International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

# Twitter Bot Detection Using Machine Learning and Deep Learning Techniques

Jyothis Joseph[*1], Vyshnavi K B[*2], Nandana Santhosh[*3], Nandhitha Binu[*4]
*College of Engineering Kidangoor, Kerala, India*
[1]jyothis@ce-kgr.org, [2]vyshnavib07@gmail.com, [3]nandanasanthosh000@gmail.com, [4]nandhithabinu2020@gmail.com

***Abstract*—The proliferation of Twitter bots poses a serious threat to the reliability of online conversations and results in disinformation, spam, and opinion manipulation. This paper presents a comprehensive examination of Twitter bot detection techniques with traditional machine learning (ML) algorithms contrasted with cutting-edge deep learning (DL) models. Key fea- tures like tweet frequency, follower-following ratios, user behavior patterns, and content features are investigated. We compare algorithms like Random Forest, Support Vector Machines (SVM), Logistic Regression, K-Nearest Neighbors (KNN), Long Short- Term Memory (LSTM), and Recurrent Neural Networks (RNN) based on accuracy, precision, recall, and F1-score metrics. Our experiments showed that Random Forest was the best with the highest accuracy and thus, it is the best-suited model for the dataset used in this experiment. We also address the issues of real- time bot detection, the limitation of single models, and suggest a hybrid approach that takes advantage of the strengths of both ML and DL approaches for better performance.**

***Index Terms*—Twitter Bot Detection, Machine Learning, Deep Learning, Social Network Analysis, Random Forest, Support Vec- tor Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN).**

## I. INTRODUCTION

Twitter has become an essential social engagement and public communications platform. However, the existence of bot accounts has also presented serious issues since bots are capable of propagating false information, shaping public opinion, and compromising platform integrity.

Bots are pieces of software intended to carry out automated tasks, typically mimicking human behavior. Some bots exist for good reasons, such as customer support or aggregating news feeds, but others are used for nefarious activities. Mis- chievous bots may send spam, artificially cause a hashtag to trend, or even orchestrate large-scale influence operations. Bots can generate tweets quickly, such as liking and retweeting posts, and following and unfollow users, making it difficult to distinguish them from legitimate users.

Bot identification and counter-measure is a significant issue for social networking websites and researchers. Twitter bots can be trained to become human-like, hence harder to iden- tify. Thus, effective detection processes using sophisticated machine learning (ML) and deep learning (DL) models are essential to fight adaptive bot strategies. This paper investi- gates different ML and DL models to create an effective bot identification system to promote online safety and authenticity.

## II. METHODOLOGY

The Twitter bot detection method applied in this research is an amalgamation of static and dynamic detection ap- proaches using machine-learning and deep-learning models. The models have been trained and tested using the Cresci 2017 dataset, which is a famous dataset with respect to bot detection research. The static detection methods used Random Forest, SVM, Logistic Regression, and KNN and the dynamic detection methods RNN, LSTM. Dynamic detection is advantageous for the service providers as it records the behavior patterns over time.The proposed system's workflow is as follows:

**Dataset Selection:** The Cresci 2017 dataset has been used having labeled Twitter accounts as bots and real users. This dataset entails features such as account metadata, tweet con- tent, and user interactions.

**Data Preprocessing:** The preprocessing of raw data was done by removing duplicates, managing missing values, and normalizing numerical features. Some pre-professional activ- ities carried out were tokenization, stopword removal, and stemming/lemmatization for text-based features.

**Feature Engineering:** Important features were derived for the static and dynamic analysis. Static features: account age, follower-following ratio, tweet frequency, and metadata at- tributes; Dynamic features: temporal patterns, sequence of activities, retweet behavior and time-based user interactions;

**Model Selection:** This study used two types of models: Static detection models such as Random Forest, SVM, Logistic Regression, and KNN all classify bots based on user metadata and content-based features. Dynamic detection models like RNN and LSTM learn from sequential user behavior and accordingly predict the impression of bots over time via behavior trends.

**Model Training and Fine-tuning:** Machine learning mod- els were built on an 80-20% train-test split with hyperparam- eter tuning to improve performance. Deep learning networks (RNN, LSTM) were trained over sequential data using opti- mized parameters of learning rate, batch size, and number of epochs.

**Evaluation:** Each model's performance was evaluated in terms of top metrics like accuracy ,precision and F1 Score.

**Deployment:** The proposed model findings could be used in deploying live bot detection systems where service providers

would employ dynamic analyses to refine the security and moderation processes. **Figure 1** shows a flow diagram representing the Twitter bot detection process, which includes data collection, feature extraction, pre-processing, model selection, training and testing, Evaluation, result visualization, and deployment.

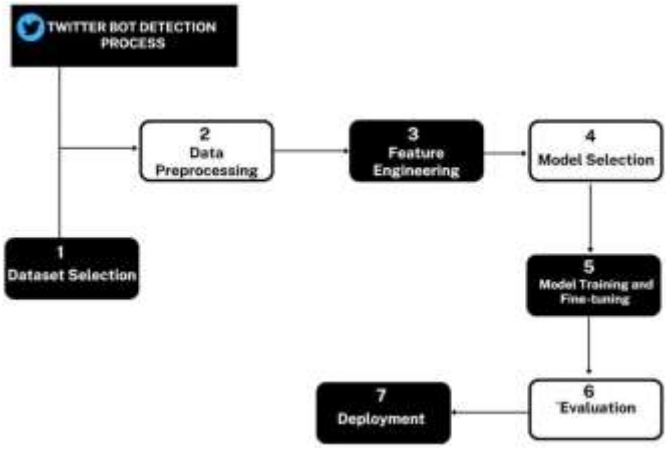


**Figure 1:** Twitter Bot Detection Process

## III. Dataset Analysis

In our analysis of the Cresci 2017 data set, we first conducted a preliminary validation test. The data set had 8,386 records with 16 attributes that included engagement measures like retweets, replies, favorites, hashtags, URLs, mentions, and account metadata like status count, followers count, friends count, and listed count.

### 1. Dataset Validation and Integrity Check

- We were able to load the data set successfully and tested for column structure.
- There were no missing values in any of the features.
- No records were verified and found to be duplicates.
- A label distribution check showed skewness: 58.57% labeled as automated accounts (label = 1), 41.42% labeled as real accounts (label = 0).

### 2. Feature Scaling and Dataset Splitting

- We extracted some relevant numerical features like 'followers_count', 'friends_count', 'ffratio', 'statuses_count', 'retweets', and 'replies' for model training.
- StandardScaler was used to normalize the numerical data.
- No records were verified and found to be duplicates.
- The data set was divided into 80% train set and 20% test set, with 6,708 train samples and 1,678 test samples.

### 3. Dataset Distribution Summary

Finally, we have made a summary of the dataset information distribution in Table I. The data set contains 8,386 records, of which 4,912 are categorized as bot accounts and 3,474 are categorized as real accounts. The data set was again split into training subset and testing subset with the use of 6,708 records for training and 1,678 for testing. The splitting offers an effectively structured data set for deep learning-based machine learning bot detection.

This methodology of big data analysis guarantees the data are well preprocessed for machine learning and deep learning- based bot detection modeling.

**TABLE I:** Dataset Information Distribution

| Name | Size |
|---|---|
| Total Dataframe | 8386 |
| Bots | 4912 |
| Genuine Accounts | 3474 |
| Train Data | 6708 |
| Test Data | 1678 |

## IV. Literature Survey

To facilitate systematic review of the Twitter bot detection literature, the papers have been categorized according to the prevailing methodologies employed. There are two prevailing paradigms: 1. Traditional Machine Learning Methodologies, employing classifiers like Random Forest, Naive Bayes, and SVM, and heuristic and feature-based ones to identify bots; and 2.Deep Learning and Graph-Based Methods, utilizing the most advanced neural networks, such as LSTMs, BERT, and graph-based models, to improve detection by relational, contextual, and linguistic embeddings.

Nguyen et al. [1] gives an overview of progress in social bot detection with focus on supervised machine learning methods because they are based on labeled data. Naive Bayes, Random Forest, SVM, and deep learning models are proposed as classifiers with Twibot-20 as the benchmark standard dataset. Relational Graph Neural Networks (R-GCNs) are discussed in the context of improved bot detection using modeling user relationships. Data quality, evasions, and feature selection are still the issues. Emergent studies need to drive the benchmark- ing models, improve feature extraction by using deep learning technology, and create real-time detection capacities to stay adaptive for new social media threats.

Bui et al. [2] compares and contrasts conventional feature-based and text-based methods with new graph-based methods of Twitter bot detection. Conventional methods categorize accounts by their extracted features or NLP but are unable to counter advanced bots that imitate users. Graph-based detection represents users as nodes and interactions as edges, utilizing network structures to distinguish bots from actual users. Machine learning combination with graph-based methods improves accuracy by considering social relations. Concerns are the bot adaptive behavior, dataset limitations, and comparative studies. Adaptive detection techniques for adap- tive social media threats need to be developed continuously through research.

Zubiaga et al. [3] presents malicious user detection in social networks from dynamic graph-based models that dynamically update user embeddings in real-time. In contrast to static models, this method considers temporal changes in behavior,

which allows for proactive detection. The model employs semantic features of hashtags and URLs to improve context comprehension. Different classifiers, such as Support Vector Classifier (SVC), Multi-Layer Perceptron (MLP), K-Nearest Neighbors (KNN), and Random Forest (RF), are compared and RF is most appropriate for early user classification. The research introduces the effectiveness of the model in terms of obtaining an F1 score of over 0.75 using constrained data, which proves that it can identify malicious users efficiently and effectively.

Gheewala et al. [4] introduce Twitter spam detection, underlining the ineffectiveness of conventional blacklisting and heuristic approaches owing to the evolvability of spammers. ML-based systems are currently a necessity that employ statistical, graph-based, and syntax-based features for classification. Methods such as SVM, Random Forest, Decision Trees, and hybrid models improve detection precision with hybrid approaches being more robust. The issues are evasive spammer behavior, data imbalance, spam drift, and computational expense of graph-based methods. The research surveys numerous classifiers and datasets and discusses limitations in feature extraction and social network representation. The future research intends to optimize feature selection, construct adaptive systems, and enhance computational efficiency.

Narayan et al. [5] reviews machine learning methods of Twitter bot detection through algorithmic performance and feature extraction. Classic heuristic methods were ineffective in detecting advanced bots, and therefore the use of ML algorithms with large datasets became prevalent. Decision Tree, Random Forest, and Multinomial Naive Bayes are the key algorithms, of which Random Forest enhances precision but is affected by data redundancy problems. Good detection relies on correctly annotated datasets and metrics like accuracy, precision, recall, and F1-score. Dynamic bot actions and real- time detection are a challenge. Models that adapt using deep learning and NLP should be what the future research targets for enhanced accuracy.

Wei et al. [6] analyze supervised, unsupervised, and deep learning approaches towards detecting Twitter bots. Supervised approaches are prevalent in using classification algorithms and statistical behavior characteristics, whereas unsupervised approaches leverage clustering to discover patterns without labels. Recent deep learning advancements bring into play neural networks, and BiLSTM models utilize word embed- dings such as GloVe and Word2Vec for better detection. The BOTLE model shows the power of linguistic embeddings in bot detection. Augmenting detection algorithms, the scope of unsupervised learning applications, and enhancing model interpretability are avenues that future work should explore to facilitate more accurate bot detection.

Feng et al. [7] present the TwiBot-20 dataset, a significant leap in Twitter bot detection, with a large-scale, multi-modal setting of 229,573 users, 33,488,192 tweets, and 455,958 follow links. It is the first publicly released dataset to contain user follow relations, with semantic, property, and neighborhood knowledge. The dataset was annotated thoroughly by crowdsourcing with manual checking to guarantee high annotation accuracy. Experiments with TwiBot-20 identify the most important bot features, including lower screen name probability and fewer tweet actions. By establishing a new benchmark for bot detection experiments, TwiBot-20 allows effective countermeasures for blocking malicious bots on social media to be created.

Uyen et al. [8] outline the progress of spambot detection from conventional heuristic to sophisticated machine learning, with a special emphasis on Long Short-Term Memory (LSTM) networks. The earlier techniques had depended on rule-based systems with user behavior under observation, while contem- porary techniques consist of supervised learning and unsu- pervised learning. Deep learning, specifically BiLSTM, has proven effective in modeling user behavior without requiring deep feature engineering. Yet, there are still challenges because spammers evolve over time and because social media is a dynamic platform. Future work must improve the robustness of models against attacks and investigate multi-modal data sources for more accurate detection.

Periasamy et al. [9] state the role of deep learning mod- els, especially transformer-based models such as BERT, in improving Twitter bot detection. These models utilize large datasets and multilingual support to attain improved detection. Transfer learning has also increased text feature extraction with improved performance. General metrics such as accuracy, precision, recall, and F1 score are used to estimate effective- ness, but false positive reduction remains an issue. Future studies should be oriented towards the hybrid methods that combine classical machine learning and deep learning methods while addressing ethical concerns related to bot detection on social media platforms.

Sadiq et al. [10] address the problem of deepfake detection on social media, in which generative language models generate sophisticated machine-generated text. Rule-based and statisti- cal methods are not equipped to handle dynamic generative text, and hence there is a requirement for advanced machine learning and NLP. Techniques like Random Forest and SVM have been used for feature extraction but are human-feature- engineering dependent. The study evaluates deep models such as CNNs and BERT using the Tweepfake dataset, and it high- lights the employment of FastText embeddings to boost text representation. However, the trade-off between false positives and false negatives is still an issue, especially in real-time scenarios with dynamic user-generated content.

Arin et al. [11] emphasize the functionality of deep learning to identify bots, the efficiency of RNNs and LSTM networks, particularly in monitoring sequence patterns on tweets. They presented in their study a novel structure that consists of three LSTMs and a fully connected layer to process tweet content, metadata, and account descriptions and did a better perfor- mance than the Mixed-2021 dataset. The transformer model like BERT enhances bot detection with its ability to high- light deep semantic nuances. But such models require large labeled datasets and are computationally expensive. Hybrid approaches involving CNN, LSTM, and attention mechanism

with embeddings like GloVe and FastText improve detection accuracy. Ensemble models and real-time detection systems must be investigated in the future to mitigate such problems as false positives and bot behavior changes.

Pramitha et al. [12] provide an overview of the development of Twitter bot detection from rule-based classifiers to sophisticated machine learning methods. Early approaches relied on metadata features such as follower counts and tweeting rates but failed to keep up with sophisticated bots that were emulating human behavior. More advanced models like Random Forest (RF) and Support Vector Machine (SVM) also enhanced detection using behavioral and text data, while tree- boosting methods like XGBoost also improved performance, particularly when combined with SMOTE for handling imbal- anced datasets. In their research, using Twitter API-scraped data, they found verification status and network features to be the important features for classification. XGBoost performed better than RF in the detection of advanced bots. Hybrid meth- ods combining account-level and content-level analysis must be investigated further and the application of deep learning methods such as GANs for simulating advanced bot behavior, but ethical considerations related to privacy and profiling still pose major challenges.

## V. RESULTS AND DISCUSSION

Experiments were performed on the Cresci 2017 dataset, a popular benchmark in Twitter bot detection. The dataset has a total of 8,386 accounts with 16 feature columns, including account metadata, behavioral features, and interaction metrics. All of the analysis from preprocessing to model training, and evaluation were done within Google Colab, taking advantage of its cloud-based environment for fast computation and experimentation. Google Colab GPU acceleration was used to accelerate training, particularly for deep learning models such as LSTM and RNN. Robust evaluation was ensured by dividing the dataset into 80% training data and 20% test data, and hyperparameter tuning for maximizing model performance. Six machine learning and deep learning models

— Random Forest, Long Short-Term Memory (LSTM), k-Nearest Neighbors (KNN), Recurrent Neural Network (RNN), Support Vector Machine (SVM), and Logistic Regression — were compared to identify both static and dynamic account behavioral patterns. Accuracy, precision, recall, and F1-score performance metrics were utilized to measure the performance of the models in detecting bot accounts. The evenly balanced ratio of bot accounts and human accounts gave a balanced plat- form for training and testing, enabling a thorough comparison of the strengths and weaknesses of each model.The detailed results of the analysis of the algorithm performance, accuracy, and other important metrics for each model are presented in **Figure 2**.

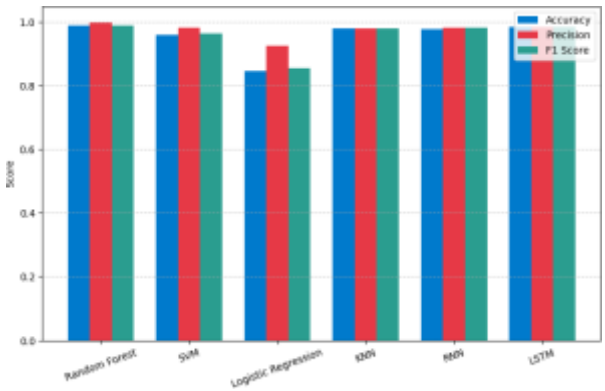


**Figure 2:** Algorithm Performance Analysis

*Performance Insights*

- Random Forest had the best performance with 98.69% accuracy and 0.9887 F1-score. The precision value of 0.9979 is very high, which means that Random Forest effectively identifies bot accounts with fewer false posi- tives. Random Forest's ensemble paradigm, which it uses to average out a collection of decision trees' decision, allows it to detect subtle patterns in the data such as uncommon interaction patterns and irregular bursts of activity commonly displayed by bots.
- LSTM also did extremely well at 98.27%. LSTMs are best with sequential data and can identify temporal pat- terns in behavior on accounts, such as timing of tweets and shifting patterns of behavior. This makes LSTMs particularly powerful for detecting sophisticated bots that simulate human behavior over a period of time.
- Both KNN and RNN performed well, with accuracies of 97.85% and 97.74%, respectively. That KNN worked indicates that bot and human accounts in the Cresci 2017 dataset have different clusters in feature space, and that RNNs learned activity sequences well, albeit not quite as well as LSTMs.
- SVM was good but trailed the best models at 95.95% accuracy. Though it attained a high precision of 0.9806, its lower F1 score (0.9640) implies that it would likely have issues identifying some bots accurately, especially in complex and overlapping feature sets, and therefore ranks lower than the best models.
- Logistic Regression, the simplest model, yielded the lowest accuracy (84.56%). Despite its good precision, its lower F1-score(0.8552) points to the failure of the model to identify a large number of bot accounts, which reflects the limitation of linear models in addressing complex, multi-feature datasets like Cresci 2017.

*Confusion Matrix Analysis*

The confusion matrices give us a better insight into the classification choices of every model, beyond common performance measures. We chose the matrices of Random Forest,

LSTM, and SVM because they embody different strengths and weaknesses in different bot detection contexts. The Random Forest matrix (**Figure 3**) indicates almost perfect classification with 963 true positives, 693 true negatives, 2 false positives, and 20 false negatives. Its ensemble method is best for recognizing simple and compound bot patterns and is the most stable model for real-world application. On the other hand, LSTM's matrix (**Figure 5**) shows how it can recognize bots with temporal patterns, such as coordinated bursts of tweets or varying interaction patterns, with 957 true positives, 691 true negatives, 4 false positives, and 26 false negatives. However, its relatively greater number of false negatives compared to Random Forest indicates that it might struggle to find bots without strong sequential patterns. In contrast, the SVM confusion matrix (**Figure 4**) indicates high accuracy but lower recall at 911 true positives, 699 true negatives, 18 false positives, and 50 false negatives. That is, SVM accurately classifies many accounts but struggles with oblique or adaptive bots that are poorly bounded within its decision boundaries. These matrices all reflect the manner in which Random Forest achieves a tradeoff between precision and recall, LSTM is better at time-oriented behavior, and SVM fails on complex or dynamic behavior — influencing a fair understanding of each model's detection.

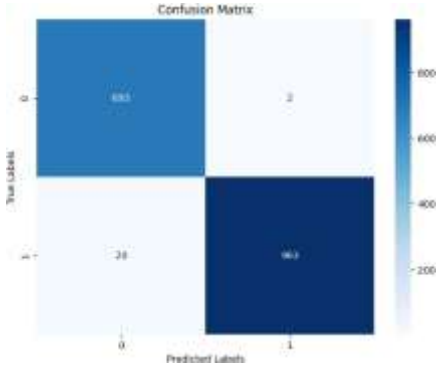


**Figure 3:** Confusion Matrix of Random Forest Model

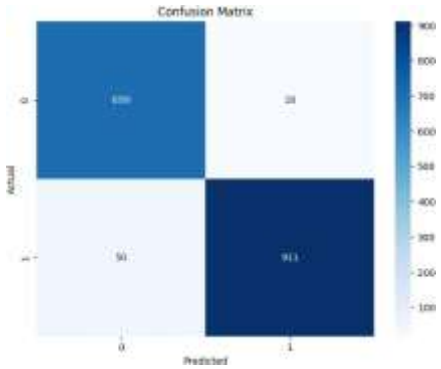


**Figure 4:** Confusion Matrix of SVM Model

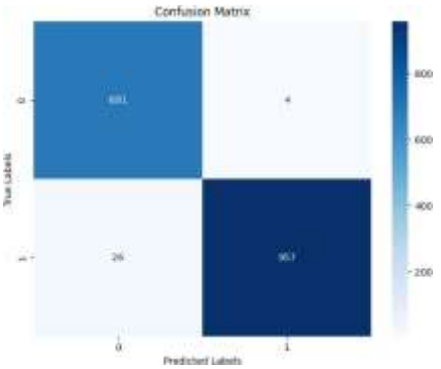


**Figure 5:** Confusion Matrix of LSTM Model

## VI. CONCLUSION AND FUTURE DIRECTIONS

The current trend in bot detection has shifted from conventional methods using predefined rules to the more advanced use of machine learning and deep learning. This study presented a comparative evaluation of the performances of different models, namely: Random Forest, SVM, Logistic Re- gression, KNN, RNN, and LSTM, in the detection of Twitter bots. Random Forest outperformed the rest of the models with the highest accuracy and precision, which were 98.69% and 99.79%, respectively, which makes this model the most suit- able for this dataset.Despite all the improvements, challenges continue to exist due to the dynamic and ever-changing nature of the problem, particularly with the strategies that bots adopt in their work. While machine learning algorithms excel at developing structured mechanisms for analyzing large volumes of data, their performance eventually suffers as the algorithms come to terms with the flexible internal environment nurtured by deep learning models such as LSTM, which have a natural suit for identifying sequential patterns and therefore have a role to play in capturing behavioral patterns.In the future, a study should focus on graph-based techniques plus multi-modal information sources, embedding user behavior, topology of the network, and content measurement, along with other features to improve the efficiency of detection. The combination of machine learning and deep learning methodologies in hybrid models can further expand their potential in improving per- formance as well as adaptability.Ethics, privacy, and fairness considerations should also take center stage to ensure that bot detection systems are truly unbiased and transparent. In the near future, a proper blend of statistical techniques along with domain knowledge is going to be critically important for developing sound and scalable solutions for the detection of bots in online environments.

## REFERENCES

[1] D. D. Nguyen, A. Nguyen-Duc *et al.*, "Supervised learning models for social bot detection: Literature review and benchmark," *Expert Syst. Appl.*, vol. 238, p. 122217, 2024.

[2] T. Bui and K. Potika, "Twitter bot detection using social network analysis," in *Proc. 4th Int. Conf. Transdisciplinary AI (TransAI)*. IEEE, 2022, pp. 87–88.

[3] R. Sa´nchez-Corcuera, A. Zubiaga, and A. Almeida, "Early detection and prevention of malicious user behavior on twitter using deep learning techniques," *IEEE Trans. Comput. Social Syst.*, 2024.

[4] S. Gheewala and R. Patel, "Machine learning based twitter spam account detection: A review," in *Proc. 2nd Int. Conf. Comput. Methodologies Commun. (ICCMC)*. IEEE, 2018, pp. 79–84.
[5] N. Narayan, "Twitter bot detection using machine learning algorithms," in *Proc. 4th Int. Conf. Electr., Comput., Commun. Technol. (ICECCT)*. IEEE, 2021, pp. 1–4.
[6] F. Wei and U. T. Nguyen, "Twitter bot detection using neural networks and linguistic embeddings," *IEEE Open J. Comput. Soc.*, 2023.
[7] S. Feng, H. Wan, N. Wang, and J. Li, "Twibot-20: A comprehensive twitter bot detection benchmark," in *Proc. 30th ACM Int. Conf. Inf. Knowl. Manag. (CIKM)*. ACM, 2021, pp. 4485–4494.
[8] F. Wei and U. T. Nguyen, "Twitter bot detection using bidirectional long short-term memory neural networks and word embeddings," in *Proc. 1st IEEE Int. Conf. Trust, Privacy, Security Intell. Syst. Appl. (TPS-ISA)*. IEEE, 2019, pp. 101–109.
[9] J. K. Periasamy, R. Srinidhi, and S. Srividhya, "A deep learning approach for twitter bot detection," in *Proc. 1st Int. Conf. Comput. Sci. Technol. (ICCST)*. IEEE, 2022, pp. 374–377.
[10] S. Sadiq, T. Aljrees, and S. Ullah, "Deepfake detection on social media: Leveraging deep learning and fasttext embeddings for identifying machine-generated tweets," *IEEE Access*, 2023.
[11] E. Arin and M. Kutlu, "Deep learning-based social bot detection on twitter," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 1763–1772, 2023.
[12] R. B. H. E. Pramitha and N. Qomariasih, "Twitter bot account detection using supervised machine learning," in *Proc. 4th Int. Semin. Res. Inf. Technol. Intell. Syst. (ISRITI)*. IEEE, 2021, pp. 379–383.