**INTRODUCTION**

Social media networks such as Face book, Twitter, Flickr, and Instagram have become the preferred online platforms for interaction and socialization among people of all ages. While these platforms enable people to communicate and interact in previously unthinkable ways, they have also led to malevolent activities such as cyber-bullying. Cyber bullying is a type of psychological abuse with a significant impact on society. Cyber-bullying events have been increasing mostly among young people spending most of their time navigating between different social media platforms. Particularly, social media networks such as Twitter and Face book are prone to CB because of their popularity and the anonymity that the Internet provides to abusers. In India, for example, 14 percent of all harassment occurs on Face book and Twitter, with 37 percent of these incidents involving youngsters [1]. Moreover, cyber bullying might lead to serious mental issues and adverse mental health effects. Most suicides are due to the anxiety, depression, stress, and social and emotional difficulties from cyber-bullying events [2]\_[4]. This motivates the need for an approach to identify cyber bullying in social media messages (e.g., posts, tweets, and comments).

In this article, we mainly focus on the problem of cyber bullying detection on the Twitter platform. As cyber bullying is becoming a prevalent problem in Twitter, the detection of cyber bullying events from tweets and provisioning preventive measures are the primary tasks in battling cyber bullying threats [5]. Therefore, there is a greater need to increase the research on social networks-based CB in order to get greater insights and aid in the development of effective tools and approaches to effectively combat cyber bullying problem [6]. Manually monitoring and controlling cyber bullying on Twitter platform is virtually impossible [7]. Furthermore, mining social media messages for cyber bullying detection is quite difficult. For example, Twitter messages are often brief, full of slang, and may include emojis, and gifs, which makes it impossible to deduce individuals' intentions and meanings purely from social media messages. Moreover, bullying can be difficult to detect if the bully uses strategies like sarcasm or passive-aggressiveness to conceal it. Despite the challenges that social media messages bring, cyber bullying detection on social media is an open and active research topic. Cyber bullying detection within the Twitter platform has largely been pursued through tweet classification and to a certain extent with topic modeling approaches. Text classification based on supervised machine learning (ML) models are commonly used for classifying tweets into bullying and non-bullying tweets [8]\_[17]. Deep learning (DL) based classifiers have also been used for classifying tweets into bullying and non-bullying tweets [7], [18]\_[22]. Supervised classifiers have low performance in case the class labels are unchangeable and are not relevant to the new events [23]. Also, it may be suitable only for a pre-determined collection of events, but it cannot successfully handle tweets that change on the \_y. Topic modeling approaches have long been utilized as the medium to extract the vital topics from a set of data to form the patterns or classes in the complete dataset. Although the concept is similar, the general unsupervised topic models cannot be efficient for short texts, and hence specialized unsupervised short text topic models were employed [24]. These models effectively identify the trending topics from tweets and extract them for further processing. These models help in leveraging the bidirectional processing to extract meaningful topics. However, these unsupervised models require extensive training to obtain sufficient prior knowledge, which is not adequate in all cases [25]. Considering these limitations, an efficient tweet classification approach must be developed to bridge the gap between the classifier and the topic model so that the adaptability is significantly proficient.

In this article, we propose a hybrid deep learning-based approach, called DEA-RNN, which automatically detects bullying from tweets. The DEA-RNN approach combines Elman type Recurrent Neural Networks (RNN) with an improved Dolphin Echolocation Algorithm (DEA) for \_netuning the Elman RNN's parameters. DEA-RNN can handle the dynamic nature of short texts and can cope with the topic models for the effective extraction of trending topics. DEA-RNN outperformed the considered existing approaches in detecting cyber bullying on the Twitter platform in all scenarios and with various evaluation metrics.

The contributions of this article can be summarized as the following

\_ Develop an improved optimization model of DEA for use to automatically tune the RNN parameters to enhance the performance;

\_ Propose DEA-RNN by combining the Elman type RNN and the improved DEA for optimal classification of tweets;

\_ A new Twitter dataset is collected based on cyber bullying keywords for evaluating the performance of DEA-RNN and the existing methods; and

\_ The efficiency of DEA-RNN in recognizing and classifying cyber bullying tweets is assessed using Twitter datasets. The thorough experimental results reveal that DEA-RNN outperforms other competing models in terms of recall, precision, accuracy, F1 score, and specificity.

The rest of this article is structured as the following: Recent related works are reviewed and analyzed in Section II. The proposed DEA-RNN model is described in Section III. Section IV discusses the experimental analysis, performance metrics, and results analysis. The discussion is introduced

in Section V. Finally, Section VI offers the conclusion and possible future directions.