Unlocking Emotions in Text: A Comprehensive Study of Computational Linguistics and Natural Language Processing Techniques

Dr. Priyanka Jibhau Bachhav*

Research Scholar, Department of English, Savitribai Phule Pune University

Abstract - The widespread use of digital communication has led to the creation of vast amounts of written content every day. Understanding the emotions expressed in this data is crucial for the development of emotionally intelligent computer systems. This article presents the development of a novel approach, known as CLBEDC-SND, which aims to recognise and classify Social networking data containing sentiments. Numerous phases of data pre-processing were performed on the CLBEDC-SND model to guarantee its suitability for subsequent processing. The CLBEDC-SND model undergoes sentiment scoring and vectorization utilising a fuzzy methodology. Vectorization is the conversion of textual information into a vector representation, which may be used with the ELM model. The ELM model was utilised by the CLBEDC-SND model to provide accurate and timely emotion categorization. The procedure selects the input weight in a random manner and computes the output weight of the Single Layer Feedforward Network (SLFN) using empirical data. Subsequently, the starting weight of the input and the bias of the hidden layer are chosen at random. Using the SFLO method, the parameters of the ELM model are optimised in the final stage. The study data demonstrated that the CLBEDC-SND method consistently yielded enhanced outcomes across all domains. A comprehensive comparison analysis will be done to improve the precision of emotion categorization results produced from the CLBEDC-SND model. The experimental evaluations have demonstrated that the CLBEDC-SND model outperforms alternative models in its ability to classify emotions. The enhanced performance of the CLBEDC-SND model can be ascribed to the integration of sentiment scoring based on fuzzy logic and the optimum parameter modification strategy based on SFLO. Thus, the proposed model is applicable.

Keywords - Emotions, Feelings, Social Networking Data, Vectorization, Single Layer Feedforward Network (SLFN), Digital Communication, Emotionally Intelligent Computer Systems

- x -

INTRODUCTION

The proliferation of digital communication has resulted in the generation of immense quantities of written information on a daily basis. Comprehending the sentiments conveyed in this data is vital for developing emotionally intelligent computer systems. This research explores the techniques used to teach computers in recognizing and understanding emotions expressed in written language. The advent of AI in 1950 marked a profound transformation in the global landscape. Artificial intelligence (AI) experienced a significant resurgence during the 20th century, inspiring scholars to conduct extensive investigations across multiple domains such as deep learning, computer vision, natural language processing (NLP), and computer vision. On the other hand, the domains of Natural Language Processing (NLP) are not well-defined, primarily because computational and linguistic approaches are used in the field. These methodologies enable computational systems to comprehend and produce human-computer dialogues expressed through written and spoken text. The purpose of this endeavour is to develop a model that can be used to a variety of processes, including perception, sentiment, beliefs, and emotions. The emotional tone of a particular text may be determined by sentiment analysis, which further categorises the content as either positive, negative, or neutral.

Emotion analysis, on the other hand, goes beyond that and is carried out by classifying the categories that fall under the umbrella of sentiment investigation. Although the application of the learning-based technique has been restricted, the keyword-based and lexical affinity approaches have been widely implemented. Despite this, the shortcomings of these methods compromise their efficacy and lead to reduced accuracy. The strategies that machine learning and deep learning use to categorise feelings are distinct from one

another. In this study, we have merged the datasets of three distinct categories, namely sentences, tweets, and dialogs, in order to get insights on three diverse variants. The sentences were in their original form; thus we have preprocessed the data to improve the usability of the text sentences. Subsequently, input the data into several machine learning (ML) and deep learning (DL) models. Ekman, 1999, classifies emotions into six distinct categories: pleasure, sorrow, fear, surprise, anger, and disgust. In addition, emotions may be categorized into several types, including love, optimism, and others, as seen in Figure 1. Facial expressions, gestures, voice, and writing are often used to convey the emotional state and mood of an individual. Unlike face expression and voice recognition, a text phrase lacks the capacity to convey its meaning via sensory perception since it lacks flavor. Due to the intricate and unclear nature of the writing, determining the emotions conveyed within it is a challenging endeavor.
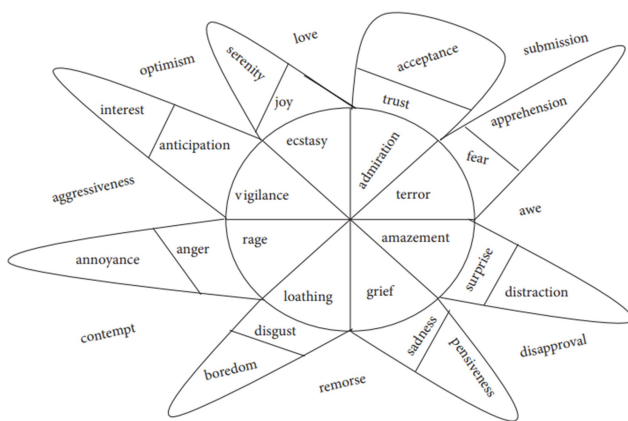

Figure 1: Various Kinds of Emotions (Plutchik, 2001)

Identifying the mood of a particular text may be challenging due to the varying meanings and morphological forms of each word. More recently, scholars have proposed a variety of methods to determine the feelings that are communicated via written language. These methods include learning-based models, hybrid models, lexical affinity models, and keyword-based models (Chopade, 2015). They initially employed a rule-based methodology comprised of two distinct strategies: keyword-based and lexical affinity-based. Conversely, an innovative methodology emerged under the name learning-based approach. The strategy exhibited more accuracy and yielded superior outcomes. Various models are used in a learning-based method to identify emotion. Several academics have been integrating several methodologies to create hybrid methods in order to achieve greater accuracy (Seal et al., 2020). According to the research, deep learning models demonstrate superior accuracy compared to machine learning models when dealing with vast amounts of text or data. However, when dealing with tiny datasets, machine learning algorithms provide us with higher levels of precision. However, none of the methods

provided a comprehensive answer for accurately identifying the emotional content of a particular text (Hasan et al., 2019).

There were many constraints in the current solutions, notably the absence of a comprehensive inventory of all emotions. The existing lists do not have an adequate vocabulary, they do not take into account certain terms, they do not have a context that is based on semantics, they only extract a limited amount of contextual information from phrases, and they do not successfully recognise certain emotions (Singh et al., 2021). In addition to this, they have poor context information extraction, a lack of semantic feature extraction, a sluggish computational speed, a disdain for feature relations, an inadequate amount of data, and they produce a significant proportion of incorrect classifications (Alnuaim et al., 2022). Certain models were ill-suited for frequently appearing emoji, inadequate extraction of semantic information, and the sentence structure. The variation exists across different models. Previous researchers have addressed several drawbacks in this technology. The suggested approach has successfully addressed many of the current restrictions (Rodriguez et al., 2022). Emotion detection is a significant benefit of human-machine interaction, since it enables nonliving entities to perceive or experience emotions similar to those of a human being. Our suggested algorithm can accurately identify emotions from text phrases that lack any discernible tone or expression (Singh et al., 2021).

In the sphere of linguistics, computational linguistics has arisen as a major discipline that has become more important. When it comes to computational linguistics, the computational methodologies that are used originate from either computer science or artificial intelligence (AI). In spite of this, the major objective of computational linguistics will continue to be the representation and simulation of human language, so firmly establishing its position within the sphere of humans. Computational linguistics focuses on the construction of language models that enable computers to better understand human language. It goes beyond studying language usage in human actions and also involves implementing specific formal techniques for accurately constructing hypotheses and evaluating them automatically using linguistic data (corpora). AI approaches are relied upon in the formal aspect of computational linguistics (Ahire & Borse, 2022).

Using a combination of computer science and linguistics, the area of sentiment analysis (SA) is a relatively new subject that aims to automatically detect the sentiment that is communicated in written text. Sentiment may be broken down into two categories: negative and positive assessments that are communicated via words. According to Mustakim et al. (2022), the most suitable sources for subjective text were social networking site posts. These sites have become more popular as a result of the many methodological approaches that are accessible for

doing sentiment analysis. Evaluation of the views made regarding prominent individuals or current events and categorization of such opinions as either positive or negative are the procedures that are used in these methods. Sentiment analysis may be used to classify text by expressing a favourable or unfavourable sentiment. Emotion recognition in text might be seen as a potential answer for these challenges. This approach does not focus on the bad or good thoughts voiced, but instead aims to identify the underlying human emotion being conveyed (Zad et al., 2021). Identifying the emotions expressed by a person may be a challenging endeavour that even people may find tough. Developing an identification method and implementing an automated system to identify expressed emotion is a difficult task. This challenge arises not only because to the lack of easily accessible training data, but also because of the restricted amount of data available in a brief text (Graterol et al., 2021). The selection of emotional categories was not only based on psychologists' study on emotion theory. Many systems modify the emotion categories based on their discoveries, either by dividing or combining emotions, without regard for scientific standards. Conversely, several accessible systems get their training data by identifying certain terms. By simply using keyword-based strategies, it is impossible to verify the accuracy of the classifier methods. The classifier is primarily taught to identify these specific keywords and categorize the text accordingly (Sailunaz et al., 2018).

In this article, a system known as CLBEDC-SND is presented. This system makes use of computational linguistics in order to recognise and categorise emotions that are present in social networking data. The CLBEDC-SND method goes through a number of different layers of data pre-processing in order to guarantee compatibility with subsequent processing measures. Furthermore, sentiment scoring and vectorization are implemented on the CLBEDC-SND model through the utilisation of the fuzzy approach. When it comes to the classification of feelings, the CLBEDC-SND model makes use of extreme learning machine (ELM). Ultimately, the Shuffled Frog Leaping Optimisation (SFLO) method is implemented to modify the ELM model's parameters in an efficient manner. Through the use of the benchmark dataset for the purpose of performance validation, it is possible to evaluate the effectiveness of the CLBEDC-SND technique.

METHODOLOGY

➢ The development of an innovative CLBEDC-SND method for identifying and classifying emotions in social networking data is described in this article. Numerous phases of data pre-processing were performed on the CLBEDC-SND model to guarantee its suitability for subsequent processing. Furthermore, vectorization and sentiment scoring are implemented for the CLBEDC-SND model via a fuzzy technique. In conclusion, the SFLO with

ELM model is implemented to identify and classify emotions. The entire operational procedure of the CLBEDC-SND algorithm under discussion is depicted in Figure 1.
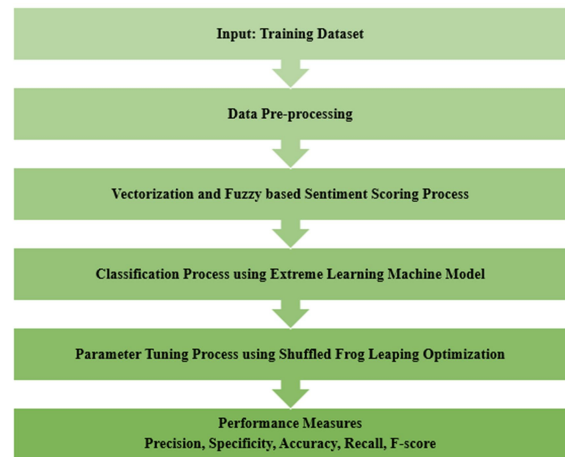

Figure 2: The CLBEDC-SND approach's operational procedure

➢ **Data Pre-Processing:** Numerous phases of data pre-processing were performed on the CLBEDC-SND model to guarantee its suitability for subsequent processing. Within social media data, there are often undesirable components that need to be removed at the pre-processing step. This removal helps enhance the overall performance. The research involves six specific processes at this stage, which include halting word removal, tokenization, punctuation removal, URL removal, and lemmatization. The pre-processing function may be executed using mathematical notation:

$$p_r = \lambda_p[I_t]$$

The output of the pre-processing function is denoted as pr in Equation (1). The input dataset is shown, and λ_p represents the pre-processing function as follows:

$$\lambda_p = \left[\lambda_{tk}, \lambda_{pr}, \lambda_{sr}, \lambda_{lm}, \lambda_{ur} \right]$$

"λ_{tk} represents the tokenization function, λ_{pr} signifies the punctuation removal function, λ_{sr} depicts the stop word removal function, λ refers to the lemmatization function, and λ_{ur} indicates the removal of URLs."

The verification of the partial dataset has been completed. However, even a little portion of missing data might lead to inaccurate prediction of emotions in the text. Addressing these concerns is a necessary step. At this point, they inspect the pre-processed dataset to determine whether or not it contains any datasets that are either irrelevant or incomplete. Whenever there is a dataset that is

Dr. Priyanka Jibhau Bachhav*

missing some information, it is replaced with a dataset that is both comprehensive and relevant, which properly summarises the information and elicits a significant amount of emotion.

- **Sentiment Scoring and Vectorization:** During this step, the CLBEDC-SND model is subjected to vectorization and sentiment scoring using a fuzzy technique. Vectorization involves converting the text information into a vector format to be used with the ELM model. In order to achieve these objectives, the Gensim module in Python is used to employ a word embedding method. This algorithm is capable of capturing context, learning word relations, and determining syntactic and semantic similarity between words in a given document. Word2Vec is used to convert the text dataset into vectors, which may then serve as the main weights for these models. Fuzzy sentiment scoring involves determining the level of sentiment expressed in a text by identifying opinionated words and linguistic hedges using a Part-Of-Speech tagger. This is followed by the application of WordNet and SentiWordNet dictionaries to extract the sentiment degree.

- **Classification of Emotion:** The CLBEDC-SND model used the ELM model to provide precise and prompt emotion categorization. The algorithm selects the input weight in a random manner and determines the output weight of SLFN based on empirical observations. Subsequently, the input weight and hidden layer bias are randomly chosen. The SLFN is considered to be a linear approach, and the output weight of the SLFN is determined experimentally using a generalized inverse function of the hidden layer output matrix. This model exhibits a rapid rate of learning in comparison to traditional feedforward network learning methods, while also achieving strong generalization performance. Furthermore, ELM consistently attains the minimum norm of weights and the least amount of training error. Figure 3 depicts the ELM structure.
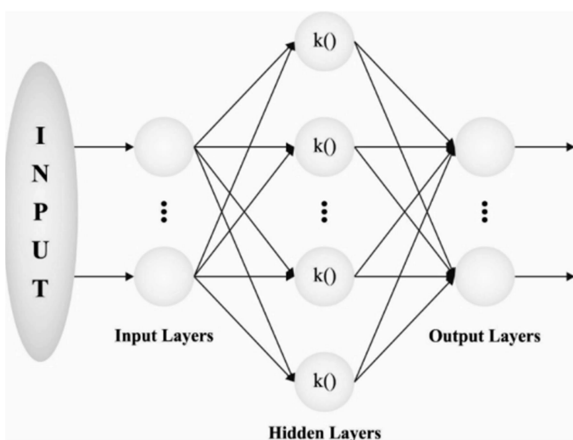

Figure 3: Layout f ELM

- **Hyper-parameter Tuning:** The parameters of the ELM model are optimized in the last step using the

SFLO method. The SFLO algorithm emulates the sub-population coevolution process seen in species of frogs as they search for food places. It combines both random and deterministic approaches, resulting in a very efficient processing power and global search capability. The wetland's frog population may be categorized into several subpopulations, each with its own unique culture. A local search strategy is used to achieve local optimization within each of these subpopulations. Initially, local search is applied to all the subpopulations. This involves performing an upgrade function on individual frogs within each sub-population that have the lowest adaptive values.

RESULTS AND DISCUSSION

- Table 1 and Figure 4 provide the comprehensive results of the CLBEDC-SND approach in classifying emotions throughout the full dataset. The trial data demonstrated that the CLBEDC-SND approach has consistently yielded enhanced outcomes across all categories.

Table 1: Analysis of the CLBEDC-SND approach's results using unique class labels for the whole dataset

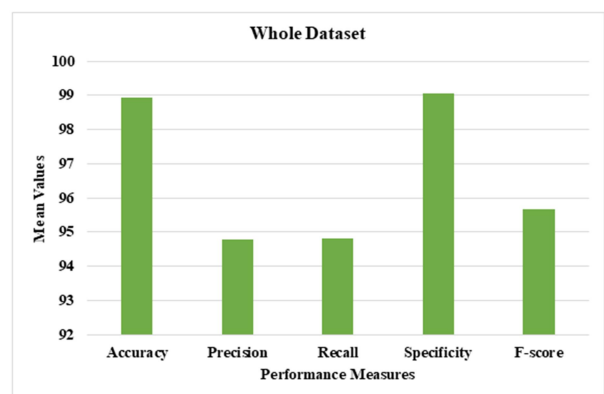| WHOLE DATASET					
Labels	Accuracy	Precision	Recall	Specificity	F-score
ANGER	99.02	95.89	96.05	99.66	94.48
DISGUST	99.76	94.77	94.77	99.56	95.15
FEAR	98.97	93.65	95.67	98.97	94.75
GUILT	98.66	93.23	93.25	99.32	96.14
JOY	98.67	94.67	94.12	98.96	96.60
SAD	98.45	96.98	95.80	98.44	96.70
SHAME	98.97	94.34	94.11	98.53	95.88
AVERAGE	98.92	94.79	94.82	99.06	95.67

Figure 4: Result analysis of CLBEDC-SND

approach under entire dataset

➢ Table 2 and Figure 5 present a comprehensive comparative analysis with the aim of providing more reliable results regarding the classification of emotions utilising the CLBEDC-SND model.

Table 2: Comparison between the current algorithm and the CLBEDC-SND method

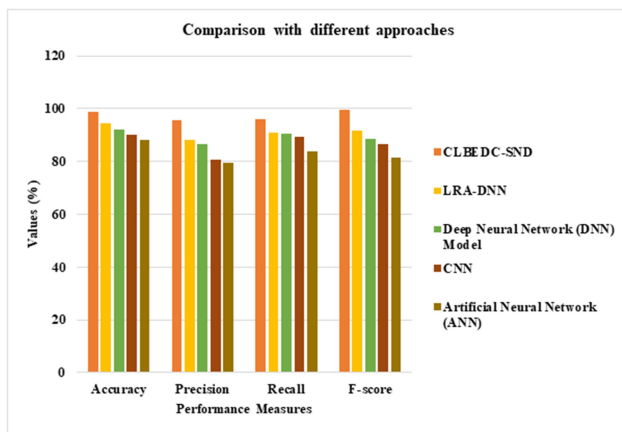| Approaches	Accuracy	Precision	Recall	F-score
Artificial Neural Network (ANN)	88.07	79.67	83.92	81.32
CLBEDC-SND	98.65	95.49	95.92	99.54
CNN	89.94	80.47	89.44	86.42
Deep Neural Network (DNN) Model	92.09	86.57	90.45	88.43
LRA-DNN	94.48	87.95	91.05	91.55

Figure 5: Comparison between the current algorithm and the CLBEDC-SND method

➢ Based on the experimental evaluations, it has been determined that the CLBEDC-SND model outperformed alternative models in its ability to classify emotions. The improved performance of the CLBEDC-SND model could potentially be attributed to the incorporation of SFLO-based optimal parameter tuning and fuzzy-based sentiment scoring. Both of these procedures were implemented. Therefore, the model that was given may be used.

CONCLUSION

As a conclusion, the increased quantity of textual material in this era of digital communication brings to light the need of incorporating emotional intelligence into computer systems. The CLBEDC-SND technique, which is a novel approach to identifying and classifying feelings in social networking data, is discussed in this article. Both of these techniques are presented in this study. After the model has been exposed to a

painstaking process of data pre-processing, it is then submitted to vectorization and emotion scoring via the use of a fuzzy computing approach. Through the use of the ELM model, the CLBEDC-SND model displays findings that are consistent and enhanced in the categorization of emotions across a number of different categories. The results of comparative studies demonstrate that it does better than other approaches in the categorization of emotions. This accomplishment may be due to the use of fuzzy-based sentiment score as well as the SFLO-based optimization of parameters. The findings of this research represent a significant contribution to the advancement of computer systems that are capable of comprehending and reacting to emotions. It offers a powerful paradigm for more accurately classifying feelings in written language because to its presentation.

REFERENCES

1. Acheampong, F. A., Wenyu, C., & Nunoo-Mensah, H. (2020). Text-based emotion detection: Advances, challenges, and opportunities. *Engineering Reports*, *2*(7), e12189.

2. Ahire, V., & Borse, S. (2022). Emotion detection from social media using machine learning techniques: a survey. In *Applied Information Processing Systems: Proceedings of ICCET 2021* (pp. 83-92). Springer Singapore.

3. Alnuaim, A. A., Zakariah, M., Shukla, P. K., Alhadlaq, A., Hatamleh, W. A., Tarazi, H., ... & Ratna, R. (2022). Human-computer interaction for recognizing speech emotions using multilayer perceptron classifier. *Journal of Healthcare Engineering*, *2022*.

4. Azam, N., Tahir, B., & Mehmood, M. A. (2020). Sentiment and emotion analysis of text: A survey on approaches and resources. *Language & Technology*, *87*.

5. Bharti, S. K., Varadhaganapathy, S., Gupta, R. K., Shukla, P. K., Bouye, M., Hingaa, S. K., & Mahmoud, A. (2022). Text-Based Emotion Recognition Using Deep Learning Approach. *Computational Intelligence and Neuroscience*, *2022*.

6. Cao, L., Peng, S., Yin, P., Zhou, Y., Yang, A., & Li, X. (2020, December). A survey of emotion analysis in text based on deep learning. In *2020 IEEE 8th International Conference on Smart City and Informatization (iSCI)* (pp. 81-88). IEEE.

7. Chopade, C. R. (2015). Text based emotion recognition: A survey. *International journal of science and research*, *4*(6), 409-414.

8. Ekman, P. (1999). Basic emotions. *Handbook of cognition and emotion*, *98*(45-60), 16.

9. Ghazi, D., Inkpen, D., & Szpakowicz, S. (2015). Detecting emotion stimuli in emotion-

Dr. Priyanka Jibhau Bachhav*

bearing sentences. In *Computational Linguistics and Intelligent Text Processing: 16th International Conference, CICLing 2015, Cairo, Egypt, April 14-20, 2015, Proceedings, Part II 16* (pp. 152-165). Springer International Publishing.

10. Graterol, W., Diaz-Amado, J., Cardinale, Y., Dongo, I., Lopes-Silva, E., & Santos-Libarino, C. (2021). Emotion detection for social robots based on NLP transformers and an emotion ontology. *Sensors*, *21*(4), 1322.

11. Hasan, M., Rundensteiner, E., & Agu, E. (2019). Automatic emotion detection in text streams by analyzing twitter data. *International Journal of Data Science and Analytics*, *7*, 35-51.

12. Mohammad, S. M., & Bravo-Marquez, F. (2017). WASSA-2017 shared task on emotion intensity. *arXiv preprint arXiv:1708.03700*.

13. Mustakim, N., Rabu, R., Mursalin, G. M., Hossain, E., Sharif, O., & Hoque, M. M. (2022, May). CUET-NLP@ TamilNLP-ACL2022: Multi-Class Textual Emotion Detection from Social Media using Transformer. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages* (pp. 199-206).

14. Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, *11*(1), 81.

15. Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, *89*(4), 344-350.

16. Rodriguez, A., Chen, Y. L., & Argueta, C. (2022). FADOHS: framework for detection and integration of unstructured data of hate speech on facebook using sentiment and emotion analysis. *IEEE Access*, *10*, 22400-22419.

17. Sailunaz, K., Dhaliwal, M., Rokne, J., & Alhajj, R. (2018). Emotion detection from text and speech: a survey. *Social Network Analysis and Mining*, *8*, 1-26.

18. Seal, D., Roy, U. K., & Basak, R. (2020). Sentence-level emotion detection from text based on semantic rules. In *Information and Communication Technology for Sustainable Development: Proceedings of ICT4SD 2018* (pp. 423-430). Springer Singapore.

19. Singh, D., Kumar, V., Kaur, M., Jabarulla, M. Y., & Lee, H. N. (2021). Screening of COVID-19 suspected subjects using multi-crossover genetic algorithm based dense convolutional neural network. *IEEE Access*, *9*, 142566-142580.

20. Suhasini, M., & Srinivasu, B. (2020). Emotion detection framework for twitter data using supervised classifiers. In *Data Engineering and Communication Technology: Proceedings of 3rd ICDECT-2K19* (pp. 565-576). Springer Singapore.

21. Zad, S., Heidari, M., James Jr, H., & Uzuner, O. (2021, May). Emotion detection of textual data: An interdisciplinary survey. In *2021 IEEE World AI IoT Congress (AIIoT)* (pp. 0255-0261). IEEE.

Corresponding Author

Dr. Priyanka Jibhau Bachhav[1]*

Research Scholar, Department of English, Savitribai Phule Pune University