Computers, Materials & Continua
DOI: 10.32604/cmc.2023.033753
Article

Tech Science Press

Firefly-CDDL: A Firefly-Based Algorithm for Cyberbullying Detection Based on Deep Learning

Monirah Al-Ajlan* and Mourad Ykhlef

Department of Information Systems, College of Computer and Information Sciences, King Saud University, Riyadh, KSA
*Corresponding Author: Monirah Al-Ajlan. Email: maalajlan@ksu.edu.sa
Received: 27 June 2022; Accepted: 02 November 2022

Abstract: There are several ethical issues that have arisen in recent years due to the ubiquity of the Internet and the popularity of social media and community platforms. Among them is cyberbullying, which is defined as any violent intentional action that is repeatedly conducted by individuals or groups using online channels against victims who are not able to react effectively. An alarmingly high percentage of people, especially teenagers, have reported being cyberbullied in recent years. A variety of approaches have been developed to detect cyberbullying, but they require time-consuming feature extraction and selection processes. Moreover, no approach to date has examined the meanings of words and the semantics involved in cyberbullying. In past work, we proposed an algorithm called Cyberbullying Detection Based on Deep Learning (CDDL) to bridge this gap. It eliminates the need for feature engineering and generates better predictions than traditional approaches for detecting cyberbullying. This was accomplished by incorporating deep learning—specifically, a convolutional neural network (CNN)—into the detection process. Although this algorithm shows remarkable improvement in performance over traditional detection mechanisms, one problem with it persists: CDDL requires that many parameters (filters, kernels, pool size, and number of neurons) be set prior to classification. These parameters play a major role in the quality of predictions, but a method for finding a suitable combination of their values remains elusive. To address this issue, we propose an algorithm called firefly-CDDL that incorporates a firefly optimisation algorithm into CDDL to automate the hitherto-manual trial-and-error hyperparameter setting. The proposed method does not require features for its predictions and its detection of cyberbullying is fully automated. The firefly-CDDL outperformed prevalent methods for detecting cyberbullying in experiments and recorded an accuracy of 98% within acceptable polynomial time.

Keywords: Firefly optimization; convolutional neural network (CNN); cyberbullying; cyberbullying detection; text classification

1 Introduction

Technology dominates our lives nowadays as we rely on it to carry out most of our daily activities. Communication is no exception to this trend. Technology has changed how people interact with one another and added a new dimension to communication. As promising as this transition has been, this major shift from the traditional to the digital world has come at a considerable cost. The anonymous nature of social networks, where users use monikers rather than their real names, makes the actions of people very difficult to trace. This anonymity has resulted in a growing number of online crimes like cyberbullying, which now constitutes one of the most pressing ethical issues related to the Internet. Another reason why cyberbullying has emerged as a critical issue is that victims are usually too afraid or embarrassed to speak out [1]. Many studies have investigated cyberbullying to assess its prevalence online. The results have shown that it is a common problem that is affecting a growing number of users [2–6].

Due to the rise of cyberbullying, methods for its detection have attracted the attention of the research community in recent years. However, most researchers have focused on using the features of text to identify cyberbullying. The use of features requires human expertise, which can easily be fabricated. Further, the use of human expertise for identifying cyberbullying requires the involvement of teenagers and young adults, who might be hesitant to participate because they live in the digital world [7]. Several issues also need to be addressed prior to the classification of cyberbullying, such as the specific features that are needed to this end, ways to extract them, and the number and types of features that need to be selected in case there are too many of them. Our previously proposed algorithm, CDDL [8], addressed the issue of human expertise by eliminating the need for feature extraction and selection, and replacing these steps with a CNN in combination with word embedding. The striking advantage of word embedding is that it captures the semantics of the given text. In other words, it can generate numerical representations of semantically related words. In this way, the meaning of the text can be incorporated into the system, rather than having to deal with it as an abstract form of data.

Although CDDL produces good predictions, the quality of prediction is highly dependent on the CNN structure (the number of filters, the number of kernels, etc.). In other words, the CNN structure is not fixed, and many structures have been introduced for different tasks. Time is a major issue as finding a structure that produces good results is manual and takes exponential time. Therefore, the main contribution that this paper makes is that it proposes a new algorithm, firefly-CDDL, which solves the problem of determining the CNN structure using the firefly-optimization algorithm. Remarkably, firefly-CDDL can determine a good (maybe optimal) CNN structure in polynomial time, making cyberbullying detection a fully automated process.

The remainder of this paper is organized as follows: Section 2 summarizes past work on cyberbullying, while Section 3 details the firefly-CDDL proposed here and highlights its strengths. Section 4 describes experiments to verify the proposed method along with their results. Finally, Section 5 summarizes the conclusions of this paper.

2 Related Work

2.1 *Cyberbullying Detection*

The first study to tackle bullying on social media was [9]. The researchers built a framework to incorporate the API of Twitter streams to collect user tweets and classify them according to content. Tweets were classified as either positive or negative, and then further classified as positive and containing bullying content, positive without bullying content, negative and containing bullying

content, and negative without bullying content. The Naïve Bayes (NB) algorithm was implemented for classification, and the method had high accuracy (70%). In [10], the researchers followed a simple approach to developing a prototype system for use by members of organisations to monitor social network sites and detect incidents of cyberbullying. This approach involved recording bullying words, storing them in a database and then incorporating the Twitter API to capture tweets and compare their contents to the bullying material recorded earlier. However, this prototype system has not yet been implemented. In [11], the researchers proposed a similar prototype system to govern website content. This system consists of three components: link filtering, age validation and comment validation. Link filtering is implemented to ensure that the returned search results have been filtered according to ranking. Age validation is used to check the eligibility of a user to view a given site. If a user does not conform to the prescribed age restriction, they are blocked from the site. When a user is logged on to a site, the content filtering component is invoked to examine their posted texts using machine learning algorithms. This system was implemented using the support vector machine (SVM) classification algorithm.

The researchers in [12] studied content features such as first- and second-person pronouns as well as the vulgarities feature and showed that they are indeed indicators of cyberbullying. Researchers followed a similar approach in [13], where they primarily depended on content features like racist words and profanity. An interdisciplinary study that followed a similar approach was conducted through the dual lens of computer science and human behaviour. As reported in [14], for their proposed method, the researchers extensively studied the content features, including URLs and hashtags. Interestingly, 64% of tweets contained URLs to external sites, while 74.2% contained hashtags. The researchers discovered that these two features were not indicators of bullying. Similarly, [15] continued to pursue cyberbullying detection from a content-based perspective; however, they introduced new features: emoticons and a dictionary of hieroglyphs. Their approach was tested using many learning algorithms, such as SVM and J 48. SVM achieved the best results, with 81% accuracy.

Subsequent research has followed a more complicated approach and has avoided simple measures in favour of complex statistical measures. For instance, [16] incorporated term frequency–inverse document frequency and latent Dirichlet allocation into topic models to extract the relevance of documents. The researchers did not rely only on statistical measures, instead extracting the features of the content, such as profanity and pronouns. In [17], researchers used document frequency, which identifies the number of documents in which a given word appears. They defined a threshold of 0.1% for words to be included, and words with a percentage of occurrence below this threshold were considered to have been misspelled. Cyberbullying was predicted in [18] using the resolution of the coreference. The researchers used stemming and lemmatisation, which involves taking a word as the input and returning its root while removing any prefixes or suffixes Coreference resolution also involves initially finding an entity and then analysing all expressions that refer to it. The method was tested using SVM and NB and recorded a low precision of 0.5%. Another statistical measure used in the literature is the bag of words (BoW), a well-known method of natural language processing. The BoW statistically creates a dictionary of words. A statistical content-based approach was also proposed in [19], where the researchers implemented a language model and a machine learning model to identify terms that reflected cyberbullying. The language model used the BoW because of its simplicity and transparency. The researchers used it to create a term-by-document matrix for 10,685 posts and to run queries. Profanity was then used in the language model as query terms. The number of posts identified as containing cyberbullying content by human annotators (true positives) was then recorded. The BoW approach returned too many words, and thus, a threshold of two was set and words that appeared only once were ignored. As this framework fell under the umbrella of information retrieval, precision and

recall were both used as evaluation metrics to verify its performance. The results showed that queries containing profanity outperformed more complex queries that contained contextual information. The researchers also proposed the essential dimensions of LSI, an approach based on machine learning that identifies additional terms. This approach is based on the generalised model of the vector space and extracts the meanings of words according to their co-occurrence with other terms in the corpus. However, the calculation of co-occurrence can result in computational overhead. To address this, singular value decomposition, a linear algebraic technique, was implemented and delivered 80% precision. The researchers in [20] used more sophisticated statistical measures to understand a given text. They applied the linguistic inquiry and word count, which is a program for textual analysis that counts the number of words and categorises messages in semantically meaningful classes. This was used to identify similarities between all of the comments posted by a given user and posts by the same user on other social media sites. The researchers reported many patterns across platforms that helped identify cyberbullying.

All past work on cyberbullying detection has been single point [21]. In other words, detection has been carried out using a single node that implements a single classification algorithm. This approach is unfeasible, given the potential for computational overheads due to the massive amount of real-time Twitter data. To address this issue, researchers have introduced a collaborative detection framework with many nodes distributed across machines, each implementing a different algorithm, which can distribute the overhead. The nodes extract features from tweets using WordTokenizer with the minimum frequency of one word, WordTokenizer with the minimum frequency of two words, Bi-GramTokenizer with the minimum frequency of one word and Bi-GramTokenizer with the minimum frequency of two words. Various classification algorithms, including SVM, were implemented with 70% accuracy.

In most cases, cyberbullying is associated with fake accounts, such that there is no accountability for abusive online behaviour. As such, [22] focused not on detecting incidents of bullying but rather on identifying fake accounts and linking them to their real counterparts. This technique has been used to monitor and control cyberbullies. The features used were mostly based on content, such as the language and the time of posting. The researchers used several supervised techniques, including random forest and J48, and reported 68% accuracy.

Recently, [23] analysed Twitter for the detection of fake news regarding COVID-19. The researchers first pre-processed the tweets and then applied semantic models to structure the data. For the sake of evaluation, they used eight machine learning algorithms, and the best results were achieved using logistic regression, decision tree, SVM and AdaBoost. Similarly, they applied several deep learning algorithms on the same dataset and BiLSTM algorithms performed the best, achieving 97% in all three measures: precision, recall and accuracy. Deep learning also performed better than classical machine learning algorithms in [24], where the researchers were interested in building domain ontology for Alzheimer's disease. They compared classical machine learning approaches, such as logistic regression and gradient boosting, with deep learning approaches such as CNN, which outperformed classical approaches and permitted future scalability and robustness. Similarly, the researchers in [25] addressed RNA sequencing, which presents a great deal of difficult tasks, given the high dimensionality of data. They used RNA sequencing to look for the independent biomarkers of different types of cancer, which necessitated multiple analyses for multiple cancer types. The researchers studied data from the Mendeley data repository, which included five cancer types, and followed a three-step framework. First, they transformed the RNA sequencing into an image, extracted useful features and performed classification. They applied several deep learning models, and the results indicate that the convolutional neural network performed the best.

2.2 Optimization Algorithms in Data Mining

The firefly algorithm (FA) has been widely used in optimisation problems, and several studies have focused on the applicability of FA in data mining. The researchers in [26] investigated the use of FA in energy conservation, which poses a critical challenge in the IoT. Specifically, the routing problem is a key issue in the IoT, and parent selection plays a role in the efficiency of the network. They considered each node in the network a firefly and calculated its location, attraction of other fireflies, random function and velocity. Comparing this approach to the state-of-the-art work in the field, this algorithm improved packet transmission by 5%. Similarly, in [27], the researchers studied the use of FA in feature selection, aiming to increase classification accuracy while reducing the number of features of the slime mould algorithm. The goal of incorporating FA was to enhance the exploration of the algorithm to find feasible regions that could potentially have an optimal solution. Their proposed algorithm was tested on a 20-UCI dataset and resulted in enhanced performance. In the same vein, the researchers in [28] studied the use of FA in chronic kidney disease diagnosis. The role of FA was to choose the best subset of features that would result in the highest prediction accuracy. They tested the algorithm and investigated its sensitivity, specificity and accuracy and concluded that the addition of FA improved prediction accuracy, compared to existing models.

Several optimisation algorithms have been used in data mining to optimise different tasks. For instance, in [29], the researchers studied the bat algorithm (BA), which is a nature-inspired metaheuristic algorithm that has been widely used to solve global optimisation problems. The goal of the research was to enhance BA to eliminate the problem of frequent capture in local optima. The researchers used the enhanced algorithm to improve the complex process of training artificial neural networks. Classical artificial neural network training suffers from several challenges, such as getting stuck in the local minima and the maximum number of iterations required. These shortcomings were surmounted by the proposed algorithm, a BA artificial neural network, which resulted in better accuracy and convergence speed. BA was also used in [30] to address the issue of balancing space's exploitation and exploration. The researchers accomplished this by introducing a modified BA algorithm that incorporated the torus walk. Their proposed work has been tested on 19 benchmark datasets, and it outperformed traditional BA, directional BA, particle swarm optimisation, cuckoo search, harmony search algorithm, differential evolution and genetic algorithm.

Metaheuristic optimisation algorithms have shown great success in solving optimisation problems; however, they are highly dependent on the population initialisation, which significantly affects the solution convergence. In [31], the researchers studied quasi random sequences in an effort to improve the convergence problem, but there were still shortcomings. Their work introduced three new population initialisation low-discrepancy sequences: the WELL sequence, the Knuth sequence and the Torus sequences. Their sequences were tested on the artificial neural network training of a well-known public dataset, and the improvements found in this study are promising.

3 Firefly-Based Cyberbullying Detection Based on Deep Learning Firefly-CDDL

The FA algorithm seeks to mimic the behavior of fireflies [32]. Fireflies move in swarms and use their flashing lights to communicate with one another and attract other fireflies. Dimmer fireflies are attracted to brighter ones. When implementing the FA, three assumptions need to be made:

- All fireflies are attracted to other fireflies regardless of their sex.
- The degree of the attractiveness of a firefly is proportional to its brightness.
- The brightness of a firefly is determined by a fitness function.

3.1 Problem Definition and Initialization

FA algorithm was built as a general method to solve optimization problems. Even though its principles are static, they must be tailored to suit the problem at hand. In the case of cyberbullying detection, we define the problem as follows:

- We define a firefly as a potential CNN that consists of four values: "filter," which represents the number of filters in the convolutional layer, "kernel," which is the size of the kernel in the convolutional layer, "pool size," which represents the size of the pooling matrix in the max-pooling layer, and "dense," which represents the number of neurons in the dense layer.
- We define the number of fireflies n as a random integer that is smaller than 10 (a constraint on complexity).
- We define the number of generations g as a random integer that is smaller than 10 (a constraint on complexity).
- We define the population of fireflies F as a set of fireflies, as defined above.
- We initialize the search with a random location of the fireflies.
- Let's initialize the search with a random location of fireflies.

3.2 Objective Function Definition

The objective function is important for optimization algorithms because the goal of optimization is to either maximize an objective function, such as the classification accuracy or to minimize it, such as the classification error. The problem in firefly-CDDL is a maximization problem because the criterion considered is the accuracy of classification as calculated by four values: true-positive tweets (TP), true-negative tweets (TN), false-positive tweets (FP), and false-negative tweets (FN). Our objective function Obj is defined as follows:

$$\text{Obj} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

3.3 Fireflies' Lightness and Attractiveness

A major principle of the movement of fireflies is their attractiveness, which is determined by the intensity of the light emitted by them. In firefly-CDDL, we relate l to the objective function Obj. However, it is not possible to say that l = Obj because the intensity of light is affected by two other factors [33]: the distance between fireflies r_{ij}, and the absorption coefficient of substances in the environment γ, which was set to one. The intensity of light I between fireflies i and j is defined as follows:

$$l_{i,j} = Obj * \ e^{-\gamma r^2} \tag{2}$$

In firefly-CDDL, the attractiveness between fireflies i and j, β_{ij}, is equal to the intensity of their light. The attractiveness β of a firefly is defined as follows

$$\beta_{ij} = l_{ij} \tag{3}$$

3.4 Distance Function

The Euclidean distance was used as the distance function in firefly-CDDL. Let d be the dimensionality of the problem; then, the distance must be calculated between every pair of fireflies. The distance ri, j between fireflies fi and fj is the difference between their locations Xi and Xj, respectively. It is calculated according to the following equation:

$$r_{i,j} = \|X_i - X_j\| = \sqrt{\sum_{n=1}^{d} \left(X_{i,j} - X_{ji} \right)^2} \tag{4}$$

3.5 Evaluation

After every iteration, all the fireflies are tested for their attractiveness, and the best firefly is marked so that the others move toward it according to a movement function. With the definition of the attractiveness between fireflies, β_{ij}, the dimmer firefly j moves to the brighter firefly i according to the distance function. However, randomization must be used so that the search is not stuck. The method of randomization used in firefly-CDDL is the one described in Ref. [34], $\propto \epsilon_i$. It represents a vector of random variables ϵ_i. The movement function from firefly i at location X_i to firefly j at location X_j is then defined as follows:

$$l_i = l_i + \beta_i * \left| X_j - X_i \right| + \propto \epsilon_i \tag{5}$$

3.6 Termination of Search

FA-based optimization searches for optimal solutions but does not always find one. In firefly-CDDL, the search is terminated when a preset number of generations has been exceeded. This number varies according to the problem at hand. The proposed algorithm is shown in Fig. 1.

Algorithm 1: *Firefly-CDDL*

Input: Objective function, *Obj* from Eq. (1)
Light function, *I* from Eq. (2)
Attractiveness function, β from Eq. (3)
Distance function, $r_{i,j}$ from Eq. (4)
Movement function, l_i from Eq. (5)
Number *n* (n=number of fireflies)
Number *max* (max=max number of generations)
Absorption coefficient γ, ($\gamma = 1$)

Output: List *fire, fire= {filter, kernel, pool size, dense}*
Number prediction (bullying=1, no bullying=0)

```
Begin
1.  Initialize a population of n fireflies Pop,  Pop={fire1,..fire_n}
2.  While ( < max) do
3.      for (i=1 :n) do
4.          for (j=1 :n) do
5.              if (I_j >I_i) then
6.                  CDDL (fire)
7.                  Calculate distance r
8.                  Calculate attractiveness β
9.                  Calculate distance r_i,j
10.                 Calculate movement li
11.                 Calculate new light intensity I
12.             End
13.         End
14.     End
15.     Evaluate fireflies by light intensity I
16.     Update best firefly best
17. End
18. Return best
End
```

Figure 1: Firefly-CDDL algorithm

4 Results and Discussion

4.1 Dataset

Our choice of data used in the experiments was not random. We used the dataset provided in Ref. [8]. It contained 39,000 tweets from Twitter and was collected through a two-step process. A Twitter timeline was randomly crawled, and a guided search for profane language was conducted to collect tweets representing bullying. The dataset was labeled by using the Figure Eight platform [35]. The classes of the dataset are illustrated in Table 1.

Table 1: Dataset

	Training	Testing
Bullying	9,000 tweets	2700 tweets
No-Bullying	21,000 tweets	6300 tweets

4.2 Evaluation Metrics

Because cyberbullying detection is a classification task, the accuracy of classification was the obvious choice of metric. However, cyberbullying detection is a problem of class imbalance. Therefore, accuracy alone is not a sufficient measure. We thus considered three other metrics: recall, precision, and the F1-measure. All four metrics were derived from a confusion matrix of four types of values: false positive (FP), false negative (FN), true positive (TP), and true negative (TN).

Accuracy refers to the percentage of correctly classified instances compared to the total instances. It is determined by the following equation.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

Recall (also called sensitivity) refers to the percentage of correctly classified positive instances compared to the instances in the actual class. It is determined by the following equation.

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

Precision (also called specificity) refers to the percentage of correctly classified positive instances compared to the total predicted positive instances. It is determined by the following equation.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

The F1-measure combines both precision and recall; it is helpful for cases where both are important. It is determined by the following equation.

$$F1\ measure = 2 * \frac{(precision * recall)}{precision + recall} \tag{9}$$

4.3 Experimental Results

The goals of our experiments were to (1) test the proposed firefly-CDDL to ensure its accuracy and robustness for cyberbullying detection, (2) compare firefly-CDDL with CDDL (without FA-based optimization) to verify that firefly-CDDL is the fully automated version of CDDL that yields better

predictions, and (3) compare firefly-CDDL with traditional machine learning algorithms to verify that it outperforms prevalent approaches to detection.

4.3.1 Experiment 1: Firefly-Based Cyberbullying Detection Based on Deep Learning (Firefly-CDDL)

The goal of this experiment was to test firefly-CDDL to ensure that it can efficiently detect cyberbullying by using optimized deep learning (specifically, CNN).

a) The Number of Fireflies

FA-based optimization depends on a population of possible solutions [30]. In firefly-CDDL, the population of fireflies represents different CNN structures. There is no standardized method for finding a suitable number of initial fireflies. We thus tested different numbers of the initial population to determine their effects on the quality of prediction. Table 2 shows the results of varying the number of fireflies while keeping all other parameters constant. It is clear that each population of fireflies was tested three times because FA-based optimization is based on randomization. Although the initial number of fireflies was the same in each step, their movement and attractiveness were different and, thus, the results were different.

Table 2: Performance with different number of fireflies

No. of fireflies	Resulting structure				Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
	Filter	Kernel	Pool-size	Neurons				
3	36	14	2	28	97.50	83.13	78.67	80.04
	24	25	3	27	96.25	83.48	79.07	81.22
	4	8	1	18	95.00	82.35	79.12	80.70
4	20	23	2	14	96.87	80.99	80.01	80.50
	34	11	2	36	97.75	84.83	77.62	81.06
	32	28	1	30	98.75	86.73	76.67	81.39
5	38	21	2	26	97.75	83.71	78.76	81.16
	21	22	1	27	96.25	78.91	82.11	80.48
	30	25	4	24	96.87	81.76	79.66	80.70
6	29	9	3	39	94.37	82.67	78.89	80.74
	22	21	3	37	95.00	83.93	78.11	80.92
	20	17	2	16	96.25	84.91	78.34	81.49
7	38	17	3	22	96.88	87.81	76.23	81.61
	22	22	1	8	98.12	72.67	81.36	76.77
	40	11	2	20	97.50	85.68	77.28	81.26
8	36	18	2	18	97.50	84.00	78.44	81.12
	22	15	2	6	96.25	80.41	80.20	80.30
	32	18	2	37	98.12	86.76	76.31	81.20

Firefly-CDDL in general yielded remarkably good results—an accuracy of detection above 95%, and precision and recall values of around 80%. No clear trend was observed to indicate that increasing the number of fireflies improved performance or vice versa. In fact, all variations led to the same conclusion: Running the algorithm three times always resulted in one outstanding structure (one with an accuracy above 97%). The worst-performing structure was obtained with populations of three and six fireflies, where the accuracy dropped to 95%. This highlights the fact that the worst structure obtained by using firefly-CDDL was still very accurate, and that firefly-CDDL yielded better results than those of most structures generated by using only CDDL.

Fig. 2 shows the average time taken by firefly-CDDL took to find a solution (a possible structure) with various numbers of fireflies. All but one of the experiments were completed in less than 10 min. Moreover, increasing the number of fireflies did not always increase the time taken by the algorithm. When we increased the size of the population over a longer search period, the time increased only slightly in our experiments with the number of fireflies, with two exceptions: in the case of six and three fireflies. This result might have been obtained because randomization, an integral part of optimization, often leads to slight variations with respect to the expected outcome.

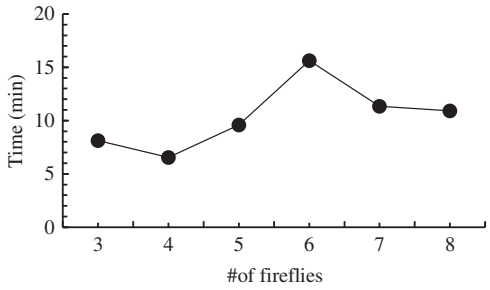

Figure 2: Time taken with different number of fireflies

b) The Number of Generations

The number of generations in firefly optimization represents the maximum number of generations that can be reached. In other words, it is the stopping criteria for the search process. When the maximum number of generations is reached, the search terminates, and the best firefly becomes the solution. In Table 3, the performance resulting from varying the number of generations is shown.

Table 3: Performance with number of generations

Max generations	Resulting structure				Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
	Filter	Kernel	Pool-size	Neurons				
2	36	6	1	33	98.12	99.00	70.00	82.01
	26	6	3	12	95.00	80.87	80.24	80.55
	32	10	1	39	98.75	82.78	80.68	80.72
3	40	17	2	19	97.50	86.36	76.82	81.31
	4	10	1	32	95.60	77.80	80.29	79.03
	35	13	2	14	96.25	88.33	75.52	81.42

(Continued)

Table 3: Continued

Max generations	Resulting structure				Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
	Filter	Kernel	Pool-size	Neurons				
4	12	14	3	26	96.87	85.00	77.56	81.11
	26	21	1	22	97.50	81.46	80.05	80.75
	40	24	1	14	98.12	99.00	70.00	82.01
5	6	12	3	35	95.00	94.52	72.77	82.23
	31	17	2	13	98.12	84.76	78.89	81.72
	8	19	1	11	96.87	83.21	78.69	80.89
6	25	20	2	39	98.75	82.24	78.97	80.57
	15	23	1	23	98.75	82.37	79.11	80.71
	28	10	1	33	97.50	80.06	81.49	80.77
7	38	18	1	20	98.12	87.38	76.33	81.48
	33	15	3	22	98.25	80.26	81.58	80.91
	26	8	4	4	95.62	83.83	78.32	80.98

Different numbers of generations resulted in different performance levels. Moreover, the same maximum number of generations performed differently each time. This is because of the randomized nature of firefly optimization, where each run results in a different search process, and two search processes are never the same. It can be noticed from the table that the number of generations did not affect the quality of the prediction. In most cases, one run out of the three conducted resulted in an outstanding performance (above 98% accuracy). On the other hand, the time significantly increased when the number of generations increased. Therefore, for the case of cyberbullying detection, it is suggested that the number of generations is kept relatively small (less than 4) to ensure that the time required is reasonable while a high quality of prediction is maintained.

Different runs using the same number of generations showed different results, as shown in the previous table. Fig. 3 shows the average performance of the runs. It is clear that the accuracy kept fluctuating and did not follow a pattern. The highest accuracy was recorded with six generations of fireflies and decreased with more generations. Thus, there appears to be no relationship between accuracy and the number of generations.

Fig. 4 provides an answer by illustrating the trend of the time taken by the proposed method as the number of generations increased. The average time taken by the method almost tripled between three and seven generations considered as population. The requisite time reached a maximum of 20 min, an impractically long duration for cyberbullying detection. There is thus a trade-off between the quality of the result and the time taken to reach it.

4.3.2 Experiment 2: Cyberbullying Detection Algorithms (Firefly-CDDLL vs. CDDL)

The goal of this experiment was to compare the algorithm proposed in Ref. [8] with the firefly-CDDL. We sought to determine whether optimization led to better predictions (by our method). We conducted experiments and compared the methods using the same metrics as before: accuracy, precision, recall, and the F1-measure. The results are shown in Fig. 5.

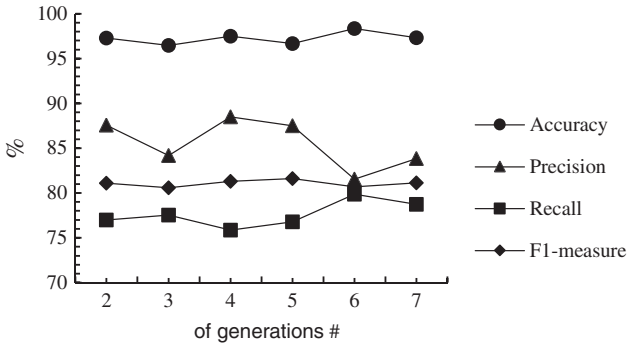

Figure 3: Performance with number of generations

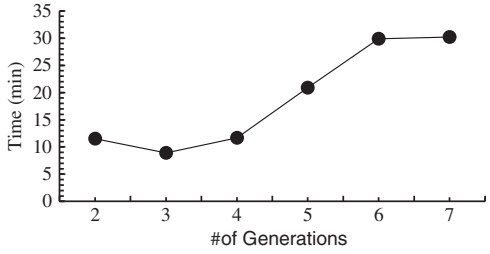

Figure 4: Time taken with different number of generations

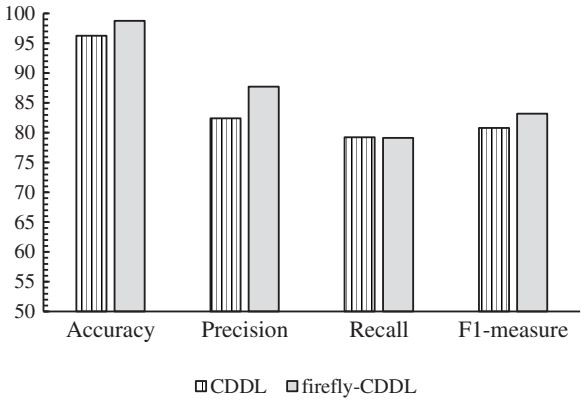

Figure 5: Performance of firefly-CDDL *vs.* CDDL

Firefly-CDDL had a slightly higher accuracy than CDDL. Although the rates of the accuracy of the two methods were close, CDDL reached this rate only through a number of trials such that firefly-CDDL was superior in terms of both accuracy and efficiency. This shows that eliminating feature engineering did not lead to a reduction in accuracy; instead, it rose to around 98%.

CDDL and firefly-CDDL had similar recall rates of 79.23% and 79.11%, respectively. Like recall, precision provides important insights into the quality of the prediction. It is a measure of how good a classification model is at finding true-positive instances in all positive predictions. As seen in Fig. 5 the precision of all algorithms was higher than their recall. Firefly-CDDL outperformed CDDL and had a precision of 87% (compared with 82% for CDDL).

Firefly-CDDL outperformed CDDL by 3% in terms of the F1-measure. This proves that the proposed algorithm is superior to previously reported methods of cyberbullying detection.

4.3.3 Experiment 3: Firefly-CDDL vs. Traditional Machine Learning

We compared the firefly-CDDL algorithm with other algorithms on the same datasets to ensure a fair comparison. The most commonly used approach in this context [15] is the content-based detection of cyberbullying. The authors of one study identified 41 papers that had focused on this technology. The content-based detection of cyberbullying focuses on textual features, such as the length of a given tweet and the presence of profanity in the text. Content-based cyberbullying detection has been implemented in Python [36] by using Spyder as the development environment. The results of applying classical machine learning-based approaches are shown in Fig. 6. It also shows the comparative performance of the firefly-CDDL algorithm to provide insights into how it differs from its classical counterparts.

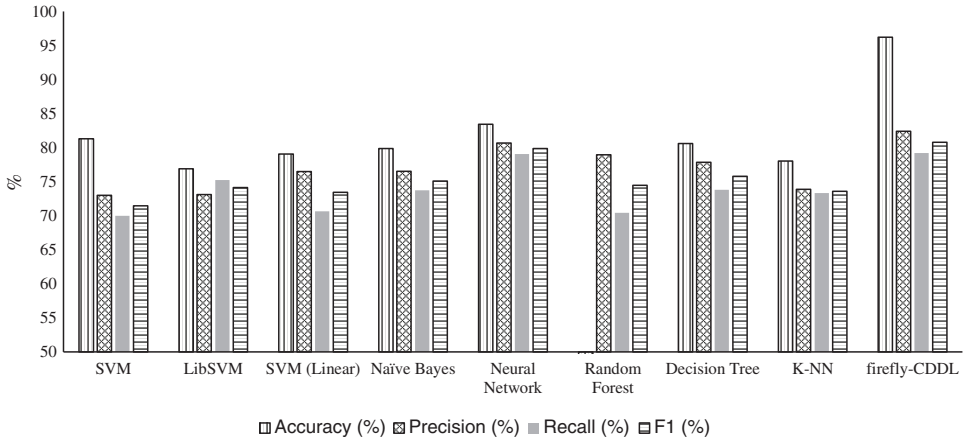

Figure 6: Performance of firefly-CDDL compared with traditional machine learning algorithms

4.3.4 Experiment 4: Semantic Inspection

We also inspected predictions of firefly-CDDL to determine whether it has the potential to understand the hidden meaning within the text. We chose to this end tweets that did not explicitly imply bullying, as shown in Table 4. Despite their hidden meanings, firefly-CDDL successfully classified them as instances of cyberbullying.

Table 4: Semantic inspection of various number of tweets

Tweet	Reason
Lion and lioness that live as mice. @XXX	Indirect
@XXX Yee with sin cast the stones.	Indirect
@XXX Sooner this sword shall plough thy bowels up.	Threat
@XXX now ur just making fake bots	Accusation
@XXX bent, crooked, and cracked lol	Sarcasm
@XXX @XXX Go back to your Special Ed class @XXX	Sarcasm/New

A closer look at the tweets revealed that they were aggressive in terms of content and, indeed, were instances of cyberbullying. However, the individual words constituting them were neither bad nor good in the absence of context. Their combination with other contexts and the intended meaning, such as in the case of sarcasm, made them cyberbullying tweets.

Therefore, classical approaches to detection failed to detect them in this case: They were heavily dependent on examining individual words to check whether they were bad such that the connoted meaning played no role in this task. Firefly-CDDL addressed this issue by examining the meaning of an entire tweet, rather than individual words in it. For instance, the first tweet implicitly means that the person in question had been living a life that was not what they should have been living. Mice were implicitly chosen because they insult and shame. The second tweet was also indirect as it invoked the saying, "let he who is without sin cast the first stone," in a manner opposite to the intended meaning of the famous dictum.

It implicitly shames and accuses the person of sin. The third tweet also exhibits implicit bullying in the form of threatening the other person. In the same vein, the fourth tweet makes a strong accusation, claiming that the relevant user's work is not truly their own. The fifth tweet was intended to humiliate the other person. However, the nature of bullying here was more sarcastic. "Special Ed class" implies that the person in question had limited mental capacity and was not capable of thinking and reasoning. This term has become common since it was used in a famous song and emphasizes the changing nature of language.

The semantic inspection of a set of tweets in which cyberbullying was challenging to detect verified the ability of firefly-CDDL. Because these tweets contained indirect threats, accusations, and sarcasm, algorithms using the classical approach to cyberbullying detection were unable to detect any bullying as they relied on directly understandable meanings. This is why firefly-CDDL could overcome this shortcoming by understanding the hidden meaning within tweets.

5 Conclusion

The technological revolution has advanced people's quality of life, but the rise of online social networks has also given predators a platform for harmful activities. Online bullying is problematic because victims can be constantly targeted and have no ability to escape it. In light of its negative effects on victims, up to and including suicidal thoughts and damaged self-esteem, controlling cyberbullying has become the focus of a considerable amount of research in psychology and computer science.

The aim of this research was to advance the current state of cyberbullying detection by shedding light on critical problems that have not yet been solved. The proposed algorithm, the firefly-CDDL, makes cyberbullying detection a fully automated process, requiring no human expertise or involvement. It addresses a problem with CDDL that was encountered in [8], namely that the many choices of network structures produce varying results. There was no guaranteed method of finding a network structure that performed well. Thus, we incorporated FA to create a suitable and well-performing automatic network structure. We concluded that cyberbullying can be detected using a CNN, and that the addition of FA to optimise the network structure shortened the detection process. In conducting fair comparisons, firefly-CDDL and CDDL were compared, and the optimised algorithm outperformed CDDL with remarkable success, reaching 98% accuracy. Additionally, the proposed algorithm was compared with several machine learning algorithms, which also reported superior results.

Funding Statement: This research project was supported by a grant from the "Research Center of the Female Scientific and Medical Colleges," Deanship of Scientific Research, King Saud University.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] M. Vyawahare and M. Chatterjee, "Taxonomy of cyberbullying detection and prediction techniques in online social networks," *Data Communication and Networks*, vol. 1049, pp. 21–37, 2020.
[2] R. Kowalski, G. Giumetti, A. Schroeder and H. Reese, "Chapter 14 cyber bullying among college students: Evidence from multiple domains of college life," *Misbehavior Online in Higher Education*, vol. 5, pp. 293–321, 2012.
[3] R. Slonje and P. Smith, "Cyberbullying: Another main type of bullying?," *Scandinavian Journal of Psychology*, vol. 49, no. 2, pp. 147–154, 2008.
[4] R. Donegan, "Bullying and cyberbullying: History, statistics, law, prevention and analysis," *The Elon Journal of Undergraduate Research in Communications*, vol. 3, no. 1, pp. 33–42, 2012.
[5] K. Royen, K. Poels, W. Daelemans and H. Vandebosch, "Automatic monitoring of cyberbullying on social networking sites: From technological feasibility to desirability," *Telematics and Informatics*, vol. 32, no. 1, pp. 89–97, 2014.
[6] P. Redmond, J. Lock and V. Smart, "Developing a cyberbullying conceptual framework for educators," *Technology in Society*, vol. 60, no. 1, pp. 101223, 2020.
[7] R. Dennehy, S. Meaney, K. Walsh, C. Sinnott, M. Cronin *et al.,* "Young people's conceptualizations of the nature of cyberbullying: A systematic review and synthesis of qualitative research," *Aggression and Violent Behavior*, vol. 51, no. 1, pp. 101379, 2020.
[8] M. Al-Ajlan and M. Ykhlef, "Deep learning algorithm for cyberbullying detection," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 9, pp. 199–205, 2018.
[9] H. Sanchez and S. Kumar, "Twitter bullying detection," *Networked Systems Design & Implementation*, vol. 12, no. 1, pp. 15, 2011.
[10] M. Dadvar, D. Trieschnigg, R. Ordelman and F. Jong, "Improving cyberbullying detection with user context," in *Proc. of 35th European Conf. on IR Research, ECIR 2013, Advances in Information Retrieval*, Moscow, Russia, pp. 693–696, 2013.
[11] H. Divyashree and N. Deepashree, "An effective approach for cyberbullying detection and avoidance," *International Journal of Innovative Research in Computers and Communication Engineering*, vol. 14, no. 1, pp. 8005–8010, 2016.
[12] M. Al-garadi, K. Varathan and S. D. Ravana, "Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network," *Computers in Human Behavior*, vol. 63, no. 1, pp. 433–443, 2016.
[13] S. Murnion, W. Buchanan, A. Smales and G. Russell, "Machine learning and semantic analysis of in-game chat for cyberbullying," *Computers and Security*, vol. 76, no. 1, pp. 197–213, 2018.
[14] S. Alim, "Analysis of tweets related to cyberbullying: Exploring information diffusion and advice available for cyberbullying victims," *International Journal of Cyber Behavior, Psychology and Learning*, vol. 5, no. 4, pp. 31–52, 2015.
[15] R. Sugandhi, A. Pande, S. Chawla, A. Agrawal and H. Bhagat, "Methods for detection of cyberbullying: A survey," in *2015 15th Int. Conf. on Intelligent Systems Design and Applications (ISDA)*, Marrakesh, Morocco, pp. 173–177, 2015.
[16] V. Nahar, X. Li and C. Pang, "An effective approach for cyberbullying detection," *Communications in Information Science and Management Engineering*, vol. 3, no. 5, pp. 238, 2013.
[17] E. Saraç and S. A. Özel, "Effects of feature extraction and classification methods on cyberbully detection," *Süleyman Demirel University Journal of Natural and Applied Sciences*, vol. 21, no. 1, pp. 190–200, 2017.

[18] B. C. Kovoor, V. Nandakumar and M. U. Sreeja, "Cyberbullying revelation in Twitter data using naïve bayes classifier algorithm," *International Journal of Advanced Research in Computer Science*, vol. 9, no. 1, pp. 510–513, 2018.
[19] A. Kontostathis, K. Reynolds, A. Garron and L. Edwards, "Detecting cyberbullying: Query terms and techniques," in *Proc. of the 5th Annual Acm Web Science Conf.*, Paris, France, pp. 195–204, 2013.
[20] H. Hosseinmardi, S. Li, Z. Yang, Q. Lv, R. Rafiq *et al.,* "A comparison of common users across instagram and ask. fm to better understand cyberbullying," in *Big Data and Cloud Computing (BdCloud), 2014 IEEE Fourth Int. Conf.*, Sydney, Australia, pp. 355–362, 2014.
[21] A. Mangaonkar, A. Hayrapetian and R. Raje, "Collaborative detection of cyberbullying behavior in Twitter data," in *Electro/Information Technology (EIT) IEEE Int. Conf.*, Mankato, MN, USA, pp. 611–616, 2015.
[22] P. Galán-García, J. G. Puerta, C. Gómez, I. Santos and P. Bringas, "Supervised machine learning for the detection of troll profiles in Twitter social network: Application to a real case of cyberbullying," *Logic Journal of IGPL*, vol. 24, no. 1, pp. 42–53, 2016.
[23] W. Bangyal, R. Qasim, N. Rehman, Z. Ahmad, H. Dar *et al.,* "Detection of fake news text classification on COVID-19 using deep learning approaches," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1–14, 2021.
[24] W. Bangyal, N. Rehman, A. Nawaz, K. Nisar, A. Ibrahim *et al.,* "Constructing domain ontology for Alzheimer disease using deep learning based approach," *Electronics*, vol. 11, no. 12, pp. 1890, 2022.
[25] L. Rukhsar, W. Bangyal, M. Ali Khan, A. Ag Ibrahim, K. Nisar *et al.,* "Analyzing RNA-seq gene expression data using deep learning approaches for cancer classification," *Applied Sciences*, vol. 12, no. 4, pp. 1850, 2022. https://doi.org/10.3390/app12041850.
[26] S. Sennan, R. Somula, A. Luhach, G. Deverajan, W. Alnumay *et al.,* "Energy efficient optimal parent selection based routing protocol for internet of things using firefly optimization algorithm," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 8, pp. 41–71, 2020.
[27] A. Ewees, L. Abualigah, D. Yousri, Z. Algamal, M. Al-qaness *et al.,* "Improved slime mould algorithm based on firefly algorithm for feature selection: A case study on QSAR model," *Engineering with Computers*, vol. 38, no. 3, pp. 2407–2421, 2021.
[28] J. Lambert and E. Perumal, "Oppositional firefly optimization based optimal feature selection in chronic kidney disease classification using deep neural network," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 4, pp. 1799–1810, 2021.
[29] W. Bangyal, J. Ahmad and H. Rauf, "Optimization of neural network using improved bat algorithm for data classification," *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 4, pp. 670–681, 2019.
[30] W. Bangyal, J. Ahmed and H. Rauf, "A modified bat algorithm with torus walk for solving global optimisation problems," *International Journal of Bio-Inspired Computation*, vol. 15, no. 1, pp. 1, 2020. https://doi.org/10.1504/ijbic.2020.105861.
[31] W. Bangyal, K. Nisar, A. Ag Ibrahim, M. Haque, J. Rodrigues *et al.,* "Comparative analysis of low discrepancy sequence-based initialization approaches using population-based algorithms for solving the global optimization problems," *Applied Sciences*, vol. 11, no. 16, pp. 7591, 2021.
[32] X. Yang, "Firefly algorithm, stochastic test functions and design optimisation," *International Journal of Bio-Inspired Computation*, vol. 2, no. 2, pp. 78–84, 2010.
[33] M. Goldanloo and F. Charehchopogh, "A hyprid OBL-based firefly algorithm with symbiotic organisms search algorithm for solving optimization problems," *The Journal of Supercomputing*, vol. 78, no. 3, pp. 3998–4031, 2022.
[34] Y. Silva, D. Hall and C. Rich, "BullyBlocker: Toward an interdisciplinary approach to identify cyberbullying," *Social Network Analysis and Mining*, vol. 8, no. 1, pp. 18, 2018.
[35] Figure Eight Inc, "Figure eight," 2018. [Online]. Available: https://www.figure-eight.com/. (accessed on 28 August 2018).
[36] "The Scientific Python Development Environment," 2018. [Online]. Available: https://www.spyder-ide.org/. (accessed on 28 August 2021).