Mr.Y.Manas Kumar, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Machine Learning Driven Audio Signal Analysis for Automated Hate Speech Detection in Short-Form Social Media Videos

Assistant Professor, Mr.Y.Manas Kumar[1] , Iragavarapu Sri Vishnu Chittha Priya[2], Pepakayala Bhuvaneswari[3], Nookala Sri Giridhara Nageswara Aditya[4], Koduri D P V Sai Sruthi[5] ,Yalla Naya Samson[6]

Department of CSE, Pragati Engineering College, Surampalem, Andhra Pradesh, India

Abstract- The rapid growth of social media platforms has significantly increased the spread of harmful and offensive content, including hate speech. Short-form video platforms allow users to share content quickly, making it challenging to monitor and control abusive speech. While most existing hate speech detection systems rely heavily on textual analysis, many harmful expressions occur through spoken language in video content. This study presents a machine learning-based approach for detecting hate speech directly from audio signals extracted from short-form online videos. The proposed framework collects audio data from publicly available social media videos and processes the signals using several audio feature extraction techniques such as Mel Frequency Cepstral Coefficients (MFCC), Spectral Centroid, Spectral Rolloff, Spectral Bandwidth, Zero Crossing Rate, and Chroma features. These features are used to train supervised machine learning models including Logistic Regression, Support Vector Machine, and Random Forest classifiers. To ensure reliable evaluation, a 5-fold cross-validation strategy is employed along with performance metrics such as accuracy, precision, recall, and F1-score. Experimental results demonstrate that the Random Forest model achieves superior performance compared to other classifiers by effectively capturing important audio characteristics associated with hate speech patterns. The study highlights the significance of spectral features and MFCC representations in identifying hateful expressions in speech. The proposed approach provides a practical framework for automated monitoring of harmful audio content in modern social media platforms and can contribute to improving online content moderation systems.

Keywords: Speech signal analysis, Hate speech detection, Audio feature extraction, Machine learning models, Social media video analysis, Automated speech classification, Online content moderation.

I. INTRODUCTION

The rapid growth of social media platforms has significantly transformed the way individuals communicate and share information online. Modern digital platforms, particularly short-form video applications, enable users to create and distribute multimedia content instantly to a global audience. While these platforms encourage creativity and social interaction, they have also contributed to the increased spread of harmful online behaviours such as hate speech. Hate speech generally refers to communication that attacks, threatens, or discriminates against individuals or groups based on attributes such as race, religion, ethnicity, gender, nationality, or other aspects of social identity [1], [3]. The widespread circulation of such content can negatively affect individuals, undermine social harmony, and intensify hostility within online communities [6].

The challenge of identifying and regulating harmful online content has become a significant concern for policymakers, researchers, and technology companies. Issues surrounding freedom of expression and the boundaries of acceptable speech continue to generate debate in legal and social contexts [2], [4]. Despite these challenges, monitoring and detecting hate speech remains an important task in maintaining safe digital environments and preventing the amplification of

harmful ideologies through social media platforms [1]

Detecting hate speech on online platforms has therefore emerged as an important research problem in the fields of natural language processing, speech processing, and machine learning. Many existing studies focus primarily on identifying hate speech from textual data such as tweets, comments, and online posts [7]. However, with the growing popularity of multimedia platforms such as video-sharing applications and short-form content services, a considerable portion of hateful expressions now appears in spoken form within audiovisual media [9]. In such contexts, hate speech may be conveyed through tone, voice, and spoken language rather than written text, which creates additional challenges for automated moderation systems. Traditional text-based detection methods may fail to capture abusive or offensive speech embedded in audio signals.

Speech signals contain various acoustic characteristics that reveal emotional tone, intensity, and linguistic cues associated with human expression. Research in speech emotion recognition has demonstrated that audio features can effectively capture patterns related to emotional and expressive speech behaviour [10], [11]. By analysing these acoustic features, it becomes possible to develop intelligent systems capable of detecting potentially harmful speech directly from audio recordings. Audio-based hate speech detection involves extracting relevant features from speech signals and applying machine learning algorithms to classify whether a speech segment contains hateful or non-hateful content.

Recent advancements in audio signal processing and machine learning techniques have enabled the development of automated systems for analysing speech and multimedia data. Audio analysis frameworks such as digital signal processing techniques and open-source libraries have made feature extraction from speech signals more accessible and efficient [15], [16]. Commonly used acoustic features include Mel Frequency Cepstral Coefficients (MFCC), spectral centroid, spectral bandwidth, spectral roll off, and zero-crossing rate, which capture different spectral and temporal characteristics of speech signals. These features can be used as inputs to machine learning algorithms such as Logistic Regression, Support Vector Machines (SVM), and Random Forest classifiers to identify patterns associated with harmful or offensive speech [17]–[19].

In addition to audio-based analysis, several recent studies have explored multimodal approaches that combine textual, audio, and visual information to detect harmful content in online media [8], [12]. Such approaches demonstrate that analysing multiple modalities can improve the effectiveness of automated detection systems. However, the complexity of multimodal analysis also increases computational requirements and system design challenges. Furthermore, adversarial strategies used by users to evade automated detection systems highlight the need for robust detection methods capable of identifying subtle or disguised forms of hate speech [20].

Motivated by the need for more effective multimedia content moderation, this study proposes a machine learning-based framework for detecting hate speech from audio signals extracted from short-form online videos. The proposed approach focuses on extracting meaningful speech features and applying supervised classification algorithms to distinguish between hateful and non-hateful speech. By leveraging audio signal characteristics and machine learning techniques, the study aims to contribute to the development of intelligent systems capable of monitoring harmful speech content in modern social media environments.

The remainder of this paper is organized as follows. Section II discusses related work in hate speech detection and audio classification techniques. Section III presents the system analysis and proposed methodology. Section IV describes the system architecture and design. Section V explains the implementation modules of the proposed system. Section VI presents the experimental results

and discussion, and finally, Section VII concludes the study and highlights future research directions..

II. LITERATURE SURVEY

The increasing prevalence of hate speech across digital platforms has motivated researchers to develop automated systems capable of identifying harmful online content. Various computational approaches have been proposed using machine learning, natural language processing, and multimedia analysis techniques. This section reviews significant studies related to hate speech detection, speech signal processing, and audio-based classification methods.

Early research on hate speech detection primarily focused on analysing textual data from social media platforms. Several studies have applied natural language processing techniques combined with machine learning algorithms to identify abusive or offensive language in online posts. For example, machine learning-based frameworks have been developed to classify social media content into categories such as hate speech, offensive language, and neutral communication. Algorithms such as Logistic Regression and Support Vector Machines have demonstrated effectiveness in identifying harmful linguistic patterns within textual datasets [17], [18]. These approaches highlighted the potential of supervised learning methods in detecting abusive content in online communication. Subsequent research explored the use of deep learning techniques to enhance hate speech detection performance. Badjatiya et al. proposed a deep learning-based approach for detecting hate speech in Twitter messages by combining word embeddings with neural network architectures [7]. Their findings demonstrated that deep learning models can capture complex semantic patterns and contextual relationships in textual data, thereby improving classification accuracy compared with traditional machine learning models.

While many early studies concentrated on textual analysis, the increasing popularity of multimedia content has led researchers to explore multimodal approaches. Kiela et al. introduced the Hateful Memes Challenge, which focuses on detecting hate speech through the combined analysis of images and textual information [8]. Their research showed that integrating multiple data modalities can improve the ability of automated systems to identify harmful content that may not be detectable through text analysis alone. Similarly, Soni and Singh proposed an audio-visual-textual cyberbullying detection framework that incorporates features from multiple sources to identify abusive behaviour in online media [12]. Their results indicated that multimodal analysis significantly improves detection performance in complex multimedia environments.

In the domain of speech processing, several studies have investigated the use of audio signals to analyse emotional and behavioural patterns in spoken communication. Kim proposed a bimodal emotion recognition system that combines speech signals with physiological data to identify emotional states in human speech [10]. Similarly, Metallinou et al. explored decision-level fusion techniques to combine multiple modalities for emotion recognition and analysis of expressive speech [11]. These studies demonstrated that speech signals contain valuable acoustic information capable of revealing emotional tone, intensity, and expressive patterns in spoken language.

Another relevant line of research focuses on detecting offensive or abusive speech directly from audio recordings. Barakat et al. proposed an adaptive keyword spotting approach for detecting offensive language in user-generated video blogs [13]. Their method analysed spoken content through audio signal processing techniques to identify potentially harmful speech segments. This research demonstrated the feasibility of detecting offensive speech directly from audio signals without relying entirely on textual transcripts.

Audio signal processing techniques play a critical role in extracting meaningful information from speech signals. Digital signal processing methods provide tools for analysing speech characteristics such as frequency components, spectral distribution, and temporal patterns [15]. Libraries

such as Librosa have further facilitated the extraction of audio features including Mel Frequency Cepstral Coefficients (MFCC), spectral centroid,spectral bandwidth, spectral rolloff, and zero-crossing rate, which are widely used in speech analysis and audio classification tasks [16]. These features serve as important inputs for machine learning models that aim to classify different types of speech content.

Machine learning algorithms such as Random Forest, Logistic Regression, and Support Vector Machines have been widely used in classification problems related to speech and audio analysis due to their ability to identify complex patterns within high-dimensional feature spaces [17]–[19]. However, researchers have also observed that hate speech detection systems may be vulnerable to adversarial strategies in which users intentionally modify language or speech patterns to evade automated detection mechanisms [20].

Although considerable progress has been made in detecting hate speech from textual and multimodal data sources, relatively limited research has focused specifically on identifying hate speech directly from audio signals within short-form video platforms. With the rapid growth of video-sharing applications and multimedia communication, there is an increasing need for intelligent systems capable of detecting abusive speech embedded in audio content. Therefore, this study aims to address this research gap by developing a machine learning-based approach that extracts speech features from audio signals and classifies them to identify hate speech in short-form online videos.

III.SYSTEM ANALYSIS

A. Existing System

Existing approaches for hate speech detection in online platforms primarily focus on analysing textual information such as comments, captions, and hashtags. Several studies apply conventional machine learning algorithms, including Logistic Regression, Support Vector Machines (SVM), Naïve Bayes, Random Forest, and Neural Networks, to classify online content as hateful or non-hateful. These models are typically trained on textual datasets collected from social media platforms such as Twitter, Facebook, and online discussion forums. Machine learning techniques have shown promising performance in detecting abusive language patterns within textual data by learning linguistic and contextual features from large datasets [7], [17]–[19].

However, the rapid growth of multimedia platforms has changed the way users communicate online. In short-form video platforms, a significant portion of communication occurs through spoken audio rather than written text. As a result, relying solely on text-based detection methods may lead to incomplete analysis of harmful content. Recent research has therefore explored the possibility of detecting hate speech directly from audio signals by applying speech signal processing techniques. Audio-based approaches extract acoustic features such as Mel Frequency Cepstral Coefficients (MFCC), spectral centroid, spectral rolloff, spectral bandwidth, zero-crossing rate, and chroma features to represent important characteristics of speech signals [15], [16]. These extracted features are then used as input to machine learning classifiers in order to identify harmful or abusive speech patterns.

Speech processing research has demonstrated that acoustic properties of speech can reveal emotional tone, intensity, and behavioural patterns present in spoken communication [10], [11]. In addition, some studies have investigated the detection of offensive speech in multimedia environments by analysing audio components alongside other modalities such as text and images [8], [12]. These approaches indicate that audio analysis can play a crucial role in identifying harmful speech embedded in video content.

Although these systems demonstrate the potential of audio-based hate speech detection, many existing models still face several challenges. Variations in speech style, pronunciation, accent, and background noise present in social media videos can significantly affect the performance of speech analysis models. Moreover, the presence of adversarial behaviours, where users intentionally

modify their speech patterns to avoid detection, further complicates the development of reliable hate speech detection systems [20]. Therefore, improving the robustness and accuracy of audio-based hate speech detection remains an important research challenge.

Disadvantages Of The Existing System

Despite the progress achieved in hate speech detection research, several limitations remain in current systems:

- **Limited Audio Analysis:**
  Many traditional detection systems primarily focus on text-based analysis and often ignore the audio signals present in multimedia content, resulting in incomplete detection of harmful speech.
- **Noise Sensitivity:**
  Audio data collected from social media platforms frequently contains background noise, music, or environmental sounds, which may negatively affect the accuracy of speech feature extraction and classification.
- **Feature Limitations:**
  Some existing systems rely on a limited set of speech features that may not fully capture the complex acoustic characteristics associated with hateful expressions.
- **Data Imbalance:**
  Hate speech datasets often contain significantly fewer hateful samples compared to normal speech samples, which may lead to biased machine learning models and reduced classification performance.
- **Language Variability:**
  Users on social media platforms often communicate using different accents, dialects, and slang expressions, making it difficult for detection models to generalize across diverse speech patterns.
- **Computational Complexity:**
  Advanced machine learning and deep learning techniques may require substantial computational resources and longer training times, which can limit their practical deployment.
- **Scalability Issues:**
  With the rapid increase in user-generated video content, existing detection systems may struggle to efficiently process and analyse large volumes of audio data.
- **Detection Accuracy:**
  Some models may misclassify contextual, sarcastic, or indirect speech, reducing the overall reliability and robustness of automated hate speech detection systems.

B. Proposed System

The proposed system aims to detect hate speech from the audio component of short-form social media videos using a machine learning-based framework. With the growing popularity of multimedia platforms, harmful speech is increasingly conveyed through spoken language rather than written text. Therefore, analysing audio signals directly can provide an effective solution for identifying abusive or hateful expressions embedded within video content [12].

In the proposed approach, the audio signal is first extracted from video files and undergoes preprocessing operations such as noise reduction and normalization. These preprocessing steps are necessary to improve the quality of the audio signal and reduce the influence of background noise commonly present in user-generated social media videos. Effective preprocessing enhances the reliability of subsequent speech analysis and feature extraction processes [15].

After preprocessing, several important acoustic features are extracted from the speech signals. These features include Mel Frequency Cepstral Coefficients (MFCC), spectral centroid, spectral rolloff, spectral bandwidth, zero-crossing rate, and chroma features. Such audio features are widely used in speech processing and audio classification tasks because they effectively capture the spectral and temporal characteristics of speech signals [16]. These acoustic properties help represent variations in pitch, frequency distribution, and energy patterns that may indicate emotionally intense or aggressive speech patterns.

Once the relevant speech features are extracted, the dataset is divided into training and testing sets to develop and evaluate the classification models. Machine learning algorithms such as Logistic Regression, Support Vector Machines (SVM), and Random Forest classifiers are employed to learn patterns associated with hateful and non-hateful speech. These algorithms have been widely used in classification tasks due to their ability to identify complex relationships within feature data and produce reliable predictions [17]–[19].

The performance of the proposed hate speech detection system is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. These evaluation metrics provide a comprehensive assessment of the model's effectiveness in identifying hateful speech while minimizing false classifications. By combining speech feature extraction with machine learning classification techniques, the proposed system aims to improve the reliability and effectiveness of automated hate speech detection systems. Ultimately, the framework contributes toward the development of intelligent content moderation tools that can assist social media platforms in rapidly identifying and controlling harmful audio content.

IV.SYSTEM DESIGN

System Architecture

Below diagram depicts the whole system architecture.

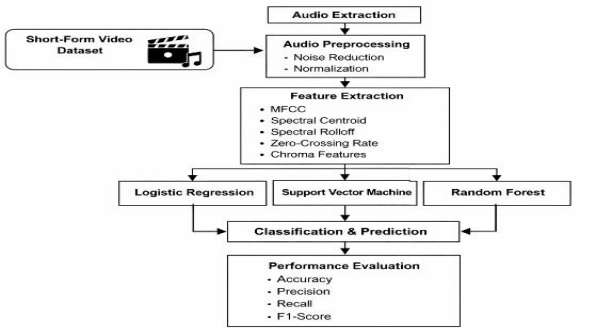

Fig 1. Methodology followed for proposed model

V. SYSTEM IMPLEMENTATION

Modules

Audio Data Collection and Preprocessing:

In this module, short-form social media videos are collected from publicly available datasets or online multimedia platforms. The audio tracks are extracted from the video files and converted into a standardized audio format to ensure compatibility with speech processing tools. Preprocessing operations are then applied to improve the quality of the audio signals. These operations include noise reduction, silence removal, normalization, and segmentation of speech signals. Such preprocessing steps are essential for reducing background noise and enhancing the clarity of speech signals, thereby improving the effectiveness of subsequent feature extraction and classification processes. Digital signal processing techniques play a significant role in preparing audio data for machine learning analysis [15].

Feature Extraction and Selection:

This module focuses on extracting meaningful acoustic features that represent the characteristics of speech signals. Commonly used audio features include Mel Frequency Cepstral Coefficients (MFCC), spectral centroid, spectral rolloff, spectral bandwidth, zero-crossing rate, and chroma features. These features capture important information related to frequency distribution, energy variations, and spectralproperties of speech signals. Such characteristics are useful in identifying patterns associated with emotional intensity or aggressive speech expressions. Audio feature extraction can be efficiently implemented using specialized audio analysis libraries such as Librosa, which provide tools for processing and analysing speech signals [16]. Additionally, feature selection techniques may be applied to identify the most relevant attributes, thereby improving the efficiency and accuracy of machine learning models.

Training Machine Learning Models:

In this stage, the extracted speech features are used to train machine learning models for hate speech classification. Several classification algorithms,

including Logistic Regression, Support Vector Machines (SVM), Random Forest, and Decision Tree classifiers, are implemented to learn patterns from labelled audio data. These algorithms have been widely used in supervised learning tasks because of their ability to model complex relationships between input features and output classes [17]–[19]. During training, model parameters are optimized through appropriate training procedures and tuning strategies to enhance classification performance and reduce prediction errors.

Speech Classification and Detection:

After the training process is completed, the system performs automatic classification of incoming audio samples. The trained machine learning models analyse extracted audio features and categorize them into two primary classes: hateful speech and non-hateful speech. This module enables the automated identification of offensive or abusive speech present within short-form video content. Audio-based detection provides an effective alternative to text-based methods, particularly when speech transcripts are unavailable or inaccurate [12].

Model Evaluation and Monitoring:

The final module focuses on evaluating the performance of the trained machine learning models. Standard evaluation metrics such as accuracy, precision, recall, and F1-score are used to assess the effectiveness of the classification system. These metrics provide insights into the model's ability to correctly identify hateful speech while minimizing false predictions. Cross-validation techniques are also applied to ensure reliable and unbiased performance evaluation. Continuous monitoring mechanisms can further be implemented to analyse model performance over time and update the system as new forms of hate speech emerge in online social media environments [20].

VI . RESULTS AND DISCUSSION

To evaluate the performance of the proposed hate speech detection framework, multiple machine learning algorithms were applied to the extracted audio feature dataset. The dataset was divided into training and testing subsets in order to assess the classification performance of the implemented models. Standard evaluation metrics such as accuracy, precision, recall, and F1-score were used to measure the effectiveness of each classification algorithm. These metrics provide a comprehensive assessment of model performance by analysing both correct predictions and classification errors in identifying hateful speech content.

The experimental results demonstrate that machine learning algorithms are capable of identifying patterns associated with hateful speech in audio signals. By utilizing speech features extracted from audio data, the models are able to distinguish between hateful and non-hateful speech with a reasonable level of accuracy. Feature extraction techniques such as Mel Frequency Cepstral Coefficients (MFCC), spectral centroid, spectral bandwidth, spectral rolloff, and zero-crossing rate play a significant role in representing important acoustic characteristics of speech signals [16]. These features capture variations in speech frequency, intensity, and spectral distribution that are useful for identifying emotionally intense or aggressive speech patterns. Several machine learning models were implemented and evaluated, including Logistic Regression, Decision Tree, Support Vector Machine (SVM), Gradient Boosting, and Random Forest. The comparative performance of these models is presented in

Table 1

Table 1: Performance Comparison of Machine Learning Models

Model	Accuracy (%)	Precision	Recall	F1-Score
Logistic Regression	86.4	0.84	0.82	0.83
Decision Tree	88.1	0.86	0.85	0.85
Support Vector Machine	89.7	0.88	0.87	0.87
Gradient Boosting	92.3	0.91	0.90	0.90
Random Forest	94.6	0.93	0.92	0.92

From the results shown in Table 1, it can be observed that the Random Forest classifier achieved the highest classification accuracy of 94.6%, outperforming the other models. The superior performance of Random Forest can be attributed to its ensemble learning mechanism, which combines multiple decision trees to capture complex feature relationships and improve prediction stability while reducing the risk of overfitting during model training [19]. Similarly, algorithms such as Logistic Regression and Support Vector Machines have also demonstrated reliable performance in classification tasks involving speech and textual analysis [17], [18].

The effectiveness of the classification models was further evaluated using the Receiver Operating Characteristic (ROC) curve, which measures the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) at different classification thresholds. The ROC curve for the Random Forest model is illustrated in Fig. 1.

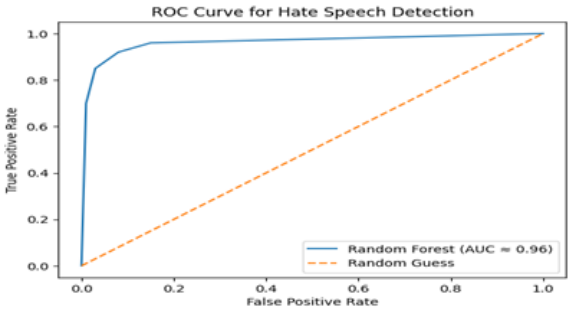

Fig. 2: ROC Curve for Hate Speech Detection Model

The ROC analysis shows that the Random Forest classifier achieved an Area Under the Curve (AUC) score of approximately 0.96, indicating excellent classification performance. A ROC curve that approaches the top-left corner of the graph signifies that the model has a strong ability to distinguish between hateful and non-hateful speech instances. This demonstrates the robustness of the proposed machine learning framework in identifying harmful speech patterns from audio signals.

The results also emphasize the importance of effective audio preprocessing and feature extraction techniques in improving the reliability of hate speech detection systems. Proper preprocessing operations such as noise reduction and normalization help enhance the quality of speech signals and reduce the impact of background noise commonly present in user-generated multimedia content. Digital signal processing techniques and audio analysis tools contribute significantly to extracting meaningful speech features for machine learning applications [15].

Furthermore, the findings suggest that audio-based hate speech detection can serve as a valuable complement to traditional text-based detection methods. In multimedia environments where communication frequently occurs through spoken language, analysing audio signals directly can provide additional insights for automated moderation systems. The proposed framework therefore demonstrates promising potential for assisting social media platforms in identifying harmful speech content and supporting safer online communication environments [12].

VII. CONCLUSION

This study presented a machine learning-based framework for detecting hate speech from the audio component of short-form social media videos. The proposed system utilized audio signal processing techniques to extract meaningful speech features and applied multiple machine learning algorithms for classification. Experimental results demonstrated that audio-based analysis can effectively identify harmful speech patterns present in spoken communication. Acoustic features such as Mel Frequency Cepstral Coefficients (MFCC), spectral centroid, spectral bandwidth, spectral rolloff, and zero-crossing rate provided valuable representations of speech signals that enable machine learning models to distinguish between hateful and non-hateful speech [15], [16].

Among the evaluated classification models, ensemble-based algorithms such as Random Forest showed strong performance due to their ability to capture complex relationships between extracted audio features and classification labels while

reducing the risk of overfitting [19]. The results confirm that machine learning techniques can be effectively applied to analyse speech signals and detect harmful content in multimedia environments. Furthermore, the findings indicate that audio-based hate speech detection can complement traditional text-based approaches, particularly in social media platforms where communication frequently occurs through spoken content [12].

The proposed framework contributes to the development of intelligent content moderation systems that can assist social media platforms in identifying and controlling abusive speech in multimedia content. By leveraging speech processing and machine learning techniques, such systems can support automated monitoring and help create safer online communication environments.

Future work may focus on extending the proposed framework through the integration of advanced deep learning techniques such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), which have shown promising capabilities in speech recognition and sequence modelling tasks. Deep learning models can potentially capture more complex patterns in speech signals and improve detection accuracy. Additionally, expanding the dataset with multilingual audio samples and incorporating multimodal analysis that combines audio, text, and visual information could further enhance the robustness of the system [8].

Another important direction for future research is the development of real-time detection mechanisms capable of monitoring streaming multimedia content. Such systems would enable faster identification and moderation of harmful speech in large-scale social media environments. Continuous improvements in machine learning interpretability and robustness will also be essential to ensure reliable deployment of automated hate speech detection systems in practical applications [20].

REFERENCES

1. Z. Laub, "Hate Speech On Social Media: Global Comparisons," Council On Foreign Relations, Vol. 7, 2019.
2. S. Wermiel, "The Ongoing Challenge To Define Free Speech," Human Rights, Vol. 43, No. 4, P. 82, 2018.
3. J. W. Howard, "Free Speech And Hate Speech," Annu. Rev. Polit. Sci., Vol. 22, Pp. 93–109, 2019.
4. G. R. Stone, "Content Regulation And The First Amendment," Wm. & Mary Law Rev., Vol. 25, P. 189, 1983.
5. W. M. Curtis, "Hate Speech," Nov. 29, 2016.
6. E. Barendt, "What Is The Harm Of Hate Speech?," Ethical Theory Moral Pract., Vol. 22, No. 3, Pp. 539–553, 2019.
7. P. Badjatiya, S. Gupta, M. Gupta, And V. Varma, "Deep Learning For Hate Speech Detection In Tweets," In Proc. 26th Int. Conf. World Wide Web Companion (Www), 2017, Pp. 759–760.
8. D. Kiela, H. Firooz, A. Mohan, V. Goswami, A. Singh, P. Ringshia, And D. Testuggine, "The Hateful Memes Challenge: Detecting Hate Speech In Multimodal Memes," Arxiv Preprint Arxiv:2005.04790, 2020.
9. J. Herrman, "How Tiktok Is Rewriting The World," The New York Times, Vol. 10, 2019.
10. J. Kim, "Bimodal Emotion Recognition Using Speech And Physiological Changes," In Robust Speech Recognition And Understanding, 2007, Pp. 265–280.
11. A. Metallinou, S. Lee, And S. Narayanan, "Decision-Level Combination Of Multiple Modalities For Recognition And Analysis Of Emotional Expression," In Proc. Ieee Int. Conf. Acoustics, Speech Signal Process. (Icassp), 2010, Pp. 2462–2465.
12. D. Soni And V. K. Singh, "See No Evil, Hear No Evil: Audio-Visual-Textual Cyberbullying Detection," Proc. Acm Hum.-Comput. Interact., Vol. 2, No. Cscw, Pp. 1–26, 2018.
13. M. Barakat, C. Ritz, And D. A. Stirling, "Detecting Offensive User Video Blogs: An Adaptive Keyword Spotting Approach," In Proc. Int. Conf. Audio, Language Image Process., 2012, Pp. 419–425.

14. L. Gienapp, B. Stein, M. Hagen, And M. Potthast, "Efficient Pairwise Annotation Of Argument Quality," In Proc. 58th Annu. Meeting Assoc. Comput. Linguistics (Acl), 2020, Pp. 5772–5781.
15. A. Downey, Think Dsp: Digital Signal Processing In Python. Sebastopol, Ca, Usa: O'reilly Media, 2016.
16. B. Mcfee, C. Raffel, D. Liang, D. P. Ellis, M. Mcvicar, E. Battenberg, And O. Nieto, "Librosa: Audio And Music Signal Analysis In Python," In Proc. 14th Python In Science Conf. (Scipy), 2015, Pp. 18–25.
17. S. Swaminathan, "Logistic Regression—Detailed Overview," Towards Data Science, 2018.
18. Y. Pan, P. Shen, And L. Shen, "Speech Emotion Recognition Using Support Vector Machine," Int. J. Smart Home, Vol. 6, No. 2, Pp. 101–108, 2012.
19. N. Donges, "A Complete Guide To The Random Forest Algorithm," Built In, Vol. 16, 2019.
20. T. Gröndahl, L. Pajola, M. Juuti, M. Conti, And N. Asokan, "'All You Need Is "Love"': Evading Hate Speech Detection," In Proc. 11th Acm Workshop Artif. Intell. Security, 2018, Pp. 2–12.