TÜRKİYE CUMHURİYETİ
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

GRADUATE SCHOOL
COMPUTER ENGINEERING DEPARTMENT

A HYBRID APPROACH FOR CYBERBULLYING DETECTION

DİLARA NİHADİOĞLU

M.Sc.

ADANA 2024

TÜRKİYE CUMHURİYETİ
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

GRADUATE SCHOOL
COMPUTER ENGINEERING DEPARTMENT

A HYBRID APPROACH FOR CYBERBULLYING DETECTION

DİLARA NİHADİOĞLU

M.Sc.

THESIS ADVISOR
Assoc. Prof. Dr. ESRA SARAÇ EŞSİZ

ADANA, 2024

DECLARATION OF CONFORMITY

In this thesis study, which was prepared following the thesis writing rules of Adana Alparslan Türkeş Science and Technology University Institute of Graduate School, I declare that I provide all the information, documents, evaluations and results in accordance with scientific ethics and moral codes without resorting to any means or assistance that would be contrary to scientific ethics and traditions. I also declare that I refer to all of the articles I used in this study with appropriate references and accept all moral and legal consequences if a situation is found contrary to my statement regarding my work.

..../..../20..

[Signature]

Dilara NİHADİOĞLU

ÖZET

SİBER ZORBALIK TESPİTİ İÇİN HİBRİT BİR YAKLAŞIM

Dilara NİHADİOĞLU

Yüksek Lisans, Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Doç. Dr. Esra Saraç EŞSİZ

Ağustos 2024, 57 sayfa

Siber zorbalık 2000'li yılların başından beri dünyada yaygınlaşmaya başlayan ve yıllar geçtikçe modern toplumun büyük sorunlarından biri haline gelen bir baskı türü olmuştur. Sosyal medyanın yaygın kullanımı siber zorbalık verilerinin gün geçtikçe artmasına sebep olmuştur. Bulgular arttıkça veri miktarı artmış ve siber zorbalık tespiti için otomatize edilmiş yöntemler aranmaya başlamıştır. Literatürde makine öğrenmesi, doğal dil işleme, derin öğrenme ve öznitelik seçimi gibi birçok yöntem ile siber zorbalığın tespit edilebilmesi için çalışmalar yapılmıştır.

Bu çalışma son yıllarda bilgisayar bilimlerinde birçok problemin çözümünde büyük başarılar elde edilmesini sağlayan Genetik Algoritma(GA) ve doğadan esinlenilen sürü optimizasyon algoritmalarından biri olan Balina Optimizasyonu Algoritması'nın(BOA) hibrit bir şekilde öznitelik seçiminde kullanılmasının siber zorbalık verisinin sınıflandırılmasında etkili bir yöntem olabileceğini ortaya koyar. Çalışmada sunulan Genetik Balina Optimizasyonu tek başına GA ve BOA'dan doğruluk (accuracy) ve F1-Score bakımından daha iyi sonuç vermiştir.

Anahtar Kelimeler: Balina optimizazyonu, genetik algoritma, hibrit optimizasyon, siber zorbalık

ABSTRACT

A HYBRID APPROACH FOR CYBERBULLYING DETECTION

Dilara NİHADİOĞLU
M.Sc., Department of Computer Engineering

Supervisor: Assoc. Prof. Dr. Esra Saraç EŞSİZ
August 2024, 57 pages

Cyberbullying has been a type of oppression that has become widespread in the world since the early 2000s and has become one of the major problems of modern society over the years. The widespread use of social media has led to an increase in cyberbullying data. As the findings increased, the amount of data increased and automated methods for cyberbullying detection began to be sought. In the literature, studies have been carried out to detect cyberbullying with many methods such as machine learning, natural language processing, deep learning and feature selection.

This study reveals that the use of the Genetic Algorithm (GA) which has achieved great success in solving many problems in computer science in recent years, and the Whale Optimization Algorithm (WOA) which is one of the swarm optimization algorithms inspired by nature, in a hybrid way in feature selection can be an effective method in the classification of cyberbullying data. The Genetic Whale Optimization (GWOA) presented in the study gave better results in accuracy and F1-Score than GA and WOA alone.

Keywords: Cyberbullying, genetic algorithm, hybrid optimization, whale optimization

Dedication

To my precious family and friends

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

GA	: Genetic Algortihm
WOA	: Whale Optimization Algorithm
NB	: Naive Bayes
KNN	: K-Nearest Neighbor
RF	: Random Forest
DT	: Desicion Tree
RNN	: Recurrent Neural Network
ANN	: Artificial Neural Network
CNN	: Conventional Neural Network
LSTM	: Long Short-Term Memory
BiLSTM	: Bi-directional Long Short-Term Memory
SVM	: Sport Vector Machine
GRU	: Gated Recurrent Unit
HAN	: Hierarchical Neural Network
BERT	: Bidirectional Encoder Representations from Transformers
PSO	: Particle Swarm Optimization
ABC	: Artificial Bee Colony

1. INTRODUCTION

Bullying is typically understood as deliberate and repetitive actions aimed at causing harm to another person, making it challenging for the victim to protect themselves (Olweus, 1999). It is characterized by an unequal distribution of power and can be described as a methodical misuse of authority (Rigby, 2002; Smith & Sharp, 1994). Cyberbullying is defined as the persistent and prolonged use of electronic methods by an individual or group to aggressively target a victim who is unable to easily protect themselves (Smith et al., 2008).

Researchers have been particularly focused on the issue of cyberbullying since the start of the 21st century. Traditional bullying often involves a victim and a perpetrator, and its identification usually relies on observation or the disclosure of one of the individuals involved. Nevertheless, cyberbullying is prevalent in situations when both the victim and the perpetrator are observed by one or more individuals. This form of harassment has increased in frequency as social media has gotten integrated into our daily routines.

The initial instances of cyberbullying emerged in the early 2000s through the use of SMS messages and anonymous phone calls. Due to the rise of social media, cyberbullying has been categorized into many forms based on the facility of the platform on which it occurs. This phenomenon initially emerges in chat rooms on social networking sites and chat tabs on gaming platforms. It may appear in posts, comments, and direct messages on popular social media platforms like Instagram. Additionally, it can be observed in quoted comments, replies, and direct messages on platforms such as Twitter. Furthermore, it can be found in inquiries on websites like Formspring.me or Curious Cat, where users have the ability to communicate anonymously.

When studying cyberbullying, researchers have found that the age group most commonly affected is adolescents aged 12-15 who have not yet reached adulthood. This group also tends to use social media frequently (Qiqi, 2024). According to research, cyberbullying is prevalent among individuals in this age exceed. A study conducted in Romania, involving a sample group of 835 individuals aged 10 to 19, demonstrated that women are typically the targets of cyberbullying. Research suggests that women have a higher prevalence of engaging in cyberbullying than men in their social lives (Lee et al., 2020). Women may experience greater

exposure and confront more actions in their jobs, particularly when compared to men (Loh et al., 2020).
Furthermore, the study found that aggressive behavior displayed by parents tends to be transmitted to their children, who are more likely to engage in cyberbullying (Iorga et al., 2022).

Certain researchers categorize cyberbullying into an overall categorization (Law et al., 2011). On the other hand, a study had been conducted that categorized cyberbullying into two types based on the setting in which it occurred (Ortega et al., 2012). The categorizing of cyberbullying into online bullying and mobile bullying has grown problematic in systematic evaluation due to the rise of smartphones and the widespread availability of Internet connection (Slonje et al., 2012).

In addition to categorizing cyberbullying based on its setting, the concept of categorizing it based on content is also introduced. Willard et al. (2006) conducted a study to analyze cyberbullying behaviors and content across 7 distinct categories. The categories encompassed in this list are flame, online harassment, cyberstalking, denigration, masquerade, outing, and exclusion. To summarize, the study has examined the many forms of cyberbullying based on the primary platforms utilized (mobile phones, Internet), more particular methods of utilizing information and communication technology (text messages, instant messaging, email, web pages), and the different types of harmful behavior involved (threats, flaming, outing, exclusion) (Slonje et al., 2012).

While study groups typically prioritize the teenage age range, cyberbullying is a prevalent issue across all age groups. This occurrence, which has the potential to impact individuals across all age demographics at some stage in their life, is anticipated to result in serious consequences ranging from various psychological problems to suicide. Hence, the detection and prevention of cyberbullying hold great significance.

For years, techniques such as natural language processing, machine learning, and deep learning have provided different approaches to detect cyberbullying. There is a multitude of studies in the literature focused on identifying cyberbullying, and several of them are highly significant in terms of both detection and prevention.

The objective of this study was to enhance the identification of cyberbullying using a dataset of text. Previously, possessing an enormous dataset was regarded as advantageous. Nevertheless, it is now important to have a reduced dataset that exhibits an equitable distribution. Feature selection is a vital technique for constructing a more streamlined and concise dataset, and it plays a significant role in effectively classifying data. Applying diverse strategies to reduce the number of features and eliminating unneeded columns from the dataset guarantees a more robust classification.

There are multiple statistical methodologies available to enhance the features in a text data set. However, metaheuristic algorithms are more effective at reducing the amount of features in larger datasets (Beştaş, 2023). GA and the WOA used in this study are effective algorithms that can solve complex optimization problems separately. WOA is used to extend the sample space, while GA is used in a hybrid way to be able to select the best features in the best extended space for an individual. It has been determined that this hybrid method increases the performance of the KNN classifier in terms of F1-Score and accuracy for cyberbullying detection.

The test scenarios are executed on a 64-bit computer equipped with an 11th Gen Intel(R) Core(TM) i7-11800H CPU running at a clock speed of 2.30 GHz. The computer has 16 GB of RAM and runs on the Windows 11 Home operating system. The data was preprocessed and the algorithm was implemented using Python in the Anaconda environment, namely Jupyter-Lab.

The first section of the study incorporates an extensive literature analysis, while the following one outlines the demographic and statistical aspects of the utilized data. The same section discusses the overall structure and mathematical principles of both the GA and the WOA. The research provides a brief description of the KNN classifier and explains the success metrics used. The novel technique introduced in Part 3 is thoroughly explained, accompanied by a comprehensive flow chart. Part 4 of the study presents the experimental results and displays them using graphs and tables. The significance of the study and the assessment of the findings were highlighted in Part 5. The final section of the report discusses possibilities for further research and development.

2. LITERATURE REVIEW

The word "cyberbullying" first defined in academic literature in 2006, as documented by Patchin and Hinduja (Patchin and Hinduja, 2006). From that day onwards, the researchers began to analyze the incidents in a statistical manner. Q Li published one of the earliest papers in the literature in 2006 (Q Li, 2006). This study defines cyberbullying as a collection of bullying actions carried out through technological means. The research suggests that out of 264 junior high school children, half of the group reported experiencing cyberbullying or knowing someone who has been cyberbullied. The study demonstrates that the perpetrator of bullying is predominantly male, and their targets are primarily female.

In 2007, Bauman classified cyberbullying into various categories, including flame, harassment, denigration, masquerading, outing and trickery (which often occur together), cyberthreats, and cyberstalking. Additionally, during the same research, the environment of cyberbullying was classified into instant messages and chat groups (Bauman, 2007).

In 2009, Vandebosch and colleagues did a comprehensive investigation with 2052 primary and secondary school children. They show that cyberbullying among young people has become a serious problem. In addition, they deliberated on the concepts of detection and prevention (Vanderbosch et al., 2009).

Given the rapid growth of cyberbullying, researchers started actively seeking methods to automatically detect and prevent such incidents. In 2010, Konstosthatis et al launched a repository for text mining datasets, specifically aimed at facilitating future study. They recommended the utilization of machine learning techniques for the detection of cyberbullying and cyber crimes (Kontostathis et al., 2010).

In the same year, Ptaszynski and his colleagues did a study applying machine learning techniques to identify instances of cyberbullying. They get data from unofficial websites and online communities related to schools. They gather data and categorize it based on whether the instance involves negative emotion or not. The researchers used the very resilient machine learning technology known as Support Vector Machine (SVM) and achieved a commendable F-score of 88.2% (Ptaszynski et al., 2010).

Dinakar et al. retrieved data from comments on YouTube and converted the textual remarks into TF-IDF vectors. They desired a response regarding the superior performance between binary classification and multi-class classification. The classification techniques applied were Naive Bayes, SVM, rule-based JRip, and tree-based J48. They demonstrated that the binary classifiers trained for individual labels outperform multiclass classifiers trained for all labels (Dinakar et al., 2011).

Reynolds and his fellow researchers conducted another investigation utilizing machine learning techniques. Data was gathered from Formspring.me, a platform that facilitates question-and-answer exchanges. Due to the anonymous nature of the platform, there is a significant amount of bullying content there. The data was labeled using Amazon Mechanical Turk, a web service, and the Weka tool's classifiers was applied to identify true positive values. The most successful classification was achieved using the J48 algorithm, a decision tree-based method, with an accuracy of 78.5% (Reynolds et al., 2011).

In 2012, Dadvar et al proposed that gender-specific discrimination in cyberbullying provides more effective solutions. Unlike other studies, this research focuses on the author's characteristics rather than the subject matter of the comments. They proved that gender-based classification provides superior solutions for detection compared to contextual classification alone (Dadvar et al., 2012).

In 2013, Dinakar and friends introduced a method for identifying bullying utilizing advanced natural language processing and a knowledge base of common sense. This method allows for the detection of bullying across a wide range of everyday themes using data from Fromspring reported by users. They examine a more restricted scope of specific topics related to bullying, such as physical appearance, intelligence, racial and ethnic insults, social acceptance, and rejection. They create BullySpace, a database that stores practical knowledge about different bullying scenarios (Dinakar et al., 2013).

Macbeth et al. introduce sophisticated interfaces for social media that utilize common knowledge bases and automated analysis of textual conversations between users. While some methods aim to simply categorize the general subject of the text, they attempted to associate

stories with more detailed "scripts" that depict typical occurrences and activities. The capabilities were included into a preliminary prototype system and subsequently evaluated using a database of actual cyberbullying narratives gathered from MTV's "A Thin Line' website (Macbeth, 2013).

Nitta et al. adopted this approach in the novel methodology they introduced. The researchers determined the relevance of the words in the document in order to classify them as harmful. During the initial phase of the process, they carefully chose specific terms and phrases. In the second phase, the words classified as harmful were further divided into three categories: offensive, abusive, and violent. During the last stage, the relevance score of the gathered words was computed using the 255-word seed word dictionary that they had established, and the objective was to maximize the relevance score. The results indicate that the Precision value achieves a maximum of 90% (Nitta et al., 2013).

Potha and her colleague proposed a time series model. They utilized both the question and the answer as the substance. The issue was addressed utilizing a sequential data modelling approach, where the predator's inquiries were framed using a Singular Value Decomposition representation. The question of bullying is modeled by quantifying the frequency of bullying incidents and representing it as a time series. The time series was analyzed using Singular Value Decomposition (SVD), and the resulting model closely resembled the class properties. The success metrics employed were MAPE and RMSFE. SVMs and Multilayer Perceptron (MLP) are employed for the purpose of classification. The Linear SVM was shown to be more successful than the MLP (Potha et al., 2014).

Huang et al. proposed a method to enhance the accuracy of cyberbullying detection by employing machine learning techniques. They incorporated the social network attributes into their dataset and subsequently conducted separate analyses for the textual features and social features. They employed the same methodology for both. They observed that the ROC curve exhibits an increase when the textual and social elements are combined. The most successful classifier was Dagging, achieving a ROC rating of 75.5%. The textual qualities that received the highest rankings were the frequency of exclamations, the use of offensive language, and the occurrence of positive bigrams, in that order. Out of the many social qualities, the number

of links, number of nodes, and number of edges were identified as the most highly ranked attributes (Huang et al., 2014).

In 2014, Nahar and fellow researchers devised a semi-supervised model aimed at identifying instances of cyberbullying. They utilized two small datasets: one including instances of bullying and another containing instances when bullying did not occur. The test was conducted without any labels. The data for the test was collected from instant messaging and comments on social media platforms. Three distinct methodologies were employed to demonstrate this study. Initially, they employ Augmented training to use enlarged data in later phases. They utilized the Kernel Fuzzy C-Mean Clustering Algorithm to produce a membership matrix for the classifier. A fuzzy classifier has been used to manage imbalanced and unstable text streams obtained from social media platforms. They evaluated this method with different scenarios in which the rates of training and evaluation were subject to change. The Naive Bayes algorithm achieved the highest accuracy of 97% across all cases (Nahar et al., 2014).

In 2015, Naghini et al introduced a novel study on the identification of cyberbullying. The Levenshtein Algorithm was employed to identify and categorize instances of cyberbullying, while the Naive Bayes algorithm was used as the classifier. Two distinct data sets were utilized, one from Formspring.me and the other from MySpace. The Formspring.me dataset, consisting of 400 posts, had an F-measure score of 94%. Similarly, the MySpace dataset, also with 400 posts, achieved an F-measure rate of 92% (Nandhini et al., 2015).

Hon and his companions developed a system to identify instances of cyberbullying on the internet. The Twitter API was utilized to collect tweets that involve instances of cyberbullying. The tweets are stored in a dataset and an index of bullying terms has been constructed. A technology was developed to identify instances of cyberbullying in tweets and provide immediate responses to the user (Hon et al., 2015).

Ptaszynski et al utilized linguistic combinatorics to establish a distinct pattern. This approach assumes that the punctuation marks and words, as well as word combinations, only occur once in the texts. Once the patterns were created, they were rearranged according to their respective weights. They created complex structured arrangements of words based on those

patterns. The model was validated using 10-fold cross-validation and had an F-score value of 79% (Ptaszynski et al., 2015).

In 2016, Al-garadi et al introduced a distinct set of variables extracted from Twitter, including network, activity, user, and tweet content. Using these data, we constructed a supervised machine learning approach to identify instances of cyberbullying on Twitter. A random selection of approximately 10,000 tweets was made, and the issue of unbalanced data was solved using SMOTE. In addition, they utilized feature selection approaches such as Information Gain, Chi-Square Test, and Pearson Correlation. The feature selection strategy had a minimal impact, but when utilizing SMOTE to handle unbalanced data with Random Forest as the classifier, it yielded state-of-the-art results in terms of F-Measure. The value exhibits an increase from 0.690 to 0.943 (Al-Garadi et al., 2016).

Zhao et al. introduced a novel technique called Semantic-Enhanced Marginalized Denoising Auto-Encoder. In this technique, researchers employ Semantic Dropout Noise to reduce the level of noise present in the data. Sparsity constraints are utilized to maintain crucial elements in textual data. The Denoising Auto-Encoder aims to rebuild the original data from the corrupted data, leading to a more resilient and concise text representation. SmSDA, a deep learning method, achieves higher accuracy rates compared to other deep learning methods. The model achieves an accuracy of 84% on the Twitter dataset and 89% on the MySpace dataset (Zhao et al., 2016).

In the same year, Zhao et al. introduced an approach that utilizes an illustration learner framework. This method applies word embeddings to augment a pre-established inventory of negative phrases, providing distinct magnitudes to these hostile characteristics. The characteristics are subsequently merged with Bag-of-Words (BoW) and latent semantic features to create the ultimate representation, and that is employed for training a linear SVM classifier. The highest F1 Score among all other Bag-of-Words systems is achieved by the Embeddings-enhanced Bag-of-Words Model, with a rate of 78% (Zhao et al., 2016).

Zhang et al. offer a novel strategy to mitigate the impact of noisy and unbalanced data sets in the field of cyberbullying. Due to the prevalence of misspelling in social network data, the researchers have introduced an original model called the Pronunciation-based Convolutional Neural Network (PCNN). This strategy applied threshold movement, revised the cost

function, and implemented a hybrid solution to address the problem of imbalanced data. This methodology corrects spelling problems that do not impact the pronunciation. The PCNN achieves superior precision and recall values compared to baseline Convolutional Neural Networks (CNNs) (Zhang et al., 2016).

In 2017, Singh et al. released a research where they used textual and visual cues to identify instances of cyberbullying. The researchers utilized an Instagram visual dataset and applied the SMOTE technique to deal with the issue of imbalanced data. They added the Classifier Subset Selection method to choose relevant features and applied Bagging as the classifier. Initially, the textual and visual data are analyzed separately, and then, the two elements are combined. The findings show that the presence of images linked to age, race, gender, pornography, etc. significantly impacts the identification of cyberbullying. By merging textual and visual data, they achieve an accuracy rate of 81.4% (Singh et al., 2017).

In contrast to most earlier studies that utilized supervised learning approaches, Raisi et al. introduced a weakly supervised approach to concurrently identify user roles and detect cyberbullying vocabulary. The participant vocabulary consistency (PVC) model assesses whether each social contact constitutes bullying by considering the individuals involved and the language used. Its objective is to optimize the alignment between these assessments, ensuring participant vocabulary consistency. The algorithm uses a process of conjecture to ascertain the potential substance of cyberbullying. Upon discovering an instance of bullying, the algorithm identifies novel vocabulary that is commonly employed by either the bully or the victim. This algorithm iterates until it reaches a point where it identifies the victim and bully accurately, and determines if newly added phrases truly belong to the vocabulary of cyberbullying. This strategy utilizes the Alternating Least Squares technique to carefully choose characteristics (Raisi et al., 2017).

In 2018, Al-Ajlan and his colleagues did a study using a deep learning technique called CNN-CB to detect instances of cyberbullying. CNN-CB is a CNN architecture that eliminates the requirement for feature engineering and yields superior outcomes compared to traditional cyberbullying detection techniques. The word embedding approach was applied to prevent the separate consideration of similar or highly related terms. The CNN-CB model outperformed other machine learning techniques with an accuracy rate of 95% (Al-Ajlan et al., 2018).

Tomkins and her colleagues have produced an original SOCIO-LINGUISTIC model that attempts to predict whether users would act as bullies or victims, as well as detect connections between users. This is accomplished through the use of collective categorization and latent variable modeling. Researchers employed N-GRAMS to establish connections between bullying tweets for the purpose of identifying instances of cyberbullying. They employed N-GRAM++ for sentiment analysis and to improve document embedding. SEED++ was used to generate seed phrases that indicate the presence of cyberbullying. LATENT-LINGUISTIC method was applied to classify text into bullying categories using latent variables in a dynamic way. The SOCIO-LINGUISTIC model improves based on the LATENT-LINGUISTIC model by determining relationships between users and combining socio-linguistic data. This research exceeds other similar studies with an F-Measure rate of 63.2% (Tomkins et al., 2018).

Cheng et al. proposed a novel model utilizing deep learning techniques. They utilized sessions of hierarchical network architecture to evaluate instances of cyberbullying. The Hierarchical Attention Networks for Cyberbullying Detection (HANCD) use a bi-directional Gated Recurrent Unit (GRU) based Recurrent Neural Network (RNN) to encode the sequence of words and comments. Each comment owner was assigned a session, in which at least one bullying phrase from their specific comment was used. They depicted the session as vectors. GRU-based RNNs are utilized to detect and identify bullying terms, rather than relying on manually labeled data. The sentences are analyzed using a GRU-based RNN to capture the true meaning, both when read forwards and backwards. The HANCD model outperforms both classic machine learning models and deep learning models, achieving an F-Measure of 0.783 and an AUC score of 0.851 (Cheng et al., 2019).

In 2019, Al-Hashedi et al conducted an empirical study to examine the effectiveness of various combinations of word embedding techniques and deep learning models. The word embeddings were generated using word2vec, gloVe, ELMo, and Reddit w2v. The data was evaluated using three types of RNNs: LSTM, BLSTM, and GRU. The BLSTM model with ELMo achieves superior performance in terms of recall, ROC curve, and MSE, with values of 0.94, 0.96, and 0.03, respectively. The combination of GRU and ELMo demonstrates superior performance in terms of time, completing in 3965 seconds. This study demonstrates that the

most effective combination of RNN for evaluating text features is the integration of ELMo with Bidirectional Long Short-Term Memory (BLSTM) (Al-Hashedi et al., 2019).

Mahlangu and Tu implemented a novel methodology by integrating BERT embeddings with GloVe embeddings and applying LSTM with stacked word embeddings. Their performance outperformed all of other machine learning methods and past utilization of CNN, achieving an accuracy rate of 94.2% (Mahlangu and Tu, 2019).

In 2020, Wang et al. presented a framework named Multi-Model Cyberbullying Detection that combines deep learning techniques. The researchers utilized data sets obtained from the visual social media platforms Instagram and Vine. This approach consists of two distinct phases: one for encoding all the data at different levels, and another for decoding the encoded data. The post was categorized based on its content, and the comments on the post were analyzed using Hierarchical Attention Networks (HAN). They assumed that each comment is influenced by the prior comment, making the HAN model the most suitable for the comment embedding process. The word level and comment level characteristics were encoded using Bidirectional GRU. The MLP is used for the decoding process. The new MMCD strategy overtook existing techniques, including deep learning and machine learning approaches, in terms of accuracy and F-1 measures. Specifically, it achieved an accuracy of 0.864 and an F-1 measure of 0.86 in the Instagram data set, and an accuracy of 0.838 and an F-1 measure of 0.841 in the Vine data set (Wang et al., 2020).

Yadav et al. utilized the pre-trained BERT model by employing a transformer neural network, which incorporates a self-attention mechanism to comprehend the word embeddings. The researchers utilized a 12-layered BERT-Based model to produce embeddings for the data. An untrained linear neural network layer has been placed on above the BERT for classification. Using a pre-trained BERT-based model, they obtained an F-1 score of 0.94 and an accuracy of 0.98 on the 3x oversampled Formspring dataset. On the Wikipedia dataset, they achieved an F1-score of 0.81 and an accuracy of 0.96. These results outperformed the other methods used on the same datasets (Yadav et al., 2020).

Reinforcement learning is a type of machine learning that focuses on achieving certain goals. Aind et al. chose to use the Q-learning algorithm, a method within this approach. For each

word in the dataset, there are two possible action types: offensive and non-offensive. The model examines each word in the sentence. If the model's prediction is incorrect, the sentences receive a significant negative reward. Conversely, if the prediction is accurate, the sentences receive a small positive reward. The researchers achieved an F1-score of 0.860 by applying the Q-Learning algorithm on a non-oversampled dataset using NLP approaches (Aind et al., 2020).

In 2021, Eşsiz et al. employed a nature-inspired optimization method to perform feature selection on a cyberbullying dataset. The researchers utilized the Artificial Bee Colony Algorithm, which is based on the foraging behavior of honey bees, to identify the most significant feature. This feature selection process resulted in a notable improvement in the macro F-measure score, increasing it from 0.659 to 0.800 (Eşsiz et al., 2021).

Kumari et al. used GA to optimize the features of the data set. They used pre-trained VGG-16 and CNN to extract features from images and texts, respectively. Their utilization of VGG-16-CNN-GA methodologies results in an F1-score of 75% (Kumari et al., 2021).

Sherly et al. utilized the Enhanced Cuckoo Search Optimization algorithm to enhance the success metrics of the detection. The data is pre-processed using the k-means technique. Feature selection is performed using ECSO, and classification is done using the Hybrid Firefly Artificial Neural Network (HFANN) algorithm. ECSO+HFANN enhances the metrics of precision, recall, F-measure, and accuracy, resulting in an accuracy of up to 0.90 (Sherly et al., 2021).

Jing et al. utilized Particle Swarm Optimization (PSO) to enhance the detection outcomes using Long Short-Term Memory (LSTM). Following the normalization and pre-processing of social media data, the researchers utilized the Particle Swarm Optimization technique to as feature selector to enhance the classification success metric. The researchers employed a deep learning technique called LSTM to train and classify the data. This method achieved an impressive accuracy of 99.1%, surpassing the previous best answer using ADABoost by 2.9 points (Jing et al., 2023).

Muneer et al. employed Deep Neural Networks and BERT-M to develop a novel model. The data set is transformed into word2vec format using Bag Of Words for feature selection. It is then inputted into a convolutional and pooling technique to identify the most important features. Finally, BERT-M is used for evaluation measures. The stacked model achieved success metrics with an F1-Score of 0.964, precision of 0.950, and recall of 0.92. Additionally, the model's execution duration was 3 minutes, making it the fastest technique to date (Muneer et al., 2023).

Table 2.1. Cyberbullying literature review with success metrics.

Researchers	Year	Techniques	Accuracy	F1-Score	Precision	Re-call	ROC/RMSE
Kontostathis et al.	2010	Dataset Collection	-	-	-	-	-
Ptaszynski et al.	2010	SVM	-	88.2%	-	-	-
Dinakar et al.	2011	TF-IDF, SVM, NB, J48	80.2%	-	-	-	-
Reynolds et al.	2011	J48	-	78.5%	-	-	-
Dadvar et al.	2012	Gender-specific classifying	-	23%	43%	16%	-
Nahar et al.	2012	Session-based approach	-	35%	32%	39%	-
Dinakar et al.	2013	BullySpace(a web environment)	-	-	-	-	-
Nitta et al.	2013	Word Revelance	-	-	90%	-	-
Potha et al.	2014	SVM,MLP	-	-	-	-	0,041(RMSFE)
Huang et al.	2014	Dagging	-	-	-	-	0.755(ROC)
Nahar et al.	2014	NB	97%	-	-	-	-
Naghini et al.	2014	Levenshtein Algorithm,NB	-	94%	-	-	-
Ptaszynski et al.	2015	linguistic combinatorics	-	79%	-	-	-
Al-garadi et al.	2016	SMOTE, RF	-	94.3%	-	-	-
Zhao et al.	2016	SmSDA	89%	-	-	-	-
Zhao et al.	2016	Embeddings-enhanced Bag-of-Words, SVM	-	78%	-	-	-

Singh et al.	2017	SMOTE, Classifier Subset Selection, Bagging	81.4%	-	-	-	-
Al- Ajlan et al.	2018	CNN-CB	95%	-	-	-	-
Tomkins et al.	2018	Socio-Linguistics	-	63.2%	-	-	
Cheng et al.	2019	RNN	-	78.3%	-	-	
Al-Hashedi et al.	2019	Elmo+BLSTM	-	-	-	94%	-
Mahlangu and Tu	2019	GloVe, BERT, LSTM	94.2%	-	-	-	-
Wang et al.	2020	HAN, Bidirectional GRU, MLP	86.4%	86%	-	-	
Yadav et al.	2020	BERT, TNN	96%	81%	-	-	-
Aind et al.	2020	Q-Learning	-	86%	-	-	-
Eşsiz et al.	2021	ABC	-	80%	-	-	
Kumari et al.	2021	VGG-16,CNN,GA	-	75%	-	-	-
Sherly et al.	2021	Enhanced Cuckoo Search Optimization	90%	-	-	-	-
Jing et al.	2022	PSO,LSTM	99.1%	-	-	-	-
Muneer et al.	2023	Word2vec, BERT-M,	-	96.4%	95%	92%	-

3. MATERIALS AND METHODS

This section provides an explanation of the approaches that were utilized. The initial section of the study focuses on providing a statistical overview of the dataset and detailing the preprocessing stages and tools used. The second part of the study provides an explanation of the KNN algorithm, which serves as the classifier. Subsequently, the following sections briefly discuss the statistical feature selection methods and the metaheuristic feature selection methods employed in the study.

3.1. Data Set

The data set employed is a publicly available Twitter Cyberbullying Dataset that has been released on Kaggle. The dataset contains a total of 8462 tweets. Out of the total number of tweets, 5080 have been classed as instances of bullying, while 3370 have been classified as 'Not-Bullying'. There are 12 tweets that have not been assigned a classification. The unlabeled tweets are identified and classified as 'Not-Bullying' to ensure that the data is not eliminated. The data set is almost evenly distributed with a 66/34 ratio of bullying versus non-bullying attributes. The data is preprocessed by removing stop words using the English stop words from the NLTK package. Punctuation is also removed as the data is already broken into sentences and punctuation is not necessary. The sole criterion for a word to be included in the data set is its primary meaning. Consequently, the only requirement is the root form of the term, which may be obtained through the process of lemmatization using WordNetLemmatizer. The words are stemmed using the Snowball stemmer. After using the CountVectorizer method, the data was transformed into a matrix with dimensions of 8452 rows and 4209 columns. Not-bullying data labeled with 0’s and Bullying data labelled with 1’s.

Table 3.1. Data set statistics.

Number of tweets	8462	100%
Bullying	5080	66.5%
Not-Bullying	3382	33.5%

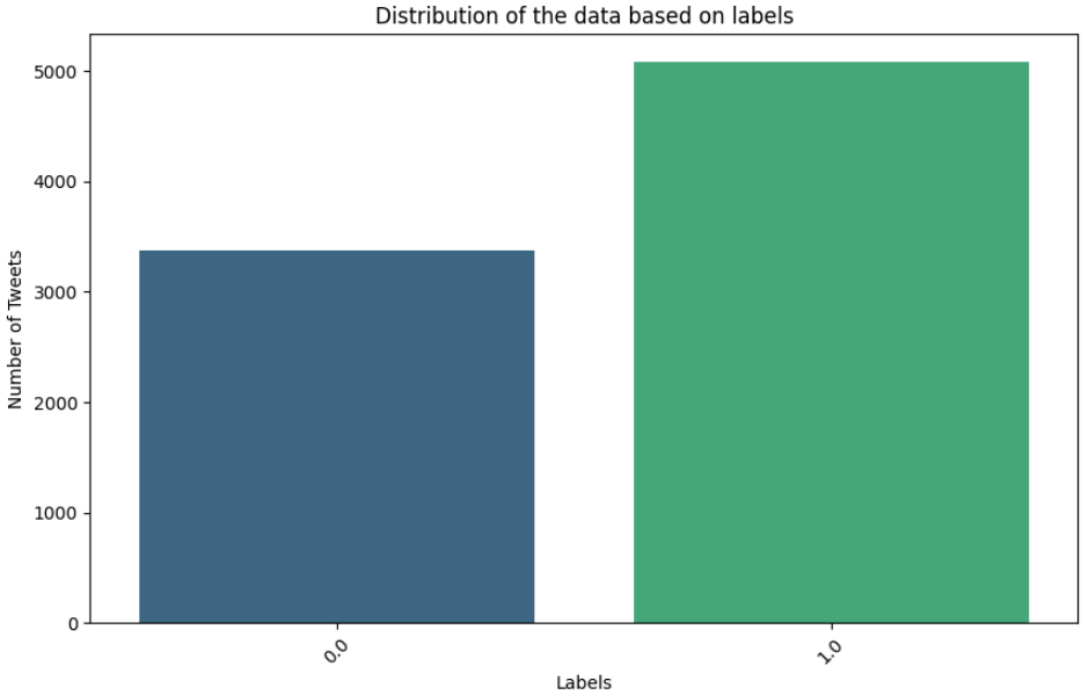

Figure 3.1. Data set distribution.

Table 3.2. Instances in the data set.

Text	Label
Ten outside soon doctor shake everyone treatment seat radio land.	0
my life has come to a standstill and at this point.	0
"Oh hey, you should be ashamed of your disgusting self"	0
I wanna fuck you	1
girl this nigga make me sick to my stomach	1
Haha isn't route running part of the position? People are dumb. But fuck edelman! He just looks like he says	1

3.2. K- Nearest Neighbor

The K-nearest neighbor algorithm is a supervised machine learning classifier that only relies on the training data and does not require any additional parameters (Taunk et al., 2019). KNN use distance measurements, such as Euclidean Distance or Manhattan Distance, to compute the distance between the training and testing data. The algorithm operates by determining the class of the test data based on the k nearest points in the training data.

3.3. Random Forest

Random forest (RF), commonly referred to as random choice forest, is a frequently used technique in studies in academia. The system has several decision trees designed to address regression or classification problems. RF utilizes a combination of random subspace methods (RSM) and bagging. It achieves this by randomly selecting a subset of features at every node of the classification and regression tree (CART) (Qiu et al., 2017). The RSM method involves building machine learning models on randomly chosen subspaces inside the primary input set. The outputs are merged via either a median or majority vote. At every node of the decision tree, a subset of m features is chosen randomly from a pool of n input features. RF is a resilient approach for effectively managing outliers and noise present in a given dataset. RF combines random sampling with ensemble processes to enhance its ability to make accurate generalizations and reliable estimations.

3.4. Logistic Regression

Logistic regression is a prevalent statistical machine learning technique that demonstrates exceptional performance when used to extensive data sets (Snakenborg et al., Barlińska et al.) Generates a hyperplane between two data sets that employing the logistic function (Patchin et al.). The output of the inputs it receives is determined based on the logistic function described in the equation…. If the value of the output exceeds 0.5, it is categorized as positive; otherwise, it is categorized as negative.

$$h_{\theta}\left(x\right)=\frac{1}{1+e^{-\theta^{T}x'}} \tag{3.1}$$

3.5. Evaluation Metrics

3.5.1.1. Accuracy

This metric quantifies the accuracy of the model by calculating the ratio of correct predictions to the total number of predictions generated from the full dataset. The calculation involves dividing the number of true positives (TP) and true negatives (TN) by the total number of data.

$$\frac{TP + TN}{TP + FN + TN + FP} \tag{3.2}$$

3.5.1.2. Precision

Precision is a metric that calculates the ratio of correct positive predictions to all positive predictions generated by the model. The calculation involves dividing the TP by the sum of the true positives and false positives (FP).

$$\frac{TP}{TP + FP} \tag{3.3}$$

3.5.1.3. Recall

Recall, which is often referred to as sensitivity or true positive rate, quantifies the ratio of correct positive predictions to the total number of positive cases. The calculation involves dividing the TP by the sum of TP and false negatives (FN).

$$\frac{TP}{TP + FN} \tag{3.4}$$

3.5.1.4. F1-Score

The F1 Score is a quantitative measure that achieves a balance between precision and recall. The calculation involves taking the harmonic mean of the precision and recall values. The F1 Score is a valuable metric for achieving a trade-off between high precision and strong recall. It effectively penalizes extreme negative values of either component.

$$\frac{2 \times precision \times recall}{precision \times recall} \tag{3.5}$$

3.6. Pearson Correlation

The Pearson correlation is a statistical technique employed to ascertain the magnitude and direction of the linear association between two variables (Nasir et al., 2020). Karl Pearson initially introduced the approach. This method is utilized to ascertain the correlation between two variables, evaluating whether it is positive or negative and the magnitude of its significance. The correlation coefficient is a scalar measure that quantifies the strength and

direction of the linear relationship between two variables. It takes values between 1 and -1, where a value of 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no correlation. A value of 1 signifies an ideal positive linear correlation between the variables. If one variable exhibits an elevation, the other variable will likewise undergo an augmentation. A correlation coefficient of 0 signifies the absence of any association between two variables, whereas a correlation coefficient of -1 implies a perfect negative linear correlation. If one variable increases, the other variable will decrease. The correlation coefficient, denoted as r, is determined using the following formula:

$$r = \frac{\sum (x_i - \bar{x}) - (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{3.6}$$

where x_i and y_i are variables and $\bar{x}$ and $\bar{y}$ are the average values of the variables.

Using Pearson Correlation for feature selection might be beneficial in identifying features that exhibit a stronger correlation with the desired feature. The Pearson Correlation approach facilitates the removal of redundant characteristics, hence enhancing the performance of the classifier by utilizing the remaining data.

3.7. Information Gain

Information Gain (IG) is a well-established method for feature selection in academic research (Azhagusundari et al., 2013). Information Gain quantifies the influence of a feature on predicting the value of the target feature. To make use of IG, it is important to calculate the entropy. Entropy is a measure used to estimate the degree of instability or unpredictability in a particular dataset. Information gain measures the reduction in uncertainty achieved by using an attribute. The equation that represents entropy, denoted as H, is:

$$H(S) = -\sum_{i=1}^{n} \left(p_i \log_2 \left(p_i \right) \right) \tag{3.7}$$

where S is the data set, p_i is the probability of the class i. The formula of the conditional entropy is:

$$H(S \mid A) = -\sum_{v \in values(A)} P(v) H(S_v) \tag{3.8}$$

where A is stand fort he feature itself, $P(v)$ is the probability of the value of feature A, $H(S_v)$ is the entropy of the feature A for v. The Information Gain is defined as the difference between the entropy and the conditional entropy of a feature.

$$IG(S, A) = H(S) - H(S \mid A) \tag{3.9}$$

3.8. Chi-Square Test

The Chi-Square Test is a method used in feature selection to quantify the extent to which the observed frequencies differ from the expected frequencies (Meesad et al., 2011). Hypothesis testing is conducted to ascertain the independence of two categorical variables. The application of a Chi-Square test is limited to data that can be categorized into distinct groups. The chi-square statistic is calculated using the following formula, which quantifies the extent of dissimilarity between the observed and expected frequencies:

$$x^2 = \sum \frac{(O_i - E_i)2}{E_i} \tag{3.10}$$

O_i represents the observed frequency, while E_i represents the predicted frequency. The p-value is calculated using the chi-squared statistic. The p-value measures the degree of statistical significance in the relationship between the attribute and the target variable. Attributes are selected according to a predetermined significance level, often set at 0.05. Features that have a p-value lower than this significance level are considered to be statistically significant.

3.9. Particle Swarm Optimization

The Particle Swarm Optimization algorithm, proposed by Kennedy and Eberhart in 1995, is a type of swarm optimization technique can be applied to solve feature selection problem

(Kennedy et al.,1995, Jing et al., 2023, Sakri et al., 2018). This methodology involves the manipulation and analysis of a collection of particles known as swarms. A specified assemblage of particles endeavors to optimize outcomes within a given search area, so emulating the collective behavior of bird flocks. During the circulation in the search space, the particles have two objectives: first, to find their own best solution (P_{Best}),and second, to find the best solution within their group (G_{Best}). They adjust their velocity and location based on the equations provided below in order to get the optimal solution.

$$\begin{aligned} v_m^{(n+1)} &= \omega v_m^n + k_1 r(P_{Best_m} - x_m^n) + k_2 r + G_{Best_m} - x_m^n) \\ p_m^{(n+1)} &= p_m^n + v_m^{(n+1)} \end{aligned} \tag{3.11}$$

In the given equations, the symbol p_m^n represents the position of the mth particle at the nth iteration, whereas v_m^n represents the velocity of the mth particle at the nth iteration. P_{Best_m} represents the optimal personal position achieved by the mth particle, while G_{Best} represents the optimal global position achieved by the entire swarm. k_1 and k_2 represent parameters that determine the rate of acceleration. r is a stochastic variable that follows a uniform distribution in the range [0,1]. The symbol ω represents the inertia weight, which regulates the equilibrium between the exploration phase and the exploitation phase.

Pseudocode for PSO:

1. Initialize a swarm of particles with random positions and velocities.
2. For each particle:
a. Evaluate its fitness based on its current position.
b. Update the personal best position (P_{Best}) if the current position is better than the previous best.
c. Update the global best position (G_{Best}) if the current position is the best among all particles.
3. Update the velocity and position of each particle using the equations
4. Repeat steps 2-3 until the stopping criterion is met.
5. Return the best solution found, which is the global best position (G_{Best}).

3.10. Genetic Algorithm

A GA is a stochastic search algorithm that operates by mimicking the process of natural selection and the principles of genetic inheritance. John Holland introduced the GA in the early 1970s as a model of biological evolution, drawing inspiration from Charles Darwin's theory of natural selection (Kumar et al., 2021). Holland is credited with establishing the use of crossover, recombination, mutation, and selection operators in the analysis of adaptive and artificial systems. The genetic operators stated above are essential components of the GA, which is used as a problem-solving approach. Since then, GAs have been applied in several domains, including graph coloring, pattern recognition, discrete and continuous systems, as well as financial markets and multi-purpose engineering optimization.

GAs are widely used evolutionary algorithms due to their diverse range of applications. GAs have proven effective in solving a variety of known optimization problems. Furthermore, it is important to note that GAs operate on a population level, and numerous contemporary evolutionary algorithms are directly derived from GAs (Yang et al., 2020). GAs are heuristic search methods that can be utilized for a diverse array of optimization challenges. To innate

drive of organisms to adapt to their environment leads to the development of complex anatomical features that enable survival in diverse habitats. Reproduction and the development of offspring are essential concepts for achieving evolutionary success. The phenomenon of organisms engaging in pairing to create more robust structures has prompted researchers to consider applying evolutionary concepts to address optimization challenges.

The search approach used in GA closely resembles natural selection in nature, with the objective of finding an optimal solution based on the principle of survival of the fittest. A population consists of potential solutions, and those members of the population that are more compatible (i.e., closer to the solution) have a higher likelihood of survival and passing on their genes to the next generation. As subsequent generations progress, the population must strive to find a resolution. In the field of GA, the search space parameters are specifically defined as sequences, often known as chromosomes. These sequences constitute a population. The directory's progression for each series is determined by the utilization of a fitness function. Applying the principle of natural selection, multiple sequences are chosen and compared based on their suitability. These processes, which are influenced by evolution, include the creation of new generations through the combination of genetic material and the introduction of genetic variations. This procedure is executed until the termination criteria are met (Topalli, 2017). GAs are used to solve optimization problems by creating solutions that continuously improve, following the concept of survival of the fittest observed in nature. Additionally, a fitness function is established to quantify the degree of "goodness". The population (Çetin, 2002) is enhanced through the utilization of operators such as mutation and selection. The evolutionary process commences with solutions that are either randomly or manually initiated. The evolutionary cycle initiates by reestablishing the connection between two or more solutions using the intersection operator. The results obtained from the intersection operator undergo mutation and persistently alternate with the selection operator. The most optimal solutions generated using this method are selected for the subsequent generation. Eventually, the evolutionary cycle evaluates if the termination condition has been met and, if the desired conditions have not been achieved, continues with the analysis of genetic optimization (Kramer et al., 2017).

Every GA consists of five primary phases. The procedure is depicted in the diagram 3.2 The mathematical expressions are as follows:

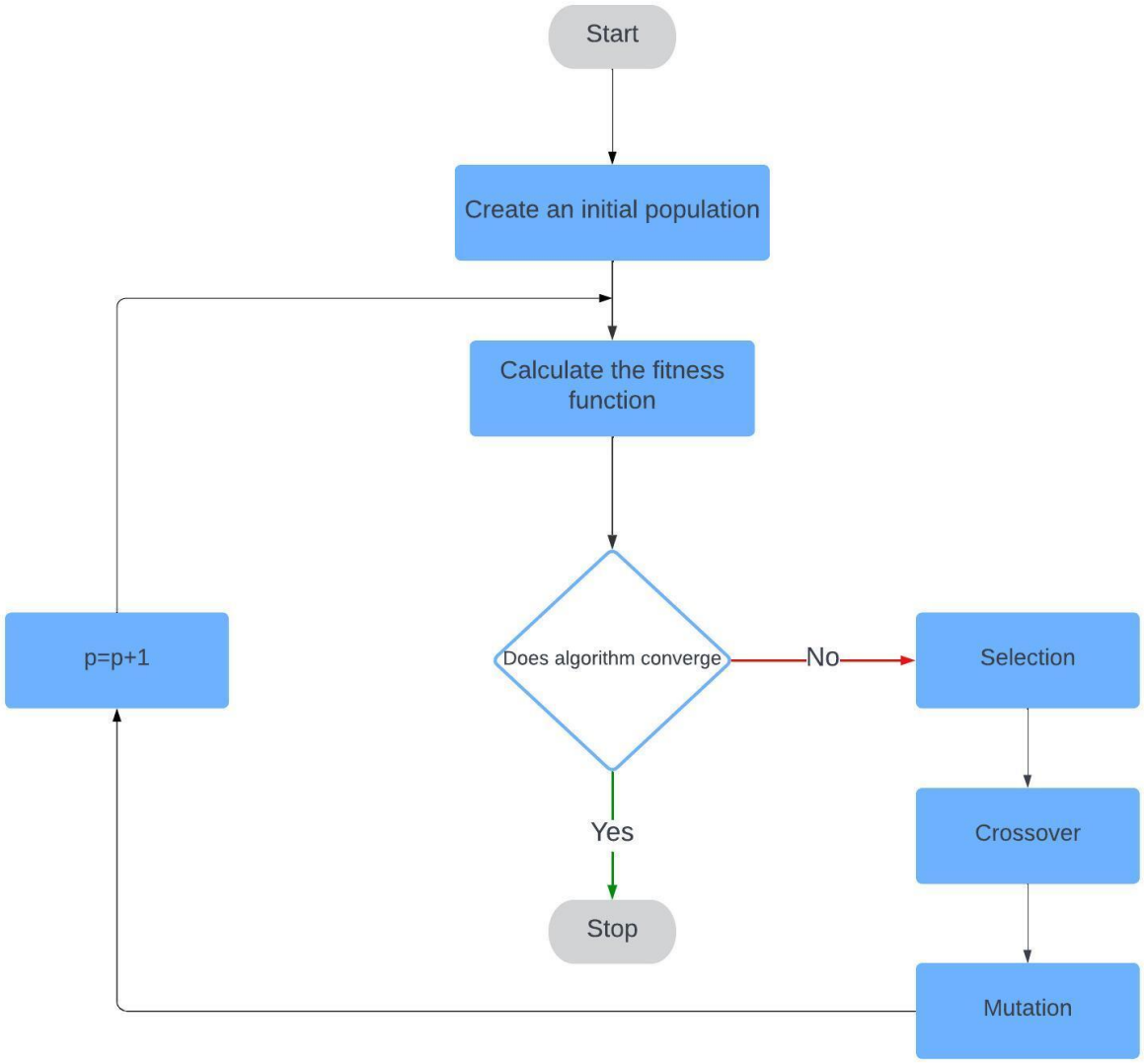

Figure 3.2. GA diagram.

1. Defining an initial population

$$P(t) = \{x_1(t), x_2(t), \ldots x_n(t)\} \tag{3.12}$$

every individual $x_n(t)$ generally represented by binary sequences or decimal number while t is number of iterations.

2. Choosing a fitness function for evaluating each individual's performance

$$f\left(x_i\left(t\right)\right) \tag{3.13}$$

3. Selection involves identifying individuals with a higher potential of success. Common methods for developing a selection process include Route Wheel Selection and Tournament Selection. The likelihood of selecting an individual via Route Wheel Selection is

$$\frac{f\left(x_i\right)}{\sum_{j=1}^{n} f\left(x_i\right)} \tag{3.14}$$

4. Crossover is the reproduction mechanism by which a new offspring is generated by the combination of genetic material from two parent individuals. There are multiple methods to create a crossover. One-way crossover is a typical method of genetic manipulation where the genetic sequences of parents are cut at a random location and then joined together.

$$\begin{aligned} child_1 &= P_1\left[1:k\right],\ P_2\left[k+1:n\right], \\ child_2 &= P_2\left[1:k\right],\ P_1\left[k+1:n\right] \end{aligned} \tag{3.15}$$

Mutation refers to the process of introducing minor alterations to the genetic material of parental organisms. Binary mutation involves the shifting of a single bit from each parent, while real mutation entails the substitution of variables with small random values.

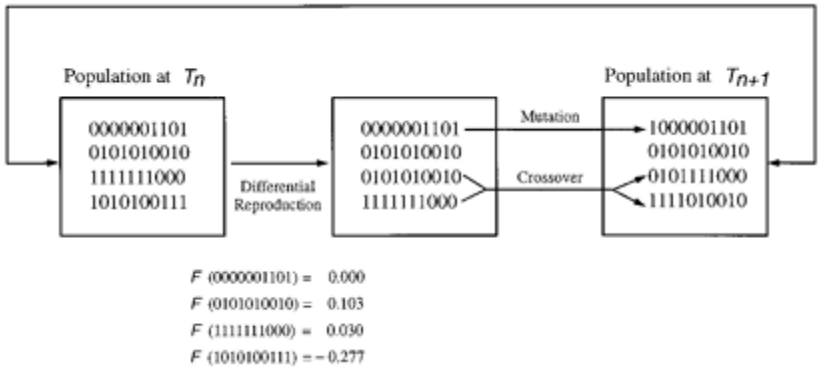

Figure 3.3. Crossover and mutation diagram.

3.11. Whale Optimization Algorithm

The Whale Algorithm is a swarm-based optimization technique that was presented by Mirjalili and Lewis in 2016. It is inspired by nature. WOA simulates the social behavior of humpback whales. Whales are fascinating creatures widely recognized as predators. They do not experience sleep in the same way as humans do. In fact, just half of their brain enters a sleep-like state, as they need to remain aware to breathe while being above the surface. They are regarded as sentient creatures and they sense emotions.

Hof and Van Der Gucht have demonstrated that certain regions of whales' brains contain spindle cells-like areas, which are similar to those seen in humans. Spindle cells are specialized cells involved in the processing of cognitive functions such as judgment, social attitudes, and emotions. Whales can be described as intelligent creatures that have intellectual abilities and experience emotions. Certain whales, particularly killer whales, exhibit a form of communication known as dialectic interaction.

Another fascinating aspect of humpback whales is their hunting behavior. Humpback whales demonstrate a preference for hunting among groups of krill or small fish at the water's surface. When they try to haunt, they create specific shapes using distinctive bubbles, such as circles or the number '9'. This particular behavior is referred to as the bubble-net feeding method. The Goldenberg et al. researchers use sensors to track the underwater behavior of whales. They observe 300 instances of bubble-net feeding, a behavior exhibited by 9 unique humpback whales, using tags. Two movements related to bubble were discovered and labeled as 'upward-spirals' and 'doubleloops'.

The WOA mathematically represents the distinctive behavior of humpback whales' bubble-net feeding mechanism and develops an optimization algorithm.

1. Encircling prey

The humpback whales can recognize the location of prey and draw a circle around them. In the first place, the optimal design of the search space is not known, therefore the current candidate assumed as the best candidate, when the best search agent is found the other agents update their positions due to best agent. The mathematical representation of this behavior is

$$D = |\overrightarrow{C}.\vec{X}*(t) - X(t)|$$
$$\vec{X}(t+1) = \vec{X}(t) - \vec{A}\vec{D} \tag{3.16}$$

Let t represent the current iteration. $\vec{A}$ and $\vec{C}$ are coefficient vectors. $\vec{X}*$ is the position vector of the best solution obtained thus far, and $\vec{X}$ is the position vector. $||$ denotes the absolute value, and $\cdot$ represents element-by-element multiplication. It is important to note that $\vec{X}*$ should be modified in every iteration if a superior solution is found. The vectors $\vec{A}$ and $\vec{C}$ are computed using the following equations:

$$\vec{A} = 2\vec{a}.\vec{r} - \vec{a}$$
$$\vec{C} = 2\vec{r} \tag{3.17}$$

The value of $\vec{a}$ is progressively reduced from 2 to 0 throughout each iteration, encompassing both the exploration and exploitation phases. Additionally, $\vec{r}$ represents a random vector within the range of 0 to 1.

The search agent's coordinates (X, Y)can be adjusted according to the coordinates of the current optimal record (X^*Y^*). By adjusting the $\vec{A}$ and $\vec{C}$ vectors, one can achieve different positions relative to the current location of the top agent. It is crucial to note that by defining the random vector $(\vec{r})$ it is possible to reach any location within the search space that is situated between the key-points. A search agent can dynamically modify its position in the region of the current ideal solution, imitating the process of surrounding the target. The same idea can be used to a search space with n dimensions, where the search agents will move within hyper-cubes that enclose the most optimal solution discovered up to that point. Humpback whales utilize the bubble-net technique to grab their prey, as mentioned above. The mathematical representation of this process is as stated:

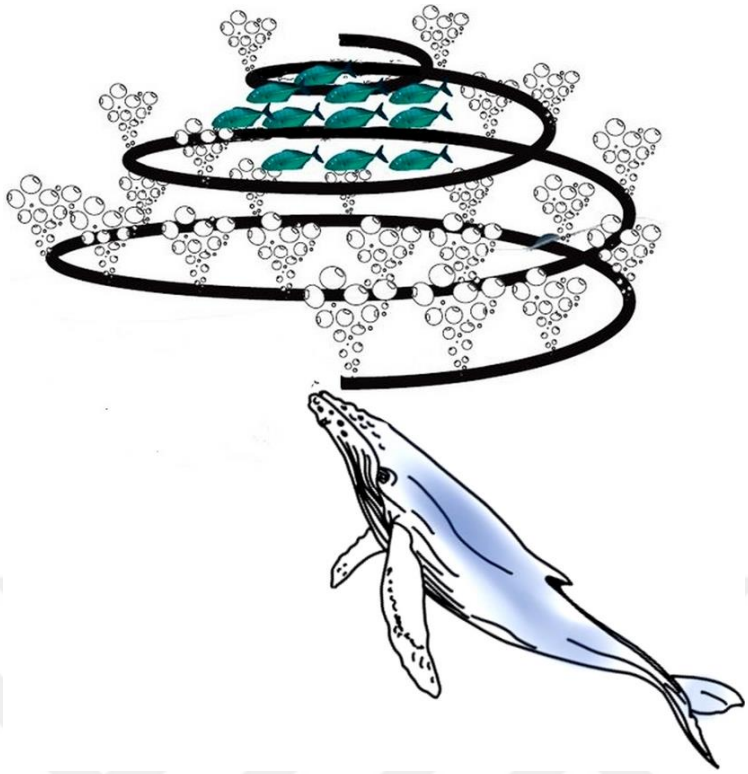

Figure 3.4. Bubble-net attacking (Wang et. al, 2022).

2. Bubble-net attacking method:

To mathematically represent the bubble-net attacking behavior of whales two approaches are created as follows

1. Shrinking encircling mechanism: This behavior is accomplished by reducing the value of $\vec{a}$ in Equation (2ra-a). It should be noted that an additionally decreases the range of fluctuation for $\vec{A}$. Put simply, $\vec{A}$ is a randomly selected value within the range of [-a,a], where the value of a gradually decreases from 2 to 0 with each iteration. By assigning arbitrary values to A within the range of [-1,1], the search agent's new position can be determined anywhere between its initial position and the location of the most optimal agent.
2. Spiral updating position: This method initially computes the distance between the whale positioned at coordinates (X, Y) and the prey positioned at coordinates (X^*Y^*). Next, a spiral equation is generated to imitate the helix-shaped motion of humpback whales. This equation is based on the positions of both the whale and its prey s follows:

$$\vec{X}\left(t+1\right)=\overrightarrow{D'}.e^{bl}.\cos\left(2\pi l\right)+\overrightarrow{X^*}\left(t\right) \tag{3.18}$$

The expression $|\overrightarrow{D'} = \overrightarrow{X^*}(t) - \vec{X}(t)|$ represents the distance between the ith whale and the prey, which is the best solution found thus far. The constant b determines the shape of the logarithmic spiral, l is a random value within the range of -1 to 1, and the symbol of dot denotes element-wise multiplication.

Humpback whales engage in a synchronous movement pattern where they swim around their prey in a diminishing circle while following a spiral-shaped path. In order to represent this concurrent behavior, we make the assumption that there is an equal likelihood of 50% to select either the shrinking encircling mechanism or the spiral model for updating the position of whales. The mathematical model is as follows:

$$\vec{X}(t+1) = \begin{cases} \vec{X}(t) - \vec{A}\vec{D}, & p < 0.5 \\ \overrightarrow{D'}.e^{bl}.\cos(2\pi l) + \overrightarrow{X^*}(t), & p \geq 0.5 \end{cases} \tag{3.19}$$

where p is a random numeric variable.

The exploration of prey can be conducted using the same strategy that relies on the modification of the $\vec{A}$ vector. Humpback whales exhibit unpredictable foraging behavior based on the relative positions of other individuals. Thus, we utilize the variable $\vec{A}$ with random values that are either larger than 1 or less than -1 in order to compel the search agent to move a significant distance away from a reference whale. In the exploration phase, we adjust the position of a search agent based on a randomly selected search agent, rather than the best search agent discovered thus far. The utilization of this method, together with the condition that $|\vec{A}|$ is greater than 1, serves to highlight the focus on exploration and enables the WOA algorithm to conduct a comprehensive global search. The mathematical model can be expressed as:

$$\begin{aligned} D &= |\overrightarrow{C}.\overrightarrow{X_{rand}} - | \\ \vec{X}(t+1) &= \overrightarrow{X_{rand}} - \vec{A}\vec{D} \end{aligned} \tag{3.20}$$

where $\overrightarrow{X_{rand}}$ is a random position vector or can said as a random whale is a candidate from the current population.

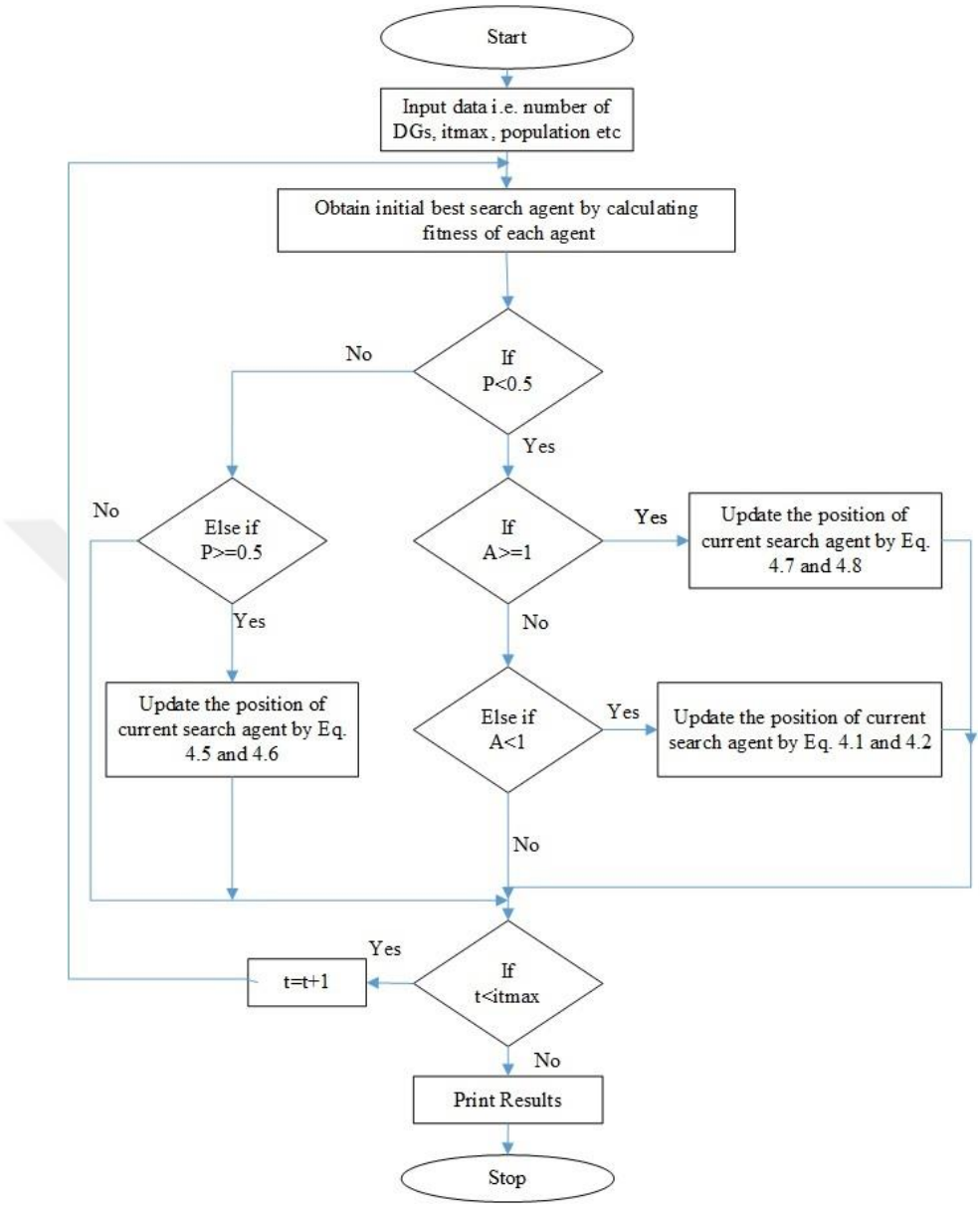

Figure 3.5. The WOA diagram.

The WOA begin by initializing a collection of solutions chosen at random. During each iteration, search agents adjust their positions based on either a randomly selected search agent or the best solution found up to that point. The value of the parameter 'a' is reduced from 2 to 0 to facilitate both exploration and exploitation, respectively. When the cardinality of $\vec{A}$ is greater than 1, a random search agent is selected to update the position of the search agents. Conversely, when the cardinality of $\vec{A}$ is less than 1, the best solution is chosen for updating the position of the search agents. WOA has the ability to alternate between a spiral or circular

movement based on the value of p. Ultimately, the WOA algorithm concludes when a termination requirement is met.

3.12. Purposed Method Genetic Whale Optimization Algorithm for Feature Selection

This section will provide an explanation of the proposed method. The WOA has demonstrated significant efficacy in addressing practical challenges, including data categorization, engineering difficulties, energy-efficient clustering, and feature selection (Sun et al., 2018; Nadimi-Shahraki et al., 2021; Liu et al., 2022; Hegazy et al., 2018; Jiang et al., 2018; Anter et al., 2022). Several studies indicate that the WOA has demonstrated significant achievements in terms of its global and local exploration capabilities, algorithm convergence, and accuracy (Binbin, 2024). In this proposed method, the WOA is applied to widen the search space and provide the GA with a fresh set of potential answers.

GAs are highly proficient in achieving a balance between exploration and exploitation in search spaces, which makes them adaptable instruments for efficiently tackling optimization problems. Scientists are always investigating new methods and combinations to improve the effectiveness of GAs while still preserving their ability to explore different possibilities (Hamblin, 2012).

This novel approach integrates the exploratory capacity of WOA with the balancing and exploitation capacity of GAs to identify the optimal characteristics for cyberbullying detection and classify the data using KNN. Utilizing the provided dataset, a bag-of-words model was constructed. During the initial stage, the algorithm generates an initial population consisting of randomly selected individuals. The WOA algorithm iterates over the population for a specified amount of time in order to improve the population and provide potential solutions within a well-researched search space. The GA is employed to modify and enhance the potential solutions of the WOA through the application of mutation and crossover techniques.

Pseudocode for G-WOA

Input: A set of all features

Output: A set of best n features

1. Initialize parameters:
    - population_size
    - generations
    - woa_iterations
    - ga_iterations
    - mutation_rate
2. Initialize population with random solutions
3. **repeat**
4. **for each** solution in population **do**
5. **WOA Phase**: Perform whale optimization algorithm exploration
6. **end for**
7. **for each** solution in population **do**
8. **GA Phase**: Perform genetic algorithm exploitation
9. **end for**
10. Update the best solution found so far
11. **until** the termination condition is met
12. **return** the best feature set

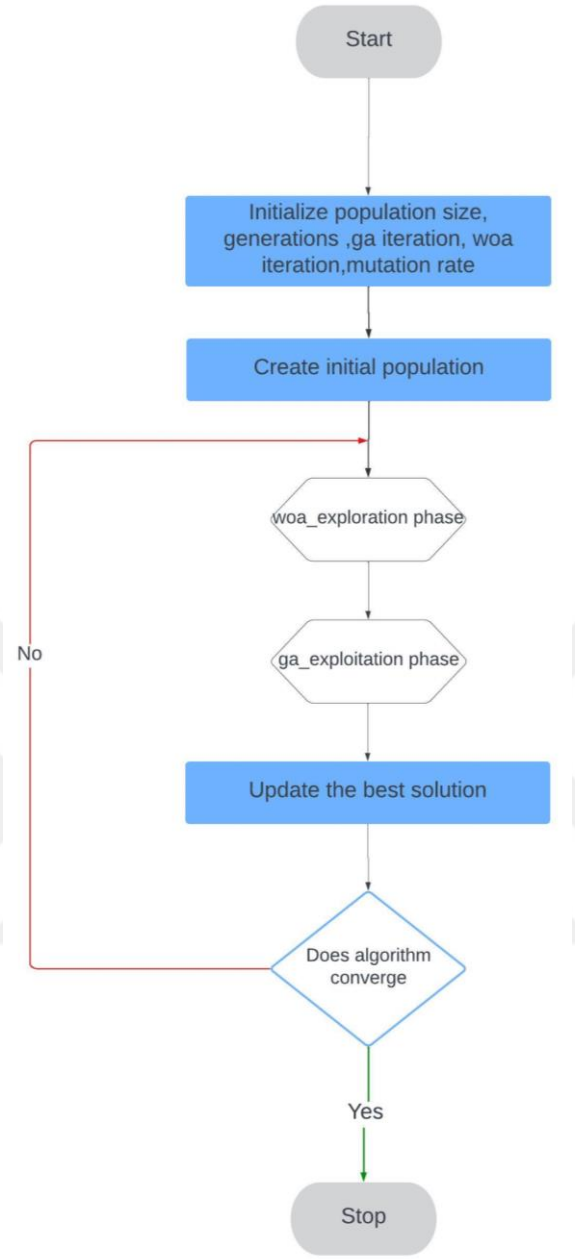

Figure 3.6. The proposed method's diagram.

4. RESULTS AND DISCUSSION

Various classifiers were employed for selecting the trial classifier. In order to select the optimal strategy, we require a classifier that already demonstrates satisfactory performance and has the potential to further improve with the assistance of GWOA. Three distinct classifiers were evaluated, and all of them were included into the fitness function. Upon examining Table 4.1, it is evident that the RF algorithm has superior performance in the absence of any feature selection approach. Logistic regression improved its performance from an accuracy of 55.24% to 83.64%. However, KNN achieves higher accuracy than LR without the need for feature selection and further boosts accuracy when combined with GWOA. Therefore, KNN has been selected to be included in the fitness function of GWOA.

Table 4.1. Classifier performances with and without GWOA.

RF	RF_GWOA	KNN	KNN_GWOA	LG	LG_WOA
%91.09	%90.58	%87.07	**%88.37**	%55.24	**%83.64**

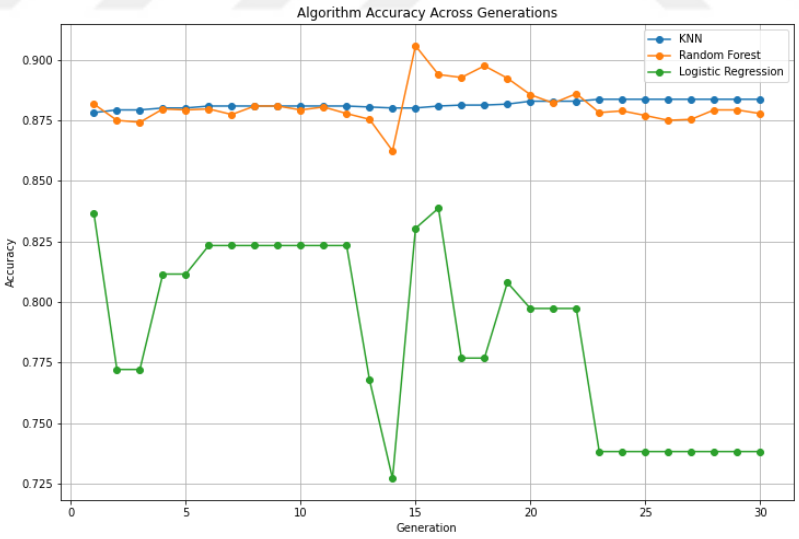

Figure 4.1. Performance of GWOA with different Classifiers.

In all the trials of GWOA and other feature selection methods we use KNN as a classifier. The data set splitted into 0.3 train and 0.7 test parts with number of neighbors k=3. Initially the algorithm was conducted with a mutation rate of 0.1, a population size of 20, 30 iterations of the WOA algorithm, 20 iterations of the GA algorithm, and a total of 50 generations. By reducing the feature size from 4209 to 1695, this approach improves the accuracy and F1-Score of the KNN classifier from 870.7 to 882.1 and from 868.9 to 880.7, respectively and converged on 33. generation.

In the second trial, the algorithm was executed with a mutation rate of 0.1, a population size of 50, 30 iterations of the WOA algorithm, 20 iterations of the GA algorithm, and a total of 50 generations. The purpose was to investigate the effect of population size on achieving the best fitness by increasing it from 20 to 50. The results indicate that a larger population size resulted in improved feature selection, as seen by higher accuracy rates and smaller feature sizes. There was one drawback to this trial. The algorithm takes approximately 11 days to complete 50 generations.

The third and fourth experiment scenarios were conducted with a total of 30 generations. The fact that the algorithm's responses in the first and second attempts indicate that 30 generations are sufficient to yield improved outcomes. In the third trial, the mutation rate remained at 0.1 and the number of iterations was set to 30. In the current trial, the number of iterations for the WOA was increased from 30 to 50 to determine if any enhancements could be observed by raising the WOA iteration count. The number of iterations for the GA remained at 20. Upon examining the findings, it was observed that the accuracy fell, however, the F-Score was the greatest among the first three attempts. The highest level of accuracy in fitness was observed in the 25th generation.

In the fourth and final trial, the mutation rate remained at 0.1. The population size was 20, and the number of iterations for the WOA algorithm reduced to 30, just like in the other two scenarios. The total number of iterations was also set to 30. Additionally, the number of iterations for the GA algorithm increased from 20 to 50. In this scenario, the accuracy rate was determined to be better than in the first and third trials, but the F1-Score was at its lowest. Upon examining Table 4.1, it is evident that Run 2 was the most successful out of all four

runs, as it had the smallest feature size and the highest accuracy rate. It can be argued that the increase in population led to improved outcomes.

Table 4.2. Results of runs table.

GEN	POP	MUT	WOA	GA	SIZE	ACC	F1-SC	BESTFIT
50	20	0.1	30	20	1731	88.09	88.07	50
50	50	0.1	30	20	988	88.6	87.86	30
30	20	0.1	50	20	1776	88.13	88.97	25
30	20	0.1	30	50	2109	88.21	87.76	20

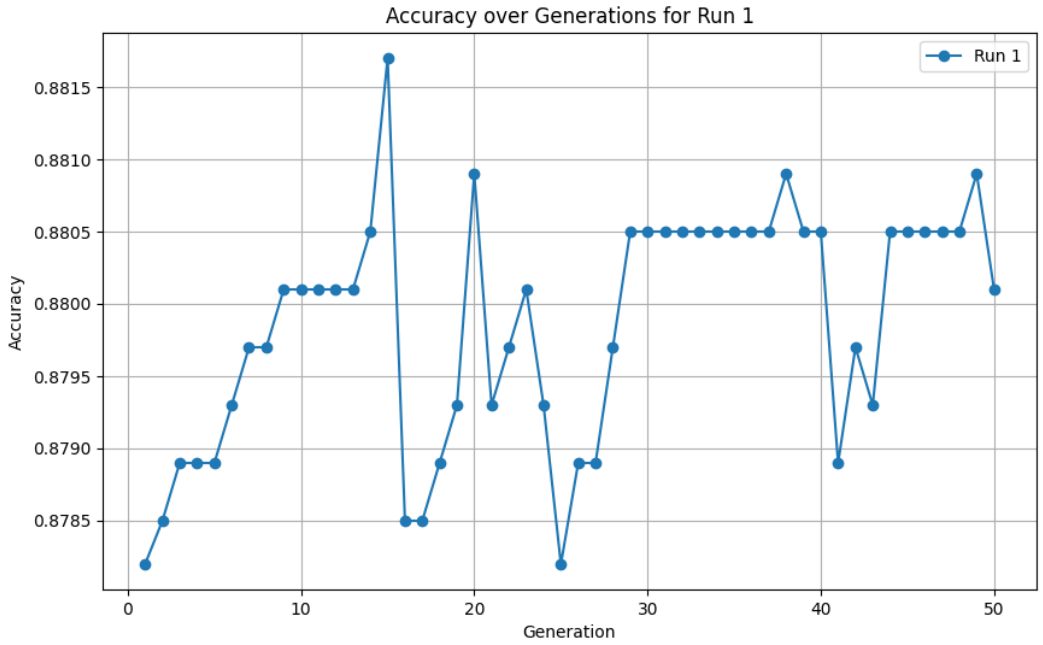

Figure 4.2. Accuracy per generations run 1.

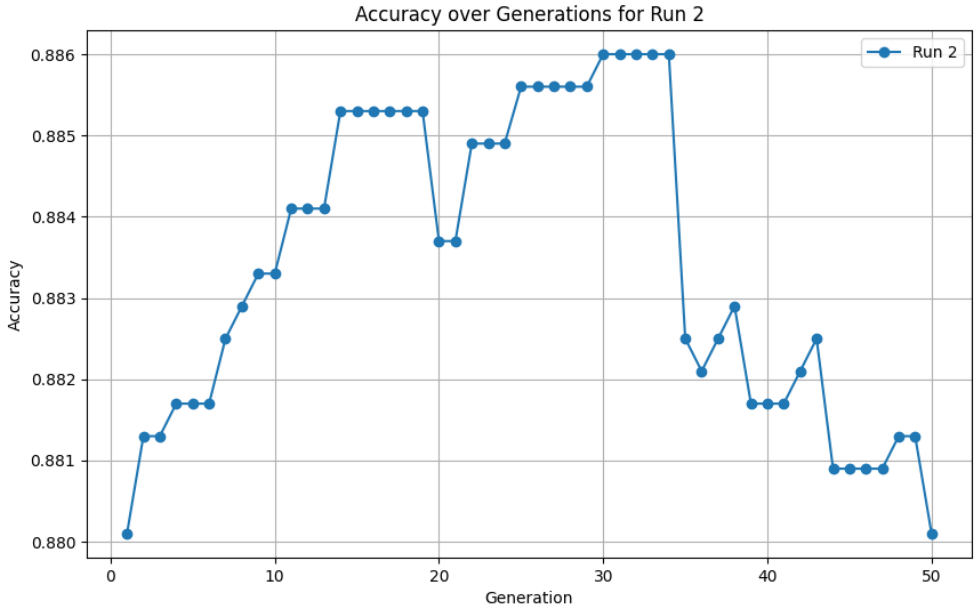

Figure 4.3. Accuracy per generations run 2.

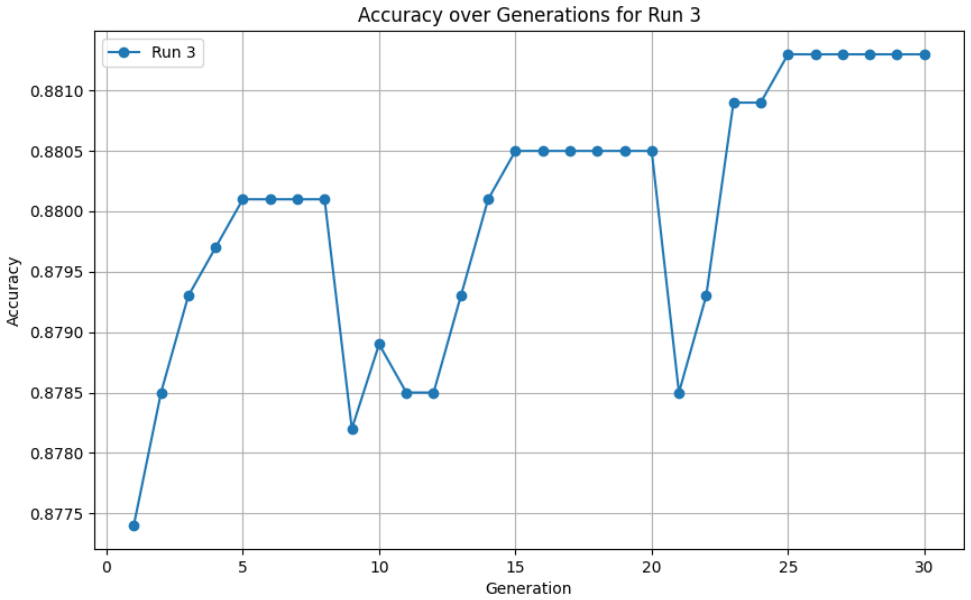

Figure 4.4. Accuracy per generations run 3.

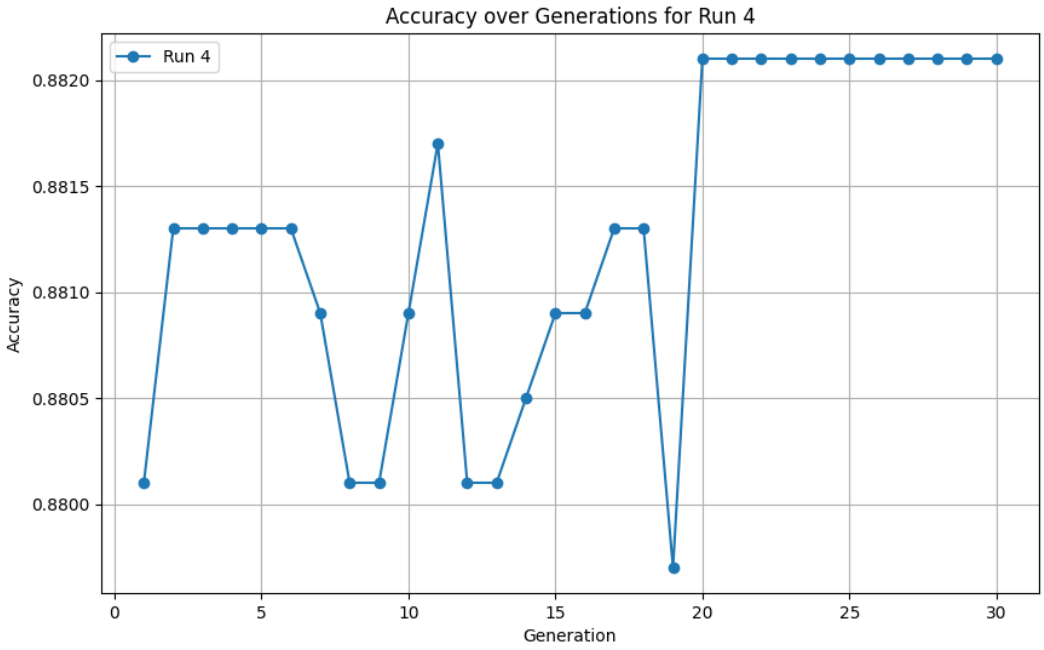

Figure 4.5. Accuracy per generations run 4.

According to Figures 4.3 and 4.4, the algorithm reaches convergence in less than 30 iterations and achieves the same level of accuracy after the 20th and 25th generations. Based on an analysis of the highest fitness rates, it can be concluded that the algorithm is capable of producing satisfactory outcomes after around 30 generations. However, due to the nature of the algorithm being a metaheuristic algorithm, its convergence may vary in terms of the amount of iterations or the success metrics achieved. However, GWOA demonstrated its efficacy by consistently outperforming other algorithms in every iteration.

Table 4.3. Results of comparison of GWOA with other algorithms.

	Without using any feature selection	**Information Gain**	**Pearson Correlation**	**Chi-Square**	**PSO**	**WOA**	**GA**	**GWOA**
Accuracy	87.07%	87.22%	87.26%	87.46%	84.57%	87.54%	87.07%	**88.60%**
F-1 Score	86.89%	87.04%	%87.04	87.28%	86.74%	86.86%	86.89%	**87.66%**

5. CONCLUSION

This thesis presents a novel hybrid feature selection strategy for detecting cyberbullying. The combination of two robust algorithms enhances the performance of GWOA, surpassing classic approaches such as Pearson Correlation, Information Gain, and Chi Square test. Furthermore, GWOA outperforms independent algorithms like GA and WOA. The KNN classifier was employed to achieve success metrics such as accuracy and F1-Score. Several experiments have been conducted to enhance the performance of GWOA. GWOA outperforms all the feature selection approaches tested in this thesis in terms of accuracy and F1-Score. The algorithm achieves its highest accuracy in the second experiment, where it uses a mutation rate of 0.1, a population size of 50, 30 iterations of the WOA algorithm, 20 iterations of the GA algorithm, and a total of 50 generations. The objective of this research was to examine whether increasing the population size has any positive impact on performance. The results indicate that increasing the population size has positive benefits on both accuracy and feature size. The population size analysis further demonstrates that the feature size can be reduced without compromising accuracy. Increasing the number of iterations for the GA and the Whale Optimization Algorithm (WOA) separately in different trials results in improved performance of the GWOA compared to the initial trial. The algorithm consistently provided its optimal outcomes throughout the 25th to 30th iterations. Based on the results, it can be concluded that increasing the number of generations would not be beneficial. This method falls under the category of metaheuristic algorithms, where the success metric is variable by changing the number of iterations.

This approach can be applied to many subjects for further feature selection work. Partitioning the algorithm into two distinct phases, namely exploration and exploitation, yielded superior outcomes for this specific challenge. This technique may be beneficial for addressing additional important problems. Various natural language processing (NLP) techniques can be employed in conjunction with algorithms to achieve improved outcomes. Additionally, several classifiers can be employed to enhance the algorithm's success metric. A drawback of this method is its lengthy response time. Each iteration of the algorithms takes nearly 8000 seconds that makes nearly 2.5 hours. Given the large size of the data collection and the inherently slow nature of metaheuristic algorithms, it is unsurprising that this outcome was

observed. By implementing various preprocessing or NLP techniques, it is possible to improve the time performance and success metrics of the method.

REFERENCES

Azhagusundari, B., & Thanamani, A. S. (2013). Feature selection based on information gain. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2(2), 18-21.

Aind, A. T., Ramnaney, A., & Sethia, D. (2020, June). Q-bully: a reinforcement learning based cyberbullying detection framework. In *2020 International conference for emerging technology (INCET)* (pp. 1-6). IEEE.

Al-Ajlan, M. A., & Ykhlef, M. (2018, April). Optimized twitter cyberbullying detection based on deep learning. In *2018 21st Saudi Computer Society National Computer Conference (NCC)* (pp. 1-5). IEEE.

Al-Garadi, M. A., Varathan, K. D., & Ravana, S. D. (2016). Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network. *Computers in Human Behavior*, *63*, 433-443.

Al-Hashedi, M., Soon, L. K., & Goh, H. N. (2019, November). Cyberbullying detection using deep learning and word embeddings: An empirical study. In *Proceedings of the 2019 2nd International Conference on Computational Intelligence and Intelligent Systems* (pp. 17-21).
Bauman, S. (2007, November). Cyberbullying: A virtual menace. In *National Coalition Against Bullying National Conference* (Vol. 2, No. 4).

Anter, A. M., Abd Elaziz, M., & Zhang, Z. (2022). Real-time epileptic seizure recognition using Bayesian genetic whale optimizer and adaptive machine learning. *Future Generation Computer Systems*, *127*, 426-434

Barlińska, J., Szuster, A., & Winiewski, M. (2013). Cyberbullying among adolescent bystanders: Role of the communication medium, form of violence, and empathy. *Journal of Community & Applied Social Psychology*, *23*(1), 37-51.

Beştaş, M. (2023). Detection of android-based applications with traditional metaheuristic algorithms. International Journal of Pure and Applied Sciences, 9(2), 381-392. https://doi.org/10.29132/ijpas.1382344

Cheng, L., Guo, R., Silva, Y., Hall, D., & Liu, H. (2019, May). Hierarchical attention networks for cyberbullying detection on the Instagram social network. In *Proceedings of the 2019 SIAM international conference on data mining* (pp. 235-243). Society for Industrial and Applied Mathematics.

Çetin, N., Genetik Algoritma, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul, 111, 2002. 74 [64] Katoch, S., Chauhan, S. S., Kumar, V., A review on

genetic algorithm: past, present, and future, Multimedia Tools and Applications, 80(5), 8091-8126, 2021.

Dadvar, M., de Jong, F. M., Ordelman, R., & Trieschnigg, D. (2012). Improved cyberbullying detection using gender information. In *Proceedings of the Twelfth Dutch-Belgian Information Retrieval Workshop (DIR 2012)* (pp. 23-25). Universiteit Gent.

Dinakar, K., Reichart, R., & Lieberman, H. (2011). Modeling the detection of textual cyberbullying. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 5, No. 3, pp. 11-17).

Hegazy, A. E., Makhlouf, M. A., & El-Tawel, G. S. (2020). Improved salp swarm algorithm for feature selection. *Journal of King Saud University-Computer and Information Sciences*, *32*(3), 335-344.

Hon, L., & Varathan, K. (2015). Cyberbullying detection system on twitter. *IJABM*, *1*(1), 1-11.

Huang, Q., Singh, V. K., & Atrey, P. K. (2014, November). Cyber bullying detection using social and textual analysis. In *Proceedings of the 3rd International Workshop on Socially-aware Multimedia* (pp. 3-6).

Iorga, M., Pop, L. M., Croitoru, I., Hanganu, E., & Anton-Păduraru, D. T. (2022). Exploring the importance of gender, family affluence, parenting style and loneliness in cyberbullying victimization and aggression among Romanian adolescents. *Behavioral Sciences*, *12*(11), 457.

Jiang, T., Zhang, C., Zhu, H., Gu, J., & Deng, G. (2018). Energy-efficient scheduling for a job shop using an improved whale optimization algorithm. *Mathematics*, *6*(11), 220.

Jing, Y., Haowei, M., Ansari, A. S., Sucharitha, G., Omarov, B., Kumar, S., ... & Alyamani, K. A. (2023). Soft computing techniques for detecting cyberbullying in social multimedia data. *ACM Journal of Data and Information Quality*, *15*(3), 1-14.

Kennedy, J. A. M. E. S. (1995, June). Eberhart, r.: Particle swarm optimization. In *Proceedings of the IEEE International Conference on Neural Networks* (Vol. 4, pp. 1942-1948).

Kontostathis, A., Edwards, L., & Leatherman, A. (2010). Text mining and cybercrime. *Text mining: Applications and theory*, 149-164.

Kramer, O., Genetic algorithm Essentials, Springer, Germany, 2017.

Kumar, M., Husain, M., Upreti, N., Gupta, D., Genetic algorithm: Review and application, International Journal of Information Technology and Knowledge Management, 2(2), 451-454, 2010.

Kumari, K., & Singh, J. P. (2021). Identification of cyberbullying on multi-modal social media posts using genetic algorithm. *Transactions on Emerging Telecommunications Technologies*, *32*(2), e3907.

Law, D. M., Shapka, J. D., Hymel, S., Olson, B. F., & Waterhouse, T. (2012). The changing face of bullying: An empirical comparison between traditional and internet bullying and victimization. *Computers in human behavior*, *28*(1), 226-232.

Lee, Y., Harris, M. N., & Kim, J. (2022). Gender differences in cyberbullying victimization from a developmental perspective: an examination of risk and protective factors. Crime &Amp; Delinquency, 68(13-14), 2422-2451. https://doi.org/10.1177/00111287221081025.

Li, Q. (2006). Cyberbullying in schools: A research of gender differences. *School psychology international*, *27*(2), 157-170.

Liu, L., Kuang, F., Li, L., Xu, S., & Liang, Y. (2022). An efficient multi-threshold image segmentation for skin cancer using boosting whale optimizer. *Computers in Biology and Medicine*, *151*, 106227.

Loh, J. and Snyman, R. (2020). The tangled web: consequences of workplace cyberbullying in adult male and female employees. Gender in Management: An International Journal, 35(6), 567-584. https://doi.org/10.1108/gm-12-2019-0242.

Macbeth, J., Adeyema, H., Lieberman, H., & Fry, C. (2013). Script-based story matching for cyberbullying prevention. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 901-906).

Mahlangu, T., & Tu, C. (2019, November). Deep learning cyberbullying detection using stacked embbedings approach. In *2019 6th International Conference on Soft Computing & Machine Intelligence (ISCMI)* (pp. 45-49). IEEE.

Meesad, P., Boonrawd, P., & Nuipian, V. (2011, January). A chi-square-test for word importance differentiation in text classification. In *Proceedings of international conference on information and electronics engineering* (pp. 110-114).

Muneer, A., Alwadain, A., Ragab, M. G., & Alqushaibi, A. (2023). Cyberbullying detection on social media using stacking ensemble learning and enhanced BERT. *Information*, *14*(8), 467.

Nadimi-Shahraki, M. H., Fatahi, A., Zamani, H., Mirjalili, S., & Oliva, D. (2022). Hybridizing of whale and moth-flame optimization algorithms to solve diverse scales of optimal power flow problem. *Electronics*, *11*(5), 831.

Nahar, V., Al-Maskari, S., Li, X., & Pang, C. (2014). Semi-supervised learning for cyberbullying detection in social networks. In *Databases Theory and Applications: 25th Australasian Database Conference, ADC 2014, Brisbane, QLD, Australia, July 14-16, 2014. Proceedings 25* (pp. 160-171). Springer International Publishing.

Nahar, V., Li, X., Pang, C., & Zhang, Y. (2013, November). Cyberbullying detection based on text-stream classification. In *The 11th Australasian Data Mining Conference (AusDM 2013)*.

Nandhini, B. S., & Sheeba, J. I. (2015, March). Cyberbullying detection and classification using information retrieval algorithm. In *Proceedings of the 2015 international conference on advanced research in computer science engineering & technology (ICARCSET 2015)* (pp. 1-5).

Nasir, I. M., Khan, M. A., Yasmin, M., Shah, J. H., Gabryel, M., Scherer, R., & Damaševičius, R. (2020). Pearson correlation-based feature selection for document classification using balanced training. *Sensors*, *20*(23), 6793.

Nitta, T., Masui, F., Ptaszynski, M., Kimura, Y., Rzepka, R., & Araki, K. (2013, October). Detecting cyberbullying entries on informal school websites based on category relevance maximization. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing* (pp. 579-586).

Olweus, D., Limber, S., & Mihalic, S. F. (1999). Blueprints for violence prevention, book nine: Bullying prevention program. Boulder, CO: Center for the Study and Prevention of Violence, 12(6), 256-273.

Ortega, R., Elipe, P., Mora-Merchán, J. A., Calmaestra, J., & Vega, E. (2009). The emotional impact on victims of traditional bullying and cyberbullying: A study of Spanish adolescents. *Zeitschrift für Psychologie/Journal of Psychology*, *217*(4), 197-204.

Patchin, J. W., & Hinduja, S. (2006). Bullies move beyond the schoolyard: A preliminary look at cyberbullying. *Youth violence and juvenile justice*, *4*(2), 148-169.

Patchin, J. W., & Hinduja, S. (2011). Traditional and nontraditional bullying among youth: A test of general strain theory. *Youth & Society*, *43*(2), 727-751.

Potha, N., & Maragoudakis, M. (2014, December). Cyberbullying detection using time series modeling. In *2014 IEEE International Conference on Data Mining Workshop* (pp. 373-382). IEEE.

Ptaszynski, M., Dybala, P., Matsuba, T., Masui, F., Rzepka, R., Araki, K., & Momouchi, Y. (2010). In the service of online order: Tackling cyber-bullying with machine learning and affect analysis. *International Journal of Computational Linguistics Research*, *1*(3), 135-154.

Ptaszynski, M., Masui, F., Kimura, Y., Rzepka, R., & Araki, K. (2015, November). Extracting patterns of harmful expressions for cyberbullying detection. In *Proceedings of 7th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics (LTC'15), The First Workshop on Processing Emotions, Decisions and Opinions* (pp. 370-375).

Qiqi, C. (2024). Reactions of adolescent cyber bystanders toward different victims of cyberbullying: the role of parental rearing behaviors. *BMC psychology*, *12*(1), 377.

Raisi, E., & Huang, B. (2018). Weakly supervised cyberbullying detection with participant-vocabulary consistency. *Social Network Analysis and Mining*, *8*(1), 38.

Reynolds, K., Kontostathis, A., & Edwards, L. (2011, December). Using machine learning to detect cyberbullying. In *2011 10th International Conference on Machine learning and applications and workshops* (Vol. 2, pp. 241-244). IEEE.

Rigby, K. (2013). Bullying in schools and its relation to parenting and family life. *Family Matters*, (92), 61. Retrieved from https://aifs.gov.au/sites/default/files/fm92f.pdf.

Rivers, I., & Smith, P. K. (1994). Types of bullying behaviour and their correlates. *Aggressive Behavior*, *20*(5), 359–368. https://doi.org/10.1002/1098-2337(1994)20:5.

Sakri, S. B., Rashid, N. B. A., & Zain, Z. M. (2018). Particle swarm optimization feature selection for breast cancer recurrence prediction. *IEEE Access*, *6*, 29637-29647.

Sarac Essiz, E., & Oturakci, M. (2021). Artificial bee colony–based feature selection algorithm for cyberbullying. *The Computer Journal*, *64*(3), 305-313.

Sherly, T. T., & Jeetha, B. R. (2021). Enhanced Cuckoo Search Optimization and Hybrid Firefly Artificial Neural Network Algorithm for Cyberbullying Detection on Twitter Dataset.

Singh, V. K., Huang, Q., & Atrey, P. K. (2016, August). Cyberbullying detection using probabilistic socio-textual information fusion. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 884-887). IEEE.

Slonje, R., Smith, P. K., & Frisén, A. (2012). Processes of cyberbullying, and feelings of remorse by bullies: A pilot study. *European Journal of Developmental Psychology*, *9*(2), 244-259.

Smith, P. K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S., & Tippett, N. (2008). Cyberbullying: Its nature and impact in secondary school pupils. *Journal of child psychology and psychiatry*, *49*(4), 376-385.

Snakenborg, J., Van Acker, R., & Gable, R. A. (2011). Cyberbullying: Prevention and intervention to protect our children and youth. *Preventing School Failure: Alternative Education for Children and Youth*, *55*(2), 88-95.

Sun, Y., Wang, X., Chen, Y., & Liu, Z. (2018). A modified whale optimization algorithm for large-scale global optimization problems. *Expert Systems with Applications*, *114*, 563-577.

Tomkins, S., Getoor, L., Chen, Y., & Zhang, Y. (2018, August). A socio-linguistic model for cyberbullying detection. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 53-60). IEEE.

Topalli, I., Modelling User Habits and Providing Recommendations Based on Hybrid Television Standards Using Artificial Neural Networks Together with Genetic Algorithms, Dokuz Eylül Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi, İzmir, 115, 2017.

Vandebosch, H., & Van Cleemput, K. (2009). Cyberbullying among youngsters: Profiles of bullies and victims. *New media & society*, *11*(8), 1349-1371.

Wang, G., Li, X., Gao, L., & Li, P. (2022). An effective multi-objective whale swarm algorithm for energy-efficient scheduling of distributed welding flow shop. *Annals of Operations Research*, 1-33.

Wang, K., Xiong, Q., Wu, C., Gao, M., & Yu, Y. (2020, July). Multi-modal cyberbullying detection on social networks. In *2020 International joint conference on neural networks (IJCNN)* (pp. 1-8). IEEE.

Willard, N. (2006). Flame Retardant: Cyberbullies Torment Their Victims 24/7: Here's How to Stop the Abuse. *School Library Journal*, *52*(4), 54.

Yadav, J., Kumar, D., & Chauhan, D. (2020, July). Cyberbullying detection using pre-trained bert model. In *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 1096-1100). IEEE.

Yang, X. S., Nature-inspired optimization algorithms, Academic Press, United Kingdom, 2020.

Zhang, X., Tong, J., Vishwamitra, N., Whittaker, E., Mazer, J. P., Kowalski, R., ... & Dillon, E. (2016, December). Cyberbullying detection with a pronunciation based convolutional neural network. In *2016 15th IEEE international conference on machine learning and applications (ICMLA)* (pp. 740-745). IEEE.

Zhao, R., & Mao, K. (2016). Cyberbullying detection based on semantic-enhanced marginalized denoising auto-encoder. *IEEE Transactions on Affective Computing*, *8*(3), 328-339.

Zhao, R., Zhou, A., & Mao, K. (2016, January). Automatic detection of cyberbullying on social networks based on bullying features. In *Proceedings of the 17th international conference on distributed computing and networking* (pp. 1-6).