International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 5, No.1 (Feb-24)

http://dx.doi.org/10.12785/ijcds/150145

Optimized Deep Neural Networks Using Sparrow Search Algorithms for Hate Speech Detection

Ashwini Kumar[1] **and Santosh Kumar**[1]

[1]*Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, India*

Received 23 Jun. 2023, Revised 23 Jan. 2024, Accepted 27 Jan. 2024, Published 1 Feb. 2024

Abstract: Deep learning has widespread use in various domains, including computer vision, audio processing, and natural language processing. The hyperparameters of deep learning algorithms have a significant impact on the performance of these algorithms. However, it can be challenging to calculate the hyperparameters of complicated machine learning models like deep neural networks due to the nature of the models. This research suggested a strategy for hyperparameter optimization utilizing the Long Short-Term Memory with Sparrow Search Algorithm (LSTM-SSA) model. The model that has been presented uses a deep neural network, which can recognize and classify instances of hate speech as either hate speech or neither. Experiments are conducted to validate the suggested technique in both straightforward and intricate network environments. The LSTM-SSA model is validated using a dataset consisting of hate speech, and an experimental investigation into the model's sensitivity, accuracy, and specificity is carried out. The outcomes of the experiments demonstrated that the suggested model might be improved upon, as it had an accuracy of 0.936.

Keywords: Hate speech, LSTM, NLP, social media, Sparrow Search Algorithm

1. INTRODUCTION

Online social networks (OSN) and microblogging websites have emerged as Internet users' most popular online platforms. The utilization of social media platforms such as Twitter, YouTube, Facebook, Gab, and Reddit has become increasingly prevalent among individuals of diverse backgrounds, cultures, and interests [1]. The contents are experiencing rapid growth, compellingly illustrating the phenomenon known as big data [2]. The sensation of big data has garnered significant interest among researchers, particularly in automated analysis of individuals' opinions and examining network structures and distributions among users. Although these online platforms provide a public forum for individuals to engage in discourse and exchange viewpoints, the sheer volume of user-generated content and the inherent characteristics of these platforms render content moderation a daunting task. Considering individuals' varying backgrounds, cultures, and beliefs, it is common for some to employ hostile and derogatory language when engaging in discourse with those who do not share similar experiences.

Deep neural networks are currently regarded as the leading neural architectures for various natural language processing tasks, such as the label of hate speech. The deep neural network is a sophisticated system for recognizing patterns. Its robust non-linear fitting capability is highly favoured among scholars. Nevertheless, when training deep neural network models with limited datasets [3], the performance acquired is generally inferior to that of simple neural networks. The process of assembling a proficient deep-learning model is typically intricate and requires a significant addition of time. It entails identifying a suitable architecture for the model and its corresponding hyperparameters. The efficiency of a deep learning algorithm is significantly influenced by its hyperparameters. Hence, the designation of hyperparameters holds significant importance in the performance of deep learning in various domains. Manual search, grid search, and random search are only some of the approaches used to set deep learning hyperparameters. But these approaches have issues, such as low precision, inefficiency, and poor performance in high-dimensional models. So, it's essential to have a reliable strategy for optimizing deep learning algorithms' hyperparameters. In deep learning, hyperparameter optimization determines the optimal values for the model's hyperparameters. A mathematical model of optimization has been employed in the search. However, when trying to find the optimal settings for a deep learning model's hyperparameters, optimization is often seen as a mysterious "black box" procedure.

Technology's rapid advancement has improved data collecting, storage, and processing in scientific research and engineering applications. Deep learning, which mimics the human brain's mechanism, has shown considerable promise

E-mail address: ashwinikumar.cse@geu.ac.in, drsantosh.cse@geu.ac.in

in data processing and analysis due to its superior learning ability, versatility, and portability. Deep learning comprises a network structure with numerous layers that execute linear and nonlinear data transformations. On the other hand, building a good deep learning model is challenging and time-consuming because it requires defining an acceptable structure and hyperparameters such as batch size, learning rate, and dropout rate. While LSTM (Long Short-Term Memory) is a powerful tool, it has limitations when applied to time-series situations [4]. First, network models like LSTM must explain their final decisions better and can't give a detailed breakdown of the parameters used to make predictions. Second, the model's hyperparameters are typically set based on prior research or experience, which introduces some subjectivity. Third, LSTM may enhance the network's architecture, generalization, and relevant skills by appropriately adjusting the model's hyperparameters. Thus, academics are interested in mitigating human variables' impact and identifying the best possible network hyperparameters.This paper contribution can be outlined as follows:

1) Construct an LSTM hate speech detection model by applying the sparrow search method to optimize LSTM networks.
2) The LSTM-SSA model demonstrates enhanced optimization efficiency compared to other state-of-the-art methods.

This research used a SSA optimization approach to determine the best settings for the LSTM network's parameters. This method solves the issue that LSTM can't describe how the model learns to make choices and is affected by human bias. This study aims to present a new LSTM-SSA model for hyperparameter optimization that demonstrates exceptional performance.

The rest of the paper is structured as follows: In Section 2, we provide a thorough analysis of the current research on identifying hate speech. Section 3 presents a comprehensive review of the investigation into and use of the proposed LSTM-SSA algorithm. Section 4 discusses the results of our experiments with the LSTM-SSA approach. Section 5 is the conclusion and future work.

2. RELATED WORK

The social web' has allowed for rapid development in NLP research. Researchers have employed machine learning, deep learning, and natural language processing to decipher the emotional tone of posts on Facebook, Twitter, and other social media sites, making the social web a dynamic field of study. Twitter, YouTube, Facebook, etc., have become multilingual forums where speakers of various languages, cultures, and socioeconomic backgrounds discuss current events, pop culture figures, and other issues of interest. As a result, the wide range of people who use these platforms increases the risk of hate speech incidents. This online behavior is challenging to manage because of the intricacy of foreign languages.

Several studies have specifically examined online hate as a form of hate speech [5] [6]. This post refers to a type of communication deemed offensive and motivated, either wholly or partially, by the writer's prejudice against a particular group or individuals [7]. The primary element under consideration in this context pertains to targeting, where the expression of hatred aims towards a particular group, including refugees, or any community [8], [9]. Waseem et al. [10] have classified different types of abuse according to the target recipient, distinguishing between those directed at individuals/entities versus groups and the degree of explicitness utilized. ElSherief et al. [11] investigated to explore the association between individuals who incite hatred and their targets and their online prominence. According to the research findings, individuals with a prominent social media presence are inclined to receive a greater degree of hatred. According to Salminen et al [8], online news commenting tends to direct hostility towards the media and police as primary targets. In general, discourse about news has been widely considered a significant breeding ground for harmful and hostile online behavior [12].

Analyzing online hate speech requires a critical examination of the relationship between groups. Numerous research studies have investigated the occurrence of online hate groups and group discrimination [13], the utilization of persuasive communication to facilitate hate influencing [14], the mechanism of polarization through being exposed to extremist content on social media [15], the societal dissemination of hatred [16], and the consequences of social exclusion [17], among other relevant subjects. The utilization of interpretive techniques is a common practice in examining subtleties due to the notable impact of contextual and subjective elements.

Wafa et al. [18] proposed an effective new technique for identifying hate speech in 2019. It focuses on a criterion-based approach to feature selection to determine which characteristics will be required by the chosen embedding strategy. After that, Basak et al. [19] built web software (block shame) to help spot and stop public shaming in cyberspace. A spammer was being silenced and blocked by the app, while the latter provided a definition of shaming that included comparison, expressing judgement on a user, sarcasm, or jokes, and whataboutery as six distinct forms of abusive behavior. The idea of pre-training a support vector machine has been presented to improve performance while decreasing computation time [19]. For modest datasets, the recurrent neural network (RNN)-based deep learning strategy has been presented [20]. The improved human sentiment categorization is a clear result of using a complex attention mechanism in conjunction with multi-task learning.

Sequeira et al. [21] utilized various neural network models, including long short-term memory and text convolutional neural networks (TextCNN), in conjunction with language embedding techniques to classify tweets about drug misuse, as reported in their study referenced [21]. The use of deep neural network AE-based hate speech identifying [22] has effectively navigated perplexing data. The SentiDiff methodology was devised to detect instances of data ambiguity by utilizing the transfer method in conjunction with deep convolutional neural networks. According to a study conducted by researchers, a CNN-LSTM model with a tree structure at the regional level has demonstrated a dependable level of accuracy in recognizing emotions. This finding is documented in reference [23]. Researchers have developed a one-class classification system that focuses on identifying rumors in online communities, as rumors have the potential to incite and propagate bigotry. This approach is novel and documented in the literature [24].

Scholars have suggested alternative methodologies for detecting hate speech, including semi-supervised [25] and unsupervised [26] techniques and the conventional supervised learning approach. Singh et al. [25] introduced the method of opinion hashtags expected to embed through performing multiple tasks learning. This approach employs an autoencoder (AE) and convolutional neural network (CNN) to classify the emotional content of a given dataset in a semi-supervised fashion. The Disaster communal (DisCom) method employs an unsupervised rule-based technique, suggested in [26], to categorize unsuitable tweets and instances of hate speech disseminated through social media. The Disaster communal (DisCom) technique, which operates without control and is based on rules, has been proposed [26] to identify and categorize hate speech and offensive tweets on online social media. The classification of sentiment at the level of sentences has been demonstrated to be effective, mainly because social media data is often characterized by a high degree of noise. Utilizing a blend of computational procedures facilitates precise recognition and perception of human sentiment. The author's Zhou et al. [27] propose a fusion technique based on deep neural networks that integrate language model embeddings (Elmo), bidirectional encoder representations from Transformers, and CNN to identify hate speech. The researcher in reference [28] explores techniques for detecting inflammatory language using pattern-based approaches, as discussed in [29]. Liu et al. utilized a mixed fuzzy rule construction technique within a fuzzy logic approach to identify hate speech from uncertain data, as described in [30].

The proliferation of deep learning in diverse application domains has led to the emergence of novel natural language processing (NLP) applications. Several deep learning strategies have been shown to outperform specific methods used for machine learning. As an illustration, a group of researchers conducted experiments to distinguish between using hate speech and using profanity. The researchers employed ensemble learning techniques, which yielded an accuracy rate of 87%. Nevertheless, the present study could benefit from optimizing the hyperparameters of the meta-classifiers. Furthermore, the study cited in reference [13] utilized Convolutional Neural Networks (CNN) to detect occurrences of hate speech in textual information

One potential method to enhance this task's effectiveness involves utilizing LSTM and Bi-LSTM models to determine the sequential structure of the data. The study referenced as [16] employed logit-boost and LSTM techniques to conduct character-level classification to obtain high-level type. Similarly, research was conducted on Bangla language text using LSTM and GRU techniques to detect events of online harassment, as reported in reference [14]. A study by [15] aimed to explore the effectiveness of deep neural networks in detecting online harassment on various social media platforms [17]. The experiments employ LSTM, BiLSTM, GRU, and RNN as learning algorithms. This section of the paper provides a detailed analysis of various academic outcomes about categorizing hate speech based on textual content.

A. *Sparrow Search Algorithm*

The Sparrow Search Algorithm is an innovative optimization algorithm that draws inspiration from the collaborative intelligence exhibited by sparrows during their foraging activities. The foraging behavior of sparrows can be categorized as a producer-scrounger model, comprising an exploratory mechanism and a pre-emptive alert mechanism. Producers demonstrate a notable degree of fitness. The producers are responsible for acquiring sustenance, designating foraging sites, and providing direction for scavenging pathways. Scroungers improve their overall fitness by selectively attending to trail producers that exhibit the most favorable fitness characteristics, thus obtaining nourishment. Moreover, designated individuals will oversee the energy status of the producers, who are viewed as individuals who get resources through unscrupulous means.

If the producer possesses a high energy level, some scavenger population will engage in active food snatching. Upon detecting a predator, the sparrow promptly emits an alarm signal and relocates to a secure location. The sparrows within the central region of the population exhibit random movements towards their conspecifics. When the safety threshold falls below the alarm value, producers must guide scavengers away from the hazardous area.

Assuming a population of N sparrows, each sparrow searches within a D-dimensional search space. the spatial coordinates of each sparrow instance can be denoted within an NxD matrix. Assuming that $X(i, j)$ represents the location of the i^{th} sparrow in the j^{th} dimension, and it can be inferred that i is an integer within the range of 0 to N, and j is an integer within the range of 0 to D. The sparrow population is subject to constraint conditions where by the SSA takes 10%-20% of the producer. The position update equation 1 has depicted as follows:

$$X^c_{i,j} = \begin{cases} X^c_{i,j} * exp(\frac{-i}{\alpha * c_{max}}), A_v < S_T \\ X^c_{i,j} + r * L, A_v \geq S_T \end{cases} \tag{1}$$

The variable "c" indicates the present iteration count, whereas the symbol "α" indicates a uniformly distributed random number ranging between 0 and 1. The symbol "c_{max}" denotes the upper limit of iterations, while "r" denotes a random variable that follows a standard normal distribution. The matrix denoted by "L" has dimensions of 1 by D and is comprised entirely of elements equal to 1. The variable "A_V" indicates the alarm value, a scalar quantity that can take values between 0 and 1. On the other hand, "S_T" represents the safety threshold, which is also a scalar quantity that ranges from 0.5 to 1. If A_V is smaller than S_T, a predator's presence is absent, thereby allowing for the execution of a comprehensive search. If A_V equals or surpasses S_T, the predator has been detected. Scroungers obtain food by following the producers, and the location of the producers is updated using the following formula in equation 2.

$$X_{i,j}^{c+1} = \begin{cases} Q * exp\left(\frac{X_{worst}^{c} - X_{i,j}^{c}}{i^2}\right), i > \frac{n}{2} \\ X_{p}^{c+1} + \left|X_{i,j}^{c} - X_{p}^{c+1}\right|.A^{+}.L, i \leq \frac{n}{2} \end{cases} \quad (2)$$

The equation, as mentioned earlier, encompasses multiple variables, where in X_p denotes the optimal position for the producer, and X_{worst} represents the global worst position. Matrix A comprises randomly generated elements that are either 1 or -1 and conforms to a particular equation. For example, suppose the value of i exceeds half of n. In that case, it can be inferred that the i^{th} individual who scavenges for sustenance is experiencing hunger, exhibiting suboptimal physical condition and diminished energy levels, and thus must relocate to alternative regions for nourishment. In this scenario, the individual who is scavenging for sustenance trails the producer who is situated in the most advantageous location.The matrix denoted by A^{+} fulfils the following equation. The formula for calculating A^{+} is $A^{T}(AA^{T})^{-1}$.

The equation 3 employed for updating monitors is as follows.

$$X_{i,j}^{c+1} = \begin{cases} X_{best}^{c} + \beta.\left|X_{i,j}^{c} - X_{best}^{c}\right|, f_i > f_g \\ X_{i,j}^{c} + K.\left(\frac{\left|X_{i,j}^{c} - X_{worst}^{c}\right|}{(f_i - f_w) + \varepsilon}\right), f_i = f_g \end{cases} \quad (3)$$

In the context provided,X_{best} denotes the optimal location within the entirety of the area. The control parameters β and K are utilized to regulate the step size. These parameters are generated randomly from a standard normal distribution and a range between -1 and 1. The variable f_i represents the sparrow's present level of fitness. Meanwhile, the fitness values of the optimal and suboptimal positions across the entire area are denoted by f_g and f_w, respectively. To avoid the occurrence of division by zero, a minute constant denoted as ϵ is employed. When the value of f_i exceeds f_g, it indicates that the sparrow is positioned at the periphery of the population, rendering it susceptible to predation. Under such circumstances, the sparrow must seek out other members of the group to ensure its safety.

The main strengths of deep learning techniques used in state-of-the-art methods are automatic feature selection, highly scalable, and transfer learning capabilities, but the weaknesses of deep learning methods are Black Box Nature and limited capacity for complex patterns. Due to this, we developed an LSTM hate speech detection model by applying the sparrow search method to optimize LSTM networks. This new LSTM-SSA approach involves the integration of SSA with deep learning, employing tuning SSA parameters, addressing its sensitivity and improving its convergence speed. Sparrow Search Algorithm is known for its simplicity, ease of implementation, and efficiency in finding optimal or near-optimal solutions.

3. PROPOSED LSTM-SSA MODEL

Sparrow Search Algorithm (SSA) is a locally robust, searchable algorithm with a minimal number of control variables. The random initialization method is employed to determine where the sparrow first arrived. However, because of some people's ideal initial sites deviating too much from the genuine optimal places, the convergence speed and accuracy of the solution are reduced, even though the initial positions are guaranteed to be random using this method. The SSA offers several options for resolving optimization problems, including rapid convergence and reliable search. Several recent studies [31] have utilized an SSA successfully in various engineering disciplines, providing valuable context for our investigation [32]. In contrast to the approach used in the cited research [33], we optimized the network model's hyperparameters to reduce the impact of humans on the model and increase its ability to make predictions. We choose to optimize the learning rate, the number of LSTM neurons, dense layer neurons, and epochs. SSA can improve these DNN target parameters to ensure the coherence of data features and model structure.

SSA's ability to quickly and effectively identify the optimal parameter values for a deep neural network's algorithms improves its interpretability. The model illustrated in Fig. 1 represents the LSTM-SSA approach suggested to achieve our goal of training deep neural networks that have been optimized. The LSTM-SSA network's architecture comprises four distinct layers: a layer for input, a layer that uses LSTM, a hidden layer, and a layer that provides output. The optimization parameters of interest to the SSA include the learning rate, epochs, and the number of neural networks in the two hidden layers. After specifying the parameter range, the population's positional information and corresponding parameters are randomly initialized. This study investigates the fitness function implemented in the Sparrow Search Algorithm (SSA) and the loss function utilized in the LSTM network, which is used in the framework of the learning model.

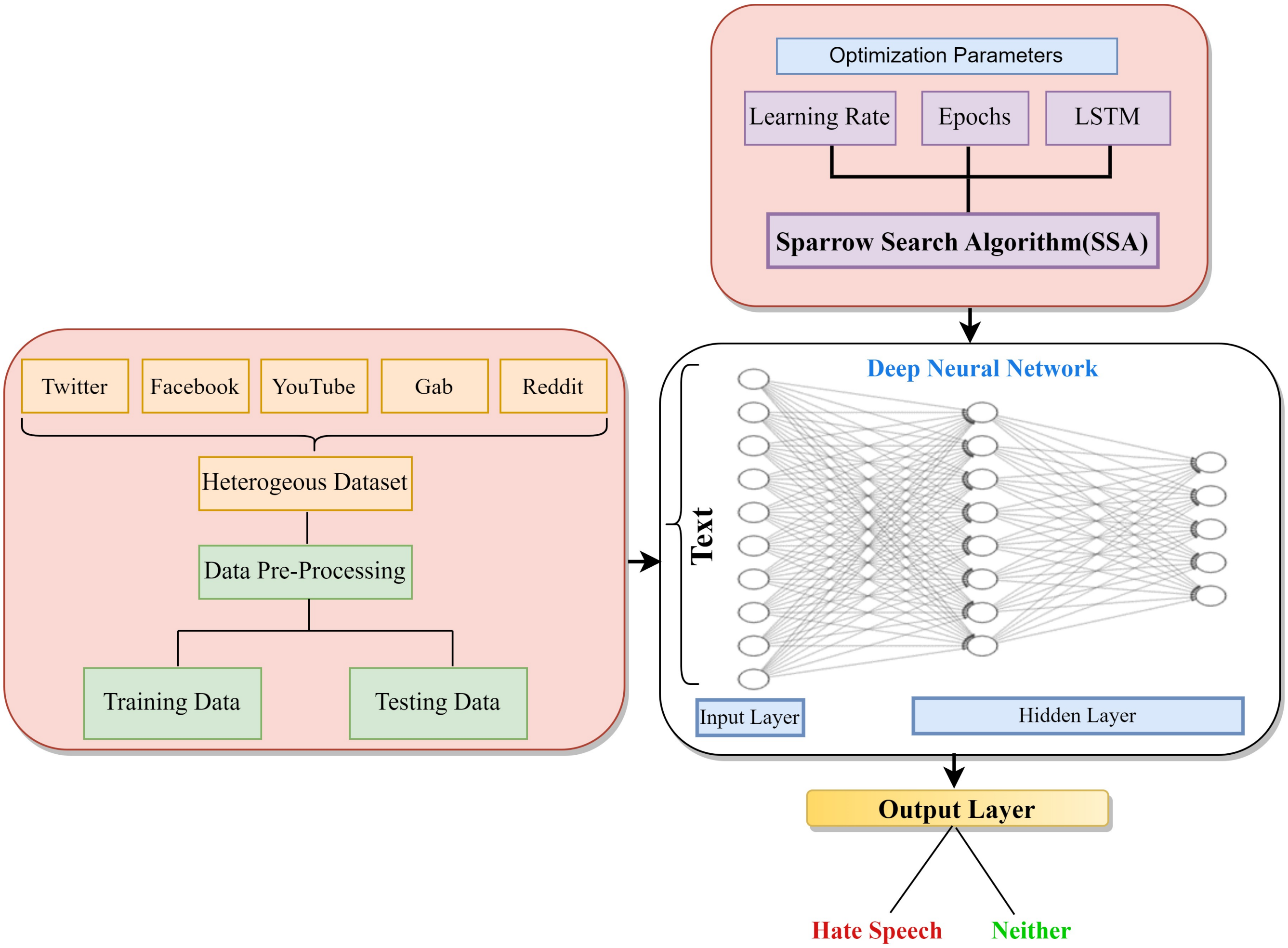


Figure 1. Architecture of the proposed LSTM-SSA model

TABLE I. Summary of existing heterogenous datasets from multi-social medias

S.No.	Data Source	Paper	Year	ML Approach	Dataset
1	Twitter	Davidson et al. [1]	2017	LR, SVM, DT, NB	24783
2	Twitter	Thomas et al. [34]	2019	LSTM	7005
3	Twitter	Zampieri et al. [35]	2019	CNN	13240
4	Twitter	Ousidhoum et al. [2]	2019	BiLSTM, BOW	5647
5	Twitter	Golbeck et al. [36]	2017	Corpus	20360
6	Twitter	Founta et al. [37]	2018	Corpus	45407
7	Facebook, YouTube	Chung et al. [38]	2019	Corpus	20186
8	Facebook, YouTube	Salminen et al. [39]	2020	SVM, LR, DT, RF, Adaboost	3222
9	Gab	Kennedy et al. [40]	2022	Gab corpus	22527
10	Reddit	Kurrek et al. [41]	2020	Reddit Corpus	40000

TABLE II. Summary of Multi-Social Media Attributes

Class	Label	Total Records
Hate Speech	0	113651
Neither	1	88726

4. EXPERIMENT RESULTS AND ANALYSIS

A. Dataset Description

In the present study, we used ten annotated datasets previously existing [[1], [2], [34]–[41]].Six of these datasets came from Twitter [[1], [2], [34]–[37]], two came from YouTube/Facebook (Chung et al. [38] and Salminen et al. [39]), one came from Gab (Kennedy et al. [40]), and the final datasets came from Reddit (Kurrek et al. [41]). Table I displays these datasets, which have been obtained from various social media platforms. These annotated datasets, collected from Twitter, YouTube, Facebook, Gab, and Reddit, include messages written in English and are classed as either hate speech or neither. Finally, we aggregated all of the datasets from the various social media platforms into a single generalized dataset with a total of 202377 items, and Table II presents the messages in this dataset categorized as either Hate Speech or Neither.

B. Data Pre-processing

Understanding and enhancing our method's efficiency requires analyzing, categorizing, and modifying the data. At first, we thought of using data analysis to extract generalizable features from the textual data. Emotional characters, passwords, URLs, other representations and noise characters are removed during the first stage of pre-processing. Then, you'll need to remove any punctuation and turn hashtags like #BanBlack into plain old text by replacing the hash sign with a regular letter. Next, all textual information should undergo stemming, lemmatization, and uppercase conversion. Finally, after tokenizing the ordinary text data, we extracted 44577 unique tokens from various social media datasets.

C. Evaluation Metrics

1) Accuracy

Accuracy is a performance measure of the correct predictions of the classifier.

$$accr = \frac{TruePos + FalsePos + TrueNeg + FalseNeg}{TruePos + TrueNeg}.$$

2) Precision

Precision (also referred to as positive predictive value): precision is the fraction of the total number of positive correctly classified among all positive classified classes.

$$p = \frac{TruePos}{TruePos + FalsePos}.$$

3) Recall

Recall (also referred to as sensitivity): The recall is the fraction of the total number of positives correctly classified among all positive classes.

$$r = \frac{TruePos}{TruePos + FalseNeg}.$$

TABLE III. Optimizing a range of hyperparameters for the LSTM-SSA model

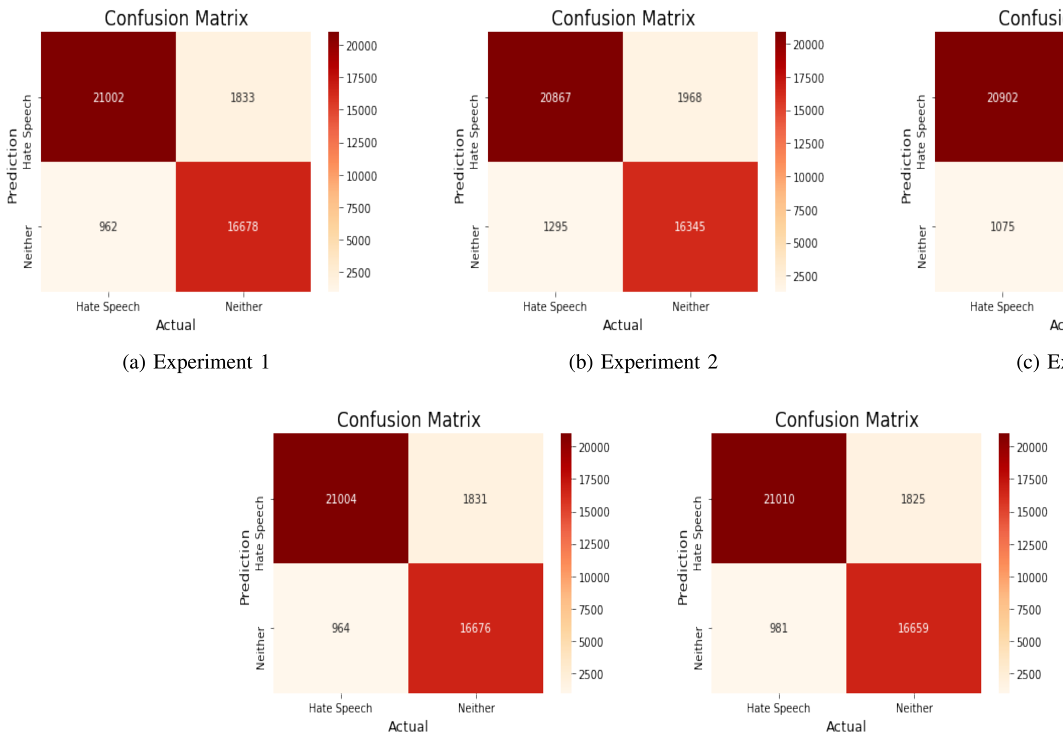
Hyper Parameters	Search Range	Value
Learning rate	0.0001 – 0.001	0.00037
Epochs	1-100	42
LSTM layer neurons	1-100	72
Dense layer neurons	1-100	95

D. Results of optimizing hyperparameters by the SSA on LSTM model

The Long Short-Term Memory (LSTM) model's hyperparameters amenable to optimization using Singular Spectrum Analysis (SSA) comprise the number of weights within the LSTM units, the learning rate, the number of weights in the dense layer, and the number of epochs. Table 3 provides a summary of the hyperparameter values. The optimal hyper-parameters of an LSTM were determined based on the data presented in Table III. These parameters include a rate of learning of 0.00037, an epoch period of 42, 72 weights in the as a layer, and 95 weights in the dense layer. A prediction model is developed to detect hate speech messages, utilizing the most significant parameter values obtained through SSA optimization.

Fig. 2 presents an ensemble of confusion matrices accomplished by the LSTM-SSA model when employed on the test dataset. The data indicates that the LSTM-SSA model has successfully distinguished the text into Hate Speech and Neither category. In Experiment-1, the LSTM-SSA model categorized 21002 texts as Hate Speech and 16678 texts as Neither. Similarly, in Experiment -2, the LSTM-SSA model successfully classified 20867 texts as Hate Speech and 16345 as Neither. In Experiment -4, the LSTM-SSA model successfully identified 21004 texts as classified under Hate Speech, while 16676 texts have classified under the Neither class. Finally, the LSTM-SSA model has successfully identified 21010 texts as Hate Speech and 16659 texts as Neither type in experiment-5.

Table IV presents an in-depth investigation of the classification results obtained from the LSTM-SSA model across five experiments regarding precision, recall, and accuracy. The experiment's findings indicate that the LSTM-SSA model effectively classified texts across multiple iterations. In Experiment 1, the LSTM-SSA model achieved an accuracy of 0.930. Experiment-2 yielded an LSTM-SSA model that achieved an accuracy of 0.919. The LSTM-SSA model has acquired an accuracy is 0.930 in experiment 4. Finally, experiment 5 resulted in an LSTM-SSA model with an accuracy of 0.936. Table



(a) Experiment 1

(b) Experiment 2

(c) Experiment 3

(d) Experiment 4

(e) Experiment 5

Figure 2. Confusion matrix of the proposed LSTM-SSA model

TABLE IV. Result Analysis of the proposed LSTM-SSA model.

Experiment	Precision	Recall	Accuracy	F1- score
1	0.919	0.956	0.930	0.937
2	0.913	0.941	0.919	0.927
3	0.915	0.951	0.925	0.932
4	0.919	0.956	0.930	0.937
5	0.920	0.955	0.936	0.937

V presents a comprehensive comparative analysis of the LSTM-SSA model with other recent deep-learning models. First, the simulation values of the methods developed by Salminen et al. [39] and Zampieri et al. [35] have used classical machine learning approach such as Logistic Regression (LR), Decision Tree (DT), and Support Vector Machine (SVM) for indicated a reduced level of accuracy, with values of 0.789 and 0.802, respectively. Subsequently, the techniques proposed by Kurrek et al. [41] and Vashistha et al. [42] have used deep learning models like Convolutional Neutral Network (CNN), Long Short-Term Memory (LSTM) for achieved good precision, precisely 0.893 and 0.913, respectively. Finally, the LSTM-SSA model achieved a peak accuracy of 0.936. The findings indicate that the LSTM-SSA model outperformed other deep learning techniques, ensuring improved outcomes.

5. CONCLUSION

We propose optimizing neural network hyper parameters with an LSTM-SSA approach to detecting hate speech. This method was developed to avoid relying too heavily on expert opinion and historical data when choosing LSTM network hyper parameters. The LSTM model's parameters have been optimized with the help of the sparrow search method, which provides an accurate account of the network architecture and parameter settings used by the model. The procedure above reduces the impact of human elements on the identification of hate speech and enhances the overall capacity of the model to generalize and predict outcomes. The empirical investigations on simple and complex network structures are highly effective hyper parameter optimization techniques. Moreover, empirical findings indicate that LSTM-SSA exhibits superior optimization efficiency to traditional cutting-edge methodologies.

The Sparrow Search Algorithm (SSA) exhibits several limitations that warrant consideration. Firstly, the algorithm demonstrates sensitivity to parameter choices, a common trait among metaheuristic algorithms. Fine-tuning these parameters for optimal performance poses a non-trivial challenge. Additionally, while SSA has proven effective in specific optimization problems, its applicability may be restricted in highly specialized or complex domains where alternative algorithms may outperform it. Convergence speed

TABLE V. LSTM-SSSA model evaluation using cutting-edge methodologies.

Paper	Methods	Precision	Recall	Accuracy
Vashistha et al. [42]	CNN, LSTM, BERT	0.937	0.929	0.913
Kurrek et al. [41]	LR, BERT	0.893	0.893	0.893
Salminen et al. [39]	LR, DT, SVM	-	-	0.789
Zampieri et al. [35]	SVM, LSTM	0.824	0.821	0.802
Proposed Model	**LSTM-SSA**	**0.920**	**0.955**	**0.936**

is another limitation, with SSA potentially exhibiting slower convergence compared to other optimization algorithms.

Future research directions for the Sparrow Search Algorithm (SSA) present promising avenues for advancing its effectiveness and applicability in optimization domains. The investigation into adaptive parameter tuning strategies seeks to enhance the algorithm's adaptability across diverse problem domains, potentially strengthening its robustness. Exploring hybrid approaches, particularly the combination of SSA with other optimization techniques, holds the promise of leveraging the strengths of each, paving the way for improved overall performance.

References

[1] W. D. M. M. Davidson, T. and I. Weber, "Automated hate speech detection and the problem of offensive language," *International Conference on Web and social media, ICWSM*, pp. 512–515, 2017.

[2] H. Z. Y. S. Nedjma Ousidhoum, Zizheng Lin and D.-Y. Yeung, "Multilingual and multi-aspect hate speech analysis," *arXiv*, pp. 1–10, 2019.

[3] D. R. Jana Kurrek, Haji Mohammad Saleem, "Towards a comprehensive taxonomy and large-scale annotated corpus for online slur usage," *Proceedings of the Fourth Workshop on Online Abuse and Harms*, pp. 138–149, 2020.

[4] S. Hochreiter and J. Schmidhuber, "Lstm can solve hard long time lag problems," *Proceedings of the 9th International Conference on Neural Information Processing Systems*, p. 473–479, 1996.

[5] Z. Waseem, "Are you a racist or am I seeing things? annotator influence on hate speech detection on twitter," *Proceedings of the First Workshop on NLP and Computational Social Science*, pp. 138–142, 2016.

[6] P. Badjatiya, S. Gupta, M. Gupta, and V. Varma, "Deep learning for hate speech detection in tweets," *Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 759–760, 2017.

[7] M. Mondal, L. A. Silva, and F. Benevenuto, "A measurement study of hate speech in social media," *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, pp. 85–94, 2017.

[8] A. H. M. M. J. S.-g. A. J. K. H. . J. B. Salminen, J., "Anatomy of online hate: Developing a taxonomy and machine learning models for identifying and classifying hate in online news media," *Proceedings of the International AAAI Conference on Web and Social Media*, pp. 330–339, 2018.

[9] R. S. S. B. Susan Herring, Kirk Job-Sluder, "Searching for safety online: Managing "trolling" in a feminist forum," *The Information Society*, pp. 371–384, 2002.

[10] Z. Waseem, T. Davidson, D. Warmsley, and I. Weber, "Understanding abuse: A typology of abusive language detection subtasks," *Proceedings of the First Workshop on Abusive Language Online*, pp. 78–84, 2017.

[11] M. ElSherief, S. Nilizadeh, D. Nguyen, G. Vigna, and E. Belding, "Peer to peer hate: Hate speech instigators and their targets," *Proceedings of the International AAAI Conference on Web and Social Media*, pp. 52–61, 2018.

[12] M. Z. Shervin Malmasi, "Challenges in discriminating profanity from hate speech," *Journal of Experimental Theoretical Artificial Intelligence*, pp. 187–202, 2018.

[13] M. A. Al-Ajlan and M. Ykhlef, "Optimized twitter cyberbullying detection based on deep learning," pp. 1–5, 2018.

[14] M. T. Ahmed, M. Rahman, S. Nur, A. Islam, and D. Das, "Deployment of machine learning and deep learning algorithms in detecting cyberbullying in bangla and romanized bangla text: A comparative study," pp. 1–10, 2021.

[15] M. Dadvar and K. Eckert, "Cyberbullying detection in social networks using deep learning based models," pp. 245–255, 2020.

[16] X. Luo, "Efficient english text classification using selected machine learning techniques," *Alexandria Engineering Journal*, vol. 60, no. 3, pp. 3401–3409, 2021.

[17] U. Khan, S. Khan, A. Rizwan, G. Atteia, M. M. Jamjoom, and N. A. Samee, "Aggression detection in social media from textual data using deep learning models," *Applied Sciences*, vol. 12, no. 10, p. 5083, 2022.

[18] W. Alorainy, P. Burnap, H. Liu, and M. L. Williams, ""the enemy among us" detecting cyber hate speech with threats-based othering language embeddings," *ACM Transactions on the Web (TWEB)*, vol. 13, no. 3, pp. 1–26, 2019.

[19] R. Basak, S. Sural, N. Ganguly, and S. K. Ghosh, "Online public shaming on twitter: Detection, analysis, and mitigation," *IEEE Transactions on computational social systems*, vol. 6, no. 2, pp. 208–220, 2019.

[20] L. Gui, J. Leng, J. Zhou, R. Xu, and Y. He, "Multi task mutual learning for joint sentiment classification and topic detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 4, pp. 1915–1927, 2020.

[21] R. Sequeira, A. Gayen, N. Ganguly, S. K. Dandapat, and J. Chandra, "A large-scale study of the twitter follower network to characterize

the spread of prescription drug abuse tweets," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 6, pp. 1232–1244, 2019.

[22] R. Zhao and K. Mao, "Cyberbullying detection based on semantic-enhanced marginalized denoising auto-encoder," *IEEE Transactions on Affective Computing*, vol. 8, no. 3, pp. 328–339, 2016.

[23] J. Wang, L.-C. Yu, K. R. Lai, and X. Zhang, "Tree-structured regional cnn-lstm model for dimensional sentiment analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 581–591, 2019.

[24] A. E. Fard, M. Mohammadi, Y. Chen, and B. Van de Walle, "Computational rumor detection without non-rumor: A one-class classification approach," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 5, pp. 830–846, 2019.

[25] L. G. Singh, A. Anil, and S. R. Singh, "She: Sentiment hashtag embedding through multitask learning," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 2, pp. 417–424, 2020.

[26] K. Rudra, A. Sharma, N. Ganguly, and S. Ghosh, "Characterizing and countering communal microblogs during disaster events," *IEEE Transactions on Computational Social Systems*, vol. 5, no. 2, pp. 403–417, 2018.

[27] Y. Zhou, Y. Yang, H. Liu, X. Liu, and N. Savage, "Deep learning based fusion approach for hate speech detection," *IEEE Access*, vol. 8, pp. 128 923–128 929, 2020.

[28] H. Watanabe, M. Bouazizi, and T. Ohtsuki, "Hate speech on twitter: A pragmatic approach to collect hateful and offensive expressions and perform hate speech detection," *IEEE access*, vol. 6, pp. 13 825–13 835, 2018.

[29] P. Shah and A. Patel, "A comprehensive study and detailed review on hate speech classification: a systematic analysis," in *2021 International Conference on Advances in Computing, Communication, and Control (ICAC3)*. IEEE, 2021, pp. 1–6.

[30] H. Liu, P. Burnap, W. Alorainy, and M. L. Williams, "A fuzzy approach to text classification with two-stage training for ambiguous instances," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 2, pp. 227–240, 2019.

[31] F. Liu, P. Qin, J. You, and Y. Fu, "Sparrow search algorithm-optimized long short-term memory model for stock trend prediction," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[32] G. I. Rajathi, R. R. Kumar, D. Ravikumar, T. Joel, S. Kadry, C.-W. Jeong, and Y. Nam, "Brain tumor diagnosis using sparrow search algorithm based deep learning model," *Computer Systems Science and Engineering (CSSE)*, vol. 44, no. 2, pp. 1793–1806, 2023.

[33] C. Zhang and S. Ding, "A stochastic configuration network based on chaotic sparrow search algorithm," *Knowledge-Based Systems*, vol. 220, p. 106924, 2021.

[34] T. Mandl, S. Modha, P. Majumder, D. Patel, M. Dave, C. Mandlia, and A. Patel, "Overview of the hasoc track at fire 2019: Hate speech and offensive content identification in indo-european languages," in *Proceedings of the 11th annual meeting of the Forum for Information Retrieval Evaluation*, 2019, pp. 14–17.

[35] M. Zampieri, S. Malmasi, P. Nakov, S. Rosenthal, N. Farra, and R. Kumar, "Predicting the type and target of offensive posts in social media," *arXiv preprint arXiv:1902.09666*, 2019.

[36] J. Golbeck, Z. Ashktorab, R. O. Banjo, A. Berlinger, S. Bhagwan, C. Buntain, P. Cheakalos, A. A. Geller, R. K. Gnanasekaran, R. R. Gunasekaran *et al.*, "A large labeled corpus for online harassment research," in *Proceedings of the 2017 ACM on web science conference*, 2017, pp. 229–233.

[37] A. Founta, C. Djouvas, D. Chatzakou, I. Leontiadis, J. Blackburn, G. Stringhini, A. Vakali, M. Sirivianos, and N. Kourtellis, "Large scale crowdsourcing and characterization of twitter abusive behavior," in *Proceedings of the international AAAI conference on web and social media*, vol. 12, no. 1, 2018.

[38] Y.-L. Chung, E. Kuzmenko, S. S. Tekiroglu, and M. Guerini, "Conan–counter narratives through nichesourcing: a multilingual dataset of responses to fight online hate speech," *arXiv preprint arXiv:1910.03270*, 2019.

[39] J. Salminen, M. Hopf, S. A. Chowdhury, S.-g. Jung, H. Almerekhi, and B. J. Jansen, "Developing an online hate classifier for multiple social media platforms," *Human-centric Computing and Information Sciences*, vol. 10, pp. 1–34, 2020.

[40] B. Kennedy, M. Atari, A. M. Davani, L. Yeh, A. Omrani, Y. Kim, K. Coombs, S. Havaldar, G. Portillo-Wightman, E. Gonzalez *et al.*, "Introducing the gab hate corpus: defining and applying hate-based rhetoric to social media posts at scale," *Language Resources and Evaluation*, pp. 1–30, 2022.

[41] J. Kurrek, H. M. Saleem, and D. Ruths, "Towards a comprehensive taxonomy and large-scale annotated corpus for online slur usage," in *Proceedings of the Fourth Workshop on Online Abuse and Harms*, 2020, pp. 138–149.

[42] N. Vashistha and A. Zubiaga, "Online multilingual hate speech detection: experimenting with hindi and english social media," *Information*, vol. 12, no. 1, p. 5, 2020.

Ashwini Kumar (ashwinikumar.cse@geu.ac.in) received his M. Tech (CSE) from BIT, MESRA, Ranchi, Jharkhand, India in 2012. He is currently working as an Assistant Professor in the Department of Computer Science and Engineering at Graphic Era Deemed to be University, Dehradun, Uttarakhand, India. He has more than 15 years teaching experience. He has presented many research papers in International conferences and journals. His area of research interests includes Machine Learning, Deep Learning, Natural Language Processing, Autonomic Computing, and Quantum Computing.

Santosh Kumar (drsantosh.cse@geu.ac.in) completed his PhD at the Indian Institute of Technology, Roorkee (India), in 2012. He earned his M. Tech. (C.S.E.) from Aligarh Muslim University, Aligarh (India), in 2007, and a B.E. (I.T.) from C.C.S. University, Meerut (India), in 2003. He has over 17 years of teaching and research experience in undergraduate (B. Tech.) and postgraduate (M.Tech.) level courses. He has held positions as a Lecturer, Assistant Professor, and Associate Professor across various academic and research institutions. Dr Kumar has supervised 4 PhD theses and 26 M.Tech. Theses. Currently, he mentors 6 PhD students and 2 M. Tech (C.S.E.) students individually and collaboratively. During 2009-2011, he completed a consultancy project titled ”MANET Architecture Design for Tactical Radios” for DRDO, Dehradun. He actively serves as a board member for reviewing various national and international journals and conferences. Dr Kumar holds memberships in A.C.M. (Senior Member), IAENG, ACEEE, and ISOC (U.S.A.). He has contributed over 90 research papers to national and international journals and conferences, focusing on Wireless Communication Networks, Mobile Computing, Grid Computing, and Software Engineering. Presently, he holds the position of Professor at Graphic Era Deemed to be University, Dehradun (India). His research interests span Ubiquitous Computing, Artificial Intelligence, Natural Language Processing, and Software Engineering.