



A Hybrid CNN-Based Approach for Detecting Machine-Generated Tweets Using Fast Text Embeddings

Gazal Begum¹, Md Ashique Hussain², MD Haseen³, Mohd Faizan Khan⁴, Sayyad Abdullah⁵

^{1,2}Assistant Professor, Department of CSE (Data Science), Lords Institute of Engineering and Technology, Hyderabad, Telangana, India.

^{3,4,5}UG Students, Department of CSE (Data Science), Lords Institute of Engineering and Technology, Hyderabad, Telangana, India.

Abstract— The rapid growth of social media platforms has led to an increase in the spread of misleading and automated content, particularly through deep fake tweets generated by bots. This project focuses on detecting such fake tweets using machine learning and deep learning techniques. The system allows users to load datasets, convert textual data into numerical vectors using FastText embedding, and train multiple algorithms to evaluate performance. Among the applied models, Convolutional Neural Network (CNN) and hybrid CNN approaches demonstrate superior accuracy in identifying fake content. The application provides a user-friendly interface for login, dataset loading, model training, and real-time prediction of tweets. By analyzing textual patterns and semantic features, the system effectively distinguishes between human-written and bot-generated tweets. This approach contributes to enhancing trust and reliability in social media communication by reducing the impact of automated misinformation.

Keywords— Deep Fake Tweet Detection, Machine Learning, Deep Learning, FastText Embedding, Convolutional Neural Network (CNN), Hybrid CNN, Social Media Analysis, Bot Detection, Natural Language Processing, Text Classification.

I. INTRODUCTION

Social media has become an essential platform for communication, information sharing, and public interaction. However, the increasing use of automated bots has introduced significant challenges, especially in the form of fake or manipulated content [2], [7]. Deep fake tweets, generated using advanced algorithms, can mimic human behavior and spread misinformation rapidly [3]. These tweets often influence public opinion,

create confusion, and impact decision-making processes [5]. Detecting such content manually is difficult due to the vast volume of data generated every second [1]. Therefore, there is a strong need for automated systems that can accurately identify fake tweets [6]. Machine learning and deep learning techniques provide efficient solutions by analyzing patterns, linguistic features, and contextual information [8], [9]. This project aims to address these challenges by developing a system that detects deep fake tweets using advanced computational methods and provides reliable predictions.

Text classification plays a crucial role in identifying fake content in social media. Traditional methods rely on manual feature extraction, which is time-consuming and less effective when dealing with large datasets [1]. With the advancement of natural language processing techniques, automated feature extraction has become more efficient [8]. In this system, FastText embedding is used to convert textual data into numerical vectors, capturing both semantic and syntactic information. These embeddings improve the performance of machine learning models by providing meaningful representations of text. Various algorithms such as Naive Bayes, Support Vector Machine, Decision Tree, and Random Forest are applied to classify tweets. Each algorithm analyzes the data differently, allowing comparison of their performance. This approach helps in identifying the most effective model for detecting fake tweets accurately and efficiently.

Deep learning models have shown significant improvements in text classification tasks compared to traditional machine learning techniques [9], [12]. Among them, Convolutional Neural Networks (CNN) are widely used due to their ability to extract important features automatically. CNN models analyze patterns in text sequences and



identify complex relationships that are often missed by conventional algorithms. In this project, both standard CNN and hybrid CNN models are implemented to enhance detection accuracy. The hybrid model combines multiple techniques to improve performance and reduce misclassification. Experimental results demonstrate that CNN-based models outperform other algorithms in detecting deep fake tweets [14], [15]. These models provide higher accuracy and better generalization, making them suitable for real-world applications where reliability is crucial.

The system is designed with a user-friendly interface that allows users to interact with the application easily. It includes functionalities such as user login, dataset loading, feature extraction, model training, and prediction. Users can upload datasets, convert text into vectors, and run multiple algorithms with a single click. The system displays results in both tabular and graphical formats, enabling users to compare the performance of different models. Additionally, users can enter custom tweet text to check whether it is fake or genuine. This interactive design makes the system accessible even to users with limited technical knowledge. The integration of backend processing with a simple frontend ensures smooth operation and effective user experience.

The detection of deep fake tweets is essential for maintaining the integrity of online platforms [3], [10]. As misinformation continues to grow, the need for reliable detection systems becomes more critical [5], [6]. This project contributes to addressing this issue by combining machine learning, deep learning, and natural language processing techniques. The proposed system not only identifies fake tweets but also provides insights into model performance and accuracy. By leveraging advanced algorithms and efficient data processing methods, the system offers a practical solution for detecting automated content. Future improvements may include real-time data analysis, integration with social media platforms, and enhancement of model accuracy. Overall, this work plays an important role in promoting trustworthy communication in the digital world.

II. RELATED WORK

Verma et al., [2016] [1] Verma et al. discussed the growing importance of big data analytics in handling various types of data such as text, audio, video, and social media content. The study highlights the challenges associated with processing large-scale unstructured data and emphasizes the need for efficient analytical techniques. The authors explain how traditional

data processing methods are insufficient for managing complex datasets. They also describe the role of advanced tools and frameworks in extracting meaningful insights from massive data sources. The paper identifies issues such as scalability, storage, and real-time processing. Furthermore, it explores applications of big data in different domains including social media analysis. The findings suggest that proper data handling techniques can improve decision-making processes. Overall, the work provides a strong foundation for understanding data-driven systems.

Siddiqui et al., [2017] [2] Siddiqui et al. focused on identifying automated bots in online environments. The study presents methods to distinguish between human users and bot-generated activities. The authors highlight the increasing presence of bots and their potential misuse in spreading misinformation. Various detection techniques are discussed, including behavioral analysis and pattern recognition. The research emphasizes the importance of security in internet-based systems. It also explores how bots can mimic human behavior, making detection more challenging. The paper suggests that combining multiple detection approaches can improve accuracy. Additionally, it discusses the limitations of existing systems in identifying sophisticated bots. The study concludes that continuous improvement in detection techniques is necessary. This work is useful for developing systems that identify fake or automated content.

Westerlund et al., [2019] [3] Westerlund et al. provided a detailed review of deepfake technology and its rapid development. The study explains how artificial intelligence is used to generate highly realistic fake media content. It discusses both the benefits and risks associated with deepfake applications. The author highlights the potential misuse of this technology in spreading false information. The paper also examines the impact of deepfakes on society, politics, and media trust. Various detection techniques are briefly discussed to counter deepfake content. The study emphasizes the need for awareness and regulation to control misuse. Furthermore, it explores future research directions in deepfake detection. The findings indicate that technological advancements must be balanced with ethical considerations. This work is important for understanding fake content generation.

Vosoughi et al., [2018] [5] Vosoughi et al. analyzed how true and false news spreads across social media platforms. The study reveals that false information spreads faster and reaches more people compared to true news. The authors used large-scale data analysis to understand user behavior and



information flow. The research highlights the role of human factors in spreading misinformation rather than automated bots alone. It explains how emotional and sensational content attracts more attention. The paper also discusses the challenges in controlling false information online. The findings suggest that misinformation can have significant societal impacts. The study emphasizes the need for improved detection and prevention mechanisms. Additionally, it highlights the importance of user awareness in reducing misinformation spread. This work provides valuable insights into online information dynamics.

Zellers et al., [2019] [9] Zellers et al. explored techniques to defend against neural fake news generation. The study introduces methods to detect machine-generated text using advanced models. The authors discuss the challenges posed by AI systems that can generate highly convincing fake content. The research presents a framework for identifying synthetic text by analyzing patterns and inconsistencies. It highlights the importance of combining detection tools with generation models. The paper also evaluates the performance of different approaches in detecting fake news. The findings show that detection systems must evolve alongside generation technologies. Furthermore, it emphasizes the need for responsible use of AI in content creation. The study contributes to improving fake news detection systems. This work is significant in addressing AI-generated misinformation.

III. DATASET DETAILS

The dataset used in this project consists of a collection of tweets gathered from social media platforms, focusing on both genuine human-generated content and bot-generated or deep fake tweets. Each record in the dataset contains textual information along with labels that indicate whether the tweet is real or fake. The data is preprocessed to remove noise such as special characters, URLs, and unnecessary symbols to ensure better model performance. Tokenization and normalization techniques are applied to standardize the text format. The dataset is structured in a tabular form where each row represents an individual tweet and its corresponding class label. This balanced dataset helps in training machine learning models effectively by providing sufficient examples of both categories. The diversity of tweet content allows the system to learn different linguistic patterns, making it suitable for detecting various types of fake or automated messages.

After preprocessing, the dataset is transformed into numerical form using FastText embedding, which converts textual data into meaningful vector

representations. This transformation captures semantic relationships between words, improving the model's ability to understand context. The dataset is then divided into training and testing subsets to evaluate model performance accurately. Multiple machine learning and deep learning algorithms are trained using this data, including Naive Bayes, Support Vector Machine, Decision Tree, Random Forest, and Convolutional Neural Network models. The inclusion of both simple and complex models enables performance comparison across different techniques. The dataset also includes sample test inputs, allowing users to manually verify predictions through the system interface. Proper handling and transformation of the dataset ensure reliable results and higher accuracy in detecting fake tweets. This structured approach supports efficient training, evaluation, and real-time prediction within the application.

IV. PROPOSED METHODOLOGY

The proposed system follows a structured approach to detect deep fake tweets using both machine learning and deep learning techniques. Initially, the dataset is collected and preprocessed by removing unwanted symbols, links, and noise from the text. The cleaned data is then transformed into numerical form using FastText embedding, which captures semantic meaning and contextual relationships between words. After feature extraction, the dataset is divided into training and testing sets. Multiple classification algorithms such as Naive Bayes, Support Vector Machine, Decision Tree, and Random Forest are applied to analyze patterns and learn from the data for accurate prediction.

In addition to traditional algorithms, deep learning models such as Convolutional Neural Networks and hybrid CNN are implemented to improve classification performance. These models automatically extract important features from text and identify complex patterns that may not be captured by conventional methods. The system evaluates all models based on accuracy and displays results in graphical and tabular formats for comparison. Finally, the best-performing model is used for real-time prediction, where users can input tweet text and receive output indicating whether it is fake or genuine, ensuring efficient and reliable detection of automated content.

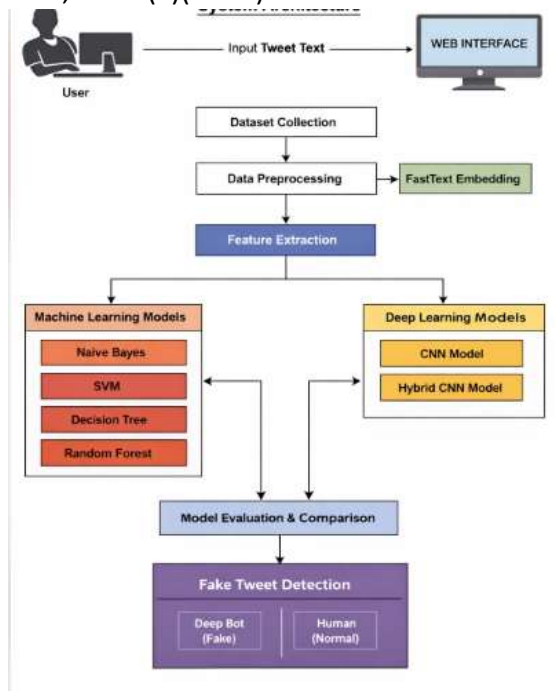


Figure [1]: System Architecture of Hybrid ML–CNN Based Deep Fake Tweet Detection

Figure [1] This architecture represents the complete workflow of the deep fake tweet detection system, beginning with user interaction through the web interface where tweet text is provided. The system processes the input data through dataset loading and preprocessing steps to clean and prepare the text. FastText embedding is then applied to convert textual data into numerical vectors, followed by feature extraction. The processed data is fed into multiple machine learning and deep learning models, including Naive Bayes, SVM, Decision Tree, Random Forest, CNN, and hybrid CNN. Finally, the system evaluates model performance and displays the prediction results, identifying whether the tweet is fake (Deep Bot) or genuine (Human) in an organized manner.

V.RESULT AND DISCUSSION

The experimental results of the proposed system demonstrate its effectiveness in detecting deep fake tweets using both machine learning and deep learning techniques. After loading and pre-processing the dataset, FastText embedding successfully converted textual data into meaningful numerical vectors, enabling efficient model training. Multiple algorithms, including Naive Bayes, Support Vector Machine, Decision Tree, and Random Forest, were applied to evaluate classification performance. Among these, traditional models showed moderate accuracy and were able to identify basic patterns in the data. However, deep learning models, particularly the

Convolutional Neural Network and the hybrid CNN, achieved significantly higher accuracy compared to other methods. The results were presented in both tabular and graphical formats, allowing clear comparison between models. The hybrid CNN model outperformed all other algorithms by capturing complex linguistic features and contextual relationships within the tweets. Additionally, the system was tested using sample tweet inputs, where it correctly classified tweets as either “Deep Bot” or “Human.” These outcomes confirm that the proposed approach is reliable and capable of producing accurate predictions in real-time scenarios.



Figure [2]: Deep Fake Tweet Detection System Interface

Figure [2] The interface provides options for user navigation such as Home and User Login, allowing users to access the system easily. It is designed to perform deep fake tweet detection by integrating data preprocessing, FastText embedding, and classification models. The interface visually represents the workflow and enables users to interact with the system for loading datasets, running algorithms, and predicting whether a tweet is generated by a bot or a human.



Figure [3]: Dataset Loading and Display Interface

Figure [3] This interface allows users to load and view the dataset used for training and testing the deep fake tweet detection system. It displays tweet data in a tabular format, including username, tweet content, and account type labels such as bot or



human. The navigation menu provides options for further processing steps like FastText embedding, running algorithms, and prediction. This screen helps users verify the dataset before applying machine learning and deep learning models.



Figure [4]: Algorithms Performance Evaluation Interface

Figure [4] This interface displays the performance results of various machine learning and deep learning algorithms used in the system. It presents evaluation metrics such as accuracy, precision, recall, and F1-score in a tabular format for easy comparison. Additionally, a graphical representation is provided to visualize the performance of each algorithm. The results clearly indicate that the proposed CNN and hybrid CNN models achieve higher accuracy compared to other methods, helping users identify the most effective model for deep fake tweet detection.



Figure [5]: Deep Fake Tweet Prediction Interface

Figure [5] This interface allows users to test and predict whether a given tweet is fake or genuine. Users can enter tweet text into the input field and submit it for analysis. The system processes the input using trained machine learning and deep learning models and displays the prediction result instantly. In this screen, the output indicates that the given tweet is detected as a bot-generated fake. This interface enables real-time classification and helps users easily verify the authenticity of tweet content.

DISCUSSION

The performance of the system highlights the importance of combining feature extraction techniques with advanced learning models for effective fake tweet detection. The use of FastText embedding played a crucial role in preserving semantic meaning, which improved the learning capability of both machine learning and deep learning models. While traditional algorithms provided acceptable results, they were limited in handling complex patterns and contextual variations in text data. In contrast, the CNN and hybrid CNN models demonstrated superior performance due to their ability to automatically extract deep features and identify subtle differences between human and bot-generated content. The graphical comparison of results clearly indicates that deep learning approaches are more suitable for this type of problem. The system's user-friendly interface also contributes to its practicality, allowing users to easily perform predictions and analyze results. However, the accuracy of the system depends on the quality and diversity of the dataset. Future improvements may include expanding the dataset, incorporating real-time data streams, and optimizing model parameters to further enhance performance. Overall, the system provides a strong foundation for detecting misinformation in social media environments.

VI. CONCLUSION

The developed system effectively identifies deep fake tweets by combining machine learning and deep learning techniques. By applying preprocessing methods and FastText embedding, the textual data is transformed into meaningful representations that improve model performance. Various algorithms were tested, and the results showed that deep learning models, especially CNN and hybrid CNN, provide better accuracy compared to traditional methods. The system is designed with a simple interface that allows users to load data, run models, and perform real-time predictions easily. It successfully distinguishes between human-written and bot-generated tweets, helping to reduce the spread of misleading information. Overall, this work demonstrates a practical and efficient approach for detecting fake content on social media platforms.

REFERENCES

1. Verma, J. P., & Agrawal, S. (2016). Challenges and applications of big data analytics across text, audio, video, and social media domains. *International Journal of Soft Computing and Artificial Intelligence Applications*, 5(1), 41–51.



2. Siddiqui, H., Healy, E., & Olmsted, A. (2017). Bot detection in online systems. In Proceedings of the 12th International Conference on Internet Technology and Secured Transactions (ICITST) (pp. 462–463).
3. Westerlund, M. (2019). A comprehensive review on the rise of deepfake technology. *Technology Innovation Management Review*, 9(11), 39–52.
4. Ternovski, J., Kalla, J., & Aronow, P. M. (2021). The impact of deepfake warnings on political video perception: Experimental evidence. Ph.D. Dissertation, Department of Political Science, Yale University.
5. Vosoughi, S., Roy, D., & Aral, S. (2018). Analysis of how true and false information spreads online. *Science*, 359(6380), 1146–1151.
6. Bradshaw, S., Bailey, H., & Howard, P. N. (2021). Organized social media manipulation: A global inventory report. Oxford Internet Institute, University of Oxford.
7. Grimme, C., Preuss, M., Adam, L., & Trautmann, H. (2017). Exploring whether social bots mimic human behavior through human control. *Big Data*, 5(4), 279–293.
8. Liu, X., Zheng, Y., Du, Z., Ding, M., Qian, Y., Yang, Z., & Tang, J. (2021). Understanding language models: Insights from GPT-based systems. *arXiv preprint arXiv:2103.10385*.
9. Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F., & Choi, Y. (2019). Methods to defend against neural fake news generation. In Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS) (pp. 9054–9065).
10. Beckman, L. (2021). Issues in internet regulation and future recommendations. *Nova Law Review*, 46(2), 277.
11. Lee, J.-S., & Hsiang, J. (2020). Generating patent claims using fine-tuned GPT-2 models. *World Patent Information*, 62, 101983.
12. Dale, R. (2021). Evaluating the applications of GPT-3 in natural language processing. *Natural Language Engineering*, 27(1), 113–118.
13. Heaven, W. D. (2020). Experiment with GPT-3 bots generating Reddit comments unnoticed. *MIT Technology Review*.
14. Gehrmann, S., Strobel, H., & Rush, A. M. (2019). GLTR: Visualization techniques for detecting generated text. *arXiv preprint arXiv:1906.04043*.
15. Adelani, D. I., Mai, H., Fang, F., Nguyen, H. H., Yamagishi, J., & Echizen, I. (2020). Generating and detecting sentiment-preserving fake online reviews using neural models. In Proceedings of the 34th International Conference on Advanced Information Networking and Applications (AINA) (pp. 1341–1354). Springer.
16. Venkata Pavan Kumar Gummadi. (2025). MuleSoft's Role in Advancing Sustainable Digital Infrastructure: An Enterprise Integration Perspective. *Journal of Information Systems Engineering and Management*, 10(62s), 1313–1321. <https://doi.org/10.52783/jisem.v10i62s.13783>
17. Shashank A. (2025). Metadata-driven data integration framework: Automating enterprise data integration through declarative approaches. *European Modern Studies Journal*, 9(4), 9.
18. Kandula, S. T. R., Boyapati, P. K., & Susarla, R. S. (2025, April). Optimized Cloud Resource Management Using Deep Dendritic Artificial Neural Networks Integrated with Kubernetes and Terraform. In International Conference on Computer Vision and Robotics (pp. 454–465). Cham: Springer Nature Switzerland.
19. Babburi, S. Lightweight Distributed Provenance Framework for Edge and IoT Data Systems.
20. Gaddam, S. From Fixed Specifications to Self-Adapting Systems: A Machine Learning Perspective on Software Engineering.
21. Immadi, S. K. (2025). Optimizing ERP for Human Capital Management. *Applied Research for Growth, Innovation and Sustainable Impact*, 377–384. <https://doi.org/10.1201/9781003684657-63>
22. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.



<https://doi.org/10.71097/IJSAT.v16.i2.3219>

23. Poojari, R. Frameworks for Data Management and Lineage in Large-Scale Healthcare Data Systems.
24. Mahimalur, R. K., Vasgam, M., & Manoharan, D. Devops Lifecycle Management And Cloud Migration Assessments: A Security-Driven CICD Perspective.
25. Viswanathan, V. Generative AI for Smarter Workforce Planning and Enterprise Resource Decisions.
26. Gajula, S. (2026, March). Two Pillars of Banking Intelligence: A Comparative Analysis of AI Techniques for Fraud Prevention and Churn Mitigation. In 2026 14th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-6). IEEE.
27. Maturi, S. Y. (2023). Crowdsourced frontier: Unveiling autonomous adversarial cybercapabilities via open AI competition. *International Journal of Intelligent Systems and Applications in Engineering*, 11(1s), 275–284.
28. Venkata Ramana, P. (2024). AI-driven predictive analytics in ERP systems for proactive supply chain optimization. *International Journal of Research in Information Technology and Computing*, 8(4).
29. Srikanth Kavuri. (2025). AI-DRIVEN TEST AUTOMATION FRAMEWORKS: ENHANCING EFFICIENCY AND ACCURACY IN SOFTWARE QUALITY ASSURANCE. *International Journal of Applied Mathematics*, 38(10s), 699–710. <https://doi.org/10.12732/ijam.v38i10s.990>
30. Venkata Pavan Kumar Gummadi. (2026). Infrastructure Optimization Techniques for Enterprise Integration Platforms: A Comprehensive Analysis. *Computer Fraud and Security*, 37–44. <https://doi.org/10.52710/cfs.875>
31. Shashank, A. (2025). AI-Enhanced ETL Processes: Leveraging Artificial Intelligence for Optimized Data Integration Systems. *Journal Of Multidisciplinary*, 5(8), 219-225.
32. Kandula, S. T. R., Susarla, R. S., & Boyapati, P. K. (2025, July). Enhanced Cyber Security Using Global Local Artificial Neural Network Based Intrusion Detection in Big Data Environment. In 2025 IEEE 4th World Conference on Applied Intelligence and Computing (AIC) (pp. 426-431). IEEE.
33. Boyapati, P. K. Building a centralized data operations hub for healthcare enterprise integration. *IJSAT-Int. J. Sci. Technol.* 16 (2).