International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net

Vol-20 Issue-01 April 2025

Multi-Modal Hate Speech Recognition

S. Shreya[1], Shaik Owais Ali[2] G. Durga Rao[3] , K. Venkat Reddy[4]
[1,2,3] UG Scholar, Dept of IT, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
[4]Assistant Professor, Dept of IT, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
shreyasallas@gmail.com

Abstract:

Because of the speedy surge in social media's expansion, the proliferation of malicious and harmful poses a substantial worry in contemporary society. The identification of hate speech on platforms like Twitter is crucial for various tasks such as controversial event extraction, AI chatterbot creation, content suggestions, and sentiment analysis. Researches have invested considerable effort in addressing the challenging task of identifying hostile content due to the rise in hate speech and harmful information. The objective is to classify tweets as Hateful, Offensive, or neither. However, this task is highly complex due to the intricate nature of natural language constructs, encompassing different manifestations of animosity directed at various demographics, and the multitude of ways the same meaning can be expressed. Previous research has predominantly relied on manual feature extraction or employed representation-learning techniques followed by linear classifiers. Nevertheless, deep learning methods have recently demonstrated significant accuracy improvements in complex problems across speech, vision, and text applications. In this study, This paper present an idea for automatic classifications of inappropriate language and expressions of hostility using transfer learning models. In this research, leverage classified tweet datasets obtained from Kaggle and conduct experiments. Findings reveal that the multilingual- BERT model and its pre-trained version deliver superior outcomes. Specifically, pre-trained BERT model notably improves classification accuracy of hateful tweets by up to 92% when compared to other algorithms.

Keywords: AI chatterbot creation , Sentiment analysis,

Hostile, intricate, deep learning methods, transfer learning models, Kaggle, multilingual-BERT model, pre-trained BERT model.

1. INTRODUCTION

An alarming global increase in racism and intolerance is being witnessed. In 2018 alone, the popularity of memes resulted in an astonishing 180 million meme posts across various social media channels [11]. Within this digital landscape, the alarming rise of Hate Speech (HS) has emerged as a pressing societal concern. Hate Speech is defined as explicit attacks on individuals grounded in attributes as racism, ethnicity, nationality, religiously, gender, or other fundamental characteristics The pervasiveness of HS on digital platforms has prompted major technology companies like Facebook, boasting millions of daily active users, to undertake substantial measures to safeguard their user base. Hate speech can take many forms, encompassing verbal, written, or nonverbal communication, all aiming the targeting every single individual or in groups on their intrinsic traits as religion, ethnicity, nationality or color. To counteract the escalating presence of the application of machine learning methods to address online hate speech, particularly DL methodologies, has become

imperative. Given the widespread adoption of social media platforms as conduits for bigotry, the real-time detection and filtration of hate speech have become indispensable.

While hate speech can manifest in a multitude of formats on social media, including text, voice, photos, and videos, research efforts have predominantly centered on language-based techniques. These encompass a spectrum of Natural Language Processing (NLP) approaches, ranging from neural networks [1], [2], [3] and n-grams
[4], to graph-based models [5, 6, 7]. However, a notable gap exists in comprehensive investigations into the analysis of multimedia data. This essay presents an innovative technique that harnesses the power of a multimodal DL architecture for toxic speech categorization. Leveraging the potency BERT and additional Transformer-based models encoder architectures, which have exhibited remarkable success across diverse NLP tasks, this approach generates vector-space representations of Natural language conducive to deep learning models. In our endeavor to capture the unmoral language and speech embedding's, we turn to pre-trained models, recognizing the constraints imposed by the dataset's limited size. To process these embedding's, we employ two distinct downstream Architectures – CNN and MLP. To facilitate the computation to create vector-space representations for our hate speech dataset, we depend on the pre-trained BERT. BERT's distinctive architecture, effectively incorporating contextual information from both preceding and succeeding content across all its layers, empowers the retraining of Profound bidirectional representations from unannotated text.

In light of the escalating challenge posed by hate speeches the digital age, this study addresses a critical research problem: the development of effective techniques for the real- time detecting and classifications of toxic speech across various forms of media. Through the integration of multimodal deep learning methods, we aim to pave the way for more comprehensive and robust hate speech

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net

Vol-20 Issue-01 April 2025

recognition solutions.

2. LITERATURE SERVEY

S. Narejo, "Generalized Epileptic Seizure Prediction using," (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 14, pp. 502-510, 2023. In recent years, the electroencephalography (EEG) signal identification of epileptic seizures has developed into a routine procedure to determine epilepsy. Since physically identifying epileptic seizures by expert neurologists becomes a labor-intensive, time-consuming procedure that also produces several errors. Thus, efficient, and computerized detection of epileptic seizures is required. The disordered brain function that causes epileptic seizures can have an impact on a patient's condition. Epileptic seizures can be prevented by medicine with great success if they are predicted before they start. Electroencephalogram (EEG) signals are utilized to predict epileptic seizures by using machine learning algorithms and complex computational methodologies. Furthermore, two significant challenges that affect both expectancy time and genuine positive forecast rate are feature extraction from EEG signals and noise removal from EEG signals. As a result, we suggest a model that offers trustworthy preprocessing and feature extraction techniques. To automatically identify epileptic seizures, a variety of ensemble learning-based classifiers were utilized to extract frequency-based features from the EEG signal. Our algorithm offers a higher true positive rate and diagnoses epileptic episodes with enough foresight before they begin. On the scalp EEG CHB-MIT dataset on 24 subjects, this suggested framework detects the beginning of the preictal state, the state that occurs before a few minutes of the onset of the detention, resulting in an elevated true positive rate of (91%) than conventional methods and an optimum estimation time of
33 minutes and an average time of prediction is 23 minutes and 36 seconds. Depending on the experimental findings' The maximum accuracy, sensitivity, and specificity rates in this research were 91 %, 98%, and 84%.

Kulsoom, F., Narejo, S., Mehmood, Z. et al. A review of machine learning-based human activity recognition for diverse applications. Neural Comput & Applic 34, 18289–18324 (2022). https://doi.org/10.1007/s00521-022-07665-9.There is an ever-present need to objectively measure and analyze sports motion for the determination of correct patterns of motion for skill execution. Developments in performance analysis technologies such as inertial measuring units (IMUs) have resulted in enormous data generation. However, these advances present challenges in analysis, interpretation, and transformation of data into useful information. Artificial intelligence (AI) systems can process and analyze large amounts of data quickly and efficiently through classification techniques. This study aimed to systematically review the literature on Machine Learning (ML) and Deep Learning (DL) methods applied to IMU data inputs for evaluating techniques or skills in individual swing and team sports. Electronic database searches (IEEE Xplore, PubMed, Scopus, and Google Scholar) were conducted and aligned with the PRISMA statement and guidelines.

International Conference on Emerging Technologies in Electronics, Computing and Communication (ICETECC), Jamshoro, Sindh,Pakistan, 2022, pp. 1-5, doi:10.1109/ICETECC56662.2022.10069760.Because of the speedy surge in social media's expansion, the proliferation of malicious and harmful poses a substantial worry in contemporary society. The identification of hate speech on platforms like Twitter is crucial for various tasks such as controversial event extraction, AI chatterbot creation, content suggestions, and sentiment analysis. Researchers have invested considerable effort in addressing the challenging task of identifying hostile content due to the rise in hate speech and harmful information. The objective is to classify tweets as Hateful, Offensive, or neither. However, this task is highly complex due to the intricate nature of natural language constructs, encompassing different manifestations of animosity directed at various demographics, and the multitude of ways the same meaning can be expressed.

Butt, A., Narejo, S., Anjum, M.R. et al. Fall Detection Using LSTM and Transfer Learning. Falls represent a significant cause of injury among the elderly population. Extensive research has been devoted to the utilization of wearable IMU sensors in conjunction with machine learning techniques for fall detection. To address the challenge of acquiring costly training data, this paper presents a novel method that generates a substantial volume of synthetic IMU data with minimal actual fall experiments. First, unmarked 3D motion capture technology is employed to reconstruct human movements. Subsequently, utilizing the biomechanical simulation platform Opensim and forward kinematic methods, an ample amount of training data from various body segments can be custom generated. Synthetic IMU data was then used to train a machine learning model, achieving testing accuracies of 91.99% and 86.62% on two distinct datasets of actual fall-related IMU data. Building upon the simulation framework, this paper further optimized the single IMU attachment position and multiple IMU combinations on fall detection. Building upon the simulation framework, this paper further optimized the single IMU attachment position and multiple IMU combinations on fall detection. The proposed method simplifies fall detection data acquisition experiments, provides novel venue for generating low cost synthetic data in scenario where acquiring data for machine learning is challenging and paves the way for customizing machine learning configurations.

Memon, Z., Turab, M., Narejo, S., & Korejo, M. T. (2023). An ensemble of CNN architectures for early detection of alzheimer’s disease using brain MRI. Early detection of Alzheimer's disease (AD) has proven to be helpful and effective in preventing the disease. If the risks and symptoms of AD are detected earlier, then it seems rather promising that the death ratio of AD might decrease as it can help a lot of patients get treated before it's too

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net

Vol-20 Issue-01 April 2025

late. Our study demonstrates promising results, achieving a remarkable accuracy of 96.52% through the utilization of the EfficientNetB2 and EfficientNetB3 models. By leveraging transfer learning, we leverage pre-trained models' knowledge to optimize the learning process, while ensemble learning further improves performance by aggregating predictions from multiple models. The integration of these methodologies provides an effective and efficient means of detecting Alzheimer's Disease at an early stage, thereby offering potential benefits to patients, caregivers, and healthcare providers alike. These findings pave the way for improved diagnostic tools and contribute to the advancement of AD research and patient care.

International Conference on Emerging Technologies in Electronics, Computing and Communication (ICETECC), Jamshoro, Sindh,Pakistan, 2022, pp. 1-5, doi:10.1109/ICETECC56662.2022.10069760.Because of the speedy surge in social media's expansion, the proliferation of malicious and harmful poses a substantial worry in contemporary society. The identification of hate speech on platforms like Twitter is crucial for various tasks such as controversial event extraction, AI chatterbot creation, content suggestions,and sentiment analysis. Researchers have invested considerable effort in addressing the challenging task of identifying hostile content due to the rise in hate speech and harmful information. The objective is to classify tweets as Hateful, Offensive, or neither. However, this task is highly complex due to the intricate nature of natural language constructs, encompassing different manifestations of animosity directed at various demographics, and the multitude of ways the same meaning can be expressed. Previous research has predominantly relied on manual feature extraction or employed representation-learning techniques followed by linear classifiers. Nevertheless, deep learning methods have recently demonstrated significant accuracy improvements in complex problems across speech, vision, and text applications. In this study, This paper present an idea for automatic classifications of inappropriate language and expressions of hostility using transfer learning models. In this research, leverage classified tweet datasets obtained from Kaggle and conduct experiments. Findings reveal that the multilingual- BERT model and its pre- trained version deliver superior outcomes. Specifically, pre-trained BERT model notably improves classification accuracy of hateful tweets by up to 92% when compared to other algorithms.

3. PROPOSED METHODOLOGY

The proposed system uses machine learning models for hate speech detection. It leverages advanced algorithms like XGBoost and logistic regression for improved accuracy. The system is designed to be integrated with multi-modal data inputs, considering text, images, and metadata. Machine learning models can process vast amounts of data quickly and accurately. These models can learn from large datasets, improving their accuracy over time and reducing false positives and negatives. Advanced machine learning models can understand the context better than simple keyword filters. They can

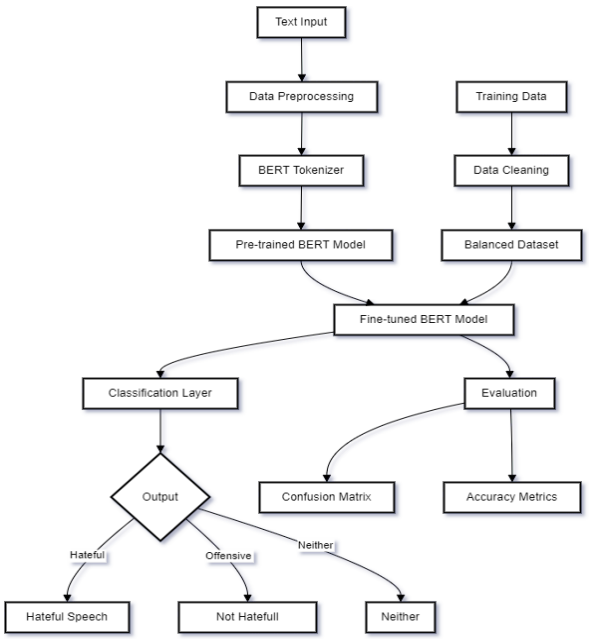


Figure 1: Proposed BERT system.

analyse the semantics of language, reducing the misidentification of hate speech. The system can scale to handle large volumes of content without the need for proportional increases in human resources. It can operate continuously, ensuring real-time monitoring and flagging of hate speech. Machine learning models provide consistent enforcement of policies without human biases. They can be trained on diverse datasets to reduce bias and improve fairness in detection. They can analyse the semantics of language, reducing the misidentification of hate speech. The system can scale to handle large volumes of content without the need for proportional increases in human resources.It can operate continuously, ensuring real-time monitoring and flagging of hate speech. Machine learning models provide consistent enforcement of policies without human biases.They can be trained on diverse datasets to reduce bias and improve fairness in detection. They can be trained on diverse datasets to reduce bias and improve fairness in detection.

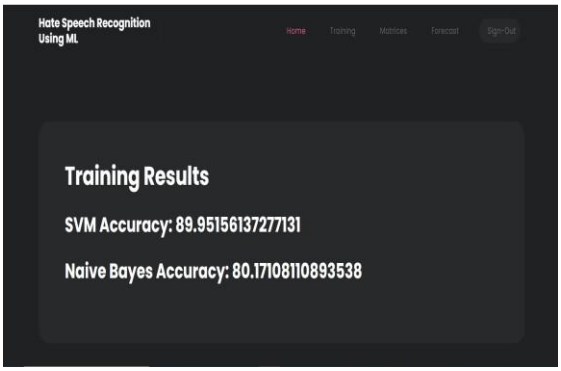
The image shows a system architecture diagram for a text classification model using BERT (Bidirectional Encoder Representations from Transformers). The system appears to be designed for detecting hateful speech, offensive speech, and neutral text. The proposed methodology typically includes the following key components:Text Input: The process begins with raw text data, which will be analyzed.Data Preprocessing: The input text undergoes preprocessing, such as tokenization, removing special characters, lowercasing, etc.BERT Tokenizer: The text is tokenized using a BERT-specific tokenizer, which converts words into numerical representations suitable for deep learning models.Pre-trained BERT Model: A pre-trained BERT model is used as a base model to understand the contextual meaning of words.Training Data Processing: Training data undergoes data cleaning to remove noise. The system appears to be designed for detecting hateful speech, offensive speech, and neutral text. The proposed

International journal of imaging science and engineering

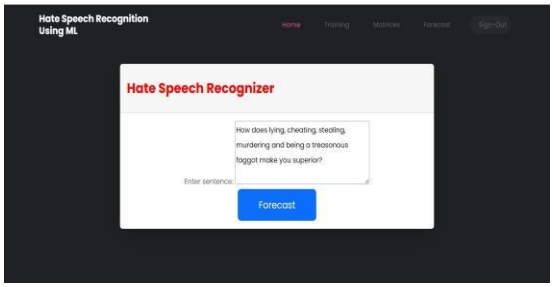
ISSN: 1934--9955 www.ijise.net

Vol-20 Issue-01 April 2025

methodology typically includes the following key components:Text Input: The process begins with raw text



data, which will be analyzed.Data Preprocessing: The dataset is balanced to ensure the model does not get biased towards a particular class. tuned BERT Model: The pre-



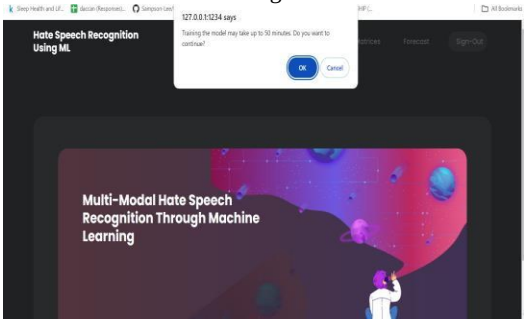
trained BERT model is fine-tuned with the cleaned, balanced dataset to specialize in th classification task. Classification Layer: The model processes input text and classifies it into different categories.Output Categories: The classification layer generates an output, which falls into one of the following: Hateful Speech (Hateful), Offensive Speech (Offensive but not necessarily hateful), Not Hateful (Neutral or non-harmful text) Evaluation Metrics: A confusion matrix is used to analyze model predictions. Accuracy metrics are used to measure the model's performance.

Applications:

The multi-modal hate speech recognition project, which combines data from multiple modalities like text, audio, and images to identify hate speech, has numerous real-world applications. Below are some potential

Advantages:

Improved Accuracy and Efficiency: Machine learning models can process vast amounts of data quickly and accurately. These models can learn from large datasets, improving their accuracy over time and reducing false positives and negatives. Contextual Understanding: Advanced machine learning models can understand the



context better than simple keyword filters. They can analyse the semantics of language, reducing the misidentification of hate speech. Scalability: The system can scale to handle large volumes of content without the need for proportional increases in human resources. It can operate continuously, ensuring real-time monitoring and flagging of hate speech. Consistency and Fairness: Machine learning models provide consistent enforcement of policies without human biases. They can be trained on diverse datasets to reduce bias and improve fairness in detection.Early Detection in Different Media: Multi-modal systems can scan a broader array of media, including social media posts (text), video content (like YouTube or TikTok), and audio

4. Experimental analysis

The figure 2 shows a webpage. The visible text includes "Training Phase Alert" and a pop-up message from "127.0.0.1:1234" stating, "Training the model may take up to 50 minutes. Do you want to continue?" with "OK" and "Cancel" buttons. there is a section labeled "Hate Speech Recognition Using ML", showing training results for two machine learning models:

Figure 2: Sample Images

Figure 3: Accuracy

Figure 4: Confusion Matrix

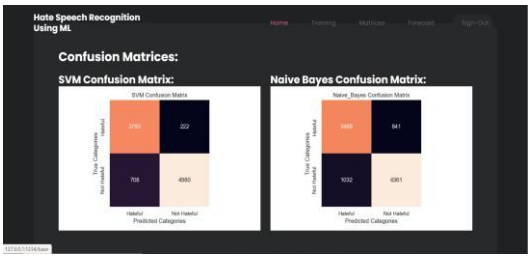


Figure 5: Testing

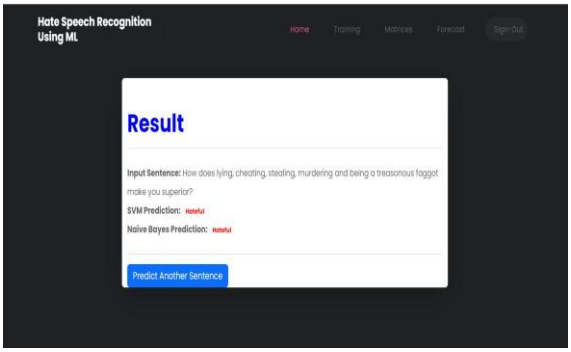


Figure 6: Result

In figure 5, we need to enter the sentence which we need to test. After entering the text we need to forecast the given text. In figure 6, It displays whether the given text is hateful or not hateful. In this way we can enter any sentence which we need to forecast and it will display

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net

Vol-20 Issue-01 April 2025

whether the sentence or text given is hateful sentence or not hateful sentence.

5. CONCLUSION

We proposed a technique for multimodal detecting hate speech in online media, which takes both audio and textual information into account. The BERT was compared with further machine learning and neural network categorization techniques. Our approach makes use of a BERT (a BERT of audio and language) that has been pre- trained and fine-tuned using a larger train dataset. We found them to significantly outperform the existing methods. In our research, we suggested a multimodal learning framework for identifying hate speech that takes into account both the text and the speaker's remarks in audio. The real- world link between these modalities, which is mutually helpful, benefits the model. In

comparison to conventional word-based machine learning techniques, our investigation demonstrated that utilizingthe pretrained BERT by fine-tuning both monolingual when employing both monolingual and multilingual BERT models for hate-speech text classification tasks, we noticed an enhancement in the macro F1 score and accuracy metrics. The dataset contains labels denoting hate speech and offensive content within identical sentences, suggesting a correlation between these tasks.

Our experiments revealed that for hate speech detection in multimedia contexts, incorporating speech features alongside text features outperformed relying solely on textual characteristics. The necessity for a new method for detecting toxic speech in social media data, which makes up a significant amount of the internet nowadays, is also highlighted by this, and it inspires new research directions for this purpose. The link between this hate speech and people in the real world is greatly agreeable, which promotes the model. The concept is beneficial in cyber security and many other industries where there is a risk of security breaches or a high likelihood of racism and abusive language.

The BERT was compared with further machine learning and neural network categorization techniques. Our approach makes use of a BERT (a BERT of audio and language) that has been pre-trained and fine-tuned using a larger train dataset. We found them to significantly outperform the existing methods. In our research, we suggested a multimodal learning framework for identifying hate speech that takes into account both the text and the speaker's remarks in audio. The real- world link between these modalities, which is mutually helpful, benefits the model. In comparison to conventional word-based machine learning techniques, our investigation demonstrated that utilizingthe pretrained BERT by fine-tuning both monolingual when employing both monolingual and multilingual BERT models for hate-speech text classification tasks, we noticed an enhancement in the macro F1 score and accuracy metrics Specifically, pre-trained BERT model notably improves classification accuracy of hateful tweets by up to 92% when compared to other algorithms.

REFERENCES

[1] S. Narejo, "Generalized Epileptic Seizure Prediction using," (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 14, pp. 502-510, 2023.

[2] Kulsoom, F., Narejo, S., Mehmood, Z. et al. A review of machine learning-based human activity recognition for diverse applications. Neural Comput & Applic 34, 18289–18324 (2022).

[3] International Conference on Emerging Technologies in Electronics, Computing and Communication (ICETECC), Jamshoro, Sindh,Pakistan, 2022,

[4] A. Noreen, M. Jawaid, S. Narejo, M. Memon and K. Kumar,"Transfer Learning Based Vascular Stenosis Detection," 2022 International Conference on Emerging Trends in Smart Technologies (ICETST), Karachi, Pakistan, 2022, 10.1109/ICETST55735.2022.9922926.

[5] Hai, Jiang, et al. "R2rnet: Low-light image enhancement via real-low to real-normal network." Journal of Visual Communication and Image Representation 90 (2023): 103712.

[6] Xiong, Wei, et al. "Unsupervised low-light image enhancement with decoupled networks." 2022 26th International Conference on Pattern Recognition (ICPR). IEEE, 2022.

[7] Zheng, Shen, and Gaurav Gupta. "Semantic-guided zero-shot learning for low-light image/video enhancement." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022.

[8] Wu, Yirui, et al. "Edge computing driven low-light image dynamic enhancement for object detection." IEEE Transactions on Network Science and Engineering (2022).

[9] Sun, Ying, et al. "Low-illumination image enhancement algorithm based on improved multi-scale Retinex and ABC algorithm optimization." Frontiers in Bioengineering and Biotechnology 10 (2022).

[10] Zhou, Jingchun, Dehuan Zhang, and Weishi Zhang. "Underwater image enhancement method via multi-feature prior fusion." Applied Intelligence (2022): 1-23.

[11] Liu, Wenyu, et al. "Image-adaptive YOLO for object detection in adverse weather conditions." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 36. No. 2. 2022.

[12] Jiang, Qiuping, et al. "Underwater image enhancement quality evaluation: Benchmark dataset and objective metric." IEEE Technology 32.9 (2022): 5959-5974.

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net

Vol-20 Issue-01 April 2025