SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V12I7P112

Volume 12 Issue 7, 148-173, July 2025

Original Article

Smart Bot Detection for Twitter/X: A Systematic Analysis of Machine and Deep Learning based Methods

Rekha Jangra[1], Abhishek Kajal[2]

[1,2]*Department of Computer Science and Engineering, Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India.*

[2]*Corresponding Author : drabhishekkajal@gmail.com*

Received: 06 May 2025 Revised: 08 June 2025 Accepted: 09 July 2025 Published: 31 July 2025

Abstract *- A significant rise in online social networks has been witnessed over the past decade, particularly on social media network platforms such as X (previously known as Twitter), Instagram, and Facebook accounts, which recorded a steep hike. This further leads to the rise of automated accounts known as bots, devised to replicate the behavior of organic user accounts automatically to disseminate false information or spread spam. For identifying these social bot accounts, many machine and deep learning methods have been proposed on X accounts' features, and based on tweets' text contents through sentiment analysis for diverse datasets; the leading works have been reviewed in this research article. However, comparing the efficacy of diverse research works in terms of problem statement, methodologies, and various evaluation parameters was extremely difficult. However, the authors put effort into a systematic literature review. It has been observed that SVM followed by RF are the most used ML algorithms for Twitter bot detection, where RF achieved the highest accuracy of 94.87% on account profile features. LSTM is also observed as the most employed DL algorithm for Twitter bot detection. While RoBERTa achieved the highest accuracy of about 98% on the COVID-19 dataset, followed by CNN on the Arabic Spam dataset. The article also summarizes the potential of methods to enhance bot detection performance and scalability over machine and deep learning methods. In addition, authors presented a demonstration with the execution of leading ML and DL methods with the combination of ReLU activation function and Adam optimizer on diverse X datasets. They presented the respective results in tabular form. During simulation with leading ML-based techniques, accuracy yielded for SVM, Naïve Bayes, and Random Forest was 81.38%, 79.47%, and 85.77%, respectively. At the same time, accuracy in the case of DL-based LSTM and Bi-LSTM approaches was 88% and 89%, respectively. Overall, this review article will provide a significant blueprint for future research on enhancing the performance of bot detection models for different online social networks.*

Keywords *- Online social networks, Bot detection, Twitter dataset, Machine Learning, Deep Learning.*

1. Introduction

In the quickly shifting digital surroundings, it has become very difficult to identify automated bots and real users. These bots may be responsible for fraudulent, spam, and other destructive actions. These programs could be designed to achieve a wide range of tasks, from simple to complex. Bots can operate in various environments, including websites, messaging platforms, social media networks, and online games. Existing bot detection research is considered to use deep neural networks [1]. People can communicate with one another on a massive scale via Online Social Networks (OSNs). Conventional research also conducted a systematic literature review on social media bots detection [2]. This research work has considered the methodology and limitations of conventional bot detection and classification approaches. Coordinated actions by groups of bots, known as botnets, may allow for more damaging and powerful assaults over Twitter [3]; thus, bot detection has been conducted with a reduced feature set. First time bots were identified by John McCarthy in 1996 by using multiple approaches to attack computers with the execution of Dendritic Cell Algorithm (DCA) to discover bots. Nowadays, bot identification has become one of the most challenging tasks due to the evolution of technology in recent years [4]. Several studies have suggested new evolving bot-detecting techniques.

An innovative method for detecting bots includes using the correlation of API calls. API calls refer to the exchanges of information between clients and servers that take place when users interact with internet services. Examining patterns and correlations in these calls makes it feasible to distinguish between human users and automated bots with enhanced precision [5]. Social bots are more likely due to social networks' increasing popularity in recent years. R. Wald et al studied a Twitter dataset composed of thousands

of tweets done by bots so as to classify the most relevant features for predicting an interaction if it is carried out by a human or bot [6].

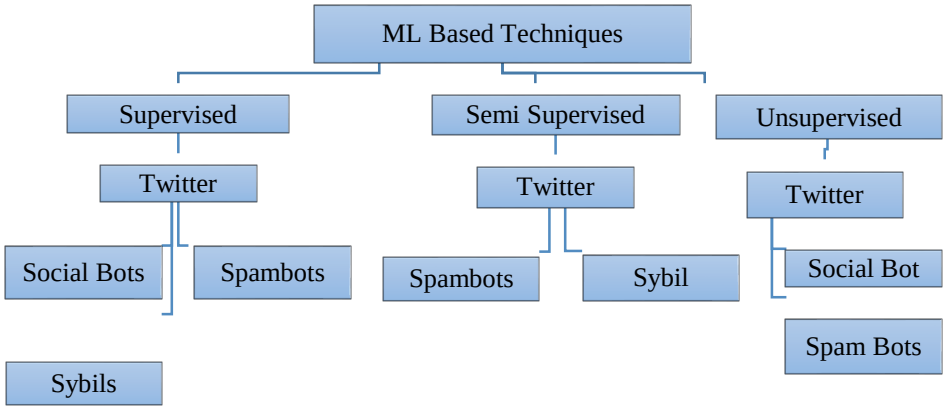
Botnets are one of the biggest cybersecurity threats to organizations in the current times. Several cybercriminals employ botnets as primary vectors. The authors suggest a botnet detection method using traffic behaviour analysis in this study. This approach classifies network traffic behaviour using various ML techniques [7]. Twitter statuses using an NB can be a valuable tool for language learning by providing insights into the emotional tone of social media content. This approach leverages the simplicity and effectiveness of the Naive Bayes algorithm to categorise text into positive, negative, or neutral sentiments, helping learners understand emotional expression and context in real-world language use [8]. Sentiment analysis may discover Twitter bots by comparing human and artificial viewpoints. Human users express [9] more diverse and stronger feelings. Sentiment is important in bot identification since bots tend to have less emotional diversity. Political bots and right-wing online activity shaped public conversation and voter impressions in Japan's 2014 General Election. Critics accused Prime Minister Shinzo Abe's government of a secret nationalist agenda, claiming internet platforms were used to promote pro-Abe emotions and crush opposition.

Political bots [10, 11] are used to flood social media with supporting remarks and influence the narrative. This technology-politics junction showed how digital techniques affect democratic involvement and the complexity of current political campaigns. A previous study proposed a contextual LSTM-DNN. These conventional research works showed that architecture can distinguish bots and individuals from a tweet with an AUC above 96%. Researchers have developed an AI-powered approach to detect and classify social bots on Twitter, using sentiment analysis, data mining, and classifiers like SVM [12]. They proposed a framework to identify autonomous entities based on user behaviour, tweet syntax, and sentiment analysis. A crawling operation was conducted on Twitter to retrieve random tweets, and user-specific tweets and characteristics were retrieved by aggregating tweets according to the senders. Contrast patterns were considered in some research for bot classification [13], and contrast guide-based classifiers provided a model that could be understood by humans.

Bot detection is also required to identify valid social posts, particularly during any pandemic [14]. The reliable method chosen by the authors is using deep learning algorithms for bot detection, which considers three layers: extracting features from tweet text, extracting time-based features from tweet metadata, and combining retrieved joint pleased traits with activist features [1]. ANNs have been used in some research studies due to their superior performance, surpassing conventional approaches by 75.7%. Deep learning has been used in meta-analysis to assess the performance of ANN, with results showing it excels in classification compared to prediction [15].

1.1. Role of ML and DL Techniques in Bot Detection

It has been observed that different machine learning (J. V. Fonseca Abreu, 2020) and deep learning techniques play a significant role in bot detection (X. Dong, 2010). Research made use of well-known machine learning techniques to identify bots on Twitter. SVM (A. Foysal, 2019), Naive bayes (Christos Troussas, 2013), and random forest methods (R. Wald, 2013) have been used to find the bots. However, these techniques did not provide the required accuracy level. However, Deep learning techniques such as LSTM (S. Kudugunta, 2018), Bert (A. Wadhawan, 2021) and Roberta have been provided to provide high-level accuracy by different authors. Some authors have proposed an algorithm-based social bot identification model for ANN (H. Ping, 2019). The performance and accuracy of algorithms used in this research are dependent on various parameters such as epochs, batch size, and optimizer. There has been significant usage of machine learning algorithms for bot detection. Any of the three ML methods could be suitable.



1 Classification of bots in ML

In contrast to unsupervised learning, which does not take labelled training data into account, supervised learning makes use of such data. One method of training machine learning models is Semi-Supervised Learning (SSL), which combines labelled and unlabelled data. Supervised Twitter account analysis identifies social bots, spambots, and sybils. It is possible to identify Sybil and spambots on Twitter using semi-supervised and unsupervised methods, respectively.

1.2. Working of Bot Detection

Bot detection process considers Feature Extraction using machine learning, where user behaviour log, IP address, typing pattern and mouse movement are analysed. Then the bot detection system performs sequence analysis using DL mechanisms such as the LSTM technique to detect bot behaviour. The classification process distinguishes between a bot and a normal tweet. Figure 2. Integration of Machine learning and an LSTM-based Bot detection and classification system.

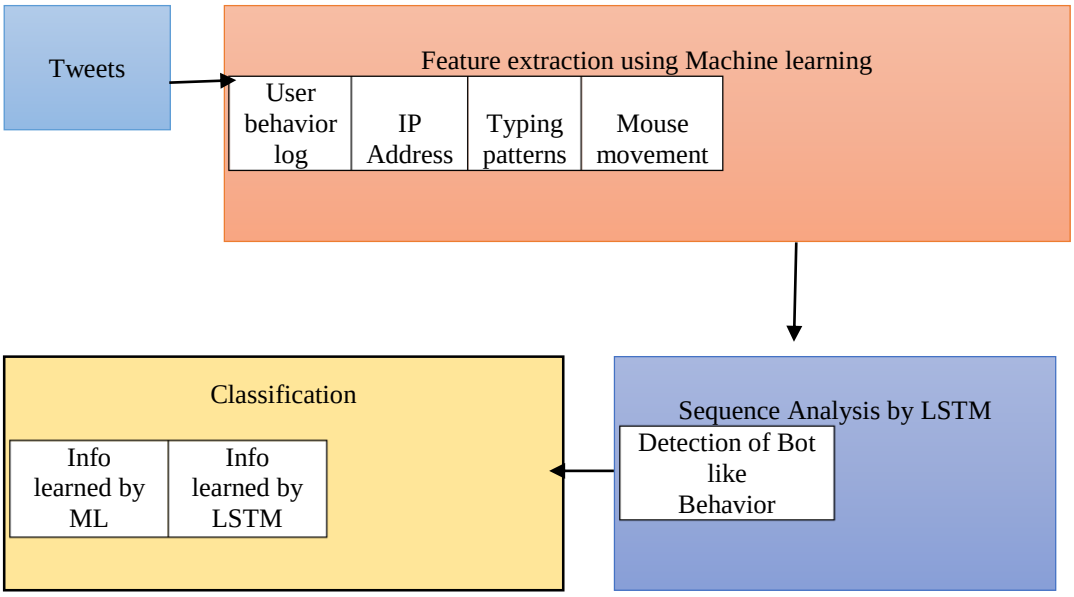


Fig. 2 Integration of machine learning with an LSTM-based bot detection and classification system

1.3. Bot Detection

Bot detection research is extensive and relevant to tackling social media platform automation issues. The following are the broad study areas in this domain. Bot Detection Techniques examines the newest machine learning, deep learning, and hybrid bot detection methods. Discover how these methods improve bot behaviour identification accuracy and flexibility. Temporal and Long-Term Patterns examine bot activity temporal dynamics and long-term trends. Understand bot strategy evolution and suggest detection tools to capture and respond to these changes. Study Other Social Media Platforms to expand the investigation beyond X (formerly Twitter). Explore the specific traits and problems of bots on Facebook, Instagram, and new platforms to expand detection methods. User-Centred Methods explore user-centric characteristics and behaviours to improve bot identification. Explore how user profiles, sentiment analysis, Detecting Propagators of Disinformation on Twitter and network analysis can identify bots and real users. Impact of Bots on Information Spread evaluates bots' effects on information spread and online communities. Research how bots affect public opinion and spread disinformation, and then offer ways to reduce their harmful consequences.

1.4. Research Gap

Research on spam profile recognition and content-based spammer detection is becoming more prominent nowadays. Several research works have been done to identify social bots on the X platform (erstwhile Twitter) using deep regression models in prediction to detect bot and Sybil accounts on social networks. Detecting bots in the Twitter environment using unsupervised learning and a multi-input deep neural network model presents significant challenges. These studies may aid in the development of resilient and flexible detection techniques and enhance our comprehension of social media bot behaviour. The present research paper answers the following questions.

RQ1: What are the main issues/challenges that were faced in conventional research for social bot detection?
RQ2: What Twitter datasets are frequently used for training and evaluating social bot detection models?
RQ3: What are the different machine learning techniques that are used for bot detection?
RQ4: What different DL mechanisms could be used for bot detection?
RQ5: What performance parameters are frequently considered for research in bot detection?

2. Literature Review

2.1. Evolution

Due to bots, ever-changing traits and techniques, detecting them on X (formerly Twitter) has become tough. Simple rule-based approaches were used to spot bots based on specified patterns or behaviors for the first time. The patterns of engagement that bots participate in are often distinct from those that human users engage in. The interrelationships between accounts may be investigated via the use of network analysis. Certain characteristics, centrality of the network, and community detection help identify suspicious conduct. Bot detection must be able to react, since bots are constantly upgrading their techniques.

Continuous monitoring and changes to detection algorithms are necessary to remain one step ahead of evolving bot techniques. When it comes to exchanging information and enhancing overall bot detection skills, collaborative efforts between academics, organizations, and social media platforms are very necessary. Common knowledge of the ever-changing bot ecosystem may be achieved via open-source technologies and shared datasets on the internet.

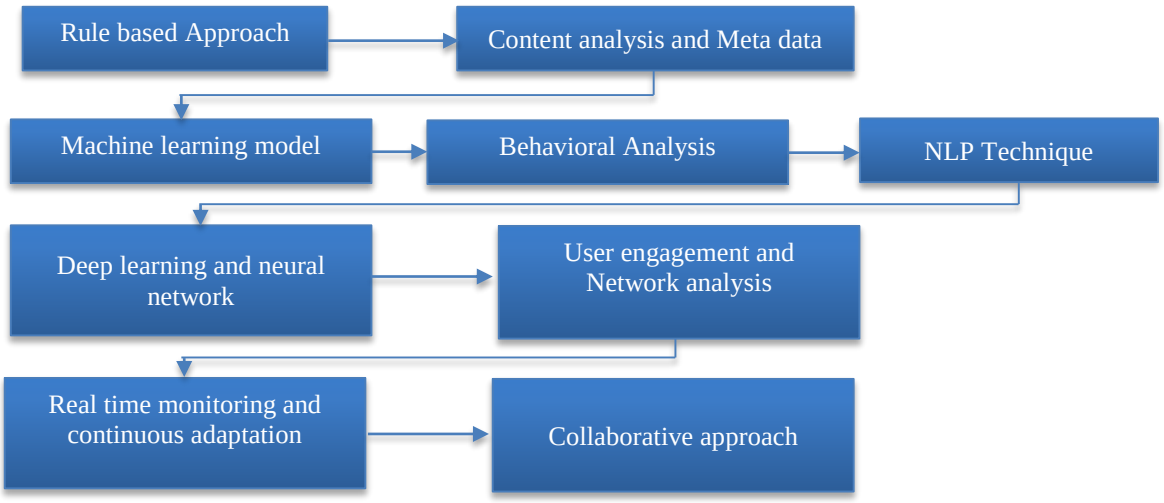


Fig. 3 Bot detection evolution

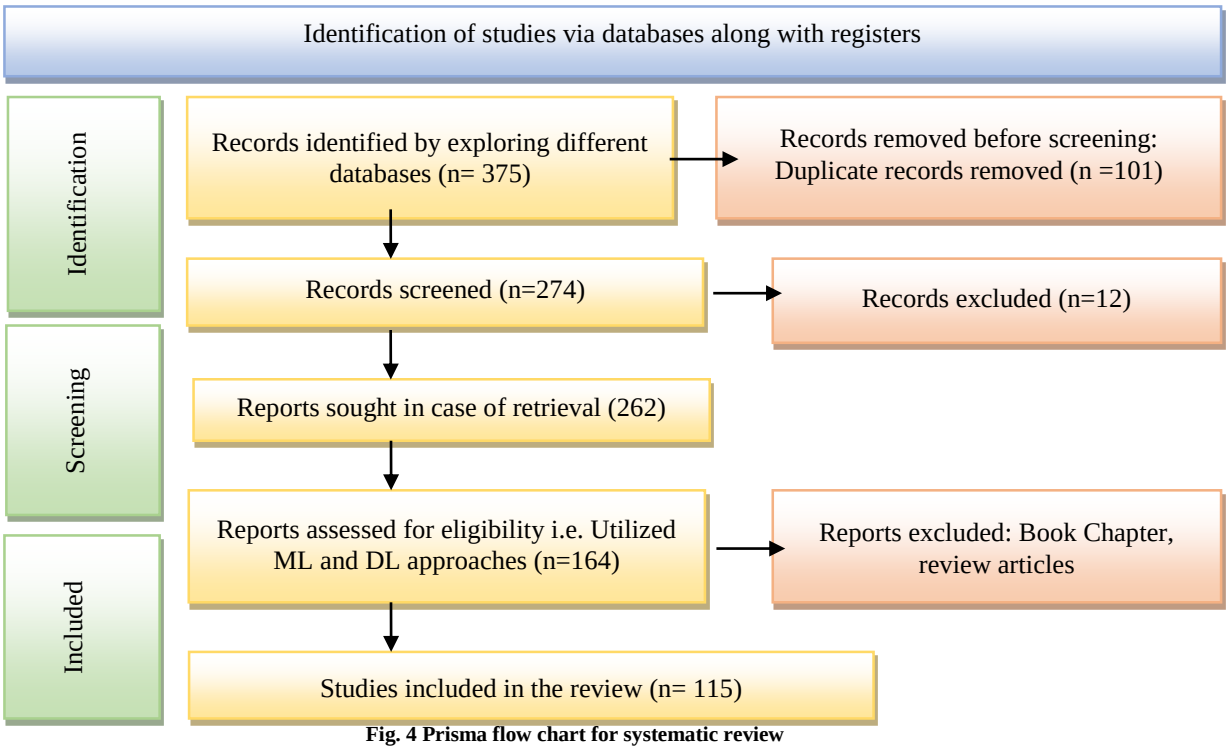


Fig. 4 Prisma flow chart for systematic review

2.2. Systematic Literature Review

To perform a literature review on the social bot detection domain, comprehensive research has been conducted to search relevant research articles from high-quality database repositories, such as ScienceDirect, IEEE, and Springer. This review included articles published in well-reputed, indexed journals that accurately documented scientific proceedings.

Selection criteria for the papers aimed to select techniques for Bot detection from the X platform using various ML and DL methods, mainly analyzing conventional ML techniques such as SVM, Naïve Bayes for classification and investigating DL techniques such as LSTM, Bert, RoBERTa for the detection of bots. Figure 4 shows a flow diagram of the PRISMA selection process, which found 375 articles using a main search strategy that focused on Bot detection on the X platform. A total of 274 papers remained after duplication was removed. This number was further reduced to 164 full-text articles after applying an additional criterion, which was further subjected to a few additional exclusion criteria. ML and DL were deciding factors in the selection of limited papers. Finally, 115 prominent studies were shortlisted in this review.

Table 1. Year-wise distribution of papers

Year	Number of papers
2008	1
2010	1
2013	3
2014	2
2015	1
2016	1
2017	2
2018	8
2019	13
2020	16
2021	13
2022	22
2023	22
2024	6
2025	4

The selected distribution of papers published in the field of research exhibits an upward trend, reflecting the growing interest in and significance of the subject. The journey began modestly in 2008 with only one paper, gradually picking up momentum with one paper each in 2010. However, from the year 2013 onwards, a more noticeable increase was observed, with the number of papers climbing up to 3 that year. During the years 2014, 2015, 2016, and 2017, research work was considered to be 2,1,1,2, respectively. Later in 2018, the number of research papers increased to 8, whereas it was 13 in 2019, 16 in 2020 and 13 in 2021. This trend continues with fluctuations over the subsequent years, reaching peaks of 22 papers in the years 2022 and 2023. Work has included 6 of the most relevant papers from 2024 and 4 papers from 2025. Notably, there's a substantial increase in publications from 2018 onwards, indicating a surge in research activity and possibly reflecting advancements, emerging technologies, or increased funding in the field. This progression underscores the evolving nature of the research landscape, with researchers contributing to the collective knowledge base through their scholarly endeavours.

Table 2. Technique-wise papers classification

Technique	No of Articles	Citation
DL Review Paper	7	[9, 19, 22, 27, 33, 63, 103]
ML Review Paper	31	[3, 10, 16, 23, 26, 36, 40, 46, 47, 48, 49, 52, 53, 56, 57, 64, 68, 73, 75, 79, 23, 78, 82, 84, 86, 89, 56, 99, 64]
Machine learning		
Random Forest	3	[18, 31, 79]
Naïve Bayes	2	[6, 8]
SVM	4	[17, 54, 65, 67]
Semi-Supervised Learning	1	[107]
Deep learning		
LSTM	7	[1, 2, 14, 1, 32, 41, 43]
BI-LSTM	2	[28, 41]
BERT	4	[28, 29, 30, 34]
RoBERTa	3	[29, 34, 39]
CNN	3	[2, 44, 88]
Hybrid	8	[3, 39, 40, 45, 45, 90, 97, 102]
Other techniques		
VADER	2	[13, 24]
Reinforcement	2	[35, 50]
One Class Classification	2	[25, 26]
NLP	2	[30, 100]
Deep Contrastive Graph Clustering	1	[104]
Multimodal Approach	1	[105]
Sentiment/Discourse Analysis	1	[106]
Evaluation of Confidence & Bot Detection Robustness	1	[108]

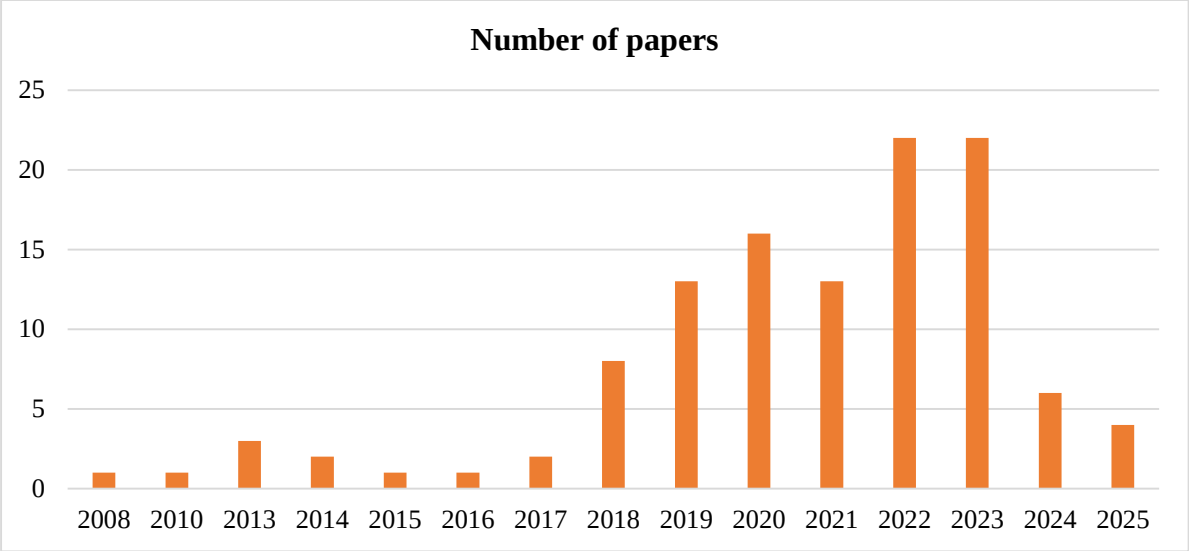


Fig. 5 Year-wise research representation

Considering the number of papers on different methods, a technique-wise tabular representation of the same has been drawn in the following table for the convenience of researchers interested in reviewing any particular method to carry on research work in bot detection.

It considers different research works related to ML and DL for the detection of bots on the X platform, as reflected in the table. This mechanism is effective in analyzing social media texts.

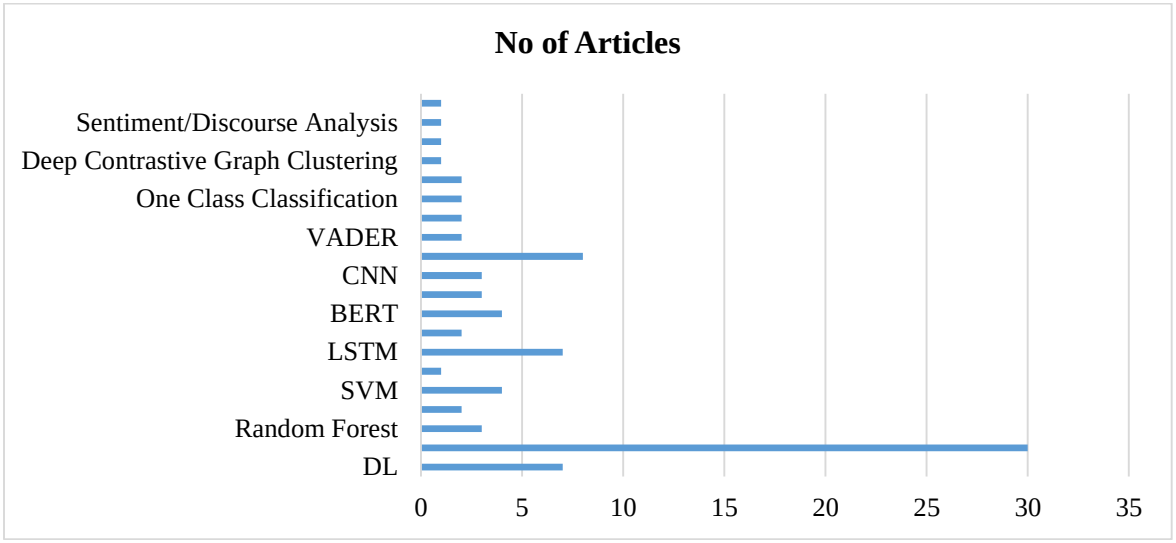


Fig. 6 Chart-based distribution presenting the percentage of articles considered for different techniques

Considering the mechanisms used for machine learning, research papers related to Random Forest, Naïve Bayes and SVM have been considered. The present work has used 3 prominent papers related to Random Forest, two papers of Naïve Bayes and four papers of SVM [17]. In the same way, DL techniques are considered in the present research work. The LSTM model, which is a popular deep learning mechanism, has been frequently used for bot detection. LSTM resolves the issue of overfitting by integrating a dropout layer. Thus, the seven most crucial papers of LSTM are considered. Among deep learning mechanisms, BERT is capable of providing better accuracy than LSTM. The present work has considered four benchmark papers on the BERT technique. Afterwards, three research papers were considered for RoBERTa. Considering major enhancements in deep learning, the three best papers related to the convolutional neural network and nine papers of the hybrid deep learning technique are considered. Moreover, there are other mechanisms that have been frequently used for bot detection and data classification, such as reinforcement learning, one-class classification and NLP. The present work has considered two prominent papers on each of these

techniques. Overall, the distribution of papers across different technologies highlights the diversity of approaches employed in research, with each technology offering unique advantages and applications within the field. Considering the above table, the following figure presents a pie chart representation of the same.

2.3. Existing Research Work

This section considers different authors' contributions in the area of bot detection using different mechanisms.

To identify bots in tweets, S. Kudugunta (2018) proposed an LSTM-based DNN that utilizes both content and metadata to detect and classify bots [1] accurately. Z. Ellaky's (2023) research aimed to identify reliable methods for identifying SMBs. The studies included in this SLR were published between 2008 and 2022 [2]. Using a publicly available dataset and a handful of expressive variables derived from basic user profile counters, J. V. Fonseca Abreu (2020) examined four ML systems for Twitter bot classification [3]. Social bot identification, with a focus on Twitter, was tackled by Y. Al-Hammadi (2008) using a pattern-based categorization technique [4]. In 2010, X. Dong put out an innovative Windows bot detection technique that relies on API call correlation [5]. To find out which characteristics of 610 Twitter users were most useful in forecasting whether these individuals would engage with the bots, R. Wald (2013) analyzes the dataset [6]. By applying ML to categorize network traffic behavior, D. Zhao (2013) suggested a novel method for detecting botnet activities [7]. To find out what makes people different from bots, J. P. Dickerson (2014) compiled a set of variables that may be featured, including ones that are network, language, and application oriented [9].

A technique for classifying hate speech on Twitter using DL was presented by B. Gambäck (2017). Each tweet is sorted into one of four predetermined categories by the classifier: non-hate speech, racism, sexism, or both [12]. S. Kudugunta (2018) proposed an LSTM-based CNN that leverages context along with information to detect Twitter bots [1]. To identify Twitter spam campaigns in real-time, Z. Chen (2018) presented an unsupervised method. Over long periods, the bot groups that they have identified tweet identical text with shortened embedded URLs [15]. Research by A. Foysal (2019) examined the actions and impact of Twitter bots on the platform. Twitter became a perfect venue for social manipulation and swaying opinions as its user base grew [16]. To identify social bots, in this case on Twitter, O. Loyola-Gonzalez (2019) used a pattern-based classification method [18]. Searching for tweets using keywords and hashtags formed the basis of the data-collecting procedure, according to L. Corti (2020). They calculated each tweet's negative to positive sentiment scores using AFiNN [19]. H. Ping (2019) proposed DL-based social bot detection (DeBD). The DeBD model was tested on three types of real-world social bot data sets, and results supported our approach [20]. ANNs have great potential for urban geography research, according to G. Grekousis (2019). Researchers discovered that ANNs for urban research conveyed data poorly [21].

Subjects such as sentiment analysis, the structure and attributes of the social network, and dangers including spam, bots, false news, and hate speech were the focus of D. Antonakaki's (2021) attempt to map the current research subjects in Twitter [22]. A new, reusable, and repeatable method for identifying Twitter bots is presented by A. Shevtsov (2020). Hundreds of characteristics retrieved from a Twitter corpus feed into the system's ML process [23]. R. H. Ali (2022) analyzed 7.6M tweets made to capture popular mood towards candidates in the 2020 US Presidential Elections [24]. To enhance Twitter bot detection, J. Rodríguez-Ruiz (2020) suggests one-class categorization [25]. A review of current research using massive publicly available Twitter datasets on these topics was provided by H. H. Chang (2020) [26]. Suppose Deep Learning models were to become the source of misprediction as a result of internal or external malicious impacts. In that case, it might cause problems in our daily lives, as M. I. Tariq (2020) pointed out, as many applications rely on these models to make choices [27]. An algorithm to automatically sort tweets into "hate," "offensive," and "neither" was suggested by B. Wei (2021) [28]. A Hinglish dataset labelled for emotion recognition was provided by A. Wadhawan (2021) [29].

In his study, A. K. Chanda (2021) compared BERT embedding to more conventional context-free word embedding techniques to determine how well they forecast disasters using Twitter data [30]. To detect cyberbullying using Twitter datasets, N. A. Azeez (2021) examined well-known classification methods and suggested an ensemble model [31]. K. N. Alamet (2021) aimed to comprehend and decipher their feelings in this study by using a few DL methods [32]. The most recent methods for detecting bots in tweets have been categorized according to a taxonomy developed by A. Derhab (2021) [33]. A citation suggestion tool was created by Z. Huang (2021) to make the process of writing opinions more efficient [34]. Using an inverse reinforcement learning (IRL) technique, D. Geissler (2022) examined the strategy of the Twitter community [35].

The Italian political class's reaction to the Russian-Ukrainian crisis on Twitter was examined by F. L. De Faveri (2022) [36]. The popularity of N. P. Shetty (2022) on platforms like Twitter, Facebook, and others skyrocketed because of the low cost of membership and the ease with which one could reach a huge audience [37]. Using cross-lingual Transformer models, A. Ryzhova (2022) investigated ways to interpret brief text messages in Eng, Hindi, and Russian, and detect such intolerance [38]. S. K. S. Joy (2022) compared transformer-based models to detect fraudulent COVID-19 news items online [39].

M. Heidari (2022) conducted research on datasets in Hindi, Bangla, and English to identify hostile remarks [40]. S. Biswas (2022) focused on the problem of identifying spam accounts by gathering an Arabic dataset that is well-suited for spam identification [41]. Real-Time Twitter Spam Detections were developed by M. M. Bailey (2022). For spam identification, these systems sort tweets using various classifiers [42]. M. S. Akter (2022) analyzed Twitter chats in English and five other widely known European languages [43]. Various deep learning models, including different models, were used by A. S. Alhassun (2022) to analyze sentiment [44]. New Twitter bot timelines exhibit more irregularities, according to P. Rodrigues's (2022) discovery of Twitter Bot Detection. The need for more research and focus is underscored by this discovery [45]. To evaluate the impact of an asymmetric reaction, V. Vrana (2023) used an innovative approach by combining two AI algorithms [46]. To successfully detect machine-generated text in various datasets, A. P. Rodrigues (2022) presented the GPT Paternity Test (GPT-Pat) [45]. Z. Lei (2022) used Reinforced Self-Supervised GNN Search for Adaptive Social Bot Detection [47]. R. A. Mendoza-Urdiales (2022) conducted an investigation on topic models to identify cyber dangers on Twitter [48]. X. Yu (2023) presented a new machine learning model for identifying bots, utilizing the advanced XGBoost approach [49]. Y. Yang (2023) predicted the intensity of hate in Twitter conversation threads [50]. Y. Wang (2023) aimed to discover significant characteristics that can distinguish between benign and dangerous bots, specifically focusing on abnormal behavior [51]. A. Shevtsov (2023) employed the utilization of MLP and LSTM. The advent of deepfake technology enables the replication of influential political and cultural personalities, facilitating the dissemination of vast volumes of misinformation [52]. Q. Meng (2023) conducted the use of turning captchas as a means of testing human capabilities [53]. Mbona (2023) examined the IoCs to ascertain their reliability and value for threat intelligence by evaluating performance factors such as accuracy, timeliness, and overlap [54]. M. M. Akhtar (2023) used state-of-the-art machine learning classifiers and content-specific feature sets to detect bots on Twitter [55].

Table 3. Literature survey

Author/ Year	Study	Techniques	Performance metrics
S. Kudugunta / 2018 [1]	Automated bot detection using deep neural networks	Contextual LSTM Model SMOTE-ENN to handle Imbalanced Data and for Over-Sampling	Precision=0.8 Recall =0.92 F1-Score =0.93 Accuracy =0.96 AUC/ROC=0.96
Z. Ellaky / 2023 [2]	A Comprehensive Literature Review on Systems for Detecting Social Media Bots	Classification of Social Media Bot Detection Models	N/A
J. V. Fonseca Abreu / 2020 [3]	Detecting Twitter Bots with a Minimal Set of Features	Reviewed ML Techniques (RF, SVM, NB) for Twitter Bot Detection	N/A
Y. Al-Hammadi / 2008 [4]	Data-driven classification algorithms for bot identification	Inspired Dendritic Cell Algorithm (DCA) Model	MAC = 0.95 and MCAV = 0.95 for the bot
X. Dong / 2010 [5]	Revolutionary API call correlation-based bot detection method	Novel Bot Detection Algorithm for API Call Correlation Analysis	Detection rate of BDA=100% and SRCD =66.7%
R. Wald / 2013 [6]	Twitter vulnerability prediction using social bots	Random Forest algorithm for the detection of vulnerability to social bots	5-NN=0.70272 LR=0.68028 MLP=0.65338 NB=0.59445 RF= 0.64017
D. Zhao / 2013 [7]	Analysis of traffic dynamics and flow intervals for botnet detection	Botnet detection considering traffic behavior analysis	True positive for Malicious 98.1% Non-malicious 97.9
Christos Troussas / 2013 [8]	Detecting and classifying social Group information	Naive Bayes Classifier	Precision: 0.77 Recall: 0.68 F-score: 0.72
J. P. Dickerson / 2014 [9]	Are people more prone to strong opinions than automated Twitter bots?	Sentiment-based bots' detection over Twitter is applied to find whether humans are more opinionated than bots.	AUROC= 0.73 ACC= 92.5%
Fabian Schafer / 2017 [10]	Detection and classification of public sentiment.	Behavioral analysis of social bots over the Japanese general election tweets	High duplication ratio of 97.0%
Maeve Duggan / 2016 [11]	Traits like respectful, focused on policy debate, and angry have been considered.	Classification of political discussion or debate considering traits	Confidence level: 95% Web component response rate: 82% Component response rate: 74%.

B. Gambäck / 2017 [12]	Identifying hate speech with the use of convolutional neural networks	Convolutional Neural Networks were used to classify hate speech	Precision (86.61%) Recall (70.42%), F-score of 77.38%
S. Kannan / 2018 [13]	The Scraper was applied to scrape a vast volume and threshold value based on negative polarity.	Twitter Scraper was applied to isolate the rumours	Accuracy: 75%
John Seymour / 2018 [14]	LSTM learns to engineer specific users using deceptive URLs	Generative Models considered an LSTM neural network	Model target: 819 Human target: 129 Post per min Model: 6.85 Human:1.075
S. Kudugunta / 2018 [1]	Automated bot detection using deep neural networks	Contextual LSTM architecture	For one single tweet, AUC > 96% And Account-level bot detection AUC > 99%
Z. Chen / 2018 [15]	An Unsupervised Method for Identifying Botnet-Reliant Twitter Spam Campaigns	Unsupervised to Detect Spam Campaigns	Detect Range 0.44-0.71, considering Accuracy = 96.28%
Chi Zhang/ 2019 [16]	To present a weakly-supervised approach that does not need annotated data to measure the impact of DoS issues by applying LDA and symmetric KLD on tweets.	Determining the Scale of Impact from DoS Attacks in Real Time	It filtered out non-attack tweets, which can be used to increase precision with recall.
A. Foysal / 2019 [17]	Sentiment Analysis and Data Mining using Support Vector Machines for Twitter AI-Powered Social Bot Classification	Classification of AI-based Social Bots by Sentiment Analysis and Data Mining	Precision = 0.75 Recall value = 0.81
O. Loyola-Gonzalez / 2019 [18]	A Twitter Bot Detection System Based on Contrast Patterns	Contrast Pattern for Bot Detection on Twitter	0.90 of AUC 0.91 of MCC
L. Corti / 2020 [19]	Topic modeling and sentiment analysis of ASD tweets from 2019 to 2020, emphasising COVID-19.	Topic modeling, sentiment analysis	In 2019, 684.032 and in 2020, 691.582 users were extracted
H. Ping / 2019 [20]	An algorithm-based social bot identification model for ANN and DL in urban geography: a comprehensive study and meta-analysis	Social Bots Detection Based on DL	Test 1, 2, 3 Precision=0.986, 0.979, 0.965 Recall=0.977, 0.989, 0.801 F1-score=0.981, 0.984, 0.875
G. Grekousis / 2019 [21]	Overview of Twitter research: data model, network architecture, sentiment analysis, attacks	ANN and deep learning in urban geography	Testing error 0.055, Correlation coefficient 0.80, Validation error 0.025, Training error 0.005
D. Antonakaki/ 2021 [22]	A transparent machine learning pipeline for identifying US presidential campaign bots on Twitter in 2020	Graph Sampling, NLP and Machine Learning	N/A
A. Shevtsov / 2020 [23]	Twitter mood research about the 2020 US presidential election on a massive scale	XGBoost model, an Explainable ML pipeline for Twitter bot detection	N/A
R. H. Ali / 2022 [24]	Twitter bot detection using a one-class classification method	Classification of Social Bots on Twitter by Sentiment Analysis	Accuracy<95% Accuracy<97%
J. Rodríguez-Ruiz / 2020 [25]	Social Media Manipulation and Social Bots in 2020: A Year in Review	One-class classification approach for bot detection on Twitter	MAC = 0.95 and MCAV = 0.95 for the bot
H. H. Chang / 2020 [26]	A Survey of DL Privacy and Security Measures	Automation Detection and Distortion in Social Media Manipulation	Accuracy above 92%
M. I. Tariq / 2020 [27]	Using DL and Transfer Learning for Detecting Offensive Language and Hate Speech	Deep Learning Security and Privacy Defensive Techniques	N/A
B. Wei / 2021 [28]	Moving Forward with Transformer-Based Emotion Recognition for Hindi-English Code-Mixed Data	To build BI-LSTM models	N/A
A. Wadhawan / 2021	How well BERT embeddings use	CNNs, LSTMs, Bi-directional	Accuracy = 71.43%

[29]	Twitter data for catastrophe prediction	LSTMs, along with transformers	
A. K. Chanda / 2021 [30]	The research was considered. Finding instances of cyberbullying on various social media sites	Efficacy of BERT embeddings on predicting disaster	Accuracy = 93%
N. A. Azeez / 2021 [31]	Analyzing Twitter Data for Sentiment on COVID-19 Vaccination Using Deep Learning	NB, KNN, LR, DT, RF, AND SVM	Detect Range 0.44-0.71considering Accuracy = 95.28%
K. N. Alamet / 2021 [32]	Identification of Bots in Tweets by Means of Big Data Analytics	Deep Learning-Based Sentiment Analysis of COVID-19	Accuracy =90.83%.
A. Derhab/ 2021 [33]	Intelligent selection of legal citations based on context with deep learning	Big Data Analytics, shallow and deep learning [34]	96.4 percent accuracy 94.5 percent accuracy score
Z. Huang / 2021 [34]	Examining the Russian Invasion of Ukraine in 2022 via the Lens of Inverse RL for Propaganda Analysis	Deep Learning and machine learning models	N\A
D. Geissler / 2023 [35]	Impact of Twitter Bots on the Russo-Ukrainian Conflict in IGE 2022	Inverse Reinforcement Learning	N\A
F. L. De Faveri / 2023 [36]	Enhancements to Sybil Guard for Social Network Bot Detection	Sybil Guard	Accuracy =81%
N. P. Shetty / 2022 [37]	Building Twitter Hate Detection Models with Multilingual and Adversarial Robustness	Sybil guard algorithm	N\A
A. Ryzhova / 2022 [38]	An Analysis of COVID-19 False News Detection Using Various Transformers	Transformer-based models.	N\A
S. K. S. Joy / 2022 [39]	Automated bot detection using deep neural networks	Five transformer-based models	Accuracy above 94%
A. K. Singha/2024 [101]	Analyzing Twitter Data for Bot detection	ANN, RF Classifier, Adaboost, MLP, Logistic Regression	Precision 0.62 to 0.66 Recall 0.66 to 1 F1-score 0.67 to 0.79 Accuracy 0.56 to 0.66
M. Vahid/2024 [102]	Effective Bot Detection in Twitter	Deep Boltzmann Machine	F1-score =0.77
R. Sánchez-Corcuera/2024 [103]	Detection and Prevention of Malicious User Behavior on Twitter Using	Deep Learning Techniques	Impressive 40.66% in F1-score

3. Limitation of Existing Research [RQ1]

During the analysis of previous research articles, it has been observed that only limited work has been done in the relevant area for the review of social bot detection. One major obstacle to reviewing this domain is that it is hard to find quality papers demonstrating results for the social bot detection using ML and DL methods. Most research in this domain did not clearly mention the datasets and results or feature selection methods, making the review process tiresome. Another major constraint is the absence of a consistent Twitter /X dataset that fairly depicts the changing bot strategies. ML and DL based Models find it difficult to generalize across many kinds of bots and adapt to new trends, as current datasets may record a restricted collection of bot behaviours or a small time span.

Furthermore, data filtering methods used to improve data quality sometimes lack consistency; different researchers apply different pre-processing methods, resulting in conflicting findings. Advancement of X (formerly Twitter) bot detection research would depend much on a comprehensive study with an eye on dataset quality, changing bot strategies, and adaptable model needs. In this article, an in-depth analysis of social bot detection methods is made to fill this gap for future research in the domain of bot detection for the leading social networking platform, Twitter.

4. Social Bot Datasets [RQ 2]

This section provides a summary of the commonly used X / X/Twitter dataset for research that is relevant to bot detection. In these datasets, significant user profile features are User ID, Username, Followers, Followings, Tweet Count, Date of Profile Creation, Last Date of Tweet, Verification Status and User Location. The number of followers assures that the Twitter handle is popular and genuine. The number of tweets and the interval between tweets confirm, to an extent, that an X account that is tweeting is a bot or human. The date of profile creation also has a significant impact on the identification of bots. Verification of the status of a Twitter account helps identify whether an X account is a bot or a genuine account. Moreover, location plays a significant role in the identification of user locality.

Table 4. Features of the most commonly used X / X/Twitter dataset for bot detection

Feature	Column Type	Description
User_ID	Numerical	Unique identifier for each Twitter user.
Username	Textual	Twitter handle or username.
Followers_Count	Numerical	Followers' user has.
Following_Count	Numerical	The user is following.
Tweet_Count	Numerical	Total tweets made by the user.
Profile_Created	Date/Time	Date and time when the Twitter account was created.
Last_Tweet_Date	Date/Time	Date of most recent tweet.
Verified	Categorical	Verification status (e.g., Verified, Not Verified).
Bio	Textual	The user's biography or description on their profile.
Location	Textual	Location specified in the user's profile.
Profile_Image_URL	Textual	URL of user's profile image.

5. Classification of ML and DL based Techniques for Bot Detection on Twitter Datasets

Conventional research works have considered various techniques for detecting and classifying bots on Twitter platforms. This section presents such techniques, results, and outcomes on different Twitter datasets and features.

5.1. Leading Machine Learning based Bot Detection Methods [RQ 3]

(J. V. Fonseca Abreu, 2020) Performed detection of Twitter bots by making use of a minimal set of features and machine learning techniques [3]. These techniques were RF, SVM, and NB. In this research, accuracy was uniform with a mean of 0.85 and an SD of 0.19. (R. Wald, 2013) Focused on Twitter vulnerability prediction. These algorithms detect social bots using the Random Forest algorithm and provide AUC of 0.70 NN, 0.68 LR, 0.65 MLP, 0.59 NB, 0.64017 RF, 0.67 SVM [6]. (D. Antonakaki, 2021) Introduced a transparent machine learning pipeline. The objective was to find the US presidential campaign bots on Twitter. Research provided accuracy below 90% [22]. (A. Shevtsov, 2020) Introduced Twitter mood research regarding the 2020 US presidential election on a massive scale. Research considered the XGBoost model and machine learning pipeline to find Twitter bot detection [23]. (F. K. Alarfaj, 2023) Focus on bot detection work by considering diverse content features and applied ML algorithms. Research provided message accuracy between 40% and 61%, special character accuracy between 39% and 62%, part of speech accuracy between 38% and 62%, and sentiment accuracy between 38% and 62% [56]. (Z. Ellaky and F. Benabbou, 2024) was focused on the detection of political social media bots. The researchers used 33 characteristics culled from the Twibot-20 dataset. Optimal characteristics were used to train several machine-learning algorithms in the research. The model performed better according to the findings. The original set yielded excellent results from a test set score of 99.50% to an AUC of 90.40% and an accuracy of 81.60%. Also, all measures utilized in the study assumed a training set size of 100% [99]. (M. Aljabri et al. 2024) considered a social media bot detection that was based on a machine learning model. Social media has been considered, and different methods have been applied for social media bot detection. A comprehensive literature review was conducted with an accuracy of between 73% and 99.54% [64].

5.2. Leading Deep Learning-based Bot Detection System [RQ 4]

The author (S. Kudugunta, 2018) introduced an automated bot detection using a deep neural network that used a contextual LSTM model in the research [1]. This research provided Precision 0.8, Recall 0.92, F1-Score 0.93, Accuracy 0.96, and AUC/ROC 0.96. (Z. Ellaky, 2023) I also did research to provide a comprehensive literature review on systems to detect social media bots. They used an SMB detection model, but research failed to provide significant accuracy [2]. (B. Gambäck, 2017) Focused on the identification of hate speech with the use of CNN. These CNNs were used to classify hate speech with 86.61% precision, 70.42% recall, and 77.38% F-score [12]. (John Seymour, 2018) Introduced, LSTM learns to engineer specific users socially. They introduced generative Models, an LSTM neural network targeting 819 models and 129 humans. 6.85 posts per minute were made by the model, and 1.075 posts per minute were made by humans [14]. (S. Kudugunta, 2018) Introduced an automated bot detection mechanism that considered deep neural networks. Here, contextual LSTM has exploited both content and metadata in order to find bots, where single-tweet AUC was more than 96% and account-level bot detection AUC was more than 99% [1].

(H. Ping, 2019) proposed algorithm based on a social bot identification model for ANN and DL in urban geography. Researchers did a comprehensive study and meta-analysis to introduce a Social Bot Detection Model based on DL. Test 1 provided Precision 0.986, Recall 0.977, and F1-score 0.981. Test 2 provided Precision 0.979, Recall 0.989, and F1-score 0.984. Test 3 provided Precision 0.965, Recall 0.801, F1-score 0.875 [20]. (H. H. Chang, 2020) did a survey of deep learning privacy and security measures, automation detection

and distortion and provided accuracy above 92% [26]. (A. K. Chanda, 2021) did research to find instances of cyberbullying over various social media sites by making use of AI efficacy.

Research considered BERT embedding for predicting disasters from Twitter data to yield an accuracy of 93%. (L. Ilias, 2023) considered Multimodal Detection of Bots, which has been made using Transformers, considering X Twitter. Research conducted extensive experiments on the Cresci'17 and Twi Bot 20 datasets. Research compared Google Net, ResNet, WideResNet, AlexNet, DenseNet, MobileNet, VGG16, and EfficientNet architectures, where precision lies between 99% and 100%, recall lies between 96% and 99.35%, and F1-score lies between 97.73% and 99.58%. Specificity lies between 99.08 and 100.00% [97].

(A. K. Singha et al., 2023) has been conducted to analyse the numerical pattern in Twitter data. This work considered unveiling fake and bot accounts. Research work utilized deep learning to address the widespread problem of phoney accounts. This work considered ANN and certain Machine Learning methods. This work considered logistic regression, adaboost, ANN, RF, GB, MLP, XG boosting, Grid search and obtained precision and accuracy between 60% and 70%, F-score between 72% and 77%, Recall between 74% and 99.3% [101]. (R. Sánchez-Corcuera, et al. 2024) considered Early Detection and Prevention of Malicious User Behaviour on Twitter Using Deep Learning Techniques. Research work focused on the early detection and prevention of malicious user behaviour. Considering juxtaposed against the identical dataset, that technique provided F1-Score of 22.90% for 10%, 44.56% for 20%, 65.21% for 30%, 78.22% for 40%, 81.85% for 50%, 79.11% for 60%, 83.04% for 70%, 70.6% for 80%, 84.20% for 90% dataset during anticipatory identification of harmful users [103].

5.3. Bot Detection in Social Networks

(A. Dehghan et al., 2024) Focused on bot detection in social networks. Profile metadata and NLP features are typically explored for bot detection. ROC AUC lies between 0.55 and 0.75 for Struc2Vec, Deep Walk, LSME, Role2Vec, Node2Vec, and Graph Wave when noise is increased [100].

5.4. Data-Driven Bot Classification System

Y. Al-Hammadi (2008) identified bots using data-driven classification techniques. Inspired DCA Model gave the bot 0.95 MAC and MCAV [4]. Suthendran (2018) utilised Twitter Scraper to scrape large amounts of tweets and information with negative polarity thresholds. Twitter Scraper isolated rumours with 75% accuracy [13]. Z. Chen (2018) proposed an unsupervised technique to detect Botnet-Reliant Twitter Spam Campaigns. The technology uses Twitter botnets and detects 0.44 to 0.71. Work was 96.28% accurate [15]. Chi Zhang (2019) presented a weakly-supervised model. This model increases accuracy but decreases recall [16]. O. Loyola-Gonzalez (2019) developed a Twitter bot detection algorithm. Twitter Bot Detection Classification has 0.90 AUC and 0.91 MCC [18]. A. Derhab (2021) chose legal citations well. This study used shallow and deep learning methods and found 94%–96% accuracy [33].

A. Sallah et al. (2024) improved social bot identification. The author improved bot detection F1-score (93%) and accuracy from 3% to 24% [98]. Effective bot identification and the Twitter dataset were examined by M. Vahid (2024). The goal was achieved using a Deep Boltzmann Machine. Different categorisation methods identified bots using specific characteristics. The proposed KNN has 0.919 precision, 0.964 recall, and 0.929 accuracy. The proposed SVM has 0.967 precision, 0.759 recall, and 0.774 accuracy. Proposed AdaBoost has 0.959 precision, 0.970 recall, and 0.962 accuracy. The proposed Decision Tree has 0.962 precision, 0.970 recall, and 0.964 accuracy [102].

5.5. API Call Correlation-Based Bot Detection Method

(X. Dong, 2010) was made use of a revolutionary API call correlation-based bot detection method. These novel bot detection algorithms for API call correlation analysis provided a detection rate of 100% BDA and 66.7% SRCD, and were about 33.3% [5].

5.6. Botnet Detection Considering Traffic Behaviour Analysis

(D. Zhao, 2013) Did an analyzis of traffic dynamics and flow intervals for botnet detection. Depending on traffic behaviour analysis, botnet detection provides true positives for 98.1% Malicious and 97.9% Non-malicious [7].

5.7. Sentiment-Based Bot Detection

Troussas (2013) detected and classified social group information. Naive bayes classifier sentiment analysis of Facebook and Twitter datasets yielded 0.77 precision, 0.68 recall, and 0.72 F-score [8]. J. P. Dickerson (2014) investigated Twitter bots for opinion categorisation. Twitter sentiment-based bot identification was used to determine whether people are more opinionated than bots. Research showed AUROC 0.73 and ACC 92.5% [9]. Fabian Schafer (2017) detected and classified public sentiment. Researchers detected 97.0% repetition in social bot behaviour across Japan general election tweets [10]. Maeve Duggan (2016) classified political conversation or argument by these features with 95% confidence, 82% web component response rate, and 74% component response rate [11]. A. Foysal (2019) combined data mining with sentiment analysis. Bot categorisation from tweets using Support Vector Machines yielded 0.75 accuracy and 0.81 recall [17]. Topic modelling and sentiment analysis of 2019–2020 ASD tweets, focusing on COVID-19, were presented by L. Corti (2020). 50.057 tweets were created by 230 bots in 2019. 188 bots tweeted

59.104 in 2020 [19]. G. Grekousis (2019) used deep learning sentiment analysis to characterise bot assaults.

Research found testing error of 0.055, correlation coefficient of 0.80, validation error of 0.025, and training error of 0.005 [21]. N. A. Azeez (2021) analysed Twitter sentiment on COVID-19 vaccination. They used deep learning. This study found 95.28% accuracy in social media cyberbullying detection [31]. K. N. Alamet (2021) used big data analytics to identify Twitter bots. Research is focusing on COVID-19 DL sentiment analysis. Twitter data yields 90.83% accuracy [32].

5.8. One-Class Classification

R. H. Ali (2022) worked on Twitter bot detection that used a one-class classification method. During the classification of social bots on Twitter by sentiment analysis research work, the provided accuracy was between 95% and 97% [24]. J. Rodríguez-Ruiz (2020) focused on social media manipulation and Social Bots. A one-class classification approach has been used for bot detection over Twitter, and a MAC of 0.95 and MCAV of 0.95 were provided for the bot [25].

5.9. Reinforcement Learning

Z. Huang (2021) considered inverse reinforcement learning for propaganda analysis. DL and ML models considered citation-list-based method and three context-based methods [34]. D. Geissler (2023) focused on the impact of Twitter bots on the Russo-Ukrainian conflict in IGE 2022 [35].

5.10. Sybil Guard Mechanism

F. L. De Faveri (2023) did enhancements to Sybil Guard to perform social network bot detection. Research work considered Twitter bots' impact on the Russo-Ukrainian War and yielded an accuracy of 81% [36]. N. P. Shetty (2022) focused on building Twitter hate detection models that considered Multilingual and Adversarial Robustness [37].

5.11. Transform Learning

(M. I. (Tariq, 2020) used deep learning and transfer learning to detect offensive language and hate speech. The DL security and privacy defensive mechanism has been discussed in research work [27]. (B. Wei, 2021) was considered transformer-based emotion recognition to perform Hindi-English code-mixed data.

Research work focused on BI-LSTM models from empty embedding [28]. (A. Wadhawan, 2021) Did a catastrophe prediction over Twitter data. Research considered CNNs, LSTMs, Bi-directional LSTMs, and transformers. Different DL models have been considered to provide an accuracy of 71.43% [29]. (A. Ryzhova, 2022) Did an analysis of COVID-19 false news detection. Research used Transformer-based models to process short text messages. Research work considered data in different languages [38]. (S. K. S. Joy, 2022) It focuses on automated bot detection, which is considered a deep neural network.

Five transformer-based models were considered in research that provided accuracy above 94% [39]. In this way, different machine learning [56, 62, 65] based on real-time implementation [60] for social media bot detection [62, 64, 26, 71] considering user profiles [67, 69], while some researchers have focused on sentiment analysis [61].

Such machine learning mechanism [66, 76] might be supervised machine learning [68, 70], semi-supervised [75] or ensemble machine learning [72]. In order to resolve accuracy and performance concerns [73] for Digital Information World's services [74], a more advanced multilayered deep learning technique [77] might be used.

5.12. Recent Research

Recent advancements in bot detection on the X platform (previously known as Twitter) have leveraged both machine learning and deep learning approaches to enhance accuracy and adaptability in identifying malicious accounts. Sánchez-Corcuera et al. [103] proposed an early detection framework using deep learning to identify and prevent harmful user behaviour on Twitter, emphasizing proactive intervention. Aljabri et al. [64] provided a comprehensive literature review, consolidating the landscape of machine learning-based bot detection techniques and outlining their strengths and limitations. Wang et al. [104] introduced an unsupervised deep contrastive graph clustering model, offering a novel approach that does not rely on labeled datasets, thus improving generalizability in real-world scenarios. Arranz-Escudero et al. [105] adopted a multimodal methodology combining textual and behavioral features to counter misinformation through more robust bot detection mechanisms. Egli et al. [106] explored influencers' nuanced use of bots in the anti-vaccine discourse, highlighting bots' ethical ambiguity and potential "benevolent" roles in controversial topics. Huang et al. [107] developed a semi-supervised framework using relational graph attention transformers, incorporating social context into detection models for improved performance in dynamic environments.

Finally, Giroux et al. [108] examined the reliability and confidence in current bot detection systems, raising critical questions about transparency, bias, and interpretability in automated detection efforts. These studies underscore the evolving sophistication of detection models and the pressing need for adaptive, explainable, and ethically grounded solutions.

Table 5. Recent research

Citation	Objective	Features Used	Accuracy / Evaluation Metric
[103] Sánchez-Corcuera et al. (2024)	Detect and prevent malicious behavior on Twitter using deep learning.	Behavioral features, user metadata, and tweet content	High precision and recall; F1-score ~0.89
[64] Aljabri et al. (2023)	Review and synthesize ML-based approaches to bot detection.	Varied: text, network, user metadata	Not applicable (survey study)
[104] Wang et al. (2024)	Detect social bots using unsupervised contrastive graph clustering.	Graph structure, embedding, and temporal patterns	Accuracy ~87%, NMI and ARI metrics used
[105] Arranz-Escudero et al. (2025)	Use multimodal data to enhance misinformation detection and bot identification.	Textual features, image content, and posting patterns	F1-score ~0.85, ROC-AUC ~0.91
[106] Egli et al. (2025)	Analyze bot use in antivaccine discourse and its ethical implications.	Text content, engagement metrics, and bot scores	Qualitative analysis, no numerical accuracy
[107] Huang et al. (2025)	Semi-supervised detection using relational graph attention transformers.	User relations, tweet content, and graph attention	Accuracy ~89%, macro-F1 ~0.87
[108] Giroux et al. (2025)	Evaluate confidence and reliability in bot detection tools.	Detection system outputs, confidence scores, and explanation methods	Focus on explainability; subjective evaluation

The commonly used performance parameters for social bot detection [75] are represented in the study. A confusion matrix, commonly used for different ML and DL based methods for predicted and actual values, is shown in the following figure.

		Actual values	
		Positive (1)	Negative (0)
Predicted values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Fig. 7 Confusion matrix and accuracy parameters

5.13. Performance Parameters [RQ 5]

Accuracy: Accuracy is the reliability of the model.

"Accuracy=" (TP+TN)/(TP+TN+FP+FN) (1)

Precision: It is required to present qualitative accuracy

"Precision=" "TP" /"TP+FP" (2)

Recall: Recall is presenting quantitative accuracy

"Recall=" "TP" /"TP+FN" (3)

F1 Score: F1 Score is the aggregated mean of precision and recall

"F1-Score=2×" "Precision × Recall" /"Precision + Recall" (4)

5.14. Hyperparameter

Based on the provided hyperparameters, here is a summary of each parameter and its description:

5.14.1. Learning Rate

One important factor in training is to find the minimum loss function, which in turn defines the step size. With 0.001, it is quite modest and might work well with the Adam optimizer.

5.14.2. Epoch

One whole training dataset, the training process will go over the whole dataset fifty times.

5.14.3. Optimizer

Optimizers alter the neural network's properties, such as learning rate and weights, to decrease losses. Here, Adam optimizer and DNN training tool are used, which are known for their adaptive learning rate optimization capabilities.

5.14.4. Batch Size

The training used in a single iteration is called the batch size. With 32 sizes, the loss and weight updates will be performed on 32 samples from the dataset.

5.14.5. Classification

This action indicates that it is a classification job. For tasks involving categorization in particular, it employs an ANN.

5.14.6. Target Size

This indicates the dimensions of the target data. In this case, it is specified as 64x64, which could mean that the input images or data are expected to be of this size. Such hyperparameters were used in research related to bot identification [23, 78], the Random Forest algorithm [79] that considered hashtags [80] over social media platforms [81]

5.15. Classification of Research Areas

There have been several social spam detections [82], some of them used machine learning-enabled post-filtering [83] while some used the Turing test [84] and Support for supervised mining methods [85]. Using ML, DL and Network Analysis to detect and categorize Twitter bots. With

the assistance of supervised and unsupervised mining approaches, several machine learning models played a key part in detecting phoney Facebook accounts [86]. Additionally, these models emphasised bots' role in polarizing stances on social media [87]. A system of this kind provided a key contribution in identifying and describing the activities of Arab spammers on Twitter. Deep learning explores how DNNs like CNNs, RNNs, and LSTM networks might improve bot identification. It also investigates how DL might capture bot behavior's complicated patterns and temporal relationships. A methodology for spam account identification on Twitter is based on both text and metadata and is a combination of the two [88]. Understands and categorizes user activity.

Table 6. Machine learning based techniques for bot detection

Reference	Technique	Accuracy	Dataset
A. Foysal /2019 [17]	SVM	75%	Twitter datasets.
Mbona/2023 [54]	SVM	76%	Twitter datasets.
F. N.Pramitha /2021 [65]	SVM	86.89%	NA
Purba KR/2020 [67]	SVM	91.76%	Fake users 'dataset
C. Troussas, M /2013 [8]	Naïve Bayes	72%	NA
Sheeba JI /2019 [79]	Random Forest	94.87%	Benchmark dataset

The identification of phishing attacks, the detection of spam on Twitter, and the detection of social bots in real time [89] are examples of situations in which conventional deep learning models have been used rather often. A hybrid feature selection technique is also used to determine characteristics of profile information in order to identify social bots on Twitter [90]. Transfer Learning has been used to examine transfer learning for bot identification. Research in these fields frequently overlaps because one discipline may help another. The following chart classifies research on such traits, considering the above Table 5.

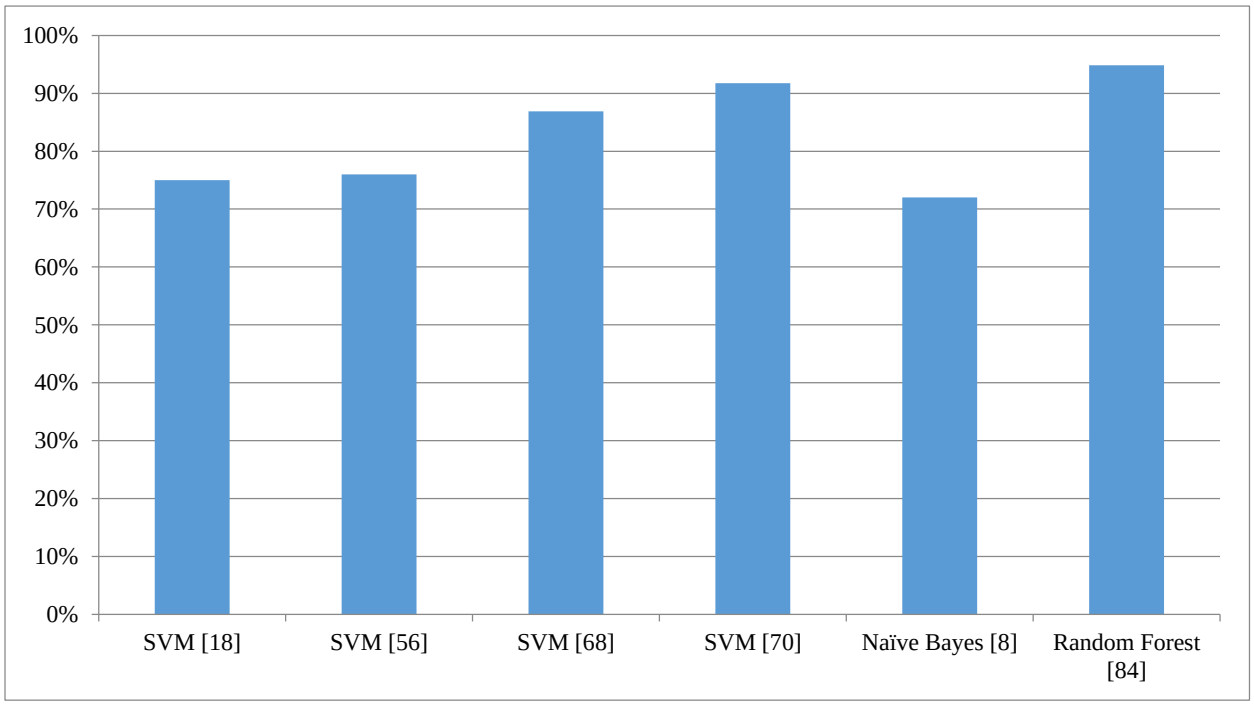


Fig. 8 Comparison of accuracy in different ML-based research

The above chart presents a comparison of the accuracy of conventional SVM, NB and RF approaches. Naïve bayes approach has provided minimum accuracy, whereas Random Forest is yielding maximum accuracy. SVM approaches have been proposed with moderate accuracy. Authors focused on Twitter social bot identification [91]. Some authors considered supervised methods [92] while others considered unsupervised machine learning [93]. Some bot identification approaches merged two channels of convolutional neural networks that run simultaneously and use fully connected

neural networks [94]. The below table presents the comparative analysis of accuracy in the case of LSTM, BI-LSTM, BERT, RoBERTa and CNN. It has been observed that Roberta achieves maximum accuracy compared to LSTM, Bi-LSTM, and BERT. A comparative analysis of all these deep learning mechanisms has been made in the following chart.

Table 7. Deep learning techniques for bot detection

Reference	Technique	Accuracy	Dataset
B. Wei, J /2021 [28]	LSTM	92%	Hate speech dataset
K. N. Alamet al /2021b [32]	LSTM	90.59%	All COVID-19 Vaccines Tweets
B. Wei, J /2021 [28]	BI-LSTM	92%	Hate speech dataset
K. N. Alamet al /2021 [32]	BI-LSTM	90.83%	COVID-19 vaccine.
S. Biswas /2022 [41]	BI-LSTM	80.05%	Twitter datasets
K. Chanda /2021 [30]	Bert	82%	Twitter dataset
A. Wadhawan /2021 [29]	Roberta	71%	Hinglish dataset
S. K. S. Joy /2022 [39]	Roberta	98%	COVID-19 data set
A. S. Alhassun /2022 [44]	CNN	94.27%	Arabic spam dataset

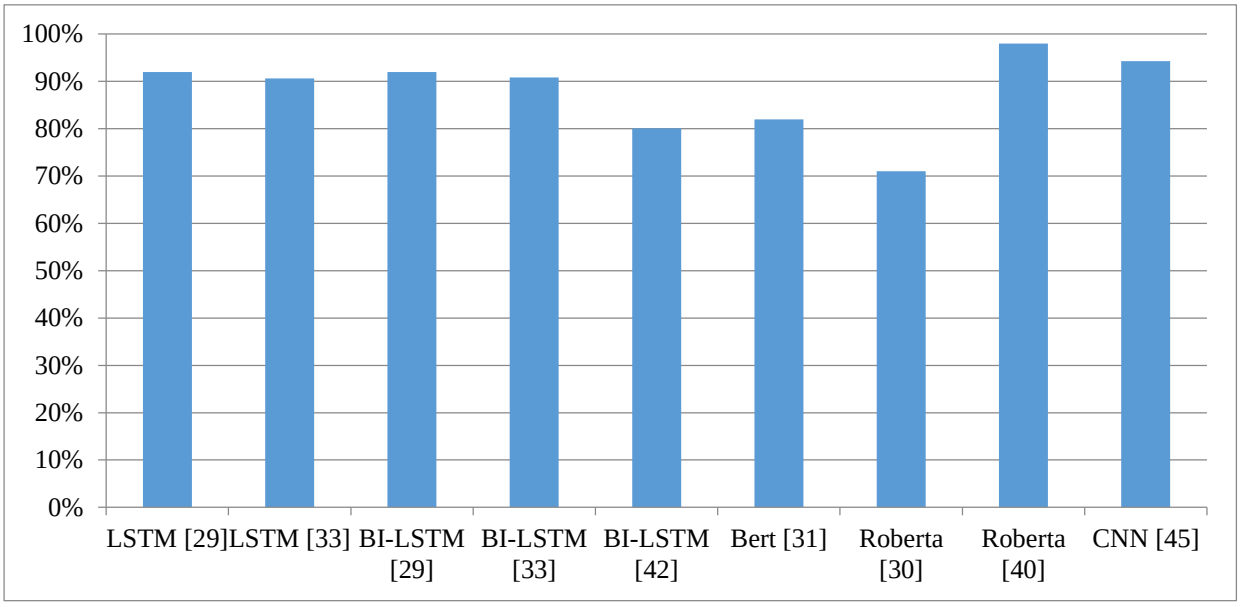


Fig. 9 Comparison of the accuracy of different deep learning approaches

6. Discussion on Review Work

This research article summarizes the SLR study and provides directions for future research. The article considered 115 shortlisted papers out of 286 relevant papers on various criteria, that are related to Bot detection for different ML and DL techniques such as SVM, Naïve Bayes, Random Forest, DT, LSTM, Bi-LSTM, CNN, BERT, RoBERTa and a few efficient hybrid methods also evaluated and reviewed.

The main aim of this review article is to analyse all important elements, such as prominent bot detection techniques, benchmark dataset collection for the X platform, feature selection, model training, and the limitations of the reviewed studies.

The existing conventional research works included in this article were demonstrated on different Twitter datasets with X account profile features and through sentiment analysis of tweet contents. During review for the same, it has been observed that Machine Learning based techniques are the most frequently used for social bot detection, followed by Deep Learning based methods. Among ML techniques, SVM followed by RF are the most used algorithms for the detection of Twitter bots, as illustrated in Table 5. RF achieved the highest accuracy of 94.87%, followed by SVM, which achieved an accuracy of 91.76% on Twitter datasets using account profile features, as illustrated in Figure 8. Among DL techniques, LSTM is the most employed algorithm for Twitter bot detection. RoBERTa achieved the highest accuracy of about 98%on the COVID-19 dataset, followed by the CNN technique with 94.27% accuracy on the Arabic Spam dataset, as illustrated in Figure 9. Graph-based methods for bot detection have also received little attention.

The literature has provided many publicly available Twitter datasets for bot detection. However, there is a lack of

big X / Twitter datasets for training of DL techniques, which can create limitations for DL methods. Further, most of the datasets represent generic names, such as 'fake', 'Twitter Dataset', 'COVID dataset', 'Hinglish dataset', 'Hate Speech dataset' or just focus on a specific type of Twitter bots, that cannot be generalized to all Twitter bots. Most of the datasets contain user account information and content features. In Table 4, the most commonly used features are listed by the researcher for X / Twitter bot detection with a brief description.

Additionally, commonly used datasets by researchers may be biased towards a specific type of bot, which complicates bot detection with such imbalanced datasets. Most datasets carry more data for one class, i.e. Organic Accounts, than the bot data, leading to weak performance of bot detection models. To overcome the above-mentioned challenges of imbalanced datasets, biased class, and dataset size, the leading ML and DL-based techniques, which use SMOTE and data augmentation methods, are demonstrated in the last section of this article. The results of these techniques have been presented further.

7. Simulation Work

This section focuses on technical simulation for training and testing benchmark X (erstwhile Twitter) datasets selected after an in-depth review analysis. This simulation has been demonstrated to understand better the workings of frequently used ML/DL techniques. Additionally, SMOTE has been used to handle imbalanced data uncertainty and data augmentation to increase the number of samples and enhance model performance. Python code was implemented over the Google Collaboratory platform to perform the simulation. The X dataset was downloaded from the Kaggle repository, the one used by the majority of the reviewed works for social bot detection. Common profile features of this dataset are Retweets, Mentions, Followers, User ID, Username, Tweet Count, Count Verified, Location, and Hashtag.

7.1. Naïve-Based Implementation

Many researchers have been working with the Naïve Bayes (NB) method for social bot detection in recent years. The simulation was demonstrated with a commonly used Twitter dataset using the NB technique after processing with the Synthetic Minority Oversampling Technique for bot detection. This yielded results in various performance parameters, which are represented in the following figure. It has been observed that overall accuracy in the case of Naïve Bayes lies between 79%-80%.

Table 8. Accuracy parameters for naïve-based implementation

Class/Accuracy parameters	Precision	Recall	F1-Score
0	0.79	0.80	0.80
1	0.80	0.79	0.79

7.2. SVM-Based Implementation

SVM, a reliable supervised learning method, is utilized for social bot identification. SVM works well in high-dimensional spaces by finding a hyperplane that optimizes class margins.

For bot identification, SVM can manage complicated data structures and outliers for social media data. SVM has good accuracy, precision, and recall, particularly when paired with feature engineering and data balance. Overall accuracy maintains between 81% and 82%.

Table 9. Accuracy parameters for the SVM-based implementation

Class/Accuracy parameters	Precision	Recall	F1-Score
0	0.82	0.81	0.81
1	0.81	0.82	0.82

7.3. Random Forest-Based Implementation

Random Forest, an ensemble learning approach, can handle enormous datasets with complicated patterns, making it ideal for social bot identification. It builds numerous decision trees during training and pooling their outputs produces robust classifications that minimize overfitting and improve accuracy. Work is yielding 0.8577.

Table 10. Accuracy parameters for naïve-based implementation

Class/Accuracy parameters	Precision	Recall	F1-Score
0	0.86	0.86	0.86
1	0.86	0.86	0.86

Random Forest can capture complex social media interactions for bot identification, delivering high accuracy and consistent performance metrics.

Researchers use Random Forest to differentiate bots from real accounts because it performs well across different metrics of accuracy. It has been observed that the overall accuracy in the random-based case is between 85% and 86%.

7.4. LSTM-based Implementation

The LSTM model can record sequential relationships in data, making it ideal for social bot identification. LSTMs can analyse time-dependent data like social media activity and user interactions since they recall long-term trends. LSTM models can identify complicated, contextual sequence patterns like tweet timing and content, which distinguish bots from humans in Bot identification tasks.

LSTM performs well in accuracy parameters when used with properly built temporal information, making it a potent social bot identification technique. It has been observed that the overall accuracy in the case of the LSTM-based model is between 85% and 86%.

Table 11. Accuracy parameters for LSTM

Class/Accuracy parameters	Precision	Recall	F1-Score
0	0.88	0.88	0.88
1	0.88	0.88	0.88

An epoch-wise comparison of training and validation accuracy for LSTM is shown in the following table.The below table chart has been plotted to compare training and validation accuracy for the LSTM model.

Table 12. Comparison of epoch-wise accuracy results for the LSTM technique

Epoch	Training Accuracy for LSTM	Validation Accuracy for LSTM
1	84.17	87.78
2	88.01	87.78
3	87.9	87.78
4	87.91	87.78
5	87.92	87.78
6	87.83	87.78
7	87.83	87.78
8	87.8	87.78
9	87.95	87.78
10	87.78	87.78

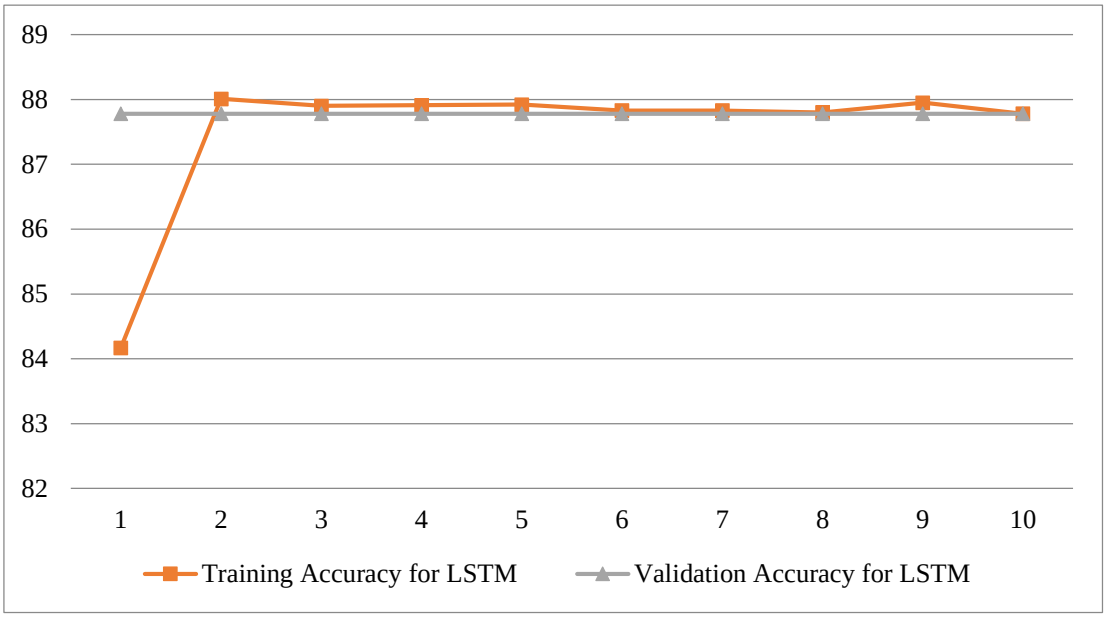


Fig. 10 Result parameters for LSTM

An epoch-wise comparison of the training and validation loss of LSTM was obtained and shown in the following table.

Table 13. Epoch-wise loss for LSTM

Epoch	Training loss for LSTM	Validation loss for LSTM
1	0.4237	0.3715
2	0.3698	0.3713
3	0.3701	0.3718
4	0.3698	0.3718
5	0.3691	0.3713
6	0.3707	0.3713
7	0.3705	0.3713
8	0.3714	0.3713
9	0.3683	0.3712
10	0.3716	0.3712

Considering the given table, the training and validation loss obtained using LSTM has been represented in Figure 11.

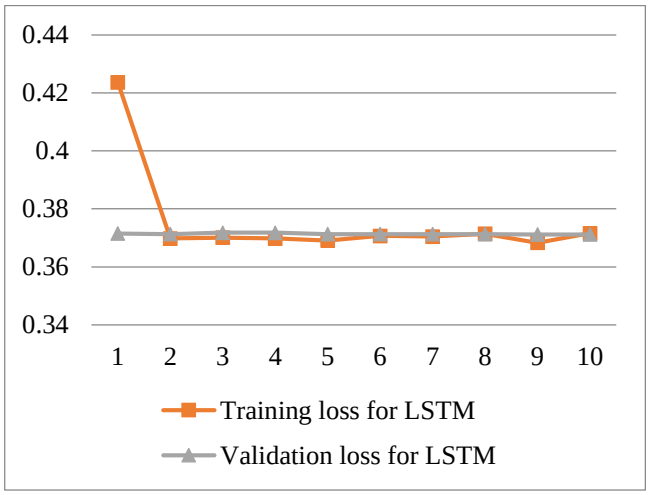


Fig. 11 Comparison of loss for LSTM

7.5. Bi-LSTM-Based Implementation

Bi-LSTM networks capture past and future contextual information and sequential data to increase bot detection. This dual-layered technique enables Bi-LSTM to discern more complicated social media patterns, which may assist in distinguishing bots from humans.

Bi-LSTM models can detect subtle patterns in time series data, such as response timing and interaction sequences, making them suitable for social bot detection. They perform well on precision, recall, and F1-score.

Bi-LSTM can capture more contextual data, making it better at detecting complex bot activities. It has been observed that the overall accuracy in the case of BI-LSTM is between 87% and 89%.

Table 14. Accuracy parameters for BI-LSTM

Class/Accuracy parameters	Precision	Recall	F1-Score
0	0.89	0.89	0.89
1	0.89	0.89	0.89

Table 15. Epoch-wise training and validation accuracy for BI-LSTM

Epoch	Training Accuracy for BI-LSTM	Validation Accuracy for BI-LSTM
1	87.51	88.82
2	89.02	88.82
3	88.94	88.82
4	88.89	88.82
5	88.91	88.82

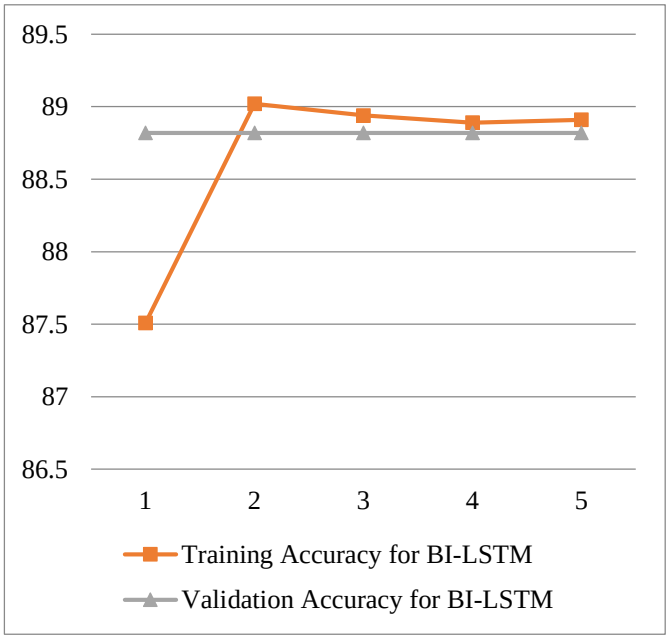


Fig. 12 Result parameters for BI-LSTM

Table 16. Epoch-wise loss for Bi-LSTM

Epoch	Training loss for BI-LSTM	Validation loss for BI-LSTM
1	0.3962	0.3504
2	0.3487	0.3503
3	0.3489	0.3506
4	0.3496	0.3505
5	0.3490	0.3506

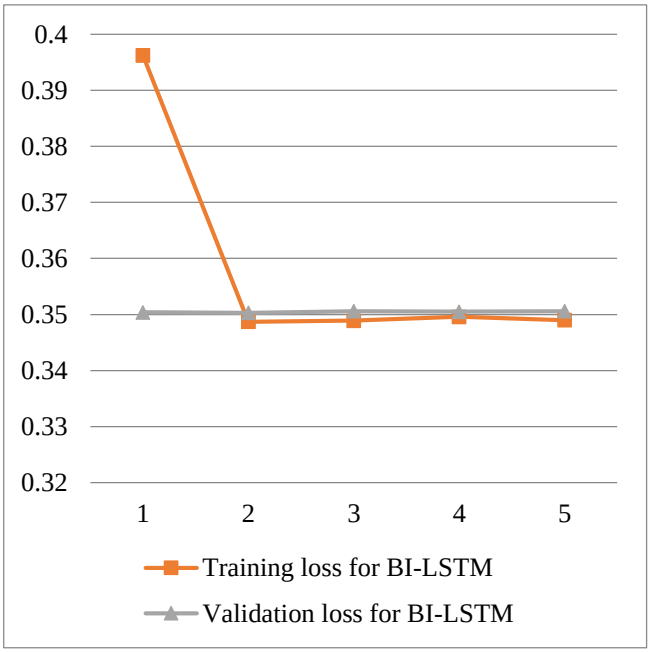


Fig. 13 Comparison of loss for Bi-LSTM

7.6. Comparative Analysis

NB, SVM, RF, LSTM, and Bi-LSTM have distinct capabilities for social bot identification, with performance varying with data complexity.

Naïve Bayes, a probabilistic algorithm, is effective in basic situations but may struggle with complex interactions. In contrast, SVM identifies an ideal hyperplane to split classes, making it suited for high-dimensional data.

In contrast, Random Forest prevents overfitting and captures complicated patterns by aggregating several decision trees.

LSTM is good at capturing time-based relationships and recognizing behavioural sequences for social media analysis, while Bi-LSTM processes data in a bidirectional fashion to grasp context better.

Complex models like LSTM and Bi-LSTM outperform standard methods for bot identification that need temporal and contextual information. Considering the above simulation, a comparative analysis of all mechanisms has been made as follows.

Table 17. Comparative analysis for ML/DL techniques

Parameters	Naïve Bayes	SVM	Random Forest	LSTM	BI-LSTM
Accuracy	79.47	81.38	85.77	87.78	88.82
Precision	80	81	86	88	89
Recall	79	82	86	88	89
F1-score	79	82	86	88	89

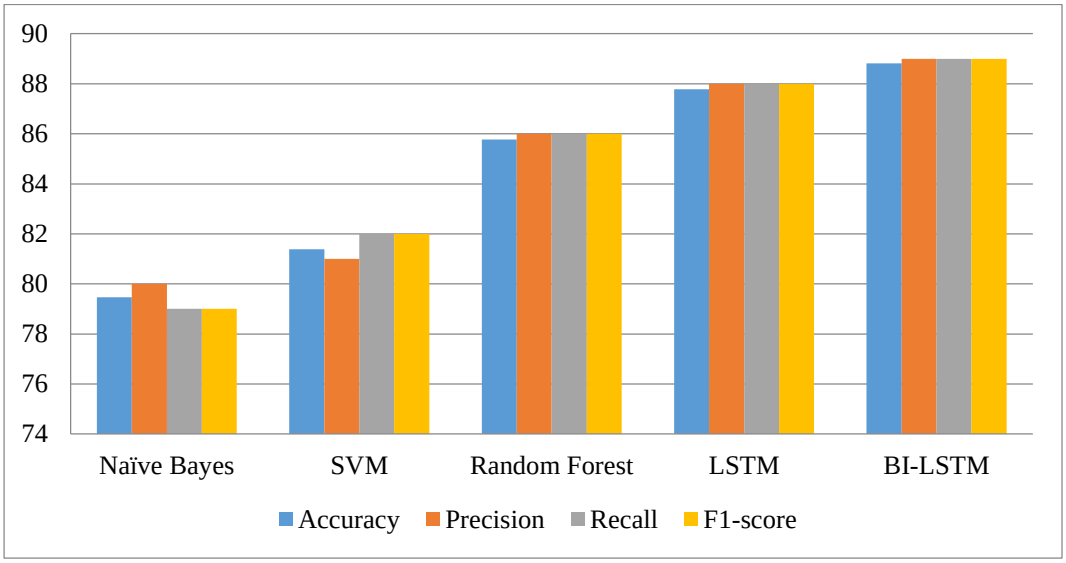


Fig. 14 Accuracy comparison for various ML/DL techniques

8. Conclusion

Today, social bot detection has become the most crucial research area, with an exponential rise in the use of online social networking platforms, such as X (formerly Twitter), which is the leading social media platform. The bot detection challenge gets further complicated with newly evolved social bots that can easily bypass the existing conventional bot detection techniques. This research article reviewed all the leading ML and DL based bot detection techniques by applying the Systematic Literature Review instructions. Work reviewed existing bot detection research works published in reputed research databases and depositories such as ScienceDirect, Scopus, Springer and IEEE. This article considers bot detection research papers published between 2008 and 2024. Further, this article excluded all papers that were not indexed or had a much smaller number of citations. Further, the PRISMA approach was used to conduct a bibliographic search for SLR. By implementing all inclusion and exclusion criteria, 115 research papers were sorted out to carry out evaluation methods on the collected research works. Research classified Twitter accounts into three classes: human, neutral and bots. The article further reviewed bot detection ML and DL based approaches by systematizing evaluation before presenting their outcomes in tabular form for the convenience of researchers. The present work provides a comparative analysis of different techniques on various parameters and summarises all outcomes in the discussion section.

In machine learning, SVM and RF are the most frequently used techniques for Twitter bot detection, while LSTM is the most used technique among deep learning-based methods. In ML-based reviewed articles, it is observed that RF yielded the best accuracy, followed by SVM, while RoBERTa yielded the best accuracy among DL based techniques. However, there is a lack of large Twitter datasets for training of DL techniques, which can create limitations for DL methods. Research works have used diverse datasets with a variety of features to differentiate between legitimate and bot accounts. However, researchers did not conduct enough exploration to analyse the textual content deeply, especially with NLP techniques.

Further, it has been concluded that commonly used datasets appeared biased towards specific bots, making bot detection complicated with imbalanced datasets. Most datasets carry huge amounts of data for one class, i.e., Organic Accounts, and then the bot data, which leads to poor performance of bot detection models.

To overcome the observed challenges of imbalanced datasets, biased class, and dataset size, conventional research works have been demonstrated with leading ML and DL based techniques in the simulation section of this article by integrating SMOTE and data augmentation methods. Research demonstrated results in various evaluation parameters in terms of F1, Precision, Recall, and Accuracy

parameters. The accuracy of SVM, RF, and NB in machine learning is 81.38%, 79.47%, and 85.77%, respectively. However, the accuracy of deep learning techniques like LSTM and Bi-LSTM is 88% and 89%, respectively. The results will provide a significant roadmap for future research on enhancing the performance of bot detection models. Overall, this study offers a useful review work, implemented work for various methods, and insights that can be used in various fields that need real-time social media analysis, helping to create a safer and more informed digital world. This study will provide an important roadmap for future studies aiming at developing an effective bot detection model for online social networks.

8.1. Future Scope

This article provides an important roadmap for future studies aiming at developing an effective bot detection model for online social networks, particularly for the X (previously known as Twitter) platform. In future research, accuracy and performance might be enhanced by including advanced pre-processing mechanisms, imbalanced data handling techniques, and optimized feature extraction and scaling methods. Further, Hybrid deep learning methods may contribute in the case of multiclass-based Twitter bot detection models. Upcoming research work may consider sensitivity and specificity parameters. One major improvement that can be made is the use of self-supervised learning over unsupervised learning to solve data labelling problems for larger datasets. Researchers need to focus on distinguishing traits to create a robust social bot detection system that can handle newly evolving bots.

Acknowledgements and Authors' Contribution

The authors would like to thank Guru Jambheshwar University of Science & Technology, India. Both authors contributed to the data analysis and background study of this paper. Rekha, the first author, contributed to the article's conceptualisation based on a review of different bot detection methods, followed by the demonstration of improved methods of handling imbalanced datasets after pre-processing through SMOTE and an augmentation technique. Dr. Abhishek Kajal, principal author, provided research directions for preparing the manuscript's conceptualization framework. He thoroughly reviewed the manuscript, made corrections, and suggested all the necessary amendments to refine it.

References

[1] Sneha Kudugunta, and Emilio Ferrara, "Deep Neural Networks for Bot Detection," *Information Sciences*, vol. 467, pp. 312-322, 2018. [CrossRef] [Google Scholar] Publisher Link]

[2] Zineb Ellaky, Faouzia Benabbou, and Sara Ouahabi, "Systematic Literature Review of Social Media Bots Detection Systems," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 5, pp. 1-31, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Jefferson Viana Fonseca Abreu et al., "Twitter Bot Detection with Reduced Feature Set," *2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*, Arlington, VA, USA, pp. 1-6, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Yousof Al-Hammadi, Uwe Aickelin, and Julie Greensmith, "DCA for Bot Detection," *2008 IEEE Congress on Evolutionary Computation (CEC 2008)*, pp. 1807-1816, 2088. [CrossRef] [Google Scholar] [Publisher Link]

[5] Xiaomei Dong et al., "A Novel Bot Detection Algorithm Based on API Call Correlation," *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery*, Yantai, China pp. 1157-1162, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[6] Randall Wald et al., "Predicting Susceptibility to Social Bots on Twitter," *2013 IEEE 14th International Conference on Information Reuse & Integration (IRI)*, San Francisco, CA, pp. 6-13, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[7] David Zhao et al., "Botnet Detection Based on Traffic Behavior Analysis and Flow Intervals," *Computer Security*, vol. 39, pp. 2-16, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[8] Christos Troussas et al., "Sentiment Analysis of Facebook Statuses using Naive Bayes Classifier for Language Learning," *2013 IEEE International Conference on Information, Intelligence, Systems and Applications (IISA)*, Piraeus, Greece, pp. 1-6, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[9] John P. Dickerson, Vadim Kagan, and V.S. Subrahmanian, "Using Sentiment to Detect Bots on Twitter: Are Humans more Opinionated than bots?" *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, Beijing, China, pp. 620-627, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[10] Fabian Schäfer Fabian, Stefan Evert, and Philipp Heinrich, "Japan's 2014 General Election: Political Bots, Right-Wing Internet Activism, and Prime Minister Shinzō Abe's Hidden Nationalist Agenda," *Big Data*, vol. 5, no. 4, pp. 294-309, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[11] Maeve Duggan, and Aaron Smith, "*The Political Environment on Social Media*," Report, Pew Research Center, pp. 1-39, 2016. [Google Scholar] [Publisher Link]

[12] Björn Gambäck, and Utpal Kumar Sikdar, "Using Convolutional Neural Networks to Classify Hate-Speech," *Proceedings of the First Workshop on Abusive Language Online*, Vancouver, BC, Canada, pp. 85-90, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[13] V. Sivasangari et al., "Isolating Rumors Using Sentiment Analysis," *Journal of Cyber Security and Mobility*, vol. 7, no. 12, pp. 181-200, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[14] John Seymour, and Philip Tully, "Generative Models for Spear Phishing Posts on Social Media," *arXiv Preprint*, pp. 1-5, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[15] Zhouhan Chen, and Devika Subramanian, "An Unsupervised Approach to Detect Spam Campaigns that Use Botnets on Twitter," *arXiv Preprint*, pp. 1-7, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[16] Chi Zhang et al, "Determining the Scale of Impact from Denial-of-Service Attacks in Real Time Using Twitter," *arXiv Preprint*, pp. 1-11, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[17] Abu Foysal, Safat Islam, and Touhidur Rahaman, "Classification of AI Powered Social Bots on Twitter by Sentiment Analysis and Data Mining through SVM," *International Journal of Computer Applications*, vol. 177, no. 25, pp. 13-19, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[18] Octavio Loyola-González et al., "Contrast Pattern-Based Classification for Bot Detection on Twitter," *IEEE Access*, vol. 7, pp. 45800-45817, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[19] Luca Corti et al., "Social Media Analysis of Twitter Tweets Related to ASD in 2019–2020, with Particular Attention to COVID-19: Topic Modelling and Sentiment Analysis," *Journal of Big Data*, vol. 9, pp. 1-17, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] Heng Ping, and Sujuan Qin, "A Social Bots Detection Model Based on Deep Learning Algorithm," *2018 IEEE 18th International Conference on Communication Technology (ICCT)*, Chongqing, China, pp. 1435-1439, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[21] George Grekousis, "Artificial Neural Networks and Deep Learning in Urban Geography: A Systematic Review and Meta-Analysis," *Computers, Environment and Urban Systems*, vol. 74, pp. 244-256, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[22] Despoina Antonakaki, Paraskevi Fragopoulou, and Sotiris Ioannidis, "A Survey of Twitter Research: Data Model, Graph Structure, Sentiment Analysis and Attacks," *Expert Systems with Applications*, vol. 164, pp. 1-25, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[23] Alexander Shevtsov et al., "Explainable Machine Learning Pipeline for Twitter Bot Detection during the 2020 US Presidential Elections," *Software Impacts*, vol. 13, pp. 1-2, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[24] Rao Hamza Ali et al., "A Large-Scale Sentiment Analysis of Tweets Pertaining to the 2020 US Presidential Election," *Journal of Big Data*, vol. 9, pp. 1-12, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[25] Jorge Rodríguez-Ruiz et al., "A One-Class Classification Approach for Bot Detection on Twitter," *Computers & Security*, vol. 91, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[26] Ho-Chun Herbert Chang et al., *Social Bots and Social Media Manipulation in 2020*, 1st ed., Handbook of Computational Social Science, Taylor & Francis, vol. 1, pp. 1-20, 2021. [Google Scholar] [Publisher Link]

[27] Muhammad Imran Tariq et al., "A Review of Deep Learning Security and Privacy Defensive Techniques," *Mobile Information Systems*, vol. 2020, pp. 1-18, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[28] Bencheng Wei et al., "Offensive Language and Hate Speech Detection with Deep Learning and Transfer Learning," *Arxiv Preprint*, pp. 1-7, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[29] Anshul Wadhawan, and Akshita Aggarwal, "Towards Emotion Recognition in Hindi-English Code-Mixed Data: A Transformer Based Approach," *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 195-202, 2021. [Google Scholar] [Publisher Link]

[30] Ashis Kumar Chanda, "Efficacy of BERT Embeddings on Predicting Disaster from Twitter Data," *Arxiv Preprint*, pp. 1-6, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[31] Nureni Ayofe Azeez et al., "Cyberbullying Detection in Social Networks: Artificial Intelligence Approach," *Journal of Cyber Security and Mobility*, vol. 10, no. 4, pp. 745-774, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[32] Kazi Nabiul Alam et al., "Deep Learning-Based Sentiment Analysis of COVID-19 Vaccination Responses from Twitter Data," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1-15, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[33] Abdelouahid Derhab et al., "Tweet-Based Bot Detection Using Big Data Analytics," *IEEE Access*, vol. 9, pp. 65988-66005, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[34] Zihan Huang et al., "Context-Aware Legal Citation Recommendation Using Deep Learning," *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*, São Paulo, Brazil, pp. 79-88, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[35] Dominique Geissler, and Stefan Feuerriegel, "Analyzing the Strategy of Propaganda using Inverse Reinforcement Learning: Evidence from the 2022 Russian Invasion of Ukraine," *Proceedings of the ACM on Human-Computer Interaction*, vol. 8, no. CSCW2, pp. 1-25 2023. [CrossRef] [Google Scholar] [Publisher Link]

[36] Francesco Luigi De Faveri et al., "Twitter Bots Influence on the Russo-Ukrainian War During the 2022 Italian General Elections," *Conference Proceedings 9th International Symposium, SocialSec 2023: Security and Privacy in Social Networks and Big Data*, Canterbury, UK, pp. 38-57, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[37] Nisha P. Shetty, "An Enhanced Sybil Guard to Detect Bots in Online Social Networks," *Journal of Cyber Security and Mobility*, vol. 11, no. 1, pp. 105-126, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[38] Anastasia Ryzhova et al., "Training Multilingual and Adversarial Attack-Robust Models for Hate Detection on Social Media," *Procedia Computer Science*, vol. 213, pp. 196-204, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[39] Sajib Kumar Saha et al., "A Comparative Study on COVID-19 Fake News Detection Using Different Transformer Based Models," *2022 IEEE Symposium on Industrial Electronics & Applications (ISIEA)*, Langkawi Island, Malaysia, pp. 1-5, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[40] Maryam Heidari, James H Jr Jones, and Ozlem Uzuner, "Online User Profiling to Detect Social Bots on Twitter," *Arxiv Preprint*, pp. 1-9, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[41] Sumana Biswas, Karen Young, and Josephine Griffith, "A Comparison of Automatic Labelling Approaches for Sentiment Analysis," *Proceedings of the 11th International Conference on Data Science, Technology and Applications*, Lisbon, Portugal, pp. 312-319, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[42] Mark M. Bailey, "Detecting Propagators of Disinformation on Twitter Using Quantitative Discursive Analysis," *Arxiv Preprint*, pp. 1-12, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[43] Mst Shapna Akter et al., "Deep Learning Approach for Classifying the Aggressive Comments on Social Media: Machine Translated Data Vs Real Life Data," *2022 IEEE International Conference on Big Data*, Osaka, Japan, pp. 5646-5655, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[44] Atheer S. Alhassun, and Murad A. Rassam, "A Combined Text-Based and Metadata-Based Deep-LearningFramework for the Detection of Spam Accounts on the SocialMedia Platform Twitter," *Processes*, vol. 10, no. 3, pp. 1-24, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[45] Anisha P Rodrigues et al., "[Retracted] Real-Time Twitter Spam Detection and Sentiment Analysis Using Machine Learning and Deep Learning Techniques," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[46] Vasiliki Vrana et al., "EU Citizens' Twitter Discussions of the 2022–23 Energy Crisis: A Content and Sentiment Analysis on the Verge of a Daunting Winter," *Sustainability*, vol. 15, no. 2, pp. 1-22, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[47] Zhenyu Lei et al., "BIC: Twitter Bot Detection with Text-Graph Interaction and Semantic Consistency," *Arxiv Preprint*, pp. 1-13, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[48] Román A. Mendoza-Urdiales et al., "Twitter Sentiment Analysis and Influence on Stock Performance Using Transfer Entropy and EGARCH Methods," *Entropy*, vol. 24, no. 7, pp. 1-17, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[49] Xiao Yu et al., "*GPT Paternity Test: GPT Generated Text Detection with GPT Genetic Inheritance*," *Arxiv*, Report, pp. 1-12, 2023. [Google Scholar] [Publisher Link]

[50] Yingguang Yang et al., "RoSGAS: Adaptive Social Bot Detection with Reinforced Self-Supervised GNN Architecture Search," *ACM Transactions on the Web*, vol. 17, no. 3, pp. 1-31, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[51] Yue Wang et al., "Exploring Topic Models to Discern Cyber Threats on Twitter: A Case Study on Log4Shell," *Intelligent Systems with Applications*, vol. 20, pp. 1-14, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[52] Alexander Shevtsov et al., "BotArtist: Generic Approach for Bot Detection in Twitter via Semi-Automatic Machine Learning Pipeline," *Arxiv Preprint*, pp. 1-10, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[53] Qing Meng et al., "Predicting Hate Intensity of Twitter Conversation Threads," *Knowledge-Based Systems*, vol. 275, pp. 1-39, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[54] Innocent Mbona, and Jan H.P. Eloff, "Classifying Social Media Bots as Malicious or Benign Using Semi-Supervised Machine Learning," *Journal of Cyber Security*, vol. 9, no. 1, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[55] Mohammad Majid Akhtar et al., "False Information, Bots and Malicious Campaigns: Demystifying Elements of Social Media Manipulations," *Arxiv Preprint*, pp. 1-21, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[56] Fawaz Khaled Alarfaj et al., "Twitter Bot Detection Using Diverse Content Features and Applying Machine Learning Algorithms," *Sustainability*, vol. 15, no. 8, pp. 1-17, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[57] Oliver Beatson et al., "Automation on Twitter: Measuring the Effectiveness of Approaches to Bot Detection," *Social Science Computer Review*, vol. 41, no. 1, pp. 181-200, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[58] Feng Wei, and Uyen Trang Nguyen, "Twitter Bot Detection Using Neural Networks and Linguistic Embeddings," *IEEE Open Journal of the Computer Society*, vol. 4, pp. 218-230, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[59] Chris Hays et al., "Simplistic Collection and Labeling Practices Limit the Utility of Benchmark Datasets for Twitter Bot Detection," *Proceedings of the ACM Web Conference*, Austin TX USA, pp. 3660-3669, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[60] Yuhan Liu et al., "BotMoE: Twitter Bot Detection with Community-Aware Mixtures of Modal-Specific Experts," *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Taipei Taiwan, pp. 485-495, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[61] Nirmalya Thakur, "Sentiment Analysis and Text Analysis of the Public Discourse on Twitter about COVID-19 and MPox," *Big Data Cognitive Computing*, vol. 7, no. 2, pp. 1-21, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[62] Tarık Talan, Adem Korkmaz, and Cemal Aktürk, "Analyzing the User'sSentiments of ChatGPT Using Twitter Data," *Iraqi Journal for Computer Science and Mathematics*, vol. 4, no. 2, pp. 202-214, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[63] Kadhim Hayawi et al., "Social Media Bot Detection with Deep Learning Methods: A Systematic Review," *Neural Computing and Applications*, vol. 35, pp. 8903-8918, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[64] Malak Aljabri et al., "Machine Learning-Based Social Media Bot Detection: A Comprehensive Literature Review," *Social Network Analysis and Mining*, vol. 13, pp. 1-40, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[65] Febriora Nevia Pramitha et al., "Twitter Bot Account Detection Using Supervised Machine Learning," *2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Yogyakarta, Indonesia, pp. 379-383, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[66] Pandu Gumelar Pratama, and Nur Aini Rakhmawati, "Social Bot Detection on 2019 Indonesia President Candidate's Supporter's Tweets," *Procedia Computer Science*, vol. 161, pp. 813-820, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[67] Kristo Radion Purba, David Asirvatham, and Raja Kumar Murugesan, "Classification of Instagram Fake Users Using Supervised Machine Learning Algorithms," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 3, pp. 2763-2772, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[68] A. Ramalingaiah, S. Hussaini, and S. Chaudhari, "Twitter Bot Detection Using Supervised Machine Learning," *Journal of Physics: Conference Series: International Conference on Mechatronics and Artificial Intelligence*, Gurgaon, India, vol. 1950, pp. 1-12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[69] F. Rangel, Francisco, and Paolo Rosso, "Overview of the 7th Author Profiling Task at PAN 2019: Bots and Gender Profiling in Twitter," *Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum*, Lugano, Switzerland, pp. 1-36, 2019. [Google Scholar] [Publisher Link]

[70] Mohsen Sayyadiharikandeh et al., "Detection of Novel Social Bots by Ensembles of Specialized Classifiers," *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, Virtual Event Ireland, pp. 2725-2732, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[71] Rachit Shukla, Adwitiya Sinha, and Ankit Chaudhary, "TweezBot: An AI-Driven Online Media Bot Identification Algorithm for Twitter Social Networks," *Electronics*, vol. 11, no. 5, pp. 1-21, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[72] Hrushikesh Shukla, Nakshatra Jagtap, and Balaji Patil, "Enhanced Twitter Bot Detection Using Ensemble Machine Learning," *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, pp. 930-936, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[73] Indira Sen et al., "Worth its Weight in Likes: Towards Detecting Fake Likes on Instagram," *Proceedings of the 10th ACM Conference on Web Science*, pp. 205-209, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[74] Ali Siddiqui, "*Facebook 2019 Q1 Earnings: The Social Media Giant Boasts 2.7 Billion Monthly Active Users on Its All Services*," Digital Information World, Report, 2019. [Publisher Link]

[75] Surendra Sedhai, and Aixin Sun, "Semi-Supervised Spam Detection in Twitter Stream," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 1, pp. 169-175, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[76] Somya Ranjan Sahoo, and B.B. Gupta, "Popularity-Based Detection of Malicious Content in Facebook Using Machine Learning Approach," *Conference Proceedings First International Conference on Sustainable Technologies for Computational Intelligence*, pp. 163-176, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[77] Sandeep Singh Sengar et al., "Bot Detection in Social Networks Based on Multilayered Deep Learning Approach," *Sensors & Transducers*, vol. 244, no. 5, pp. 37-43, 2020. [Google Scholar] [Publisher Link]

[78] Chengcheng Shao et al., "The Spread of Low-Credibility Content by Social Bots," *Nature Communications*, vol. 9, pp. 1-9, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[79] J.I. Sheeba, S. Pradeep Devaneyan, and G. Velvizhi "Detection of Spambot Using Random Forest Algorithm," *Proceedings of International Conference on Advancements in Computing & Management (ICACM)*, Jaipur, India, pp. 746-753, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[80] Surendra Sedhai, and Aixin Sun, "HSpam14: A Collection of 14 Million Tweets for Hashtag-Oriented Spam Research," *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Santiago Chile, pp. 223-232, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[81] Elisa Shearer, and Amy Mitchell, "*News Use Across Social Media Platforms in 2020*," Pew Research Center, Report, pp. 1-18, 2021. [Google Scholar] [Publisher Link]

[82] Arafatur Rahman et al., "SPY-BOT: Machine Learning-Enabled Post-Filtering for Social Network-Integrated Industrial Internet of Things," *Ad Hoc Networks*, vol. 113, pp. 1-20, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[83] Fatih Cagatay Akyon, and M. Esat Kalfaoglu, "Instagram Fake and Automated Account Detection," *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Izmir, Turkey, pp. 1-7, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[84] Erdem Beğenilmiş, and Suzan Uskudarli, "Organized Behavior Classification of Tweet Sets Using Supervised Learning Methods," *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics*, Novi Sad Serbia, pp. 1-9, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[85] Mohammed Basil Albayati, and Ahmad Mousa Altamimi, "An Empirical Study for Detecting Fake Facebook Profiles Using Supervised Mining Techniques," *Informatica*, vol. 43, no. 1, pp. 77-86, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[86] Mohammed Basil Albayati, and Ahmad Mousa Altamimi, "MDFP: A Machine Learning Model for Detecting Fake Facebook Profiles Using Supervised and Unsupervised Mining Techniques," *International Journal of Simulation Systems, Science & Technology*, vol. 20, no. 1, pp. 1-10, 2020. [Google Scholar] [Publisher Link]

[87] Abeer Aldayel, and Walid Magdy, "Characterizing the Role of Bots' in Polarized Stance on Social Media," *Social Network Analysis and Mining*, vol. 12, pp. 1-24, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[88] Atheer S. Alhassun, and Murad A. Rassam, "A Combined Text-Based and Metadata-Based Deep-Learning Framework for the Detection of Spam Accounts on the Social Media Platform Twitter," *Processes*, vol. 10, no. 3, pp. 1-24, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[89] Eiman Alothali et al., "Real Time Detection of Social Bots on Twitter Using Machine Learning and Apache Kafka," *2021 5th Cyber Security in Networking Conference (CSNet)*, Abu Dhabi, United Arab Emirates, pp. 98-102, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[90] Eiman Alothali, Kadhim Hayawi, and Hany Alashwal, "Hybrid Feature Selection Approach to Identify Optimal Features of Profile Metadata to Detect Social Bots in Twitter," *Social Network Analysis and Mining*, vol. 11, pp. 1-15, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[91] Eiman Alothali et al., "Detecting Social Bots on Twitter: A Literature Review," *2018 International Conference on Innovations in Information Technology (IIT)*, Al Ain, United Arab Emirates, pp. 175-180, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[92] Panagiotis Andriotis, and Atsuhiro Takasu, "Emotional Bots: Content-Based Spammer Detection on Social Media," *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, Hong Kong, China, pp. 1-8, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[93] Ahmed Anwar, and Ussama Yaqub, "Bot Detection in Twitter Landscape Using Unsupervised Learning," *Proceedings of the 21st Annual International Conference on Digital Government Research*, Seoul Republic of Korea, pp. 329-330, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[94] Sameh M. Attia, Ahmed M. Mattar, and Khaled M. Badran, "Bot Detection Using Multi-Input Deep Neural Network Model in Social Media," *2022 13th International Conference on Electrical Engineering (ICEENG)*, Cairo, Egypt, pp. 71-75, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[95] S. Barhate et al., "Twitter Bot Detection and their Influence in Hashtag Manipulation," *2020 IEEE 17th India Council International Conference (INDICON)*, New Delhi, India, pp. 1-7, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[96] Muhammad Bazm, and Masoud Asadpour, "Behavioral Modeling of Persian Instagram Users to Detect Bots," *Arxiv Preprint*, pp. 1-8, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[97] Loukas Ilias, Ioannis Michail Kazelidis, and Dimitris Askounis, "Multimodal Detection of Bots on X (Twitter) Using Transformers," *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 7320-7334, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[98] Amine Sallah et al., "Fine-Tuned Understanding: Enhancing Social Bot Detection with Transformer-Based Classification," *IEEE Access*, vol. 12, pp. 118250-118269, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[99] Zineb Ellaky, and Faouzia Benabbou, "Political Social Media Bot Detection: Unveiling Cutting-Edge Feature Selection and Engineering Strategies in Machine Learning Model Development," *Scientific African*, vol. 25, pp. 1-20, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[100] Ashkan Dehghan et al., "Detecting Bots in Social-Networks using Node and Structural Embeddings," *Journal of Big Data*, vol. 10, pp. 1-37, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[101] Amrit Kumar Singha et al., "Analyzing Numerical Patterns in Twitter Data: Unveiling Fake and Bot Accounts During Telangana State Elections," *2024 IEEE 13th International Conference on Communication Systems and Network Technologies (CSNT)*, Jabalpur, India pp. 407-412, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[102] Maryam Vahid, and Reza Ravanmehr, "Effective Bot Detection in Twitter using Deep Boltzmann Machine," *2024 10th International Conference on Web Research (ICWR)*, Tehran, Iran, Islamic Republic of, pp. 303-308, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[103] Rubén Sánchez-Corcuera, Arkaitzs Zubiaga, and Aitor Almeida, "Early Detection and Prevention of Malicious User Behavior on Twitter Using Deep Learning Techniques," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 5, pp. 6649-6661, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[104] Xiujuan Wang et al., "Unsupervised Twitter Social Bot Detection Using Deep Contrastive Graph Clustering," *Knowledge-Based Systems*, vol. 293, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[105] Olmar Arranz-Escudero, Lara Quijano-Sanchez, and Federico Liberatore, "Enhancing Misinformation Countermeasures: A Multimodal Approach to Twitter Bot Detection," *Social Network Analysis and Mining*, vol. 15, pp. 1-22, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[106] Antonia Egli et al., "Bad Robot? The Benevolent Use of Automated Software and Social Bots by Influencers in the #Antivaxx Discourse on Twitter," *Online Information Review*, vol. 49, no. 8, pp. 44-61, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[107] Di Huang, Jinbao Song, and Xingyu Zhang, "Semi-Supervised Social Bot Detection with Relational Graph Attention Transformers and Characteristics of the Social Environment," *Information Fusion*, vol. 118, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[108] James Giroux et al., "Unmasking Social Bots: How Confident Are We?," *EPJ Data Science*, vol. 14, pp. 1-19, 2025. [CrossRef] [Google Scholar] [Publisher Link]