classical ML classifiers like RF, SVM, and Gradient Boosting Machines (GBM).Recent studies (2022–2025) have focused on optimizing these models for structured data. For instance, research has shown that GBM-based models can outperform basic classifiers when analyzing engagement metrics[12], while hybrid optimization techniques have improved precision and reduced false alarms. However, metadata-based approaches are limited by the ease with which these attributes can be manipulated by bad actors[14]. Furthermore, they often lack the textual, temporal, and visual context necessary to detect advanced or AI-generated profiles[11,14].

2.3 Behavioral and Engagement Modeling: Behavioral modeling shifts the focus from static profile features to time-based patterns of user interaction, such as posting bursts and session clickstreams. Recurrent Neural Networks (RNNs), specifically LSTMs, have demonstrated high efficacy in capturing the temporal dynamics that differentiate human interaction from automated bot rhythms [15, 16].

Recent developments include hybrid sequence models combining LSTM and Gated Recurrent Units (GRU) to better detect complex temporal anomalies [16, 17]. Unsupervised sequence models, such as LSTM autoencoders, have also been explored for anomaly detection without the need for extensive labeled data [18, 19]. Despite their success, these models face challenges when bots successfully mimic human-like engagement rhythms, and they often lack integration with multimodal semantics [16, 20].

2.4 Text-Based Detection Using NLP: NLP-based detection leverages semantic clues within user-generated text—including biographies, captions, and hashtags—to identify deceptive accounts. The emergence of transformer-based models like BERT, RoBERTa, and XLNet has allowed for deep contextual analysis of language patterns and emotional aberrations [21, 22].

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 50-55

Lightweight variants like DistilBERT have been proposed to reduce computational overhead [23, 24]. While these models are effective at detecting linguistic anomalies, text-only systems are intrinsically limited by their lack of behavioral and visual cues[25]. This is particularly problematic against adversarial profiles that adopt a convincing linguistic style but maintain fraudulent activity in other domains.

2.5 Image-Based Detection Strategies: Image-based systems apply computer vision to detect artifacts in profile pictures and media posted that are indicative of automated generation or manipulation. The prevailing structure of this paradigm is Deep Convolutional Neural Networks (CNNs), which are used to extract fine-grained visual clues that distinguish genuine images and artificial (e.g., GAN-generated) images [25, 26].

Recent studies have investigated the combination of image analysis with metadata and NLP characteristics to enhance accuracy [27, 18]. Nevertheless, the growing realism of generative AI is a major challenge to image-only detection. In addition, the field experiences the lack of large-scale, dedicated datasets directly on fake profile images, which impedes the standardized benchmarking [29].

2.6 Multimodal Fusion Frameworks: Multimodal systems overcome the limitations of single-modal systems by incorporating disparate data streams, including text, images, metadata, and behavior, into a single system. Initial efforts to be multimodal were based on basic feature concatenation, which in most cases was unable to model profound semantic associations among data types [31].

Recent trends (20222025) have moved on to attention-based and transformer-based cross-modal alignment. The 2021 use of Vision-Language Models (LVLMs) and CLIP-based encoders to detect inconsistencies between a profile textual narrative and its visual representations have been significant. Similarity maximization techniques to learn to coherently map modalities and penalize discrepancies have also been proposed [34, 35].

2.7 Synthesis and Research Gaps Identified.

Although the set of heuristic rules has been replaced by multimodal deep learning, there are some critical gaps in research:

- Over-Reliance on Unimodal Detection: There is still a tendency to focus on a single domain, and it is unable to detect the cross-modal inconsistencies that are characteristic of advanced fake profiles [36].
- Shallow Fusion Approaches: Most multimodal systems do not have deep semantic convergence, but operate on simple combination of features [37].
- Weak Focus on Class Imbalance: In reality, the data is highly imbalanced, but many models are concerned with the overall accuracy of the model, regardless of whether it recalls minority-class suspects [37].
- Inadequate False-Negative Analysis: There is insufficient systematic analysis as regards to false-negatives in security sensitive environments [38].
- Scalability, Dataset Scarcity: A small number of studies justify performance on large-scale datasets or give an evaluation on Instagram-specific multimodal data [39]].

It is these gaps that give the immediate push behind the proposed research, which is to develop a scalable, imbalance-conscious, and multimodal framework that is capable of integrating structured metadata, temporal behavior.

3. Proposed Method: The study uses an experimental approach based on a progressive modeling strategy that incrementally incorporates different modalities of Instagram profile data. The systematic methodology allows a stepwise assessment of the contribution of each data stream to the identification of suspicious profiles. The research avoids complex multimodal architectures at first and instead uses simpler unimodal and hybrid models, which has several advantages: it allows setting baseline performance benchmarks, clarifies the contribution of individual feature modalities, identifies weaknesses of unimodal detection strategies and provides empirical support for multimodal fusion.

Dataset and Preprocessing: The experiments utilized a curated dataset of 10,000 Instagram profiles (7,500 Trustworthy and 2,500 Suspicious), manually labeled for binary classification. The dataset was split in a stratified manner: 70% training, 15% validation, and 15% testing.

Preprocessing steps included:

Metadata: Feature scaling (StandardScaler), creation of derived features (follower-following ratio, engagement rate).

Sequential Engagement: Extraction of the most recent 9 posts per profile, normalization of timestamps, and conversion into fixed-length sequences.

Class imbalance handling: Application of SMOTE oversampling on the training set and class-weighted loss during training.

The experimental pipeline proceeds in four main stages:

1. Experiment 1 – Detection Using Metadata: This phase is based on only structured profile attributes to estimate the baseline performance.
2. Experiment 2 - Simulating the Engagement Sequence: In this experiment, engagement is evaluated through the temporal interaction patterns modelled by Long Short-Term Memory (LSTM) networks.
3. Experiment 3 – Hybrid Fusion Model: This step fuses metadata and engagement features with a late fusion architecture.

The first three experiments are baseline evaluations and the fourth experiment introduces the proposed multimodal detection framework. As the performance evaluation metrics, we use accuracy, precision, recall, F1-score and false negative count for consistent comparison among these experiments.

Among these metrics, recall is particularly important, as undetected suspicious profiles pose significant risks to platform integrity.

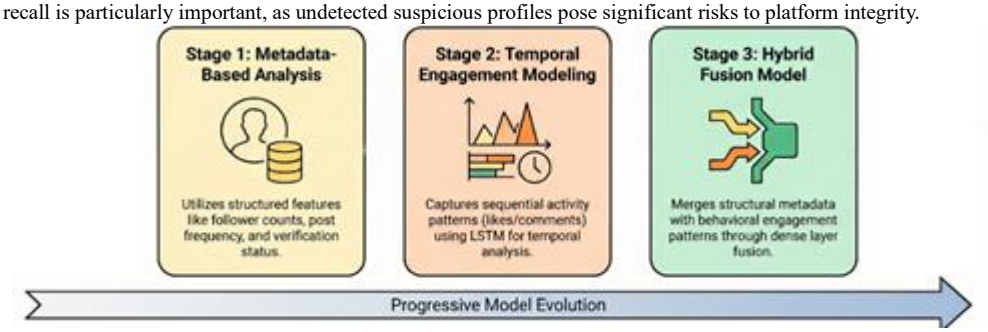

Figure 4.1 Experimental Evolution Framework for Suspicious Profile Detection

The figure shows the gradual experimental advance of the detection structure. The evolution starts with a metadata-based base model, which is then followed by sequential engagement modelling based on LSTM networks, a hybrid fusion model between structural and behavioural features, and the last technology is a multimodal framework that incorporates metadata, engagement behaviour, textual semantics, and visual embeddings.

The experimental design utilized in the study is progressive and modular-based where each step is characterized with a progressive increase in the complexity of the model. The design is specially designed to assess the contributions of individual and combinatorial form of various feature modalities in the detection of suspicious profiles.The paper does not attempt to apply directly to a complex multimodal framework and instead starts with a simple baseline model to add more dimensions of features over time. This gradual development allows one to see how each of the modalities of metadata, behavioural engagement and multimodal content contributes towards the overall performance of detection.The initial experiment is based on metadata-based classification, based on structured profile attributes, a baseline performance is determined. Although computationally efficient, this model does not have contextual understanding and behavioural understanding.The second experiment presents engagement-based modeling with LSTM networks, which attends to dynamic patterns of interaction (likes, comments, and posting behavior). This phase improves detection by adding dynamic user activity.The third experiment combines metadata and engagement features based on a hybrid fusion. The combination minimizes false negatives because it makes use of structural and behavioral information. Lastly, the fourth experiment hypothesizes that there could be a purely multimodal approach that integrates text semantics and visual characteristics on top of metadata and behavior indicators. This step identifies cross-modal discrepancies, and it can be trained to identify advanced and AI-generated bogus accounts very effectively.This is a progressive experimental design that effectively allows systematic assessment of model improvement at a stage and the development of multimodal learning is related to unimodal learning. The Figure 4.1 shows that the successive stages are based on the constraints of the last one, which leads to a better detection capacity, higher recall, and lower false-negative rates.The result of such a structured evolution not only confirms the effectiveness of multimodal integration, but also makes it interpretable, isolating the importance of each of the feature domains.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 50-55

4 Overview of Experimental Design

4.1 Experiment 1: Metadata-Based Model

The initial experimental phase is devoted to the assessment of the efficiency of structured metadata features to detect suspicious Instagram profiles. Metadata features are the attributes that are part of the user profile and that are generally used in traditional systems of fake account detection.

Examples of such features include:

- Number of followers
- Number of accounts followed
- Total posts
- Verification status
- Business account indicator
- Privacy status

These characteristics help to send convenient messages based on account authenticity because suspect accounts can frequently have a disproportionate follower-following ratio, strange posting behaviour, or lack of verification indicators. However, metadata features are inherently static and easily manipulable, making them insufficient for detecting sophisticated fake accounts. The metadata-based classifier is implemented using a Dense Feedforward Neural Network (FNN).

Architecture configuration:

Input Layer: Metadata feature vector
Hidden Layers: Fully connected layers with ReLU activation
Regularization: Dropout layers to prevent overfitting
Output Layer: Sigmoid activation for binary classification

Table 4.1 Metadata Features Used for Baseline Model

Feature	Description
followersCount	Number of followers
followsCount	Number of accounts followed
postsCount	Total number of posts
igtvVideoCount	Number of IGTV videos
highlightReelCount	Number of highlight reels
isBusinessAccount	Business account indicator
Verified	Verification status
Private	Privacy indicator
businessCategoryName	Business category

The structure of the architecture shown below depicts the baseline classification model where the structured metadata of Instagram profiles is used as input features. The metadata vector is then fed through fully connected dense layers which is powered by ReLU activation, and regularization with dropout is then applied to minimize overfitting. The last output layer employs a sigmoid activation function to carry out binary classification, and this defines profiles as trustworthy or suspicious.

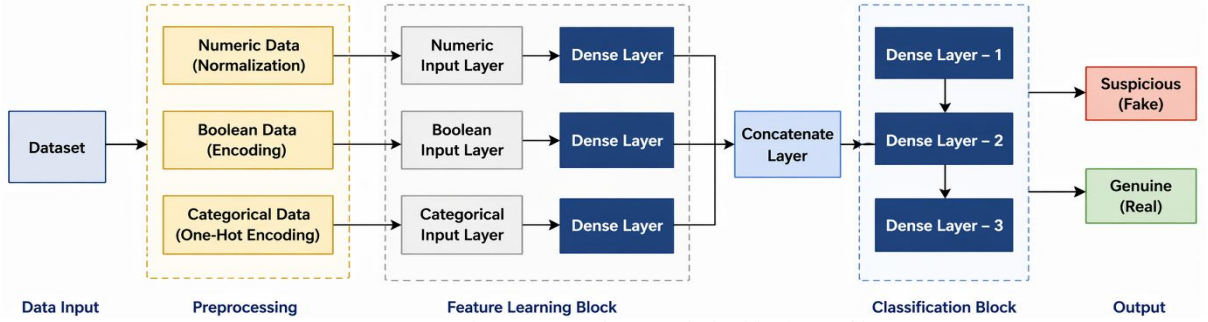

Figure 4.2 Metadata-Based Classification Architecture

The metadata-only model establishes the baseline performance for suspicious profile detection.

Table 4.2 Performance of Metadata-Based Model

Metric	Value
Accuracy	78%
Precision	76%
Recall	74%
F1-score	75%

The findings show that metadata characteristics are able to detect moderately. The model is effective at detecting some of the structural anomalies existing in suspicious accounts. The following limitations are however observed:

- Metadata features can be easily manipulated by adversaries.
- The model fails to capture behavioural dynamics.
- Content-level signals are ignored.

Such constraints drive the necessity of the addition of the temporal engagement behaviour into the following experimental phase.

4.2 Experiment 2: LSTM Engagement Model

The second experimental phase entails the modelling of temporal engagement behaviour by sequential deep learning structures.

Whereas metadata are some of the attributes of a statical nature, the engagement characteristics are of a dynamic nature as they are the patterns of user interaction in the form of likes, comments, and the time of posting.

Suspicious accounts often exhibit abnormal interaction behaviors including:

- Engagement spikes
- Repetitive interaction patterns
- Artificially inflated engagement metrics

Sequential modeling enables one to capture these patterns.

Input Features

Each profile contains engagement metrics extracted from the most recent nine posts:

- *likesCount_i*
- *commentsCount_i*
- *timestamp_i*

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 50-55

These features are arranged chronologically to form engagement sequences.
Model Architecture
The engagement model is implemented using a Long Short-Term Memory (LSTM) network.
Architecture components:
Input Layer: Sequential engagement tensor
LSTM Layer: Captures temporal interaction patterns
Dense Layer: Feature projection
Output Layer: Binary classification using sigmoid activation

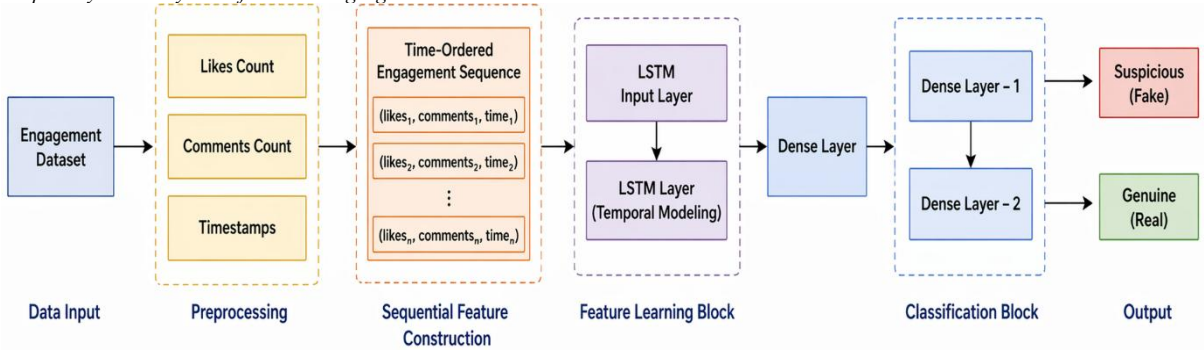

Figure 4.3 LSTM-Based Engagement Modeling Architecture for Suspicious Profile Detection

This architecture is a model of sequential deep learning of temporal engagement behavior across Instagram posts. The engagement data such as likes, comments, and times of posting data are arranged in chronological order and fed through LSTM network to learn interaction patterns over time. This obtained sequential representation is then subjected to a dense layer and ultimately classified with the help of a sigmoid activation function to determine the profiles as credible or suspicious.

Table 4.3 Performance of Engagement-Based Model

Metric	Value
Accuracy	82%
Precision	81%
Recall	79%
F1-score	80%

Engagement-based modeling presents better detection performance as compared to metadata-only modeling. The LSTM network is able to capture the patterns of temporal interactions and anomalies of engagement.

However, several limitations remain:

- Advanced bots can simulate realistic engagement behavior.
- Sequential modeling fails to take semantic content into consideration.
- Visual information remains unused.

These observations motivate combining metadata and engagement features in a hybrid architecture.

4.3 Experiment 3: Hybrid Fusion Model

The third experiment is a phase where the hybrid architecture incorporating metadata and engagement capabilities is brought in.

The motive of hybrid modeling is that suspicious profiles have a tendency of showing inconsistency in more than one type of feature. The model can be used to define more anomalies through the integration of behavioral and structural indicators.

The hybrid model consists of two parallel processing branches, metadata branch processes structured profile attributes using dense neural layers. Engagement branch processes temporal engagement sequences using an LSTM network. Both branches generate feature embeddings that are merged through a fusion layer.

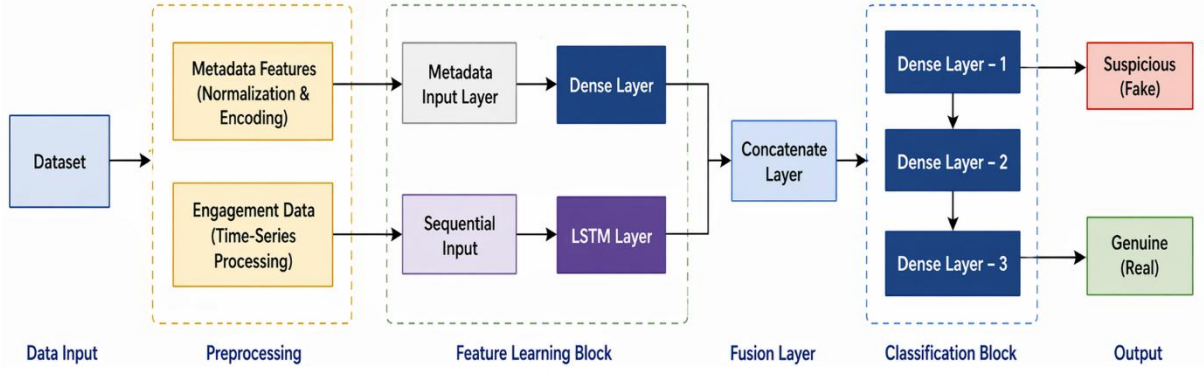

Figure 4.4 Hybrid Metadata–Engagement Fusion Architecture

The architecture is a hybrid between structured metadata capabilities and temporal engagement sequences in order to enhance suspicious profile detection. The metadata features are trained on a dense neural network, whereas engagement patterns are trained on an LSTM network to learn temporal interaction patterns. A fusion layer between the feature representations of the two branches then move to a final classifier, which gives an indication on whether a profile is trustworthy or suspicious.Late fusion is applied by concatenating embeddings from both branches before passing them to the final classification layer.

Table 4.4. Performance of Hybrid Model

Metric	Value
Accuracy	87%
Precision	86%
Recall	84%
F1-score	85%

The hybrid model has a high level of classification that is much better than unimodal models. The integration of metadata and engagement features allows the model to detect inconsistencies between structural and behavioral attributes.

4.4 Comparative Analysis of Baselines

In order to assess the effectiveness of every stage of the experiment, the baseline models are described comparatively.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 50-55

Table 4.5 Performance Comparison across Experiments

Experiment	Modalities Used	Accuracy	F1-score
Exp 1	Metadata	78%	75%
Exp 2	Engagement	82%	80%
Exp 3	Metadata + Engagement	87%	85%

The figure compares the effectiveness of the detection in three experimental settings, namely metadata-based classification, sequence modeling of engagement with LSTM and the hybrid fusion model combining the two modalities. The findings show how the performance increases gradually with the inclusion of more behavioural information, whereby the hybrid model has the best accuracy and F1-score. This tendency indicates the advantage of combining several feature modalities in suspicious profile detection.

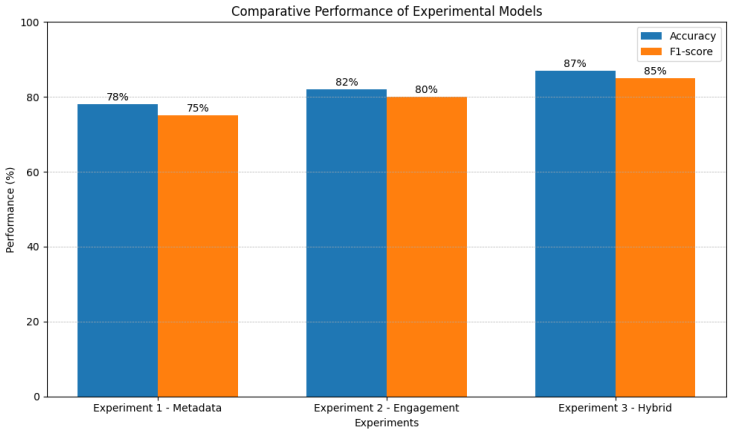

Figure 4.5 Comparative Performance across Experimental Stages

5. Conclusion:

The experimental evolution presented in this study demonstrates a clear performance trajectory, proving that the integration of diverse data modalities significantly enhances the detection of suspicious Instagram profiles. Initial benchmarks established that static metadata-based models, while computationally efficient, achieve a limited accuracy of 78% due to their vulnerability to adversarial manipulation. However, the transition to sequential engagement modeling using LSTM networks improved accuracy to 82%, confirming that temporal interaction patterns are superior to static attributes in identifying behavioral anomalies. The hybrid fusion of these two modalities further elevated accuracy to 87%, illustrating that combining structural and behavioral indicators provides a more robust defense against modern fraudulent accounts and progressively reduces false negatives

6. Future Scope

Though it is true that hybrid models have gained more in terms of performance, it is indisputable that they have very serious limitations in their inability to conduct deep textual and visual semantic analysis. Nowadays, suspicious accounts tend to use repetitious promotional messages, automated messages, and AI generated or stock images that circumvent systems that prioritize metadata and engagement rhythms. Also, unimodal and hybrid architectures do not detect cross-modal discrepancies, where a profile may maintain a realistic follower-following ratio and history of interactions but at the same time deploy misleading or inconsistent content both in text and image modalities. This does not provide sufficient scrutiny of the content of the advanced accounts of adversarial, and, therefore, the advanced adversarial account does not appear, even when the indicators of its behavior are organic. Future studies ought to look beyond the current binary classification to the multi-class classification, differentiating between certain types of malicious actors such as automated bots, cyborgs, and coordinated inauthentic networks. The urgent requirement is also to explore cross-platform detection and real-time system integration to empower the scalability and practical implementation of these frameworks in live production set-ups. Lastly, as generative AI continues to democratize advanced tools of deception, future versions will need to refine adversarial robustness and adaptive learning mechanisms to ensure high accuracy in detecting more sophisticated deception tools and more realistic synthetic identities and more dynamic behavioral camouflage.

References:

1. Terumalasetti, S., & Reeja, S. R. (2024). Enhancing Social Media User's Trust: A comprehensive framework for detecting malicious profiles using Multi-Dimensional Analytics. IEEE Access, 13, 7071–7093. https://doi.org/10.1109/access.2024.3521951
2. Button, M., Shepherd, D., & Jung, J. (2025). Economic espionage via fake social media profiles in the UK: professional workers awareness and resilience. Security Journal, 38(1). https://doi.org/10.1057/s41284-025-00476-2
3. Sarfraz, A., Ahmad, A., Zeshan, F., Hamid, M., & Alshalali, T. a. N. (2025). Unmasking deception: detection of fake profiles in online social ecosystems. Journal of Big Data, 12(1). https://doi.org/10.1186/s40537-025-01254-y
4. Fulzele, P., Jiss, M. M., Fulzele, S., & Das, T. (2025). LIMFADD: LLM-enabled Instagram Multi-Class Fake Account Detection Dataset. TechRxiv, 1–6. https://doi.org/10.1109/istas65609.2025.11269636
5. Yazıcı-Kabadayı, S., Mercan, O., & Öztemel, K. (2025). Cyberbullying roles and psychosocial dynamics: a latent profile analysis of loneliness, resilience, and self-regulation in adolescents. BMC Public Health, 25(1), 1480. https://doi.org/10.1186/s12889-025-22745-w
6. Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151. https://doi.org/10.1126/science.aap9559
7. Shah, A., Varshney, S., & Mehrotra, M. (2024). Detection of fake profiles on online social network platforms: Performance evaluation of artificial intelligence techniques. SN Computer Science, 5(5). https://doi.org/10.1007/s42979-024-02839-9
8. Mu, G., Chen, C., Li, X., Chen, Y., Dai, J., & Li, J. (2025). CLAAF: Multimodal fake information detection based on contrastive learning and adaptive Agg-modality fusion. PLoS ONE, 20(5), e0322556. https://doi.org/10.1371/journal.pone.0322556
9. Kesharwani, M., Kumari, S., Niranjan, V., & IRJET. (2021). Detecting Fake Social Media Account Using Deep Neural Networking. International Research Journal of Engineering and Technology, 08(07), 1191. https://www.irjet.net
10. Li, J., Jiang, W., Zhang, J., Shao, Y., & Zhu, W. (2024). Fake user detection based on Multi-Model Joint Representation. Information, 15(5), 266. https://doi.org/10.3390/info15050266
11. Pokharna, S., Sharma, P., Taneja, T., Verma, S., & Ojha, H. (2025). Secure & Reliable Fake Profile Detection on Recruitment Platforms using Machine Learning and Blockchain. Procedia Computer Science, 259, 1228–1238. https://doi.org/10.1016/j.procs.2025.04.078
12. Alharbi, N., Alkalifah, B., Alqarawi, G., & Rassam, M. A. (2024). Countering social media cybercrime using deep learning: Instagram Fake accounts detection. Future Internet, 16(10), 367. https://doi.org/10.3390/fi16100367
13. Li, J., Jiang, W., Zhang, J., Shao, Y., & Zhu, W. (2024b). Fake user detection based on Multi-Model Joint Representation. Information, 15(5), 266. https://doi.org/10.3390/info15050266
14. Bussu, A., Pulina, M., Ashton, S., & Mangiarulo, M. (2023b). Exploring the impact of cyberbullying and cyberstalking on victims' behavioural changes in higher education during COVID-19: A case study. International Journal of Law, Crime and Justice, 75, 100628. https://doi.org/10.1016/j.ijlcj.2023.100628

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 50-55

15. Chattaraj, D., S, V., Nayak, V. R., Hegde, V. V., A, K. K., & Tadkal, S. (2025). Spurious Social Network Profiles Identification Through Hybrid ML Techniques: Analysis and Observation. ICCSP Conference Proceedings, i–vi. https://doi.org/10.1109/iccsp64183.2025.11088748
16. 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
17. Unni, M. V., S, J., Kalapurackal, J. J., & Fatma, S. (2024). Enhancing authenticity and trust in social media: an automated approach for detecting fake profiles. Indonesian Journal of Electrical Engineering and Computer Science, 35(1), 292. https://doi.org/10.11591/ijeecs.v35.i1.pp292-300
18. Kenny, R., Fischhoff, B., Davis, A., Carley, K. M., & Canfield, C. (2022). Duped by Bots: Why Some are Better than Others at Detecting Fake Social Media Personas. Human Factors the Journal of the Human Factors and Ergonomics Society, 66(1), 88–102. https://doi.org/10.1177/00187208211072642
19. Varshitha, K., Talada, S. V., & Mitra, A. (2025). Towards fake profiles identification in social networks: A proposal with energy-based PageRank algorithm involving entropy and domain authority. Risk Sciences., 1, 100013. https://doi.org/10.1016/j.risk.2025.100013
20. Padmavathi, A., & Vaisshnavi, K. (2024). Comparative Analysis of Fake Account Detection Using Machine Learning Algorithms. IEEE International Conference on Advanced Information and Smart Processing, 1–7. https://doi.org/10.1109/aisp61711.2024.10870733
21. Cresci, S., Roberto, D. P., Petrocchi, M., Spognardi, A., & Tesconi, M. (2017). The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. Technical University of Denmark, DTU Orbit (Technical University of Denmark, DTU). https://doi.org/10.48550/arxiv.1701.03017
22. Zhou, Y., Yang, Y., Ying, Q., Qian, Z., & Zhang, X. (2023). Multi-modal Fake News Detection on Social Media via Multi-grained Information Fusion. International Conference on Multimedia Retrieval, 343–352. https://doi.org/10.1145/3591106.3592271
23. Deep learning technique to detect fake accounts on social media. (2024, March 14). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/10522400
24. Varol, O., Ferrara, E., Davis, C., Menczer, F., & Flammini, A. (2017). Online Human-Bot Interactions: Detection, Estimation, and Characterization. Proceedings of the International AAAI Conference on Web and Social Media, 11(1), 280–289. https://doi.org/10.1609/icwsm.v11i1.14871
25. Sri Harsha, A., Safoora, T. S., Nithisha, N., Krishna, V., & Shaik Karimulla. (2025). Advanced detection of fake social media accounts using machine learning algorithms. In International Journal of Advance Scientific Research and Engineering Trends (Vol. 9, Issue 3, pp. 65–66).
26. Karamu, M. B., Araka, E. N., & Department of Computing and Information Science, School of Pure and Applied Science, Kenyatta University, Nairobi, Kenya. (2024). A Hybrid Machine Learning Model for Detection of Fake Profile Accounts on Social Media Networks. International Journal of Engineering Research & Technology (IJERT). http://www.ijert.org
27. Verma, S., Warsi, S. Y. A., & Kumar, R. (2025). Enhancing Online security: Detection of fake profiles on Instagram using GBM. International Journal of Scientific Research in Science Engineering and Technology, 12(2), 176–184. https://doi.org/10.32628/ijsrst25122234
28. Yulia, Gunawan, H., Budhi, G. S., & Kartawidjaja, K. G. (2025). Machine Learning-Based Fake Account Detection System: Instagram Case study. Journal of Information and Communication Convergence Engineering, 23(2), 94–100. https://doi.org/10.56977/jicce.2025.23.2.94
29. Goyal, B., Gill, N. S., & Gulia, P. (2024). Securing social spaces: machine learning techniques for fake profile detection on instagram. Social Network Analysis and Mining, 14(1). https://doi.org/10.1007/s13278-024-01399-3
30. M. S. Kumar, J. Sabeena, K. M. Veena, K. Pavan, M. Sukavya and K. Sravanthi, "Fake Profile Detection on Social Networking Websites using Machine Learning," 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2023, pp. 119-122, doi: 10.1109/ICSCSS57650.2023.10169168.
31. Nayak, A. (2025). Fake Profile Detection using Machine Learning Algorithms. Journal of Information Systems Engineering & Management, 10(16s), 391–401. https://doi.org/10.52783/jisem.v10i16s.2624
32. Ahmad, S., & Tripathi, M. M. (2023). A review article on detection of fake profile on Social-Media. International Journal of Innovative Research in Computer Science & Technology, 11(2), 44–49. https://doi.org/10.55524/ijircst.2023.11.2.9
33. Kuruvilla, A., Daley, R., & Kumar, R. (2023, November 12). Spotting fake profiles in social networks via keystroke dynamics. arXiv.org. https://arxiv.org/abs/2311.06903
34. Shukla, P. K., Veerasamy, B. D., Alduaiji, N., Addula, S. R., Pandey, A., & Shukla, P. K. (2025). Fraudulent account detection in social media using hybrid deep transformer model and hyperparameter optimization. Scientific Reports, 15(1), 38447. https://doi.org/10.1038/s41598-025-24326-8
35. Mane, R. B. V. (2025). A hybrid model for detecting fake profiles in online social networks: enhancing user trust. Journal of Information Systems Engineering & Management, 10(10s), 170–185. https://doi.org/10.52783/jisem.v10i10s.1364
36. Singh, N., Sharma, T., Thakral, A., & Choudhury, T. (2018). Detection of fake profile in online social networks using machine learning. Journal of Computational Analysis and Applications, 231–234. https://doi.org/10.1109/icacce.2018.8441713
37. Deng, M. (2025). Early detection of malicious accounts on social platforms based on temporal graph feature learning. ACM Conference Proceedings, 1320–1328. https://doi.org/10.1145/3773365.3773574
38. Mannocci, L., Cresci, S., Monreale, A., Vakali, A., & Tesconi, M. (2022). MuLBot: Unsupervised bot Detection based on multivariate time series. 2022 IEEE International Conference on Big Data (Big Data), 1485–1494. https://doi.org/10.1109/bigdata55660.2022.10020363
39. Sarala, V., & Patapalla, S. G. (2023). Fake Account Detection Using Machine Learning and Data Science. International Journal of Engineering Science and Advanced Technology (IJESAT), 23–23(9), 341–342. https://www.ijesat.com/ijesat/files/V23I9038_1694754948.pdf
40. Sharma, D., & Singh, N. (2025). A review of deep learning approaches for fake profile detection on social networking sites. International Journal of Scientific Research in Science Engineering and Technology, 12(4), 432–445. https://doi.org/10.32628/ijsrset2512523