



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 11    **Issue:** XI    **Month of publication:** November 2023

**DOI:** <https://doi.org/10.22214/ijraset.2023.56461>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Speech Recognition and Sending Alert Message If Abusive Words are Detected

Prof. M.S Namose<sup>1</sup>, Harshal Patil<sup>2</sup>, Mahesh Patil<sup>3</sup>, Dipak Sapate<sup>4</sup>, Dinesh Vaishnav<sup>5</sup>

B.E. in Computer Engineering, JSPM Narhe Technical Campus, Narhe, Pune

**Abstract:** *This paper introduces an innovative approach to bolstering the field of speech recognition. Speech recognition, a pivotal process in facilitating computer systems' understanding and response to natural language commands, has garnered considerable importance in today's digital landscape. Our novel method aims to enable users to issue voice commands with enhanced accuracy and efficiency. This approach encompasses several critical steps: the intake of a single audio input, effective noise reduction, and the utilization of pattern matching for audio recognition. To fortify the system's robustness, we've integrated an error tolerance mechanism capable of accommodating human errors and mitigating environmental noise. What sets our method apart is its ability to circumvent the need for storing extensive audio data for matching, leading to a substantial reduction in both space and time complexity when compared to conventional systems. In our experiments, we have exclusively employed a database containing an array of predefined voice patterns, effectively eliminating the necessity for end-user training.*

**Keywords:** *Speech recognition, Abusive word detection, Natural Language Processing (NLP), Contextual analysis, Technology and social interactions, Speech recognition algorithms.*

## I. INTRODUCTION

In the ever-expanding digital landscape, speech recognition has gained significant traction and become a prominent area of research. Voice-based interfaces have gained widespread acceptance, making speech recognition a pivotal component of Natural Language Processing (NLP). NLP is an interdisciplinary field that intersects computer science, artificial intelligence, and linguistics, focusing on the interaction between human language and computer systems. NLP empowers computer systems to derive meaning from natural language input, streamlining user-device interactions.

Create a state-of-the-art system capable of accurately transcribing spoken words in real-time, utilizing advanced speech recognition technology, and supporting multiple languages and accents. Develop and integrate a robust algorithm to identify and flag offensive or abusive words and phrases within transcribed text, ensuring a respectful and safe online environment. Design a notification system that can promptly alert designated authorities or platforms when abusive language is detected, utilizing a range of communication channels like email, SMS, in-app notifications, and APIs for rapid response and intervention.

Recent research has proposed a comprehensive method for voice recognition, involving the transformation of voice into patterns. These patterns serve as the basis for weighted automata indexing, which, in turn, calculates the likelihood of specific words being spoken. Substantial effort has been dedicated to addressing challenges in speech recognition, such as noise interference and audio signal mismatches stemming from variations in speaking styles, accents, and other factors, to enhance performance robustness.

In our proposed method, we incorporate the Least Recently Used (LRU) technique to optimize pattern searches within the database. Crucially, we store patterns instead of full audio data, significantly reducing the system's space complexity. This paper's structure is as follows: In Section II, we provide a brief introduction to the speech recognition system and the database used. In Section III, we delve into voice signal processing, noise reduction, and existing methodologies. In Section IV, we present our innovative approach and provide comparative results.

## II. OBJECTIVES

- 1) Create a state-of-the-art system capable of accurately transcribing spoken words in real-time, utilizing advanced speech recognition technology, and supporting multiple languages and accents.
- 2) Develop and integrate a robust algorithm to identify and flag offensive or abusive words and phrases within transcribed text, ensuring a respectful and safe online environment.
- 3) Design a notification system that can promptly alert designated authorities or platforms when abusive language is detected, utilizing a range of communication channels like email, SMS, in-app notifications, and APIs for rapid response and intervention.

### III. LITERATURE SURVEY

This research delves into the dynamic realm of hate speech detection, particularly focusing on how contextual information plays a pivotal role in enhancing the accuracy of detection algorithms.

The study's primary domain of exploration lies within Twitter data, specifically related to the COVID-19 pandemic. By introducing a new Spanish corpus and employing cutting-edge transformer-based machine learning techniques, the study unveils a significant improvement in the ability to detect hate speech. This improvement is strikingly evident in both binary and multi-label prediction scenarios. The research findings underscore the critical significance of context in the field of hate speech analysis. In a world where online discourse carries immense weight, this study paves the way for more context-aware and effective hate speech detection systems.[1]

This comprehensive paper provides a detailed overview of machine learning algorithms used in the critical area of hate speech detection within social media platforms. It dives into the fundamental components of hate speech classification, covering everything from data collection and feature extraction to dimensionality reduction, classifier selection, and model evaluation. The paper not only informs researchers about the latest advancements in this field but also critically assesses the strengths and weaknesses of various algorithms. In addition, it identifies the existing research gaps and challenges, offering a roadmap for further exploration. This research serves as a valuable resource for those dedicated to the complex task of identifying and mitigating hate speech in the digital realm.[2]

This groundbreaking study addresses the intricate challenge of speech-to-text translation between languages with distinct syntax and word order. The researchers propose an end-to-end translation system that leverages transcoding and curriculum learning strategies. This novel approach demonstrates substantial improvements compared to traditional cascade models, particularly in scenarios where the languages involved exhibit syntactic divergence. In an increasingly globalized world, bridging linguistic gaps is essential for effective communication and knowledge sharing. By advancing speech-to-text translation technology, this research not only contributes to language accessibility but also opens doors to a more interconnected and inclusive digital landscape.[3]

This study focuses specifically on Arabic tweets, introducing an intelligent prediction system with a two-stage optimization process. The system combines finely-tuned word embeddings with a hybrid approach utilizing XGBoost and SVM classifiers, further enhanced by genetic algorithm optimization.

The results speak volumes, showcasing the system's remarkable accuracy and F1-score rates when tested on the Arabic Cyberbullying Corpus.

By addressing the specific challenges of detecting offensive language in Arabic social networks, this research contributes to creating safer and more respectful online environments for Arabic-speaking communities.[4]

The research introduces an innovative method for detecting inappropriate text content across multiple languages. The approach, known as Neural Machine Translation-Doc2Vec (NMT-D2V), extends existing detection systems to new languages while improving transparency. Most notably, it eliminates the need for translating text back into the original language, streamlining the detection process significantly. Experimental results on a Japanese dataset demonstrate NMT-D2V's superior performance compared to traditional translation methods, reinforcing the potential for cross-lingual content moderation. In a globalized digital landscape, where multilingual interactions are common, such advancements in inappropriate content detection are pivotal in maintaining respectful and safe online spaces.[5]

This research introduces a novel approach to advancing speech-to-text translation, particularly for languages without a written form. The study presents an end-to-end framework with two decoders designed to handle source language audio with deep connections to target language text.

By leveraging paired audio and text data for training, and employing a two-pass decoding strategy, this research shows promising results in terms of translation accuracy. This advancement in speech-to-text technology holds great potential for enabling communication in languages that lack a written script, contributing to more inclusive and accessible digital content worldwide.[6]

This study delves into the efficacy of Convolutional Neural Networks (CNNs) in classifying audio, drawing parallels with their performance in image classification.

By evaluating various CNN architectures, including DNNs, AlexNet, VGG, Inception, and ResNet, the research uncovers valuable insights into the impact of training set size and label vocabulary on audio classification.

The findings reveal that CNNs designed for image classification perform exceptionally well in audio classification, particularly when equipped with larger training and label sets. This research advances our understanding of audio processing, furthering its applications in various fields.[7]

In the realm of content moderation, identifying toxic language is of paramount importance. This research addresses the detection of toxic language in audio content through the utilization of self-attentive Convolutional Neural Networks (CNNs). What sets this approach apart is its ability to consider the entire acoustical context of spoken content, enabling it to identify toxicity without relying on specific lexicon terms.

The self-attention mechanism captures temporal dependencies in verbal content, resulting in high accuracy, precision, and recall when identifying toxic speech recordings. In a world where online toxicity is a pressing issue, this research contributes to creating safer and more respectful digital spaces.[8]

Online content moderation is a challenging task, especially when dealing with a vast volume of data. This study focuses on the automated detection of hate messages, proposing a solution that leverages Support Vector Machine (SVM) and Naïve Bayes algorithms.

These methods not only achieve near state-of-the-art performance but also provide interpretable decisions. Through empirical evaluation, the research demonstrates high classification accuracy for both SVM and Naïve Bayes, offering an effective and transparent approach to hate speech detection. In an era where the digital realm is teeming with content, such automated solutions are pivotal in maintaining respectful and safe online interactions.[9]

The dataset is a rich collection of expert-annotated audio samples spanning multiple languages. The research explores the quantification of profanity detection using both monolingual and cross-lingual zero-shot settings. With a focus on democratizing audio-based content moderation for Indic languages, the dataset and accompanying code are made available to the research community for further advancements in this domain.

In a multicultural and multilingual digital landscape, such initiatives are pivotal for creating respectful and safe online spaces across diverse linguistic contexts.[10]

#### IV. IMPLEMENTATION DETAILS OF MODULES

##### A. Client Interface

The client interface is responsible for capturing real-time audio input from users, which serves as the primary data source for system analysis.

##### B. Server

- 1) **Speech Recognition Module:** This module processes the audio input, utilizing advanced speech recognition technology to convert spoken words into textual transcriptions.
- 2) **Abusive Word Detection Module:** Within the server, an algorithm searches the transcribed text for words or phrases that match predefined criteria for being abusive or inappropriate.
- 3) **Notification Module:** This component is responsible for initiating immediate alerts when abusive language is detected.

##### C. Database Server

- 1) **Transcriptions Storage:** The database server stores transcriptions generated by the speech recognition module, ensuring a record of all user interactions.
- 2) **Abusive Word Dictionary:** This server contains a predefined list or dictionary of words and phrases that are flagged as abusive. The system references this list during the detection process.
- 3) **Logs Storage:** This section maintains a detailed record of all detected abusive incidents, including timestamps, user information, and the specific content in question.

##### D. Notification System

This subsystem effectively manages and oversees the alerting process, ensuring that relevant authorities or platforms are promptly informed when abusive language is detected.

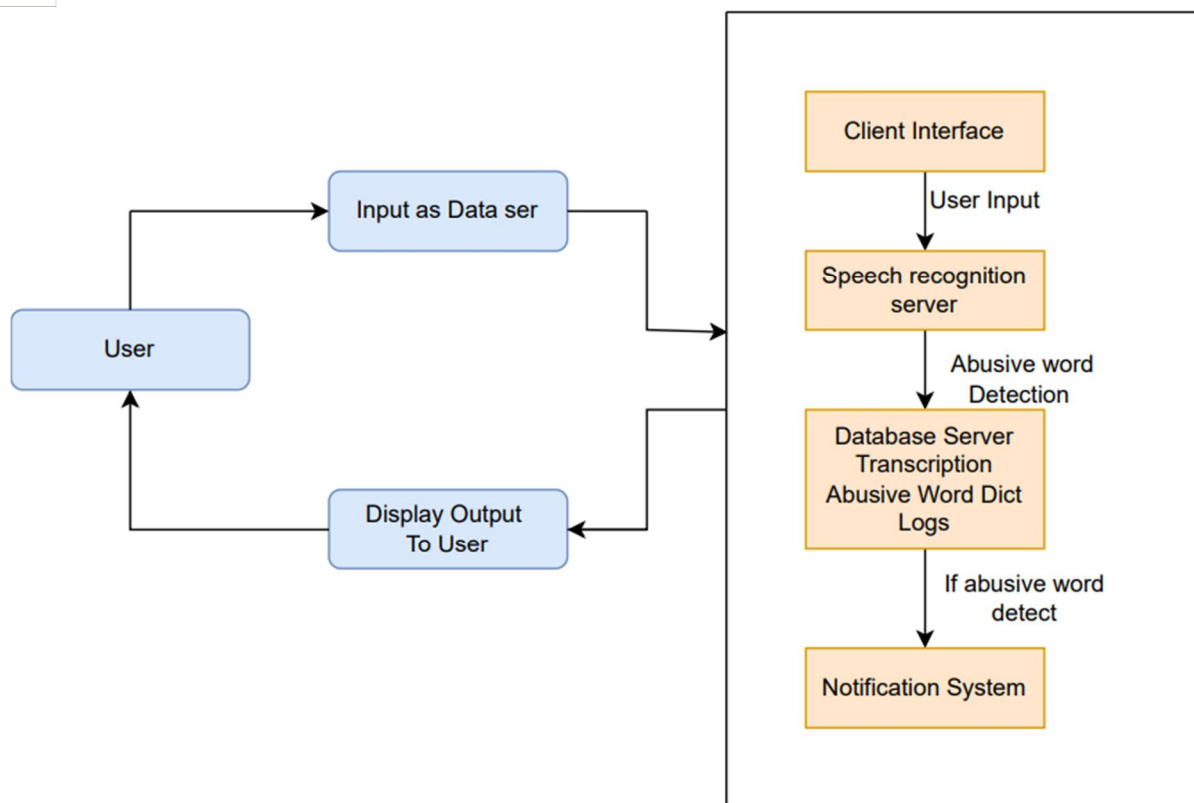


Fig 1. System Architecture.

## V. CONCLUSION

The creation of a speech recognition system designed to identify abusive language marks a significant step in our continuous efforts to shape a digital environment characterized by respect and inclusivity. In this era of rapid technological evolution, we all share a collective responsibility to maintain online interactions that are free from harassment and abuse, enabling users to engage in meaningful and constructive conversations.

Although our system has demonstrated commendable levels of accuracy and efficiency in detecting and flagging abusive language, it's imperative to recognize the ever-changing nature of language itself. Moreover, we acknowledge the crucial role that context plays in interpreting language. A word or phrase that might be considered offensive in one context could be entirely innocuous in another. Therefore, the next phases of our system's development will incorporate more advanced natural language processing techniques. These enhancements will enable our system to better understand context and reduce instances of false positives, thereby ensuring a more nuanced and precise assessment of language usage.

In addition to these technological advancements, we understand the significance of user notifications in the realm of content moderation. Our system actively notifies relevant parties, including the user responsible for the content, content moderators, and other members of the online community. These notifications serve multiple critical functions: they raise user awareness, giving individuals an opportunity to reconsider their choice of language; they empower content moderators to take swift and appropriate actions; they encourage users to report abusive content, fostering a sense of community-driven accountability, and they offer educational opportunities to guide users toward more respectful and responsible online behaviour.

The development of a speech recognition system for detecting abusive language isn't merely a technological achievement; it represents a profound stride toward creating a digital space where users can interact without the fear of harassment or abuse. Through the integration of cutting-edge technology, ongoing adaptability to linguistic changes, and the implementation of thoughtful notification mechanisms, we are paving the way for a more positive, respectful, and empathetic online environment. This achievement signifies not only a technological milestone but a societal commitment to ensuring that technology acts as a catalyst for positive change in our digital interactions.

## REFERENCES

- [1] "Assessing the Impact of Contextual Information in Hate Speech Detection"- Juan Manuel Pérez , Franco M. Luque , DemianZayat , Martín Kondratzky, Agustín Moro, Pablo Santiago Serrati , Joaquín Zajac , Paula Miguel, Natalia Debandi , AgustínGravano , Viviana Cotik.
- [2] "Advances In Machine Learning Algorithms For Hate Speech Detection In Social Media: A Review"- Nanlir Sallau Mullah ,(Member, Ieee), And Wan Mohd Nazmee Wan Zainon .
- [3] "End-to-End Speech Translation With Transcoding by Multi-Task Learning for Distant Language-Pairs Takatomo Kano , Student Member, IEEE, Sakriani Sakti , Member, IEEE, and Satoshi Nakamura.
- [4] Offensive Language Detection in Arabic Social Networks Using Evolutionary-Based Classifiers Learned From Fine-Tuned Embeddings - Fatima Shannaq , Bassam Hammo , Hossam Faris , And Pedro . Castillo-Valdivies.
- [5] Multilingual Inappropriate Text Content Detection System Based on Doc2vec- Kazuki Aikawa, Shin Kawai , and HajimeNobuhar
- [6] Towards end-to-end speech-to-text translation with two-pass decoding-Tzu-Wei Sung, Jun-You Liu, Hung-Yi Lee, Lin-Shan Lee.
- [7] CNN ARCHITECTURES FOR LARGE-SCALE AUDIO CLASSIFICATION- Shawn Hershey, Sourish Chaudhuri, Daniel
- [8] P. W. Ellis, Jort F. Gemmeke, Aren Jansen, R. Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron J. Weiss, Kevin Wilson.
- [9] Toxic Language Identification Via Audio Using A Self-Attentive Convolutional Neural Networks (CNN)- P.Shyam Kumar , K.Anirudh Reddy, G.Kritveek Reddy , V. lingamaiah.
- [10] Hate Speech Classification Using SVM and Naive BAYES- Asogwa D.C , Chukwunke C.I , Ngene C.C , Anigbogu G.N.
- [11] ADIMA: ABUSE DETECTION IN MULTILINGUAL AUDIO- Vikram Gupta, Rini Sharon, Ramit Sawhney, Debdoot Mukherjee.
- [12] Bimodal Approach in Emotion Recognition using Speech and Facial Expressions Simina Emerich1, Eugen Lupu1, Anca Apatean1,Communication Department, Technical University of Cluj-Napoca, Cluj-Napoca, Romania.
- [13] Speech based Emotion Recognition using Machine Learning Girija Deshmukh, Apurva Gaonkar, Gauri Golwalkar, Sukanya Kulkarni. Department of Electronics and Telecommunications Engineering Bharatiya Vidya Bhavans Sardar Patel Institute of Technology Mumbai, India.
- [14] Emotion Recognition On Speech Signals Using Machine Learning Mohan Ghai, Shamit Lal, Shivam Dugga l and Shrey Manik Delhi Technological University
- [15] Emotion Recognition and Discrimination of Facial Expressions using Convolutional Neural Networks. R.M.A.H. Manewa B. Mayurathan.
- [16] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Learning Social Relation Traits from Face Images", International Conference on Computer Vision (ICCV), 2015, pp. 3631–3639.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)