Social Media Forensics: An Adaptive Cyberbullying-Related Hate Speech Detection Approach Based on Neural Networks with Uncertainty

ADITYA KUMAR YADAV[1], Prof ROOP RANJAN[2]
[1]Department of computer science and engineering, A.P.J Abdul Kalam Technical University,Lucknow

ABSTRACT Cyberbullying is a worldwide crisis that affects victims and society as a whole. It is a social media network issue. Because social media platforms are complex and use complex vocabulary, it has become very difficult to automatically detect cyberbullying on these platforms. It might be difficult to precisely grasp the intended meaning of a document because of its casual and concise style, which frequently leads to ambiguous or imprecise language. When confronted with unclear or contextually ambiguous content, identifying cyberbullying becomes even more difficult. There are now many methods for detecting cyberbullying, but they still struggle to differentiate between different types of hate speech linked to cyberbullying because it is unclear and ambiguous and because they are not very accurate. Through the integration of Neutrosophic Logic into the Multi-Layer Perceptron (MLP) model, this paper suggests a novel method for fine-grained cyberbullying classification. By reducing the difficulties caused by the vagueness and overlapping borders between different forms of cyberbullying, the suggested model improves cyberbullying types. By addressing the ambiguity, indeterminacy, and uncertainty inherent in classification decisions, Neutrosophic Logic provides a more thorough and adaptable method for managing intricate categorization scenarios. Because there are overlaps and unclear instances with other types of cyberbullying, the model, which uses the one-against-one technique in MLP classification, captures complex interactions between multiple types of cyberbullying. This model's testing phase highlights the importance Using class probabilities from several one-against-one classifiers, Neutrosophic Logic offers a thorough understanding of classification results.The results of the proposed model demonstrate the performance enhancement of incorporating Neutrosophic Logic for fine-grained cyberbullying classification tasks.

INDEXTERMS Cyberbullying,hatespeechdetection,one-against-one,multiclassclassification, neutrosophic sets, social media forensics.

I. INTRODUCTION

With the progression of digital technologies and the widespread adoption of social media, bullying has escalated in its threat to individuals, as it now can be carried outusing internet technologies [1]. Threats, online harassment, disgrace, fear, and other forms of cyberbullying are characterized as new forms of violence or bullying that are perpetrated through technical gadgets and the World Wide Web [2]. Socialmedia forensics involves the collection, analysis, and investigation of digital data gathered from diverse social media platforms to uncover evidence pertinent to legal or criminal inquiries.

Within the realm of digital forensics, social media evidence represents a novel area of study [3]. In order to detect cyberbullying, social media evidence analysis is essential. Because language ambiguity can vary widely depending on the speaker, the audience, the context, the informality of the language, and the diversity of cultures and situations, detecting cyberbullying is a difficult task [4], [5]. There are two primary methods for detecting cyberbullying speech [6]: ensemble and machine learning-based methods. The machine learning (ML) method learns the linguistic patterns linked to hate speech related to cyberbullying using statistical models. Additionally, the machine learning-based methodologies are combined in the ensemble approach. This method verifies whether the post qualifies as cyberbullying speech using machine learning models. This may increase the classification accuracy of cyberbullying speech. The MLP is a commonly used method in the field of machine learning[7].

MLP classification is a machine learning technique that can be used to categorize cyberbullying in textual content. Multiple layers of artificial neurons coupled in a particular way make up MLP classifiers. Each layer's neurons are in charge of learning distinct aspects from the input text, and the final layer's output is used to categorize the input text into various groups, such whether or not cyberbullying is occurring. MLP has the advantage of being able to learn intricate nonlinear relationships between features, which makes it a good fit for jobs involving text classification. Large data sets can be handled by MLPs, which makes text categorization tasks easier. MLPs are a suitable option for applications when data is restricted since they are relatively easy to train [7]. The intrinsic subjectivity of language in verbal communication makes it difficult to recognize and classify the many forms of cyberbullying. This intricacy stems from the fact that how a text is interpreted can vary based on a number of circumstances, including the audience's cultural background, the speaker or writer's goals, and the context in which it is used [8]. A statement that might be seen as cyberbullying in one setting might not be in another. Usually, a dataset of annotated text is used to train models, however this dataset might not be representative of the various ways that cyberbullying might manifest itself. Because of this, machines may mistakenly label cyberbullying as either victimization or non-cyberbullying.

Neutrosophic logic (NL)[9] is an extension of classical logic that represents indeterminacy by adding a third truth value in addition to true and false. NL enables thinking and decision-making processes to deal with ambiguity and uncertainty. Neutrosophic logic offers a more thorough method of handling faulty or insufficient data and has applications in a number of fields, such as artificial intelligence, decision support systems, and pattern recognition. NL has a lot of advantages over conventional methods of classification. The first is its capacity to represent and rationalize within ambiguous and deterministic information. Conventional classification techniques frequently produce erroneous or insufficient findings because they are unable to manage data uncertainty. NL, on the other hand, offers a structured framework for representing and reasoning with ambiguous data, enabling more robust and flexible classification. The capacity of NL to collect and model intricate relationships between variables in a nuanced manner is another benefit. Less accurate classifications may result from the oversimplification or neglect of subtle relationships between elements in traditional classification systems.

With NL, on the other hand, levels of truth, falsity, and indeterminacy are represented using a three-valued system. Indeed, ML usually simply accounts for truth and falsehood, whereas deep learning depends on probability. According to Fuzzy Logic [10], [11], however, uncertainty is represented by degrees of membership and non-membership. The explicit depiction of indeterminacy and membership functions in NL's method makes it a more effective tool for handling the complexities of detecting hate speech related to cyberbullying than typical fuzzy tools and machine learning.

.

A.CONTRIBUTIONANDMETHODOLOGY

A fine-grained categorization model based on eutrosophic neural networks for cyberbullying is presented in this research. The proposed model uses neutrosophic reasoning to present a novel technique to classifying cyberbullying types. use an MLP classifier and a one-against-one strategy for multiclass classification. The probability of each class are also subjected to neutrosophic categorization.

One of this article's primary contributions is the introduction of a novel fine-grained neutrosophic neural network classification model. (2) Using the One-Against-One approach, building and training a group of binary classifiers to address multiclassification. (3) Class probabilities for various forms of cyberbullying are predicted using an MLP classifier. (4) Using a set of binary classifiers, generate probabilities for each class, and then use these probabilities to determine the prevailing class for the specified data kinds. (5) Using interval neutrosophic sets as the basis for the final classification decision, probabilities are converted into neutrosophic sets.

The rest of the paper is organized as follows: Section II presents some of the recent related work. Section III describes the proposed cyberbullying classification model. In Section IV the results, and discussions on the cyberbul- lying dataset. Finally, conclusions are drawn in Section V.

II. RELATEDWORK

Cyberbullying detection has been widely researched, begin- ning with user studies in the social sciences and psychology sectors, and more recently shifting to computer science with the goal of building models for automated identification. There are many kinds of ML techniques, however, the most was used in almost every study on cyberbullying prediction on social media. But there isn't a single best machine learning technique for every problem. As a result, most research selects and assesses a range of supervised classifiers to determine which one best suits their problem. To choose classifiers, the most popular predictors in the field are employed, along with the data variables that are available for trials. However, only after completing a comprehensive practical trial can researchers decide which algorithms to employ when developing a cyberbullying detection model [6].

Despite obtaining strong indicators, the author of the work[12] evaluates machine learning approaches against the lexical method, acknowledging limitations in distinguishing verbally stated emotions. The authors suggest sentiment identification techniques that make use of knowledge bases linked to particular emotions in order to get around this restriction. The study presents three different approaches to cyberbullying recognition: supervised machine learning that examines different linguistic features, deep machine learning using neural networks like convolutional neural networks, and a rules-based method that detects explicit cyberbullying through keyword combinations and lexical resources. Every strategy has its own benefits. Deep learning captures intricate patterns and correlations, supervised learning offers flexibility with a variety of linguistic variables, and the rules-based approach offers interpretability for explicit cyberbullying identification. Limitations, however, include the necessity for considerable labeled data in supervised learning, the computational intensity of deep learning, the potential oversight of subtle cyberbullied kinds in the rules-based system, and potential difficulties with excessively long texts.

Additionally, the authors in [13]suggested an automated cyberbullying detection model to deal with imbalanced short text and diverse dialects appears in the Arabic text. The simulated annealing optimization algorithm is used to find the optimal set of samples from the majority class to balance the training set. The work employed a comprehen- sive evaluation by testing the model with both traditional machine learning algorithms and deep learning algorithms. This approach ensures a robust assessment of the frame- work's performance across different methodologies. The authors mentioned that the limitation of this work is associ- ated with the complexities introduced by linguistic diversity and regional variations in the Arabic language, particularly when applied to cyberbullying detection.

Furthermore, the authors in [14]introduced a strategy for social media cyberbullying detection. They employed four machine learning models: Support Vector Machine (SVM), Naïve Baise (NB), Decision Tree (DT), and K-Nearest Neighbor (KNN) to categorize texts into cyberbullying and non-cyberbullying categories. The training of these mod-els involved the application of various features, including badwords, negativeemotion, positiveemotion,links,proper nouns, and pronouns. The work didn't deal with cyberbul- lyingsub-types.Similarly,theauthorsin[15]developed an ensemble model for cyberbullying detection. They used

Long Short-Term Memory(LSTM) and Convolutional Neural Network(CNN), which have demonstrated effectiveness in detecting cyberbullying. The results demonstrate the method's efficacy in identifying and categorizing offensive language on social media platforms. The authors admit there's still a lot of work to do in making more depend-able methods for spotting cyberbullying. They point out challenges, especially the difficulty of making the method work well in different situations. They suggest looking into advanced techniques and new ways of using technology to improve cyberbullying detection, emphasizing the need for ongoing innovation.

The authors of [16] used six learning algorithms and three deep learning algorithms to classify cyberbullying. According to the findings, LSTM has the best accuracy and recall for detecting cyberbullying. However, class imbalance data and fine-grained classification to categorize cyberbullying types were not addressed. Furthermore, a framework for identifying cyberbullying in texts was developed by the work in [17]. It uses a fuzzy logic system that uses the outputs of SVM classifiers as its inputs to identify cyberbullying. The findings indicate that in order to assess the degree of bullying using fuzzy logic, SVM classifier accuracy must be increased. Determining how bullying instances were based on the collected tweets was another issue with the work. The authors found it difficult to consistently determine the severity of bullying episodes even with a fuzzy logic system. Because each author had a different perspective, even when they used the same criteria to create the fuzzy rules, they found that determining how severe a bullying episode was became challenging. Because of this, the authors' opinions on how serious a bullying situation was varied, which made it a subjective and difficult part of their research.

As an extension of fuzzy logic, neutrophilic sets [18] offer a more flexible strategy for successfully managing uncertainty. These studies offer important insights into the various uses and advantages of neutrosophic sets. In [19], the authors described the use of neutrosophic sets in multi-attribute group decision-making, emphasizing their capacity to manage uncertainty in complex evaluations, especially when assessing math teachers. The study emphasizes the flexibility and resilience of neutrosophic sets in scenarios involving multi-attribute group decision-making by using single-valued trapezoidal neutrosophic numbers. Additionally, the authors of [20] developed a novel method for classifying skin cancer by fusing deep features into a neutrosophic framework. This study demonstrates how neutrophilic sets improve medical diagnostic accuracy and reliability while demonstrating their adaptability in this setting. Additionally, a generalized linguistic neutrophilic cubic aggregation operator was used in the work presented in [21] to introduce an image processing procedure, demonstrating its efficacy in addressing image processing challenges during uncertainty.

These varied studies underscore the increasing interest and promise of neutrosophic sets in diverse domains. By integrating uncertainty into decision-making and analytical procedures, neutrosophic sets emerge as a valuable tool for improving the precision and resilience of complex systems.
The conducted survey showed that the cyberbullying social media detection systems have the following limitations:

(a) Handling Class Imbalance: Many cyber bullying datasets suffer from class imbalance, where the number of instances of cyberbullying types maybe significantly lower than other cyberbullying types of instances. Failure to address this issue can lead to biased models and decreased performance in detecting cyberbullying accurately.

(b) Fine-Grained Classification: While some works focus on binary classification of cyberbullying versus non-cyberbullying, there’s a need for more fine-grained classification to differentiate between various types or severity levels of cyberbullying. Ignoring this aspect may lead to oversimplified models that can not effectively address the degrees of cyberbullying behavior.

(c) Subjective Determination of Bullying: The subjec- tive process of assessing the severity of bullying incidents presents a hurdle in reliably identifying and classifying instances of cyberbullying. This subjectivity can result in discrepancies in data labeling and model evaluation, ulti- mately affecting the dependability of cyberbullying detection systems.

(d) Fuzzy approach is dealing with uncertainty, but it has some disadvantages that can make it unsuitable in fine grained cyberbulying as fuzzy logic systems are typically designed by human experts, who must specify the membership functions for the fuzzy sets. This can be a time-consuming and error-prone process. Also, fuzzy logic systems are based on fuzzy sets, which are inherently imprecise. This can lead to in accurate results, especially in applications where high accuracy is like cyberbullying detection.

(e) Inaccurate classification results are found when using MLP classification method separately, where as combining it with neutron sophic improves the classification results.To the best of our knowledge, little attention has been paid to devising a new neutron sophic technique for cyber bullying fine-grained classification.

III. METHODOLOGY

In order to make accurate fine-grained classifications for cyberbullying types, initially, the model utilizes an MLP classifier employing the One-Against-One strategy, enabling it to discernintricate patterns in the data. Subsequently, probabilities for each cyberbullying class are extracted through this process, providing a rich understanding of the likelihood of different types of cyberbullying occurrences. These probabilities are then converted into neutron sophic sets, leveraging the flexibility and adaptability of neutrosophic logic to capture un certainties and complexities in the classification task. Finally, utilizing neutron sophic intervals, the model makes the ultimate classification decision, offering a refined approach to cyberbullying detection that accounts for the inherent ambiguities and intricacies of online communication. Fig. 1illustrates the key components of the model and their interconnected relationships. Subsequent sections elaborate on these major model elements.

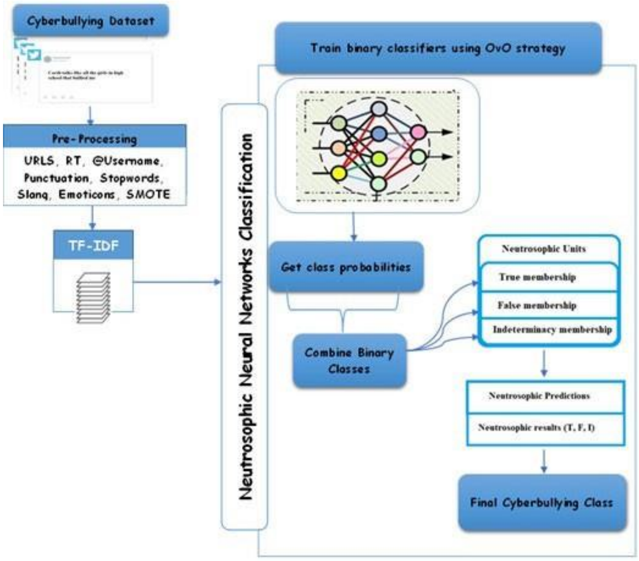


FIGURE 1.The proposed Neutrosophic cyberbullying fine-grained classification.

A. DATACOLLECTIONPHASE

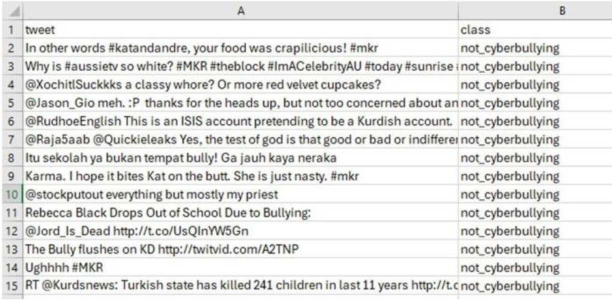
This phase is concerned with collecting the data essentialfor validating the proposed neutrosophic neural network model. Twitter was chosen to apply the model. There is a cyberbullying dataset [22], this dataset contains more than 47000 tweets labelled according to the class of cyberbully- ing that contains cyberbullying, gender, other_cyberbullying, age, not_cyberbullying, and ethnicity. The dataset contains twocolumns:(tweet_text,cyberbullying_type),‘tweet_text’ contains tweets. cyberbullying_type contains four types of cyberbullying, age, gender, ethnicity, religion in addition to other_cyberbullying column and not cyberbullying. Fig. 2shows a sample of the cyberbullying dataset.

B. PRE-PROCESSINGPHASE

In text preprocessing, texts are cleared by stripping emoji fromtext,removingstopwords,removepunctuations,links, mentions and new line characters, clear hashtag and special characters to represent the main body of the text.

C. FINE-GRAINEDCLASSIFICATION

Fine-grained classification is a more challenging ML type where the goal is to predict the specific subcategory or class with in a larger category. Fine-grained cyberbullying classification is the task of classifying cyberbullying incidents into specific sub categories like cyberbullying, age,



	A	B
1	tweet	class
2	In other words #katandandre, your food was crapilicious! #mkr	not_cyberbullying
3	Why is #aussietv so white? #MKR #theblock #ImACelebrityAU #today #sunrise	not_cyberbullying
4	@XochitlSuckkks a classy whore? Or more red velvet cupcakes?	not_cyberbullying
5	@Jason_Gio meh. :P thanks for the heads up, but not too concerned about an	not_cyberbullying
6	@RudhoeEnglish This is an ISIS account pretending to be a Kurdish account.	not_cyberbullying
7	@Raja5aab @Quickieleaks Yes, the test of god is that good or bad or indifferer	not_cyberbullying
8	Itu sekolah ya bukan tempat bully! Ga jauh kaya neraka	not_cyberbullying
9	Karma. I hope it bites Kat on the butt. She is just nasty. #mkr	not_cyberbullying
10	@stockputout everything but mostly my priest	not_cyberbullying
11	Rebecca Black Drops Out of School Due to Bullying:	not_cyberbullying
12	@Jord_Is_Dead http://t.co/UsQInYW5Gn	not_cyberbullying
13	The Bully flushes on KD http://twitvid.com/A2TNP	not_cyberbullying
14	Ughhhh #MKR	not_cyberbullying
15	RT @Kurdsnews: Turkish state has killed 241 children in last 11 years http://t.c	not_cyberbullying

FIGURE2. Cyberbullyingdatasetsample.

[23], and one-against-all (OvA) [24]classification are related techniques that can be used to improve the accuracy of ML models for text classification tasks. In multiclass classification, the go a list classify text into one of multiple categories. One-against-one classification works by training a separatebinaryclassifierforeachpairofclasses.Forexample,ifthere arefourclasses,thensixbinaryclassifierswouldbetrained.Eachbinaryclassifierwouldbetrainedtodistinguishbetweenone pair of classes. To classify a new text instance, all sixbinaryclassifierswouldbeusedtopredicttheprobabilitythattheinstancebelongstoeachclass.Theclasswiththehighest predicted probability is assigned to the instance.

One-against-all classification works by training a separate binary classifier for each class against the rest of the classes. For example, if there are four classes, then four binary classifiers would be trained. Each binary classifier would be trained to distinguish between one class and the rest of the classes. To classify a new text instance, all four binary classifiers would be used to predict the probability that the instance belongs to each class. The class with the highest predicted probability is assigned to the instance [22].

D. *BINARYCLASSIFICATION*

The binary classification problem aims to find a linear function able to correctly classify an input vector between two classes. Given a training set$Z=(x_i,y_i):i\in 1,...,l$
With points $x_i\in R^d$and classes$y_i\in -1,+1\}$,where
$Z^+$is$(x_i,y_i)\in Z:y_i=+1\}$, the positive classset, and $Z^-$is
$(x_i,y_i)\in Z:y_i=-1\}$,the negative classset, the objective is to utilize an MLP to delineate complex decision boundaries between these classes represented by a normal vector $w\in R^d$ and a bias $b\in R$. The training process involves adjusting the weights and biase site ratively based on the error observed in The training set.The objectiveist of in the optimal weights
$(w)$and bias(b)that minimize the classification error[25].

E. *MULTI-CLASSCLASSIFICATION*

Inmanyreal-worldapplications,aclassifiermustbeableto classify an input vector between n classes, n $\in$N. Given a trainingset$Z=(x_i,y_i):i\in 1,...,l$with points $x_i\in R^d$and classes,$y_i\in 1,2,...,n$,theobjectiveistobuildafunction capable of assigning the correct class to an input vector.

The optimization problem that emerges from the expansion of the original binary formulation for large margin classi- fiers to work with more than two classes becomes highly complex according to the increase of the number of classes. Solving a binary classification problem is faster than solving a multiclass classification with the same amount of data. Therefore, instead of expanding the formulation of the binary classification, it is more common to break the multiclass classification problem into binary ones and combine the answers from the binary classifiers to assign the correct class[23].In this section, we present the two main approaches for solving multiclass classification: one-against-one and one-against-all [24].

1) ONE-AGAINST-ALL

The one-against-all approach takes into consideration each classjagainsttheothers,where$j\in 1,2,...,n$, for breaking the multiclass problem into a binary classification problem. For each class j, the full training set Z is taken into con- sideration, but the class j is seen as the positive class and
the other classes are seen as the negative class. A decision boundary$(w,b)_j$is then generated for each class j following a MLP and stored in a decision boundary set H. At the end of the process, n decision boundaries are generated, where n is the number of classes. Each decision boundary tells whether an input is likely to be of class j or not. The final class of an input is decided by finding out which decision boundary is the closest to the input. The class related to this decision boundary is then assigned to the input. Decision boundary
equation (w_j,b_j) for each class j:$w^Tx+b_j=0$,Thenumberof decision boundaries grows ^{j}linearly with regard to the number
of classes[23],[24].

2) ONE-AGAINST-ONE

The one-against-one approach takes into consideration pairs of classes(j,k),where$j,k\in 1,2,...,n$and$j<k$,for breaking the multiclass problem into a binary classification. For each pair(j,k),a subset Z of the original training set Z consisting of points with classes j and k is created, where j can be seen as the positive class and k can be seen as the negative class. The subset Z is used to generate a decision boundary $(w,b)_{j,k}$following MLP, and the decision boundary is added to the decision boundary set H. This process generates in total $n(n-1)/2$decisi on boundaries, where n is the number
Of classes. A decision boundary$(w,b)_{j,k}$ predicts the class j if
the input is classified as positive and k otherwise. A new input must be classified by every decision boundary and the class with the highest predicted probability is finally assigned to the input. The decision boundary$(w,b)$for each pair(j,k)can be represented by the equation: w$^Tx+b_{(j,k)}=0$,wherew^T is the$_{(j,k)}$transpose of the weight vector,x is the input instance,and b is the bias term. The sign of this equation determines the classification result for class j and class k. The number of decision boundaries grows quadratically with regard to the number of classes [23],[24].

3) NEUTROSOPHICCLASSIFICATION

The fine-grained classification procedure shown in Fig. 1 uses the NL concept. A more thorough and adaptable method for handling complex classification scenarios where there may not be clear boundaries between classes is provided by NL [9], which permits the representation of uncertainty, ambiguity, and indeterminacy in classification decisions. In the proposed model, the OvO approach with multiple binary MLP classifiers is used to introduce the neurosophical aspect. Each binary classifier in the OvO approach is constrained to distinguish between a particular pair of classes, capturing their relationship and subtleties. A frequent occurrence in fine-grained classification tasks, this method acknowledges and takes into account the possibility of overlapping or ambiguous instances that may fall between two classes. The OvO strategy is used to train the binary MLP classifiers on the dataset during the training phase [24]. By choosing samples and labels that correspond to each pair of classes, this strategy creates a distinct classifier for each pair of classes.

By training multiple binary classifiers, the model gains a deeper understanding of the intricate boundaries and relationships between the classes. This approach enables the model to capture the complex decision boundaries necessary for fine-grained classification tasks. In the testing phase, the neutrosophic concept is further emphasized. For each instance in the testing set, predictions are obtained from all the one-against-one classifiers.

During the testing phase, the predicted probabilities [26]for each class can be ob tained by evaluating the input instance x with each decision boundary $(w,b)_{j,k}$and applying a soft- max function, given by Eq. 1:

$$P(j|x)=\exp(-w\_((j,k))\wedge Tx+[b]\_((j,k)))/$$

$$\times((\exp(-w\_((j,k))\wedge Tx+[b]\_((j,k)))$$

$$+\exp(w\_((j,k))\wedge Tx+[b]\_((j,k))))))(1)$$

where $P(j|x)$ and represent the probabilities of the input instance x belonging to class j and class k, respectively.

The predicted probabilities from each classifier are col- lected and sorted in descending order. By considering the collective knowledge from all the classifiers, the model can make more informed decisions, taking into account the uncertainty and ambiguity associated with each instance's class assignment. The class with the highest predicted prob- ability is assigned to the instance, representing the most likely classmembership, considering the input's neutrosophic nature. By incorporating the concept of neutrosophic logic, the fine-grained classification model becomes more robust and capable of handling complex classification scenarios.It allows for the representation of uncertainty and ambi- guity, enabling the model to make nuanced decisions even in situations where crisp class boundaries do not exist. This approach enhances the model's accuracy and performance by considering the relationships between classes and capturing the inherent unuertainty presentinfine-grained classification tasks.

Converting neutrosophic [27]: converting class probabilitiestoNeutrosophicSets(NS):$N(P)=(T,I,F)$, Truth-Membership (T) represents the degree to which the samplebelongstotheclass.Wesetathreshold$T$andassign Tif$P\geq$T,and0otherwise.Indeterminacy-Membership(I) representsthedegreeofuncertaintyorambiguity.Weseta threshold$I$andassign$I$if$I\leq P<T$,and0otherwise.Falsity-Membership(F) represents the degree to which the sample doesnotbelongtotheclass.Youcanassign$F$if$P$ $<I$,and 0 otherwise.

Final Classification Decision based on Interval Neutrosophic Sets (INS) [28]: INS are an extension of traditional NSthatprovideamoreflexiblerepresentationofuncertainty. In INS instead of specifying precise values for the degrees of truth, indeterminacy, and falsity, Intervals is defined for theseparameters.Theseintervalsallowforarangeofpossible values, capturing the inherent uncertainty and imprecisionin the data more effectively. The use of INS offers several advantages including decision-making, classification, and risk assessment.

IV. ILLUSTRATIVEEXAMPLE

In this section, an illustrative example is provided for solving the proposed neutrosophic classification with five classes: Age, Ethnicity, Gender, Religionand Other types of cyberbullying.

Tweet text example =''Hey loser, why don't you go cry to your mommy? You' repathetic''.Following the proposed model'ssteps:

Step 1: Tokenizing the tweet into individual words or tokens.

Tokens:[''Hey'',''loser'',''why'',''don't'',''you'',''go'', ''cry'',''to'',''your'',''mommy'',''You're'',''pathetic''].

Step 2: Converting the tokens into numerical vectors using TF-IDF technique.

TF-IDF assigns weights to each token based on its frequency in the tweet and rarity across the dataset. Each token is represented by a vector of numerical values.

TF-IDFvaluesforeachtoken):''Hey'':[0.1,0.0,0.05, 0.0,...,0.02],''loser'':[0.0,0.2,0.0,0.0,...,0.03],''why'': [0.05,0.0,0.08,0.0,...,0.0],''don't'':[0.0,0.0,0.0,0.1,..., 0.0], ...

Step3:One-vs-OneClassificationusingMLP.

After obtaining the TF-IDF vectors, we feed them into multiple MLP classifiers trained for each pair of classes.We train multiple MLP classifiers, each focusing on dis- tinguishing between a pair of classes. These classifiers are trained using the one-vs-one strategy. Each MLP classifier takes the TF-IDF vectors as input and produces predictions for each class pair. For example, we have classes Age, Ethnicity,Gender,Religion,andOther_cyberbullying,wetrain classifiersforpairslike(Agevs.Ethnicity),(Agevs.Gender), (Agevs.Religion),(Agevs.Other_cyberbullying),(Ethnicity vs. Gender), and so on.

Classifier for (Age vs. Ethnicity) predicts Age: 0.2, Ethnicity:0.1,Classifierfor(Agevs.Gender)predictsAge:0.1, Gender: 0.3, Classifier for (Age vs. Religion) predicts Age: 0.4,Religion:0.5,Classifierfor(Agevs.Other)predictsAge: 0.2,Other:0.4,Classifierfor(Ethnicityvs.Gender)predicts Ethnicity: 0.2, Gender: 0.3.

Step4:Extractingprobabilitiesforeachclasspair.

For example: Probability of Age: 0.9, Probability of Ethnicity:0.3,ProbabilityofGender:0.3,ProbabilityofReligion: 0.5, Probability of Other_Cyberbullying: 0.4.

Step5:Fine-grainedClassification:

ConvertingprobabilitiestoNeutrosophicSets:converting the probabilities into neutrosophic sets for each class using specifiedthresholds(T,I,F).Assuming$T$=0.9,I= 0.3,F=0.2:

Example(forAge):

GivenprobabilityforAge:0.9(since0.9>=T) T(Age) = 0.9
I(Age)=|0.9-0.5|=0.4(sinceT>P(Age)>=F) F(Age) =1 - 0.9 =0.1 (since P(Age) <F)

Example(forEthnicity):

GivenprobabilityforEthnicity:0.3(F<0.3<T) T(Ethnicity) = 0.0
I(Ethnicity)=|0.3-0.5|=0.2(sinceT>P(Ethnicity)>= F)
F(Ethnicity)=1-0.3=0.7(sinceP(Ethnicity)<T) Example (for Gender):

GivenprobabilityforGender:0.3(F<0.3<T) T(Gender) = 0.0
I(Gender)=|0.3-0.5|=0.2(sinceT>P(Gender)>=F)
F(Gender)=1-0.3=0.7(sinceP(Gender)<T) Example (for Religion):

GivenprobabilityforReligion:0.5(F<0.5<T) T(Religion) = 0.0
I(Religion)=|0.5-0.5|=0.0(sinceP(Religion)=T)
F(Religion)=1-0.5=0.5(sinceP(Religion)<T) Example (for Other):

GivenprobabilityforOther_Cyberbullying:0.4(F<0.4 <T)

T(Other_Cyberbullying)=0.0
I(Other_Cyberbullying)=|0.4-0.5|=0.1, (since T >P(Other) >= F)

F(Other_Cyberbullying)=1-0.4=0.6(sinceP(Other) <T)

Step 6: Final Classification Decision Using Interval Neutrosophic Set.

Interval NeutrosophicSet for:Age:0.9,0,0.1}, Ethnicity: 0.0,0.2,0.7},Gender:0.0,0.2,0.7},Religion:0.0,0,0.5}, andOther_Cyberbullying:0.0,0.1,0.6}.Theseintervalneutrosophicsetsrepresentthetruth,indeterminacy,andfalsity memberships for each class, calculated based on the given probabilities and thresholds. for Age, the truth mem- bership (T) is highest (0.9) among all classes, and the indeterminacy membership (I) is also relatively low (0.4). Therefore, according to the neutrosophic classification, the final decision is to classify the tweet as related to ‘‘Age.’’In this case, the tweet is classified as Cyberbullying_type: Age.

V. EXPERIMENTALRESULTS

In this section, the performance of the proposed model is validated on two data sets focused on cyberbullying classification in social media, an arena where its prevalence and impact have grown considerably. The experiment was conducted using an Intel (R) Core (TM) i3 processor with 8.00 GB RAM and implemented in Anaconda. Herein, we utilize the evaluation metrics used in[6]:Precision,Recall,andF1Score as evaluation metrics [19].

$$Precision = T_P{}^{,}(T_P + F_P) \tag{2}$$

$$Recall = T_P{}^{,}(T_P + F_N) \tag{3}$$

$$F1Score = 2*(Precision*Recall)/(Precision+Recall) \tag{4}$$

Cyberbullying dataset 1 [22]contains over 47,000 labeled tweets specifically classified according to various categories related to cyberbullying: Age, Ethnicity, Gender, Religion, Other types of cyberbullying, and not classifiedas cyberbullying.

Cyberbullying dataset 2 [29]contains over total of approximately 100,000 tweets classified according to many categories related to cyberbullying: Race/Ethnicity,Gender/ Sexual, Religion, Other types of cyberbullying, and not_cyberbullying.

A. EXPERIMENT 1: MODEL PERFORMANCE FOR FINE GRAINEDCYBERBULLYINGCLASSIFICATION

In this experiment we dropped the (not_cyberbullying) data form classification type in data set to make a fineg rain classification between cyberbullying types: age, gender, ethnicity, religion, and other cyberbullying. We got 95% accuracy of our proposed model using dataset 1 and97%u sing dataset 2. We defined thresholds (T,I,F) for each class probability and convert to neutrosophicsets;T=0.9 (Truththreshold), I=0.3 (Indeterminacy threshold), F =0.2 (Falsity threshold).Table1 shows the neutrosophic classification report,There as on of this result is because of the combination of extensive text cleaning operations, including removing emojis, handling contractions, eliminating punctuation and non-ASCII characters, along with handling URLs and men- tions, significantly refines the data. These steps are crucial to ensure uniformity, relevance, and consistency within the dataset, thus enhancing the model’s ability to extract mean- ingful patterns from text data.

Also, using SMOTE [30]plays a pivotal role in addressing the class imbalance problem by artificially generating synthetic instances for the minority class. This technique essentially bridges the gap between classes by oversampling the underrepresented class, thus avoiding bias towards the majority class. By creating synthetic examples of the minority class, SMOTE prevents the model from favoring the more dominant class and allows it to learn effectively from both classes. Consequently, this leads to a more

TABLE1. Evaluationresultsoffine-grainedcyberbullyingclassification.

Data	Class	Precision	Recall	F1
Dataset 1 [22]	age	0.98	0.99	0.98
	ethnicity	0.99	0.98	0.98
	gender	0.89	0.92	0.91
	Other_cyberbullying	0.88	0.87	0.87
	religion	0.99	0.97	0.98
	Accuracy			0.95
Dataset 2 [29]	ethnicity/race	0.99	0.99	0.99
	gender/sexual	0.99	0.98	0.98
	religion	0.89	0.92	0.91
	Accuracy			0.97

The utilization of the one-vs-one strategy with MLP clas- sifier in our model played a pivotal role in achieving the high accuracy observed in our results. By training multiple MLP classifiers, each focusing on distinguishing between a pair of classes, we were able to capture intricate relation- ships and nuances between different cyberbullying types. This approach allowed the model to learn discriminative patterns specific to each class pair, leading to more precise and refined classification decisions. Further more, the extrac- tion of probabilities from the predictions of the classifiers provided valuable insights into the model's confidence levels for each class. These probabilities served as the basis for converting the classification outputs into neutrosophic sets, which enabled a more representation of uncertainty and ambiguity in the classification process. The conversion of probabilities to neutrosophic sets using predefined thresh- olds(T,I,F) further enhanced the model' sability to handle uncertainty and imprecision inherent in cyberbullying classification tasks. By setting appropriate thresholds for truth, indeterminacy, and fals it memberships, we ensured that the model could make in formed decisions while considering the inherent uncertainty in the data.

Furthermore, combining the one-on-one approach with Our model successfully navigated the complexities of cyberbullying classification with the help of the MLP classifier and the converse onto neutrosophic sets. The conversion to neutrosophic sets allowed for a more flexible and high-quality representation of classification outputs, while the one-vs-onest strategy offered a strong framework for capturing fine-grained distinctions between various forms of cyberbullying. Lastly, the thorough methodology used in our model—which combines cutting-edge machine learning methods with sophisticated logic principles—helped to achieve the high classification accuracy for cyberbullying that was noted. Utilizing the advantages of both approaches, our model produced accurate and trustworthy classification results by exhibiting a superior capacity to manage the ambiguity, uncertainty, and over-lapping characteristics present in cyberbullying data.

B. EXPERIMENT 2: COMPARISON BETWEEN THE PROPOSEDMODELANDOTHERMACHINE LEARNING MODELS

TABLE2. Comparisonresultsofdifferentmachinelearningmethodson cyberbullyingdataset.

Algorithm	Class	Precision	Recall	F1
RF [32]	age	0.96	0.98	0.97
	ethnicity	0.98	0.97	0.98
	gender	0.92	0.84	0.88
	Other_cyberbullying	0.78	0.90	0.84
	religion	0.98	0.93	0.96
	Accuracy			0.92
LR [33]	age	0.99	0.96	0.97
	ethnicity	0.95	0.97	0.96
	gender	0.97	0.74	0.84
	Other_cyberbullying	0.71	0.94	0.81
	religion	096	0.90	0.93
	Accuracy			0.90
SVM [31]	age	0.98	0.98	0.98
	ethnicity	0.98	0.98	0.98
	gender	0.83	0.81	0.82
	Other_cyberbullying	0.77	0.82	0.79
	religion	0.99	0.96	0.97
	Accuracy			0.91

This group of experiments was carried out to compare the efficiency of the suggested model and machine learning algorithms in the area of fine-grained classification of cyberbullying using a mix of a few machine learning algorithms. These machine learning algorithms include the Random Forest (RF) Algorithm [32], the Logistic Regression (LR) Algorithm [33], and the Support Vector Machine (SVM) [31]. These machine learning algorithms were selected due to their widespread use and performance in a range of classification tasks. Every algorithm has advantages and disadvantages, and the objective was to evaluate how well the suggested neutrosophic model performed in contrast. The suggested neutrosophic model outperformed the other machine learning algorithms in terms of classification accuracy, according to the results shown in Table 2. The proposed combination resulted in a 3% increase in the accuracy of cyberbullying classification.

The neurosophic model outperformed the other algorithms in terms of accuracy for a few reasons: The Neutrosophic model integrates the ideas of indeterminacy-membership and falsity-membership, which enables it to handle uncertain and conflicting information more skillfully. Complex relationship modeling: Cyberbullying fine-grained classification may incorporate intricate patterns and relationships within the data. These intricate relationships may have been better captured and modeled by the neutrosophic model in conjunction with the aforementioned machine learning algorithms, improving classification accuracy. By taking into account various membership levels, the neutrosophic model is better able to differentiate between the various forms of cyberbullying speech. The model can identify and categorize instances that may have conflicting or uncertain characteristics by incorporating indeterminacy-membership and falsity-membership, which improves accuracy.

Furthermore, SVM, known for its robustness in linear classification tasks, may falter when confronted with the nonlinear intricacies inherent in cyberbullying text data. Conversely,ourmodel,harnessing the power of OvO strategy and MLP classifiers, excels in capturing nonlinear patterns and subtlelinguisticcues,leading to more accurate

TABLE3. Comparisonaccuracywithdifferentactivationfunctions.

Activation Function	Accuracy
ReLU	95
Tanh	92
Sigmoid	

classification results. Similarly, even though RF is good at identifying intricate relationships in data, it can be swayed by majority class bias in datasets that are unbalanced. On the other hand, our model's use of interval-neutral sets guarantees robust decision-making and balanced representation across all forms of cyberbullying, thereby reducing the drawbacks of RF. Furthermore, LR may not be able to capture complex feature relationships due to its linear nature, despite its efficiency and simplicity. On the other hand, our model's incorporation of MLP classifiers allows it to learn complex patterns, successfully circumventing the drawbacks of LR and attaining greater accuracy. In order to overcome the drawbacks of conventional machine learning algorithms, the suggested model combines OvOstrategy, MLPclassifiers, and interval neutrosophic sets, exhibiting superior performance in fine-grained cyberbullying classification tasks.

C. EXPERIMENT3:THEEFFECTOFUSINGDIFFERENT ACTIVATIONFUNCTIONFORMLPUSING CYBERBULLYINGDATASET

This set of experiments was performed to compare the accuracy of the proposed model that employs different activation functions like ReLU (RectifiedLinearUnit), Sigmoid(Logistic), and Tanh (HyperbolicTangent)[34]. The results shown inTable3 revealed that the use of ReLU activation function improve of 3% for the same method with other activation functions. The performance improvement comes from many factors; non-saturation of gradients: Unlike activation functions like sigmoid or tanh, ReLU does not saturate in the positiveregion, allowing the gradient to flow smoothly during back propagation. This facilitates faster convergence during training. Also, Sigmoid and functions suffer from vanishing gradient problems for extremely large or small input values, which can hinder learning in deeper networks. ReLU helps mitigate this issue.

D. EXPERIMENT 4: COMPARISON BETWEEN THE PROPOSEDMODELANDFUZZYSETSUSING CYBERBULLYINGDATASET

This set of experiments was performed to compare the efficiency of the proposed model and fuzzy logic in thefield of fine-grained cyberbullying classification. We apply threshold = 0.7 for fuzzy classification. The results pre-sentedinTable4confirmedthattheproposedneutrosophic Model out performed the fuzzy logic interms of classification accuracy. The suggested model achieved a 2% increase in accuratelyclassifyingcyberbullyingtypes.The justification of this result is that fuzzy sets usea single membership Grade to handle uncertainty, while neutrosophicsets use three independent membership grades (truth, indeterminacy, and falsity) to provide a more comprehensive representation of uncertainty, especially insituations where truthand falsity are not mutually exclusive and there is room for indeterminacy or ambiguity.

TABLE4. Evaluationresultsoffine-grainedcyberbullyingclassification afterusingfuzzysets.

Class	Precision	Recall	F1
age	0.98	0.97	0.98
ethnicity	0.99	0.97	0.98
gender	0.90	0.87	0.89
Other_cyberbullying	0.82	0.89	0.85
religion	0.99	0.96	0.98
Accuracy			0.93

TABLE5. Evaluationresultsoffine-grainedcyberbullyingclassification after using Bert.

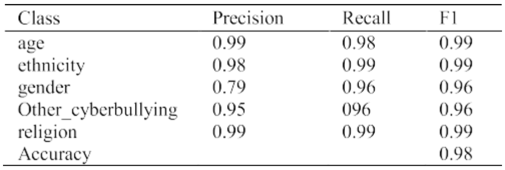
Class	Precision	Recall	F1
age	0.99	0.98	0.98
ethnicity	0.98	0.99	0.98
gender	0.90	0.92	0.91
Other_cyberbullying	0.89	0.88	0.88
religion	0.99	0.97	0.98
Accuracy			0.96

E. EXPERIMENT5:BERTINTEGRATI ONIN PREPROCESSINGFORTHEPROPOSE D MODEL USING CYBERBULLYING DATASET

This experiment was conducted to explore the effectiveness of integrating BERT [15](Bidirectional Encoder Represen- tations from Transformers) into the preprocessing pipeline for cyberbullying detection. BERT, known for its excep- tional language understanding capabilities, was incorporated to enhance contextualanalys is of speech and capturenuanced changes in key word meanings, there by improving the overall detection accuracy. We applied BERT as part of the text preprocessing step before feeding the data into the classifi- cationmodel. BERT was utilized to tokenize and encode the input text data, ensuring that the semantic context and word meanings were preserved effectively.

Table 5confirm that the results of this experiment have NLP techniques in this domain. The results of this experiment were indeed promising. The integration of BERT led to a noticeable increase in classification accuracy compared to previous experiments. This improvement can be attributed to severalfactors,includingBERT'sability to capture semantic representations of text, its contextual understanding of language nuances, and its robustness to noise and variations in languageusage.Overall,theinclusion ofBERT in the preprocessing pipeline represents asignificant enhancement to the proposedmodel's performance, aligning

TABLE6. Evaluationresultsoffine-grainedcyberbullyingclassification afterusingdataaugmentation.



Class	Precision	Recall	F1
age	0.99	0.98	0.99
ethnicity	0.98	0.99	0.99
gender	0.79	0.96	0.96
Other_cyberbullying	0.95	096	0.96
religion	0.99	0.99	0.99
Accuracy			0.98

F. EXPERIMENT 6: ENHANCING CYBERBULLYING DETECTIONTHROUGHDATAAUGMENTATION

The purpose of this experiment was to evaluate how data augmentation methods affected the effectiveness of cyberbullying detection models. A common strategy to overcome dataset inadequacy is data augmentation [35], particularly in situations when the size of the existing dataset may not be adequate for extensive analysis. We supplemented the initial cyberbullying dataset using a variety of data augmentation methods. These methods included replacing terms in the text samples with synonyms, adding words at random, and deleting words at random. In order to build a bigger, augmented dataset for training and evaluation, the augmented dataset was then combined with the original dataset. Table 6 demonstrates that the experiment's outcomes showed that the data augmentation strategies considerably enhanced the cyberbullying detection model's performance. The classification accuracy increased from 95% to 98% with the enhanced dataset, demonstrating how well data augmentation works to address dataset insufficiency problems.

The improvement in results after applying data augmenta- tion techniques can be attributed to three key factors. Firstly, the augmented dataset size provided the model with a more extensive and diversese to examples to learn from, enhancing its ability to generalize and capture complex patterns. Secondly, by introducing variations in the training data, the model was exposed to a wider range of linguistic scenar- ios, leading to improved generalization to unseen instances. Lastly, data augmentation helped address class imbalance issues by generating additional samples, ensuring a more balanced representation of minority classes, and thereby improving overall predictive performance.

VI. CONCLUSIONANDFUTUREWORK

This paper suggested an accurate model for fine-grained cyberbullying classification. The proposed model uses the integration of NL within the MLP classification model and offers an innovative approach toward handling fine-grained classification scenarios. NL allows the representation of uncertainty, ambiguity, and indeterminacy within classifi- cation decisions, thereby enhancing the model's ability to handle complex classification tasks where clear bound-aries between classes might be lacking. In this work, we successfully incorporated the principles of Neutrosophic Logic through the utilization of the one-against-onestrategy in the training phase. The model, built upon a series of binary MLP classifiers, each discriminating between specific pairs of classes, effectively captured the intricate relationships and nuances between different classes. This approach acknowl- edges and accounts for potential overlapping or ambiguous instances, addressing the common challenge of intricateclass boundaries in fine-grained classification tasks. During the testing phase, the significance of the Neutrosophic concept became further pronounced. The predictions from multiple one-against-one classifiers collectively provided a compre- hensive insight into classification outcomes. The extracted dominant class from the Neutrosophic class probabilities showcased the adaptability of the model in handling com- plex classification scenarios. The results, as evidenced in the comparative analys is of the accuracy between the traditional MLP and the Neutrosophic-empowered MLP, demonstrated the utility and potential performance enhancements offered by incorporating Neutrosophic Logic in the classification process. The model, leveraging Neutrosophic Logic, stands as a flexible and comprehensive solution for fine-grained classification tasks, fostering deeper understanding of intricate boundaries and relationships between different classes. Future work includes using different languages like Arabic and utilizingGPU with deep learning techniques to discover and enhance the model accuracy. We have also planned to explore the integration of Large Language Models (LLMs) in our future work.

REFERENCES

[1] J. R. W. Yarbrough, K. Sell, A. Weiss, and L. R. Salazar, ''Cyberbul-lyingandthefacultyvictimexperience:Perceptionsandoutcomes,'' *Int. J. Bullying Prevention*, vol. 5, no. 2, pp.1–5, Jun. 2023, doi:10.1007/s42380-023-00173-x.

[2] A.Bussu,S.-A.Ashton,M.Pulina,andM.Mangiarulo,''Anexplo-rative qualitative study of cyberbullying and cyberstalking in a highereducationcommunity,''*CrimePreventionCommunitySaf.*,vol.25,no.4, pp.359–385,Oct.2023,doi:10.1057/s41300-023-00186-0.

[3] A. K. Jain, S. R. Sahoo, and J. Kaubiyal, ''Online social networkssecurity and privacy: Comprehensive review and analysis,'' *ComplexIntell. Syst.*, vol. 7, no. 5, pp. 2157–2177, Oct. 2021, doi: 10.1007/s40747-021-00409-7.

[4] G.Fulantelli,D.Taibi,L.Scifo,V.Schwarze,andS. C. Eimler,‘‘Cyberbullyingandcyberhateastwointerlinkedinstancesofcyber-aggressioninadolescence:Asystematicreview,’’*FrontiersPsychol.*,vol.13,May2022,Art. no. 909299, doi: 10.3389/fpsyg.2022.909299.

[5] M.S.JahanandM.Oussalah,‘‘Asystematicreviewof hate speech automatic detection using natural language processing,’’*Neurocomputing*, vol. 546, Aug. 2023, Art.no.126232, doi:10.1016/j.neucom.2023.126232.

[6] G.Kovács,P.Alonso,andR.Saini,‘‘Challengesofhate speechdetectionin socialmedia,’’*SocialNetw.Comput.Sci.*,vol.2,no.2,pp.1–15, Feb.2021,doi:10.1007/s42979-021-00457-3.

[7] M.ShyamsunderandK.S.Rao,‘‘ClassificationofLP Iradarsignals usingmultilayerperceptron(MLP)neuralnetworks,’’in*Proc.ICASPACE*,Singapore, Dec. 2022, pp. 233–248.

[8] F.M.Plaza-del-Arco,D.Nozza,andD.Hovy,‘‘Respectfulortoxic?Using zero-shot learning with language models to detect hate speech,’’ in *Proc.7th WOAH*, Toronto, ON, Canada, Jul. 2023, pp. 60–68.

[9] V.ChristiantoandF.Smarandache,‘‘Areviewofseven applicationsof neutrosophiclogic:Inculturalpsychology,economicstheorizing,conflictresolution,philosophyofscience,etc.’’*J.Multidiscip.Res.*,vol.2,no.2, pp.128–137,Mar.2019,doi:10.3390/j2020010.

[10] F. Smarandache, ‘‘Neutrosophic logic—A generalization of the intuition-istic fuzzy logic,’’ *SSRN Electron. J.*, vol. 4, p.396, Jan. 2016, doi:10.2139/ssrn.2721587.

[11] S.Das,B.K.Roy,M.B.Kar,S.Kar,andD.Pamučar,‘‘Neutrosophicfuzzy set and its application in decision making,’’ *J. Ambient Intell. HumanizedComput.*,vol.11,no.11,pp.5017–5029,Mar.2020,doi:10.1007/s12652-

[12] R. Zhao and K. Mao, ‘‘Cyberbullying detection based on semantic-enhanced marginalized denoising auto-encoder,’’ *IEEE Trans. Affect.Comput.*, vol. 8, no. 3, pp. 328–339, Jul. 2017.

[13] M.Alzaqebah,G.M.Jaradat,D.Nassan,R.Alnasser,M.K.Alsmadi, I. Almarashdeh,S.Jawarneh,M.Alwohaibi,N.A.Al-Mulla,N.Alshehab,and S. Alkhushayni, ‘‘Cyberbullying detection framework for short andimbalancedArabicdatasets,’’*J.KingSaudUniv.Comput.Inf.Sci.*,vol.35,no. 8, Sep. 2023, Art. no. 101652, doi: 10.1016/j.jksuci.2023.101652.

[14] L.J.Thun,P.L.Teh,andC.-B.Cheng,‘‘CyberAid:Areyourchildrensafe fromcyberbullying?’’ *J.King Saud Univ.Comput. Inf. Sci.*,vol. 34, no.7, pp.4099–4108,Jul.2022,doi:10.1016/j.jksuci.2021.03.001.

[15] A. Muneer, A. Alwadain, M. G. Ragab, and A. Alqushaibi, ‘‘Cyberbul-lying detection on social media using stacking ensemble learning andenhanced BERT,’’ *Information*, vol. 14, no. 8, p.467, Aug. 2023, doi:10.3390/info14080467.

[16] D.Sultan,A.Toktarova,A.Zhumadillayeva,S.Aldeshov, S.Mussiraliyeva,G. Beissenova,A.Tursynbayev,G.Baenova,and A.Imanbayeva,‘‘Cyberbullying-relatedhatespeechdetectionusingshallow-to-deeplearning,’’*Comput.,Mater.Continua*,vol.74,no.1, pp.2115–2131,Apr.2023,doi:10.32604/cmc.2023.032993.

[17] C. R. Sedano, E. L. Ursini, and P. S. Martins, ‘‘A bullying-severityidentifierframeworkbasedonmachinelearningandfuzzylogic,’’in*ArtificialIntelligenceandSoftComputing*,vol.10245,1sted.Cham,Switzerland:Springer,2017,pp. 315–324,doi:10.1007/978-3-319-59063-9_28.

[18] F.Smarandache,M.Ali,andM.Khan,‘‘Arithmeticoperationsof neutrosophic sets, interval neutrosophic sets and rough neutrosophicsets,’’ in *Fuzzy Multi-criteria Decision-Making Using Neutrosophic Sets*,vol. 3, 1st ed. Cham, Switzerland: Springer, 2019, ch. 2, pp.25–42, doi:10.1007/978-3-030-00045-5_2.

[19] I.Irvanizam and N.Zahara, ‘‘An extendedEDAS based onmulti-attribute group decision making to evaluate mathematics teachers with single-valued trapezoidal neutrosophic numbers,’’ in *Handbook of Research onthe Applications of Neutrosophic Sets Theory and Their Extensions inEducation*,S.Broumi,Ed.Hershey,PA,USA:IGIGlobal,Jun.2023, pp.40–67,doi:10.4018/978-1-6684-7836-3.ch003.

[20] A. Abdelhafeez, H. K. Mohamed, A. Maher, and N. A. Khalil, ‘‘A novelapproachtowardskincancerclassificationthroughfuseddeepfeaturesandneutrosophic environment,’’ *Frontiers Public Health*, vol. 11, pp. 1–15,Apr. 2023, doi: 10.3389/fpubh.2023.1123581.

[21] G.KaurandH.Garg,‘‘Anewmethodforimageprocessingusinggeneral-ized linguistic neutrosophic cubic aggregation operator,’’ *Complex Intell.Syst.*,vol.8,no.6,pp.4911–4937,Dec.2022,doi:10.1007/s40747-022-00718-5.

[22] J.Wang,K.Fu,andC.-T.Lu,‘‘SOSNet:Agraphconvolutionalnetwork approach to fine-grained cyberbullying detection,’’ in *Proc.IEEEInt.Conf.BigData(BigData)*,Atlanta,GA,USA, Dec.2020, pp.1699–1708.

[23] S.Kang,S.Cho,andP.Kang,‘‘Constructingamulti-classclassifierusingone-against-oneapproachwithdifferentbinaryclassifiers,’’*Neurocomput-ing*,vol.149,pp. 677–682,Feb.2015,doi:10.1016/j.neucom.2014.08.006.

[24] W.A.SilvaandS.M.Villela,‘‘Improvingtheone-against-allbinary approachformulticlassclassificationusingbalancingtechniques,’’*Int. J. Speech Technol.*, vol. 51, no. 1, pp.396–415, Aug. 2020, doi:10.1007/s10489-020-01805-1.

[25] W.Wang,L.Feng,Y.Jiang,G.Niu,M.-L.Zhang,andM. Sugiyama, ‘‘Binary classification with confidence difference,’’ 2023, *arXiv:2310.05632*.

[26] J. Ma, T. Li, X. Li, S. Zhou, C. Ma, D. Wei, and K. Dai, ‘‘A probabilityprediction method for the classification of surrounding rock quality oftunnelswithincompletedatausingBayesiannetworks,’’*Sci.Rep.*,vol.12,no. 1, p. 19846, Nov. 2022, doi: 10.1038/s41598-022-19301-6.

[27] R.Essameldin,A.A.Ismail,andS.M.Darwish,‘‘Anopinionmining approach to handle perspectivism and ambiguity: Moving

toward neu-trosophic logic,'' *IEEE Access*, vol. 10, pp.63314–63328, 2022, doi:10.1109/ACCESS.2022.3183108.

[28] H.Wang,P.Madiraju,Y.Zhang,andR.Sunderraman,''Intervalneutro-sophicsets,''2004,.

[29] M.Ahmadinejad,N.Shahriar,L.Fan.(2023).*ABalancedMulti-Labeled Dataset for Cyberbully Detection in Social Media*. [Online].Available: https://www.kaggle.com/datasets/momo12341234/cyberbully-detection-dataset/data

[30] D.Elreedy,A.F.Atiya,andF.Kamalov,''Atheoreticaldistribution analysis of synthetic minority oversampling technique (SMOTE) forimbalanced learning,'' *Mach. Learn.*, vol. 2023, pp.1–21, Jan. 2023, doi:10.1007/s10994-022-06296-4.

[31] N.Novalita,A.Herdiani,I.Lukmana,andD.Puspandari,''Cyberbul-lying identification on Twitter using random forest classifier,'' *J. Phys.,Conf. Ser.*, vol. 1192, Mar. 2019, Art. no. 012029, Art.no.012029, doi:10.1088/1742-6596/1192/1/012029.

[32] J.M.Ortiz-Marcos,M.Tomé-Fernández,andC.Fernández-Leyva, ''Cyberbullying analysis in intercultural educational environments usingbinarylogisticregressions,''*FutureInternet*,vol.13,no.1,p. 15,Jan. 2021, doi: 10.3390/fi13010015.

[33] A.Apicella,F.Donnarumma,F.Isgrò,andR.Prevete,''Asurveyon moderntrainableactivationfunctions,''*NeuralNetw.*,vol.138,pp. 14–32,Jun. 2021, doi: 10.1016/j.neunet.2021.01.026.

[34] A.MumuniandF.Mumuni,''Dataaugmentation:Acomprehensivesurvey of modern approaches,'' *Array*, vol. 16, Dec. 2022, Art.no.100258, doi:10.1016/j.array.2022.100258.