A Comprehensive Review of Fake Profile Detection Techniques Using Machine Learning & Deep Learning

Ruchi J. Dalsaniya[1], Ankit Kalariya[2]
Atmiya University

***Abstract*— The development of social media websites has led to the creation of many fake profiles for different uses such as spreading spam emails, deceiving users, misinformation, and many others. Furthermore, developments in the field of artificial intelligence have enabled the creation of realistic images by use of GANs. This paper provides an analysis on the existing methods of detecting fake profiles with machine learning, deep learning, and image forensics techniques. In addition to this, the existing literature will be compared based on their efficiency and drawbacks. It needs to be mentioned that deep learning has high accuracy than machine learning, but on the other hand, it is highly complex. On the contrary, machine learning has low accuracy but less complexity.**

***Index Terms*— Fake Profile Detection, Machine Learning, Deep Learning, GAN, Social Media, Image Forensics**

I. INTRODUCTION

Social networking websites have become very important in communication and exchange of data in recent years. On the other hand, the increase in the number of individuals accessing the social networking websites has also led to the alarming trend of an increasing number of fake user profiles on these social websites. Fake user profiles can be very damaging and can be used to manipulate others by means of deception and cheating.

The advancement of technology and the development of AI systems have made it difficult for any rule-based or manual approach to detect such fake user profiles. The images generated through these AI systems can be quite authentic-looking so much so that the task of distinguishing between fake and real profiles becomes challenging using rule-based and manual methods.

Thus, various machine learning and deep learning models have been suggested for this purpose. The machine learning models make use of structured data containing user behaviour and profile data, while deep learning approaches are designed to learn complex patterns existing within the images. Image forensics have been another approach utilized to detect synthetic images.

In this paper, a detailed analysis of all the techniques that have been developed to detect fake profiles will be provided. Furthermore, a comparative study of results from various studies carried out on the subject matter will be presented in this paper. This will help in determining the pros and cons of these techniques.

II. LITERATURE REVIEW

A number of studies have been conducted recently on identifying different methods of detecting fake profiles using machine learning, deep learning, and image forensic technologies. With the increasing capabilities of generative models like GANs, discerning real from fake images has become quite challenging.

Other methods consider analysis at the level of images themselves. For example, camera-fingerprinting aims to detect the presence of fingerprints left while generating images. Though this technique ensures reliable results, it consumes a lot of computational power, which may affect its ability to handle sophisticated fake images.

Approaches employing machine-learning techniques take advantage of structured data, for instance, behaviour, profile attributes, and statistics. The approaches include Random Forests, SVMs (Support Vector Machines), and KNN (K-Nearest Neighbours). They are generally efficient yet rely on the chosen features to a great extent. New fake account profiles can make their work inefficient.

The deep learning models, specifically the Convolutional Neural Networks, have been found to

exhibit better results as these automatically learn the complex features present in the image data sets. ResNet is one such deep learning model that has further advanced the detection process.

Moreover, there are some recent researches that incorporate several types of data sources, such as images, text, and behavior, to enhance the performance of detection systems. Despite their increased performance rates, these multi-modal techniques add complexities to the system. Generally, the present detection systems indicate the necessity for improved methods.

III. COMPERATIVE ANALYSIS

A comparative study of different Algorithm reveals differences in Performance & Accuracy.

Paper	Author's	Title of Paper	Objective of Paper	Method Used	Description & Accuracy
P1	Manisha et al.[1]	AI-Synthesized Image Detection: Source Camera Fingerprinting to Discern the Authenticity of Digital Images	Forensic Technique is used to detect AI-generated images.	Camera Fingerprinting	High detection accuracy is achieved with strong generalization across multiple datasets, effectively differentiating real and AI-generated images even under post-processing conditions.
P2	Kai-Cheng Yang et al.[2]	Characteristics and Prevalence of Fake Social Media Profiles with AI-Generated Faces	Identifying GAN-generated profile images in social media	GAN Feature-Based Detection	Provides moderate detection performance by identifying visual inconsistencies in GAN-generated faces, showing effectiveness on real-world social media data.
P3	R.R. Arunprakaash et al.[3]	Leveraging Machine Learning Algorithms for Fake Profile Detection on Instagram	ML model is used to detect fake profile.	Machine Learning (RF, SVM, KNN)	Achieves approximately 92.5% accuracy, demonstrating reliable performance for fake profile detection by using structured profile features.
P4	P. Asha Bhat et al.[4]	Fake Instagram Profile Detection	Detect fake profiles using multi-modal data	Multi-modal Approach	Shows high detection accuracy by combining multiple data sources, improving overall classification performance and detection reliability.
P5	R.R. Arunprakaash et al.[5]	GAN Generated Fake Human Face Image Detection	Detect GAN-generated human faces	CNN-Based Detection	Achieves very high accuracy of approximately 99.42%, demonstrating strong capability in identifying GAN-generated fake face images.

Paper	Author's	Title of Paper	Objective of Paper	Method Used	Description & Accuracy
P6	Rimsha Rafique et al.[6]	Deep fake detection and classification using error-level analysis and deep learning	Detect manipulated images	ELA + CNN	Achieves around 89.5% accuracy, effectively detecting manipulated images, though performance is influenced by image quality variations.
P7	Vera Wesselkamp et al.[7]	Misleading Deep-Fake Detection with GAN Fingerprints	Detect fake images using frequency patterns	GAN Fingerprint Analysis	Demonstrates high detection performance by capturing unique patterns in synthetic images, enabling effective identification of GAN-generated content.
P8	Yasir Hamid et al.[8]	An Improvised CNN Model for Fake Image Detection	Classify fake images using deep learning	ResNet50 CNN Model	Achieves approximately 99% accuracy, showing strong performance in distinguishing real and fake images through deep feature representation.
P9	Akihisa Kawabe et al.[9]	Fake Image Detection Using an Ensemble of CNN Models Specialized for Individual Face Parts	Detect fake images using facial regions	Ensemble CNN (Face Parts)	Provides high accuracy by analyzing multiple facial regions, improving detection performance especially for partially manipulated images.
P10	Aayush Sunil Chamria et al.[10]	Detecting Fake Profiles in Online Social Networks using EnsemStack Classification Algorithm	Detect fake profiles using ensemble ML	Stacked Ensemble Learning Model	Shows good overall accuracy by combining multiple classification outputs, leading to improved decision-making compared to individual models.

IV. DISCUSSION

For the task of identifying fake profiles, image processing is commonly applied to find out the hidden patterns within the images in the fake profiles. The frequency domain transforms like FFT and DCT can be utilised to uncover the artifacts within the images generated by the GAN models. Machine learning algorithms leverage handcrafted features that are created using those transforms or based on the user's interaction with the profiles. Such features enable classifying the profiles as

genuine or fake. While this approach is highly successful, it depends heavily on the feature engineering process.

Deep learning algorithms, particularly CNNs, automatically detect useful patterns in the raw data without any pre-processing. In other words, the deep learning models perform well in spotting the artifacts within the fake profile images. Nevertheless, they demand substantial computing power and data for training.

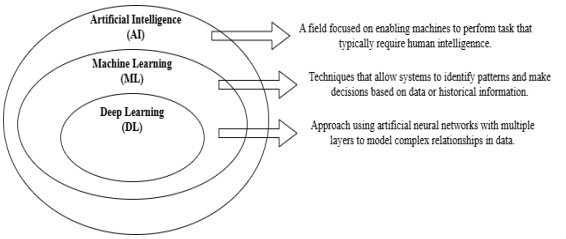

Figure 1: Hierarchical Relationship among AI, ML, and DL

A. Traditional Method

Conventional techniques used for identifying fake accounts typically rely on manual analysis and rule-based approaches. Manual analysis involves humans checking if there is any suspicious activity. Rule-based approaches involve creating rules that may be used to detect fake accounts based on some guidelines. Even though these approaches can easily be implemented, they require a lot of time and cannot be scaled for social networking sites.

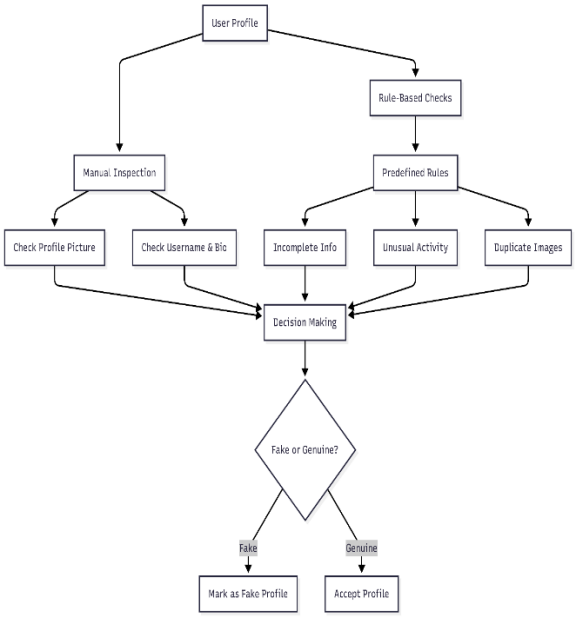

Figure 2: Working of Traditional Method

B. Machine Learning Approach

Machine learning techniques have contributed significantly to detecting false social media identities through automated classification systems. These algorithms analyze user behaviour, social profile characteristics, and statistical features in order to identify fraudulent users. Some examples include random forest, support vector machine, and K-nearest Neighbors. However, their effectiveness depends a lot on feature selection and the quality of input data. They might not be able to catch up with evolving trends of fake social media identities.

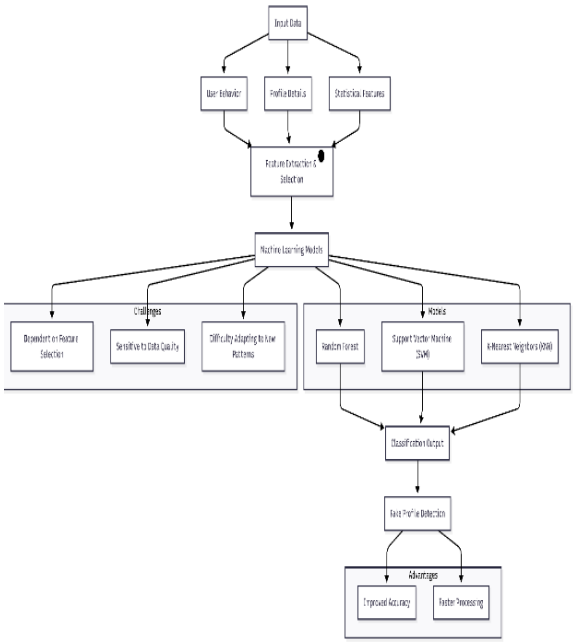

Figure 3: Working of Machine Learning Approach

C. Deep Learning Approach

Deep learning techniques, such as Convolutional Neural Networks (CNNs), have shown great results in detecting fake accounts by identifying complex features automatically through machine learning. They are highly efficient in analyzing image-based data such as pictures used on social media accounts and detecting discrepancies that exist in images generated by AI technology. Deep learning techniques provide high accuracy and strong feature extraction capabilities but require large data sets and substantial computational resources.

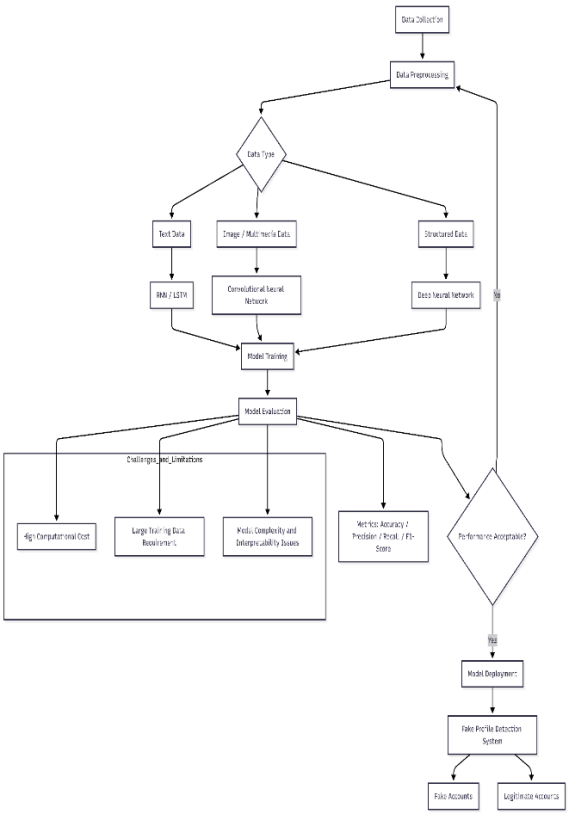

Figure 4: Working of Deep Learning Approach

V. RESEARCH GAP

However, despite the many developments that have taken place in this area, there are several problems that remain unsolved in fake profiles detection approaches. Most of the techniques available today are not capable of identifying highly realistic images created by GANs since GANs continue to advance and improve.

Machine learning approaches are prone to feature selection errors that make them inflexible and unable to recognize new patterns. On the other hand, deep learning algorithms have demonstrated impressive results but require a lot of time and computing power.

What is more, most researchers concentrate on using one particular data source only instead of considering different data types. Finally, many of the techniques proposed in the literature are prone to adversarial attacks in the form of GAN fingerprints.
In summary, it can be stated that further research into this topic is required to develop more reliable and scalable detection techniques.

VI. CONCLUSION

In conclusion, this paper has critically examined various approaches to detect fake profiles, such as machine learning, deep learning, and image forensics techniques. It is evident from this analysis that traditional techniques are inadequate when dealing with highly realistic fake profiles created using advanced techniques.

Machine learning techniques are effective since they utilize predetermined features to make a detection process efficient. On the other hand, deep learning approaches are more efficient in their performance than the former. Nonetheless, there are drawbacks associated with each technique, especially when dealing with advanced techniques of creating fake profiles.

It is advisable that future detection systems combine several approaches to increase the effectiveness of the process.

VII. FUTURE WORK

Future works will entail the need for hybrid models that combine the use of machine learning and deep learning techniques in order to enhance efficiency and effectiveness.

Also, there is need to design models that can provide timely detection by dealing with the vast amount of data that exists within social media networks.

Moreover, integration of several types of data including imagery, text and behavioural data can go a long way to enhance the performance of detection.
Robustness of models against adversarial attacks can be another area that future work can explore.
Lastly, development of lightweight models can lead to decreased computational cost as well as efficient performance.

REFERENCES

[1] Li, C. T., & Kotegar, K. A. (2025). AI-Synthesized Image Detection: Source Camera Fingerprinting to Discern the Authenticity of Digital Images. IEEE Access.
[2] Yang, K. C., Singh, D., & Menczer, F. (2024). Characteristics and prevalence of fake social media profiles with AI-generated faces. arXiv preprint arXiv:2401.02627.

[3] Arunprakaash, R. R., & Nathiya, R. (2024, August). Leveraging Machine Learning algorithms for Fake Profile Detection on Instagram. In 2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT) (Vol. 1, pp. 869-876). IEEE.

[4] Bhat, P. A., Chaitra, M., Thyli, S. R., & Anitha, N. (2024, November). Fake Instagram Profile Detection. In 2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS) (pp. 1-6). IEEE.

[5] Shilaskar, S., Talewar, M., Tak, S., & Goud, S. (2024, January). GAN generated fake human face image detection. In 2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE) (pp. 1-6). IEEE.

[6] Rafique, R., Gantassi, R., Amin, R., Frnda, J., Mustapha, A., & Alshehri, A. H. (2023). Deep fake detection and classification using error-level analysis and deep learning. Scientific reports, 13(1), 7422.

[7] Wesselkamp, V., Rieck, K., Arp, D., & Quiring, E. (2022, May). Misleading deep-fake detection with GAN fingerprints. In 2022 IEEE Security and Privacy Workshops (SPW) (pp. 59-65). IEEE.

[8] Hamid, Y., Elyassami, S., Gulzar, Y., Balasaraswathi, V. R., Habuza, T., & Wani, S. (2023). An improvised CNN model for fake image detection. International Journal of Information Technology, 15(1), 5-15.

[9] Kawabe, A., Haga, R., Tomioka, Y., Okuyama, Y., & Shin, J. (2022, December). Fake image detection using an ensemble of CNN models specialized for individual face parts. In 2022 IEEE 15th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC) (pp. 72-77). IEEE.

[10] Chamria, A. S., Mane, A. D., Dambal, P. V., & Bharne, S. (2022, August). Detecting fake profile in online social networks using ensemstack classification algorithm. In 2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA) (pp. 1-6). IEEE.