2022 25th International Conference on Computer and Information Technology (ICCIT)
17-19 December, Cox's Bazar, Bangladesh

An Integrated Embedded System Towards Abusive Bengali Speech and Speaker Detection Using NLP and Deep Learning

Syed Taha Yeasin Ramadan[1], Tanjim Sakib[2], Md. Ahsan Rahat[3], Md. Mushfique Hossain[4],
Raiyan Rahman[5], Md. Mahbubur Rahman[6]
Department of Computer Science and Engineering
Military Institute of Science and Technology (MIST), Dhaka-1216, Bangladesh
Email: tahayeasin11@gmail.com, tsakib77@gmail.com, ahsanrahat11@gmail.com, mushfique3214@gmail.com,
raiyan.cse@gmail.com, mahbub@cse.mist.ac.bd

2022 25th International Conference on Computer and Information Technology (ICCIT) | 979-8-3503-4602-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCIT57492.2022.10054785

***Abstract*— Intelligible speech, while it provides an excellent means of communication for humans and sets us apart from other lifeforms, our abuse of speech creates deep and lasting issues in our society. The use of derogatory language has a significant impact not only on children's mental health but also on adults, for instance, in an abusive work environment. Accountability for such actions is one of the key steps toward maintaining a healthy atmosphere or at least making it less frequent. In this paper, we describe our work on detecting abusive or hate speech in Bangla in real time. Our system converts the speech to text and then uses NLP and deep learning to detect such occurrences in real-time. Also, if the voice is registered on our system, it identifies the person engaging in abusive words, opening ways to greater workplace accountability. We also describe our mobile application and the microcontroller-based standalone embedded system that can be deployed in target places (for instance, daycare centers, schools, workplaces, etc.) to record audio and detect the abusive speech and the speaker in real-time. Several datasets have been deployed on the LSTM, Bi-LSTM, GRU, and BERT models to assess the system's efficacy. Identification of the individual speaking the words is done using the audio signal extraction feature MFCC. The experimental results show that the BERT model provides the highest accuracy compared to other algorithms.**

***Index Terms*—BERT, Bi-LSTM, Deep Learning, GRU, LSTM, MFCC, Natural Language Processing, Speech Detection, Voice Recognition**

I. Introduction

Abuse is a pattern of conduct demonstrating inappropriate or severe use of power to enforce, obtain, and uphold authority and control that violates and harms the victim. Abuse and Violence are health risks among Children. It would have the tendency to affect persistent and crippling mental health concerns such maladaptive behaviors, anxiety disorders, and relationship or personality issues [1].

Children are protected and stabilized by a secure family, school, and community environment. The growth of a well-adapted child depends on this. Early childhood is a crucial time when children are most vulnerable to exposure to traumatic events. An adverse childhood experience that is characterized by child maltreatment has enormous immediate and long-term repercussions. Beyond death, physical injury, and disability, abuse, and violence can lead to toxic stress that impairs well-being from childhood to adulthood in all aspects of life [2].

Higher-ranking employees frequently use abusive language in the workplace because they believe it is smart and reflects their superiority. In certain offices, it has become so common that no one questions it either. Additionally, changing the workplace setting is a difficult task. Therefore, even if some employees are not happy with this environment and frequently feel unpleasant, it is typically overlooked. Therefore, this issue may benefit from a monitoring system as a solution.

However, much of the work on abusive words has been done in text fields, which are usually used to detect abuse on social networking sites in different languages. Many machines and deep learning models have been used for detecting abusive and hateful words and comments and taking further action after that. Previously users had to manually detect and block the account of these users which have been automated using these methods [3] [4] [5]. These are carried out using datasets for several languages [6] [7] [8] [9]. The Bangla language also reflects this. Many methods have been explored to detect and eliminate abusive Bengali words from social media and other internet platforms [10] [11] [12] [13]. However, there is a shortage of work on real-time word detection from voice recognition in Bengali, which is the area that has been explored here. Another point is to identify the person using those words that were not required in text-based detection because the person has a separate account from which he can be identified. An approach to recognizing the person can be storing the voice of each individual in a company or class at the start and then matching with that voice pattern if an abusive word is detected. A speech feature can be extracted from an audio source using Mel-frequency cepstral coefficients (MFCC) [14].

To detect abusive words in a real-time setting, a deep learning-based model has been suggested. This model can identify the speaker as well as the abusive word that has been said by that person. Here, voice input from the workplace has been used to determine whether or not a word is abusive. If it determines that the sentence's use of that word is unsuitable in its context, it flags the sentence and adds it to the database along with the speaker's name. A hardware device is used to process the voice received by microphones set in different places in a closed environment. Speech separation is used to separate the voices of multiple people. Mel-frequency cepstral coefficients (MFCC) are used to extract features from the person's audio input, and a Deep Learning (DL) model is then used to identify the person.

The same word might be abusive and not abusive in different situations, making it difficult to identify abusive Bengali words. Normal English translation is insufficient to determine context because it differs significantly from one language to another. Therefore, datasets from the Bangla language have been used to train a variety of deep learning models, enabling them to detect these words more accurately. Models like LSTM, BI-LSTM, GRU, and BERT have been used to achieve the desired result. The experiment showed that the BERT (Bidirectional Encoder Representations from Transformers) model correctly detects abusive words with the highest accuracy compared to other models.

The rest of the paper is organized as follows: Section II presents the related works in the relevant field and demonstrates the present scenario. While III describes the features, conceptual design and implementation along with the work flow of our proposed approach. The over all results has been added in Section IV. Finally, Section V concludes the article.

II. Literature Review And Present System

Finding offensive and hateful words is not a novel concept. For years, it has been done utilizing a variety of approaches. It was primarily conducted to lessen the growing cyberbullying and hate speech directed at celebrities on online social media platforms. The strategy involves detecting specific hostile comments and taking appropriate action on that basis. Social media is now a lot safer to use than it was in the past due to numerous enhanced algorithms that have been utilized to successfully detect context-based words in many online platforms over time [15] [3] [16] [4] [5]. Later, it was used in several online games as well as other platforms to create a safe and pleasant environment for all users [17].

These methods weren't just confined to the English language because there are many other languages in the world. Various languages have been used over time to detect abusive language. To identify hateful words on Indonesian social networking sites, Ibrohim MO and Budi I. constructed an Indonesian abusive words dataset and suggested a machine learning-based technique [6]. Similar methods are adopted in the same language to lessen toxicity in online comment sections [18]. Surzhyk is a hybrid language that combines Russian and Ukrainian words. A unique dataset in that language is developed and then used to detect abusive phrases on Russian-Ukrainian social networks [7]. It was done in Hindi in conjunction with English using six different data sets [19]. Similar experiments were conducted in other languages such as Urdu, Arabic and Tamil [20] [21] [22] [23] [24]. The goal is to identify racist remarks in French on Twitter, where immigrants and Muslims frequently receive hostile

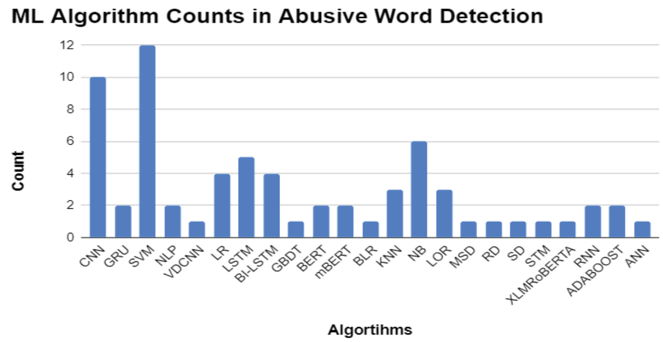

Fig. 1: Machine Learning Algorithm Counts in Abusive Word Detection

comments. This platform successfully creates a new dataset for eventual use in abusive word detection [25]. Plaza-Del-Arco FM, Molina-González MD, Ureña-López LA, Martín-Valdivia MT. [26] proposed a model in the Spanish language which successfully detects racial abuse and hurtful comments made to women on social platforms.

As one of the most widely spoken languages in the world, Bangla has been used extensively in this subject. MG Hussain, T Al Mahmud, and W Akthar developed a fairly sophisticated algorithm for identifying the abusive language in Bengali [10]. Later, utilizing data from Facebook, a lot more extensive effort is performed in which pre-trained techniques such as BERT and ELECTRA are utilized to recognize Bangla abusive words [11]. English and Bangla are frequently mixed on social media sites, and it is currently fairly widespread. To identify offensive terms in Bangla, Bangla-English, and transliterated Bangla, a machine learning-based detection model is developed [27]. Other similar studies using other machine and deep learning models have been conducted on Bangla datasets [28] [29].

One major deficiency in this sector is the lack of real-time voice recognition technology that can identify the offensive language. An automatic method that can do this in broadcasting channels and censor those terms was proposed by Endah SN, Nugraheni DM, and Adhy S [30]. The same, however, has not been discovered for Bengali, where our method can contribute to the detection of abusive words from audio signals and determine whether they are abusive or not. Another consideration is identifying the speaker from the audio that has been received. It is required to keep people accountable for their actions. Since it creates coefficients from the user's voice that are specific to each user, the Mel Frequency Cepstral Coefficients (MFCC) algorithm is typically preferred as a feature extraction technique for voice recognition [31]. This technique has been applied in this case to identify the individual.

In summary, numerous studies have been conducted in the area of identifying offensive or abusive words in a variety of languages, including Bangla as well. Several Algorithms have been used in identifying offensive or abusive words (see Figure 1) and machine learning and deep learning has been

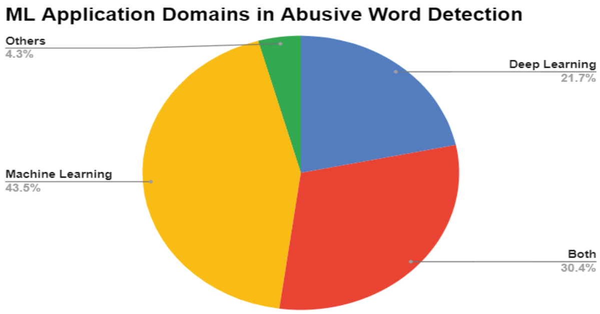

Fig. 2: Machine Learning Domains in Abusive Word Detection

largely used for identifying speech (see Figure 2). However, the absence of an automated system that can identify the same for voice or audio is extremely uncommon. As a result, in a real-time setting, our system can identify it from audio signals. Additionally, it will identify the speaker of the words.

III. Proposed System

In this section, we have discussed how our proposed system is designed and developed. We have represented this with four sequential steps that include- (A) Experimental Design, (B) Work-flow Design, (C) System Architecture and (D) Development of the system.

A. Experimental Design

The experimental design (see Figure 3) of the system has been defined based on the user requirements. By collecting the speech from the observed area through hardware device, speech has been processed for two different functionalities, one is for classifying the speech and another is for speaker recognition. Dataset for classifying speech [32] [33] [34] [35] has been classified into two categories, abusive and non-abusive.

In the data preprocessing step all Non-Bengali letters have been removed as only the Bengali language has been considered. All punctuations, digits, and emoticons are removed. Stemming is an approach to finding the word to its root word and stop words are regularly used in languages. Stemming and removing stop words also has been done in data preprocessing. For word embedding, the Word2Vec model has been used using the gensim module. After that LSTM, Bi-LSTM, GRU, and pre-trained BERT model using ktrain library has been used for training and building the model and these models have been deployed into the hardware device. Speaker dataset [36] has been used for speaker identification and extracted features from audio data using MFCC (Mel-frequency Cepstrum Coefficients), and train and build deep learning model for speaker identification then deployed the trained model into the hardware device. The hardware device sends the notification to the software application if abusive speech is detected.

B. Work-flow Design

The work-flow diagram (see Figure 4) depicts the overall process of the system. Firstly, two deep learning model

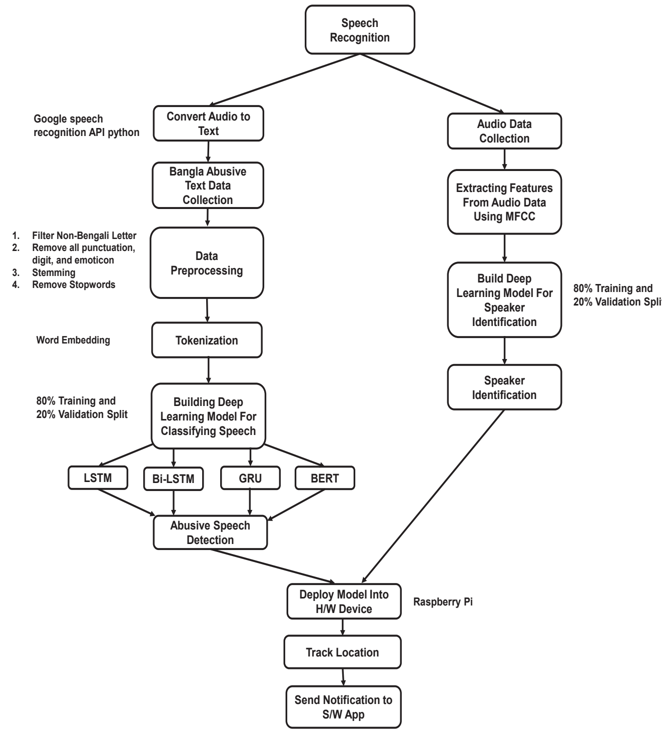

Fig. 3: Experimental design of the proposed system

has been designed and developed for classifying speech and speaker identification. Then collect the voice data from the observed area. If abusive speech is detected then the speaker's voice is identified by using the deep learning model of speaker identification. Then abusive speech, speaker name, date, time, and location are sent to the Firebase real-time database and monitors can access this information through the mobile application.

C. System Architecture

The proposed system is suggesting a microphone for collecting the speech from the observed area or person. All information will be stored and can be accessed from the mobile application that will be used by the monitor or concerned person. This system will ensure whether abusive speech is being used or not in the observed area and identifies the speaker and sends the notification to the mobile application that will be accessible by the monitor. Firebase real-time database has been used for sending the information of the observer area. In mobile applications, users can filter data by using several options that have been designed for user convenience to get desired data quickly. The overview of the system architecture is presented in Figure 5.

D. Development of the System

The suggested system consists of both software and hardware components.

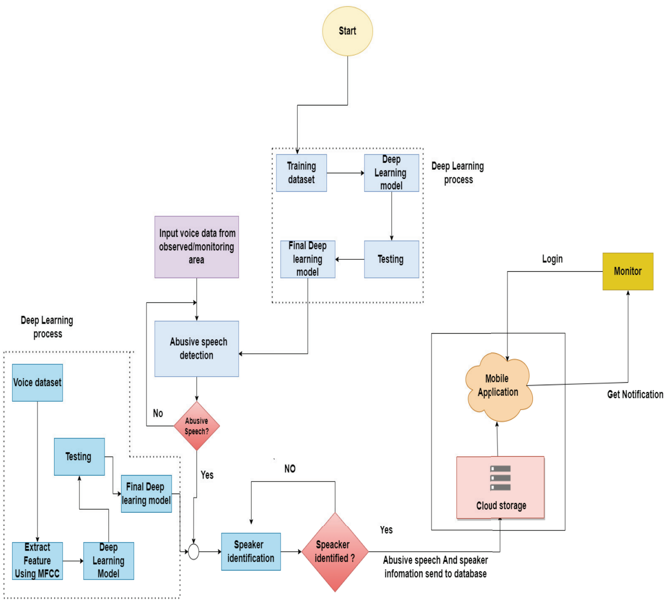

Fig. 4: Work-flow diagram of the initial system

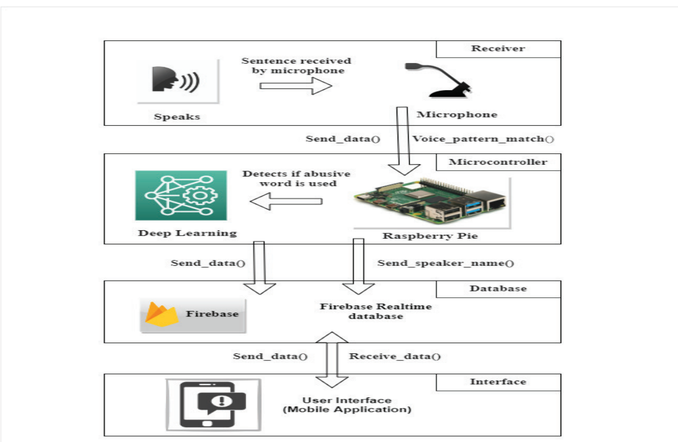

Fig. 5: An overview of the system architecture

1) *Hardware Development:* The microphone is integrated with the raspberry pi 4 microcontroller. First, the deep learning model for classifying speech and speaker identification has been trained in google colab, and built models are deployed into the microcontroller. BERT model gives the best accuracy. Since BERT model is computationally expensive LSTM model is deployed into the microcontroller. Deep learning model for speaker identification has also been deployed into the microcontroller to identify the speaker if abusive speech is detected. Table I shows the overall hardware implementation part.

2) *Software Development:* Making the entire system accessible and usable for the monitor is the primary goal of the mobile application. React native is used for building the software application for both Android and iOS mobile devices. The software application is connected to the Firebase real-time database and gets the necessary information from the hardware device through Firebase real-time database. Table 2 shows the overall software implementation part.

TABLE I: Hardware Implementation

Device Used	Device Type	Application
Microprocessor	Raspberry pie	Deep learning model is deployed here
Sensor	GPS module	Track location
Sensor	Microphone	Receive audio signal as input

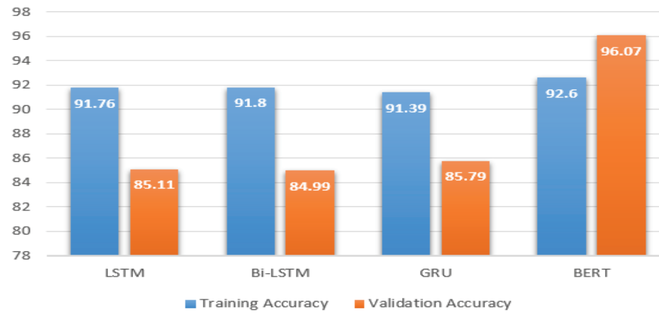

Fig. 6: Accuracy of Bangla Abusive Speech Detection

TABLE II: Software Implementation

Software	Framework	Application
User Application	React native js	Mobile application to check any actions
Database	Firebase	Storing data concerning abusive words

IV. Result Analysis

In the accuracy and usability evaluation segment of our work, we evaluated each segment of our work separately. Firstly, Table III shows the comparative accuracy of the deep learning models that we used to detect abusive Bengali words. As wee see, Pre-trained BERT model gives the best accuracy 92.6% training accuracy and 96.07% validation accuracy which is the highest accuracy among other models (see Figure 6).

TABLE III: Accuracy of The Bangla Abusive Word Detection

Deep Learning Algorithm	Training Accuracy(%)	Validation Accuracy(%)
LSTM	91.76	85.11
Bi-LSTM	91.8	84.99
GRU	91.39	85.79
BERT	92.6	96.07

TABLE IV: Comparative Accuracy of The Bangla Abusive Word Detection

Deep Learning Algorithm	Other Papers Accuracy(%)	Our Accuracy(%)
LSTM	80	85.11
Bi-LSTM	79	84.99
GRU	83	85.79
BERT	84	96.07

TABLE V: Accuracy of Speaker Identification

Training Accuracy (%)	Validation Accuracy (%)
93.36	92.05

We also compare the performance of our implementation with similar previous work in table IV, Paper [29] got 80% accuracy in LSTM and 83% accuracy in GRU model and paper [13] got 79% accuracy in Bi-LSTM model. Paper [37] used mBERT(cased) got 84% accuracy. Whereas in our approach, LSTM and Bi-LSTM give around 85.11% and 84.99% validation accuracy respectively and the GRU model gives 85.79% validation accuracy which is an improvement over other papers' models.

As for the abusive speaker identification portion of our system through voice recognition, we achieved a 93.36% and 92.05% training and validation accuracy respectively, which is shown in table V.

Table VI shows the system usability evaluation of our companion mobile application. 10 people were asked to complete 3 central tasks using the app. Mean and standard deviation were calculated for 3 parameters (number of attempts, task completion time and number of times asking for help) for each of the tasks and all of them show satisfactory results.

V. Conclusion

Our system will contribute to an environment for people who face verbal abuse or onslaught. Verbal abuse is a huge concern for parents and also in an office environment where everyone wants to be harmonious with each other but sometimes in these environments, people use abusive speech to humiliate someone. To thwart this our system will help to classify abusive speech and for that, LSTM, Bi-LSTM, GRU, and BERT models have been used. BERT gives better accuracy than other models with a validation accuracy of 96.07%. Speaker is also identified by a deep learning model and deployed these models into a microcontroller and it sends the information to the mobile application which will be accessible by the monitor or concerned person so that they can be aware of such situations and take necessary steps to prevent these people.

TABLE VI: Results of the evaluation study (system usability)

Task	Number of Attempts (M ± SD)	Task Completion time (second) (M ± SD)	Number of times Asking Help (M ± SD)
T1: Find the latest 5 abusive speech that has been detected.	1.375 ±1.875	10.875 ± 38.875	0.5 ±1.875
T2: Find the list of abusive speech of a specific person that has been detected. (e.g. list of abusive speech of Mushfique)	1.125 ±0.875	13.625 ± 7.875	0.167± 0.875
T3: Find the list of abusive speech on a specific date. (e.g. list of abusive speech on 5 September , 2022)	1.25 ± 1.5	17.375 ± 23.875	0.333 ± 1.5

References

[1] M. L. A. de Vera, "Long term effects of abuse and violence on children's behavior," *Am J Biomed Sci & Res*, vol. 4, pp. 347–53, 2019.

[2] J. Brown, P. Cohen, J. G. Johnson, and E. M. Smailes, "Childhood abuse and neglect: specificity of effects on adolescent and young adult depression and suicidality," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 38, no. 12, pp. 1490–1496, 1999.

[3] C. Nobata, J. Tetreault, A. Thomas, Y. Mehdad, and Y. Chang, "Abusive language detection in online user content," in *Proceedings of the 25th International Conference on World Wide Web*, ser. WWW '16. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2016, p. 145–153. [Online]. Available: https://doi.org/10.1145/2872427.2883062

[4] J. H. Park and P. Fung, "One-step and two-step classification for abusive language detection on twitter," *arXiv preprint arXiv:1706.01206*, 2017.

[5] U. Naseem, I. Razzak, and I. A. Hameed, "Deep context-aware embedding for abusive and hate speech detection on twitter." *Aust. J. Intell. Inf. Process. Syst.*, vol. 15, no. 3, pp. 69–76, 2019.

[6] M. O. Ibrohim and I. Budi, "A dataset and preliminaries study for abusive language detection in indonesian social media," *Procedia Computer Science*, vol. 135, pp. 222–229, 2018.

[7] B. Andrusyak, M. Rimel, and R. Kern, "Detection of abusive speech for mixed sociolects of russian and ukrainian languages." in *RASLAN*, 2018, pp. 77–84.

[8] N. U. Haq, M. Ullah, R. Khan, A. Ahmad, A. Almogren, B. Hayat, and B. Shafi, "Usad: an intelligent system for slang and abusive text detection in perso-arabic-scripted urdu," *Complexity*, vol. 2020, 2020.

[9] R. Stanković, J. Mitrović, D. Jokić, and C. Krstev, "Multi-word expressions for abusive speech detection in serbian," in *Proceedings of the Joint Workshop on Multiword Expressions and Electronic Lexicons*, 2020, pp. 74–84.

[10] M. G. Hussain, T. Al Mahmud, and W. Akthar, "An approach to detect abusive bangla text," in *2018 International Conference on Innovation in Engineering and Technology (ICIET)*. IEEE, 2018, pp. 1–5.

[11] T. T. Aurpa, R. Sadik, and M. S. Ahmed, "Abusive bangla comments detection on facebook using transformer-based deep learning models," *Social Network Analysis and Mining*, vol. 12, no. 1, pp. 1–14, 2022.

[12] T. Islam, N. Ahmed, and S. Latif, "An evolutionary approach to com-

parative analysis of detecting bangla abusive text," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 4, pp. 2163–2169, 2021.
[13] M. T. Ahmed, M. Rahman, S. Nur, A. Islam, and D. Das, "Deployment of machine learning and deep learning algorithms in detecting cyberbullying in bangla and romanized bangla text: A comparative study," in *2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*. IEEE, 2021, pp. 1–10.
[14] A. Bala, A. Kumar, and N. Birla, "Voice command recognition system based on mfcc and dtw," *International Journal of Engineering Science and Technology*, vol. 2, no. 12, pp. 7335–7342, 2010.
[15] H.-S. Lee, H.-R. Lee, J.-U. Park, and Y.-S. Han, "An abusive text detection system based on enhanced abusive and non-abusive word lists," *Decision Support Systems*, vol. 113, pp. 22–31, 2018.
[16] S. B. Bodapati, S. Gella, K. Bhattacharjee, and Y. Al-Onaizan, "Neural word decomposition models for abusive language detection," *arXiv preprint arXiv:1910.01043*, 2019.
[17] A. Ekiciler, İ. Ahioğlu, N. Yıldırım, İ. Ajas, and T. Kaya, "The bullying game: Sexism based toxic language analysis on online games chat logs by text mining," in *Conference on Gender Studies and Sexuality*, 2014.
[18] D. R. K. Desrul and A. Romadhony, "Abusive language detection on indonesian online news comments," in *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*. IEEE, 2019, pp. 320–325.
[19] N. Vashistha and A. Zubiaga, "Online multilingual hate speech detection: experimenting with hindi and english social media," *Information*, vol. 12, no. 1, p. 5, 2020.
[20] M. Humayoun, "Abusive and threatening language detection in urdu using supervised machine learning and feature combinations," *arXiv preprint arXiv:2204.03062*, 2022.
[21] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. Abdelmajeed, and M. T. Sadiq, "Automatic detection of offensive language for urdu and roman urdu," *IEEE Access*, vol. 8, pp. 91 213–91 226, 2020.
[22] R. Rajalakshmi, A. Duraphe, and A. Shibani, "Dlrg@ dravidianlangtech-acl2022: Abusive comment detection in tamil using multilingual transformer models," in *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*, 2022, pp. 207–213.
[23] H. Mubarak, K. Darwish, and W. Magdy, "Abusive language detection on arabic social media," in *Proceedings of the first workshop on abusive language online*, 2017, pp. 52–56.
[24] E. A. Abozinadah, A. V. Mbaziira, and J. Jones, "Detection of abusive accounts with arabic tweets," *Int. J. Knowl. Eng.-IACSIT*, vol. 1, no. 2, pp. 113–119, 2015.
[25] N. Vanetik and E. Mimoun, "Detection of racist language in french tweets. information 2022, 13, 318," 2022.
[26] F.-M. Plaza-Del-Arco, M. D. Molina-González, L. A. Ureña-López, and M. T. Martín-Valdivia, "Detecting misogyny and xenophobia in spanish tweets using language technologies," *ACM Transactions on Internet Technology (TOIT)*, vol. 20, no. 2, pp. 1–19, 2020.
[27] M. Jahan, I. Ahamed, M. R. Bishwas, and S. Shatabda, "Abusive comments detection in bangla-english code-mixed and transliterated text," in *2019 2nd International Conference on Innovation in Engineering and Technology (ICIET)*. IEEE, 2019, pp. 1–6.
[28] E. A. Emon, S. Rahman, J. Banarjee, A. K. Das, and T. Mittra, "A deep learning approach to detect abusive bengali text," in *2019 7th International Conference on Smart Computing & Communications (ICSCC)*. IEEE, 2019, pp. 1–5.
[29] A. K. Das, A. Al Asif, A. Paul, and M. N. Hossain, "Bangla hate speech detection on social media using attention-based recurrent neural network," *Journal of Intelligent Systems*, vol. 30, no. 1, pp. 578–591, 2021.
[30] S. Endah, D. Nugraheni, S. Adhy *et al.*, "The automation system censor speech for the indonesian rude swear words based on support vector machine and pitch analysis," in *IOP Conference Series: Materials Science and Engineering*, vol. 190, no. 1. IOP Publishing, 2017, p. 012039.
[31] K. Chakraborty, A. Talele, and S. Upadhya, "Voice recognition using mfcc algorithm," *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, vol. 1, no. 10, pp. 2349–2163, 2014.
[32] M. R. Karim, B. R. Chakravarti, J. P. McCrae, and M. Cochez, "Classification benchmarks for under-resourced bengali language based on multichannel convolutional-lstm network," in *7th IEEE International Conference on Data Science and Advanced Analytics (IEEE DSAA,2020)*. IEEE, 2020.
[33] M. R. Karim, S. K. Dey, T. Islam, S. Sarker, M. H. Menon, K. Hossain, M. A. Hossain, and S. Decker, "Deephateexplainer: Explainable hate speech detection in under-resourced bengali language," in *2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2021, pp. 1–10.
[34] N. Romim, M. Ahmed, H. Talukder, S. Islam *et al.*, "Hate speech detection in the bengali language: A dataset and its baseline evaluation," in *Proceedings of International Joint Conference on Advances in Computational Intelligence*. Springer, 2021, pp. 457–468.
[35] aimansnigdha, "aimansnigdha/bangla-abusive-comment-dataset." [Online]. Available: https://github.com/aimansnigdha/Bangla-Abusive-Comment-Dataset
[36] M. A. Islam and A.-N. Sakib, "Bangla dataset and mmfcc in text-dependent speaker identification," *Engineering and Applied Science Research*, vol. 46, no. 1, pp. 56–63, 2019.
[37] M. F. Mridha, M. A. H. Wadud, M. A. Hamid, M. M. Monowar, M. Abdullah-Al-Wadud, and A. Alamri, "L-boost: Identifying offensive texts from social media post in bengali," *Ieee Access*, vol. 9, pp. 164 681–164 699, 2021.