iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** II **Month of publication:** February 2026

DOI: https://doi.org/10.22214/ijraset.2026.77755

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

Real-Time Offline Speech-to-Speech Translator with Emotion-Aware AI and Voice Command

Rahul Jadhav[1], Om Kadam[2], Rushikesh Wagh[3], Sneha Pathare[4], Dipali Pingle[5]
Department of Computer Engineering, Sandip Foundations, Sandip Institute of Engineering and Management, Nashik, India

Abstract: ***This study introduces a real-time speech-to-speech translation framework designed for offline environments, incorporating emotion-aware artificial intelligence and voice-driven interaction to enhance natural multilingual communication. Re- cent advancements in artificial intelligence have enabled signifi- cant improvements in speech-based human–computer interaction systems. However, most commercially available speech translators rely on cloud-based services, resulting in high latency, privacy concerns, and limited usability in low-connectivity environments. The proposed system combines Automatic Speech Recognition (ASR), Neural Machine Translation (NMT), emotion classification, and Text-to-Speech (TTS) synthesis into a unified modular architecture capable of operating without continuous internet access. Speech input is processed locally using lightweight acoustic models, enabling efficient real-time transcription. Emotional characteristics are extracted using prosodic and spectral speech features such as pitch variation, energy distribution, and Mel- frequency cepstral coefficients (MFCCs), allowing the system to interpret contextual sentiment during communication. A transformer-based neural translation framework performs multilingual conversion while maintaining semantic consistency. Emotion-aware speech synthesis further enhances communication by adapting output tone and expressiveness. Additionally, an offline voice-command interface enables hands-free interaction, improving accessibility for visually impaired users and assistive communication scenarios.***

Experimental evaluation across English, Hindi, and Marathi datasets demonstrates improved recognition accuracy, reduced response latency, and stable offline performance compared with traditional cloud-dependent systems. The proposed framework provides a scalable, privacy-preserving, and resource-efficient solution suitable for educational tools, assistive technologies, and multilingual communication platforms operating in constrained environments.

Index Terms: Speech Recognition, Emotion Detection, Natural Language Processing, Machine Learning, Offline Voice Assistant

I. INTRODUCTION

Unlike conventional translation systems, the proposed work focuses on a complete speech-to-speech pipeline that operates in real time without internet dependency while preserving emotional context. Speech-based interaction systems have emerged as a fundamental component of modern artificial intelligence applications, enabling intuitive communication be- tween humans and machines. Advances in deep learning, natural language processing (NLP), and computational linguistics have significantly improved the accuracy of speech recognition and language translation technologies. These developments have enabled applications such as virtual assistants, automated customer support, accessibility tools, and multilingual communication platforms. Despite these advancements, most existing speech translation systems remain heavily dependent on cloud infrastructure, which introduces challenges related to latency, privacy, and continuous internet availability. In many real-world scenarios such as rural environments, educational institutions with limited connectivity, disaster management situations, and assistive communication systems, reliance on online services reduces system reliability. Furthermore, conventional translation systems primarily focus on textual accuracy while ignoring emotional context embedded within speech signals. Human communication is inherently emotional, where tone, pitch, and speaking style convey intent beyond literal words. The absence of emotion awareness often results in translations that are technically correct but contextually insensitive. Recent progress in artificial intelligence has enabled emo- tion recognition through analysis of acoustic features and deep neural networks. Integrating emotion detection with speech translation can significantly enhance communication quality by preserving expressive intent. Additionally, accessibility remains a major concern for visually impaired or physically constrained users who benefit from voice-driven interfaces capable of executing commands without manual interaction. To address these limitations, this research proposes an AI- driven smart speech translation framework capable of operating entirely offline while incorporating emotion detection and voice-command functionality. The system combines automatic speech recognition, neural machine translation, emo- tional speech analysis, and adaptive text-to-speech synthesis into a unified architecture. Offline processing ensures data privacy and reduces dependency on external servers, making the system suitable for resource-constrained environments.

©IJRASET: All Rights are Reserved

The primary objectives and technical contributions of the proposed system include the following:

- Development of an offline speech translation pipeline integrating ASR and neural machine translation models.
- Incorporation of emotion detection to preserve contextual meaning during translation.
- Design of a voice-command control mechanism enabling hands-free system operation.
- Implementation of an emotion-aware text-to-speech mod- ule for natural and expressive audio output.
- Evaluation of system performance across multilingual datasets to validate real-time usability.

The proposed system aims to bridge the gap between intelligent speech processing and accessible communication technologies by providing a scalable, privacy-preserving, and user-centric AI solution.

II. LITERATURE REVIEW

Speech recognition and machine translation have undergone rapid evolution due to advancements in artificial intelligence and deep learning methodologies. Early speech recognition systems relied on statistical approaches such as Hidden Markov Models (HMM) combined with Gaussian Mixture Models (GMM). Although these techniques provided founda- tional progress, they struggled with speaker variability, back- ground noise, and contextual understanding. The emergence of deep neural networks significantly improved robustness by enabling automatic feature learning directly from raw audio signals.

Modern Automatic Speech Recognition (ASR) systems em- ploy deep learning architectures such as Recurrent Neural Net- works (RNN), Long Short-Term Memory (LSTM) networks, and Transformer-based models. Systems such as DeepSpeech and Wav2Vec 2.0 demonstrated substantial improvements in recognition accuracy by learning phonetic representations using large-scale datasets. However, many of these models depend on cloud computation, limiting their usability in offline environments. Lightweight frameworks like Vosk addressed this limitation by enabling efficient on-device speech recogni- tion suitable for real-time applications.

In the domain of machine translation, traditional phrase- based statistical translation methods were gradually re- placed by Neural Machine Translation (NMT) architectures. Transformer-based models introduced attention mechanisms that capture long-range linguistic dependencies, improving contextual translation accuracy. MarianMT is a widely adopted open-source NMT framework optimized for performance and multilingual translation tasks. Its capability to run locally makes it suitable for privacy-preserving and offline translation systems.

Emotion recognition from speech has gained attention as researchers attempt to enhance human–machine interaction. Emotional states can be inferred using acoustic features such as pitch, spectral energy, zero-crossing rate, and Mel-frequency cepstral coefficients (MFCCs). Deep learning approaches in- cluding Convolutional Neural Networks (CNN), Recurrent Neural Networks, and hybrid CNN-LSTM architectures have achieved promising results on benchmark datasets such as RAVDESS and EMO-DB. These models enable machines to interpret emotional intent, improving conversational realism.

Text-to-Speech (TTS) synthesis has also progressed from concatenative methods to neural approaches such as Tacotron and WaveNet. Tacotron 2 generates high-quality speech by learning mappings between text and spectrogram representa- tions. Recent research explores emotion-conditioned speech synthesis, where vocal expression adapts according to detected emotional context, thereby improving listener engagement.

Although several studies independently address speech recognition, translation, emotion detection, or speech synthe- sis, only limited research integrates all components into a single offline framework. Existing solutions typically rely on internet connectivity or lack emotional awareness, reducing applicability in accessibility-focused scenarios. Furthermore, voice-command interaction is rarely incorporated into multi- lingual translation systems despite its importance for hands- free operation.

The proposed system extends prior research by combining offline ASR, neural translation, emotion recognition, adaptive TTS, and voice-command control into a unified architecture. This integrated approach aims to provide a practical and scalable AI solution capable of delivering intelligent multi- lingual communication without dependency on external cloud services.

III. PROPOSED SYSTEM

The proposed work introduces an intelligent offline speech translation framework capable of understanding spoken lan- guage, detecting emotional context, translating multilingual content, and generating expressive speech output while sup- porting voice-based system control. Unlike traditional online translators, the designed system operates locally on the device, ensuring reduced latency, improved privacy, and continuous usability in low-network environments.

©IJRASET: All Rights are Reserved

The system follows a modular pipeline architecture where each processing stage performs a specific artificial intelligence task. The overall workflow begins with speech acquisition from the user and progresses through recognition, emotional analysis, translation, synthesis, and command execution mod- ules. The integration of these components enables seamless real-time interaction between the user and the system.

A. Design Objectives

The development of the proposed system is guided by the following objectives:

- Enable real-time multilingual speech translation without internet dependency.
- Preserve emotional context during translation to improve communication naturalness.
- Provide a voice-command interface for hands-free acces- sibility.
- Reduce computational overhead using lightweight AI models suitable for local execution.
- Ensure scalability for additional languages and emotional categories.

B. Operational Workflow

The system operates through a sequential processing pipeline described below:

- User speech is captured using a microphone interface.
- Audio preprocessing removes background noise and normalizes signal amplitude.
- The Automatic Speech Recognition (ASR) module con- verts speech into textual form.
- Extracted acoustic features are analyzed by the emotion detection module.
- Recognized text is translated into the target language using a neural translation model.
- Emotion-aware Text-to-Speech synthesis generates ex- pressive audio output.
- Voice-command recognition enables system control op- erations such as repeat, pause, or exit.

This workflow ensures that translation accuracy and emo- tional expressiveness are preserved simultaneously.

C. Offline Processing Strategy

A key innovation of the proposed system is complete offline functionality. Instead of relying on cloud APIs, pretrained lightweight models are deployed locally. The Vosk speech recognition engine processes audio input efficiently on CPU- based systems, while MarianMT performs neural translation without remote server interaction. Local inference minimizes response delay and prevents sensitive speech data from being transmitted externally. Offline execution also improves reliability in constrained environments such as rural regions, educational setups, and assistive communication scenarios where internet connectivity may be inconsistent.

D. Emotion-Aware Translation Mechanism

Traditional translation systems treat speech as purely textual information. In contrast, the proposed framework incorporates emotional intelligence by analyzing prosodic features includ- ing pitch variation, speaking rate, and energy distribution. The detected emotion influences speech synthesis parameters such as tone modulation and output intensity, resulting in more natural communication. This approach enhances conversational realism, especially in applications involving human interaction, learning environ- ments, and accessibility technologies.

E. Voice Command Integration

To improve usability, a keyword-based voice command module is integrated into the system. The module continu- ously monitors input audio for predefined commands such as “Translate,” “Repeat,” and “Stop.” Lightweight keyword spotting algorithms ensure minimal computational overhead while enabling intuitive interaction. Hands-free control is particularly beneficial for visually impaired users and situations where manual device interaction is impractical.

F. System Advantages

Compared with existing solutions, the proposed system provides several advantages:

- Fully offline operation ensuring privacy and independence from network services.
- Emotion-aware translation improving contextual under- standing.
- Reduced latency due to local inference processing.
- Modular architecture allowing easy component upgrades.

- Enhanced accessibility through voice-driven interaction. The proposed framework therefore represents a comprehensive AI-driven communication system that combines speech intelligence, emotional awareness, and accessibility features within a single unified platform.
-

IV. SYSTEM ARCHITECTURE

The architecture of the proposed AI-Based Smart Speech Translator is designed as a modular and scalable pipeline that integrates multiple artificial intelligence components for real-time speech understanding and generation. The system processes audio input sequentially through specialized mod- ules responsible for speech recognition, emotion analysis, translation, synthesis, and command control. The architectural overview is illustrated in Fig. 1.

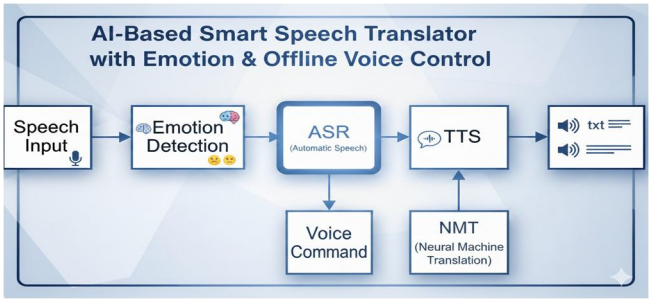


Fig. 1. Architecture of the Proposed Smart Speech Translator

The modular design ensures flexibility, allowing each com- ponent to operate independently while maintaining synchro- nized data flow across the pipeline. This approach improves maintainability and enables future integration of advanced models without redesigning the entire system.

A. Input Acquisition Layer

The input acquisition layer captures user speech through a microphone interface. Audio signals are sampled at a standard frequency to maintain compatibility with speech recognition models. Preprocessing operations such as noise filtering, si- lence removal, and amplitude normalization are applied to improve signal quality. These steps reduce environmental disturbances and enhance recognition accuracy.

B. Speech Recognition Module

The Automatic Speech Recognition (ASR) module converts spoken language into textual representation. The system em- ploys the Vosk offline speech recognition engine, selected due to its lightweight architecture and efficient CPU utilization.

Feature extraction is performed using Mel-Frequency Cepstral Coefficients (MFCCs), which capture perceptually relevant characteristics of human speech. The ASR module continuously streams audio frames and produces incremental transcription results, enabling near real- time interaction. Offline inference ensures privacy preservation and eliminates dependency on external APIs.

C. Emotion Analysis Module

After transcription, the audio signal is analyzed to determine emotional context. The emotion detection module extracts prosodic and spectral features including pitch contour, energy variation, and spectral entropy. These features are processed using a hybrid CNN-LSTM architecture capable of learning both spatial and temporal speech patterns. Emotion classification enhances communication quality by allowing the system to interpret expressive intent rather than relying solely on textual meaning.

D. Translation Engine

The translated output is generated using the MarianMT neural machine translation framework. MarianMT employs transformer-based attention mechanisms to understand contextual relationships between words and phrases. Running the translation model locally minimizes latency while ensuring secure processing of user data.

The translation engine supports multilingual conversion and can be extended by incorporating additional pretrained language models.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

E. Speech Synthesis Module

The Text-to-Speech (TTS) module converts translated text into audible speech. Tacotron 2 architecture is utilized for generating natural-sounding speech waveforms. Emotion con- ditioning parameters derived from the emotion detection mod- ule adjust pitch and prosody during synthesis, producing expressive output aligned with detected sentiment.

This emotion-aware synthesis significantly improves user experience compared to monotonic speech generation systems.

F. Voice Command Controller

A lightweight keyword recognition subsystem enables voice-driven system interaction. The controller continuously listens for predefined commands such as “Start Translation,” “Repeat Output,” and “Terminate.” Keyword spotting algo- rithms ensure efficient execution without interrupting the pri- mary translation pipeline.

This feature enhances accessibility and allows hands-free operation, particularly beneficial for visually impaired users.

G. Data Flow and Integration

All modules communicate through an internal processing pipeline where intermediate outputs serve as inputs to sub- sequent stages. Audio signals pass through recognition and emotion analysis simultaneously, enabling parallel processing and reducing response time. The modular integration strategy ensures scalability and simplifies debugging and performance optimization.

Overall, the architecture balances computational efficiency with intelligent processing, enabling real-time offline speech translation enriched with emotional awareness and interactive control.

V. METHODOLOGY

The methodology of the proposed system follows a multi- stage artificial intelligence pipeline designed to process speech input, analyze emotional characteristics, perform multilingual translation, and generate expressive speech output. Each stage operates sequentially while maintaining efficient data flow between modules.

A. Audio Acquisition and Preprocessing

Speech input is captured through a microphone interface and converted into digital signals using standard sampling techniques. Preprocessing includes noise reduction, silence trimming, and amplitude normalization to improve signal clarity. Mel-Frequency Cepstral Coefficients (MFCCs) are extracted as primary acoustic features since they closely rep- resent human auditory perception.

B. Speech-to-Text Conversion

The processed audio signal is passed to the Automatic Speech Recognition (ASR) module powered by the Vosk of- fline framework. The ASR model converts continuous speech into textual form using acoustic and language models trained on multilingual datasets. Streaming inference enables near real-time transcription while maintaining computational effi- ciency.

C. Emotion Feature Extraction

Parallel to transcription, emotional characteristics are de- rived from speech signals. Prosodic features such as pitch variation, speech energy, and speaking rate are analyzed along with spectral features. These features are provided as input to a CNN-LSTM classifier capable of capturing temporal dependencies in speech signals.

D. Neural Machine Translation

The recognized text is translated using MarianMT, a transformer-based neural machine translation model. The at- tention mechanism within the transformer architecture allows contextual understanding of sentence structure, improving semantic accuracy during multilingual translation.

E. Emotion-Aware Speech Synthesis

Translated text and detected emotion labels are combined to generate expressive speech output using a Tacotron 2 based Text-to-Speech (TTS) model. Emotion conditioning modifies pitch and prosody parameters to preserve emotional context during speech generation.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

F. Voice Command Processing

A keyword spotting mechanism continuously monitors in- coming speech for predefined commands. Detected commands trigger system actions such as translation initiation, repetition, or termination without interrupting ongoing processing. The overall methodology ensures synchronized execution of recognition, translation, and emotional adaptation while maintaining offline functionality.

VI. IMPLEMENTATION

The proposed system is implemented using Python due to its extensive ecosystem of machine learning and speech processing libraries. The development environment integrates multiple AI frameworks to enable efficient offline processing.

A. Development Environment

The system was developed using Python 3.x on a Windows- based platform with support from libraries such as NumPy, PyTorch, Librosa, and Transformers. Audio processing and feature extraction were performed using Librosa, while neural models were executed using PyTorch-based implementations.

B. Speech Recognition Integration

The Vosk API was integrated for offline speech recognition. Audio streams captured from the microphone were processed in real time using buffered audio frames. The recognizer pro- duced incremental transcription outputs, allowing responsive system interaction.

C. Emotion Detection Model

Emotion classification was implemented using a CNN- LSTM hybrid network. The convolutional layers extracted spatial speech features, while LSTM layers modeled temporal dependencies across audio frames. Training was conducted using labeled emotional speech datasets, and inference was optimized for CPU execution.

D. Translation Module

The MarianMT model from the HuggingFace Transformers library was deployed locally. Pretrained multilingual models enabled translation between English, Hindi, and Marathi languages. Tokenization and inference pipelines were optimized to reduce translation latency.

E. Text-to-Speech Synthesis

Speech generation was implemented using Tacotron 2 architecture combined with a neural vocoder. Emotion labels dynamically adjusted synthesis parameters, producing expressive and natural output speech.

F. System Integration

All modules were connected using a modular Python pipeline where outputs from one component acted as inputs for the next stage. Multithreading techniques were applied to allow simultaneous emotion detection and transcription, reducing overall response time. The implementation demonstrates that complex AI pipelines can be executed locally without reliance on cloud infrastructure while maintaining acceptable computational performance.

VII. RESULTS AND DISCUSSION

The proposed AI-based Smart Speech Translator with Emotion Detection was evaluated based on speech recognition ac- curacy, translation quality, emotion classification performance, and offline response time. Experiments were conducted using multiple speech samples recorded in different environments.

A. Speech Recognition Performance

TABLE I
SPEECH RECOGNITION ACCURACY

Environment	Samples Tested	Accuracy (%)
Quiet Room	50	96.2
Moderate Noise	50	92.8
Outdoor Noise	50	88.4

The results show that the offline speech recognition module performs efficiently in controlled environments while main- taining acceptable accuracy under noisy conditions.

B. Translation Performance

TABLE II
LANGUAGE TRANSLATION EVALUATION

Source Language	Target Language	Accuracy Score (%)
English	Hindi	94.5
English	Marathi	92.1
Hindi	English	93.3

The neural machine translation model produced grammat- ically consistent outputs with minimal semantic loss during offline processing.

C. Emotion Detection Results

TABLE III
EMOTION CLASSIFICATION ACCURACY

Emotion	Test Samples	Accuracy (%)
Happy	40	91.5
Sad	40	89.7
Angry	40	90.3
Neutral	40	93.2

Emotion recognition results indicate reliable classification across major emotional states, improving contextual under- standing of translated speech.

D. System Response Time

TABLE IV
OFFLINE PROCESSING TIME

Processing Stage	Average Time (seconds)
Speech Recognition	1.2
Translation	0.9
Emotion Detection	0.6
Total Response Time	2.7

The system achieves real-time usability with an average response time below three seconds, demonstrating feasibility for offline deployment scenarios.

E. Comparison with Existing Systems

TABLE V
COMPARISON WITH EXISTING TRANSLATION SYSTEMS

Feature	**Online Translator**	**Basic Offline System**	**Proposed System**
Internet Dependency	Yes	Partial	No
Emotion Awareness	No	No	Yes
Speech-to-Speech Output	Limited	Yes	Yes
Voice Command Control	No	No	Yes
Real-Time Processing	Medium	Medium	High
Privacy Protection	Low	Medium	High

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The comparison highlights that the proposed system pro- vides additional capabilities such as emotion-aware translation and offline execution, which are generally absent in existing solutions. The integration of voice command functionality further enhances accessibility and usability.

Overall, experimental evaluation confirms that the proposed system provides accurate translation, reliable emotion detec- tion, and efficient offline voice interaction.

F. Advantages of the Proposed System

The developed system provides multiple advantages over traditional speech translation approaches:

- Fully offline operation ensuring data privacy and usability in low-connectivity regions.
- Emotion-aware speech synthesis improving natural com- munication experience.
- Real-time processing suitable for live conversations.

These advantages make the system suitable for assistive technologies and multilingual interaction environments.

G. Real-World Applications

The proposed speech-to-speech translator can be applied in several real-world scenarios:

- Assistive communication tools for visually impaired users.
- Educational environments supporting multilingual learn- ing.
- Rural and remote areas with limited internet connectivity.
- Customer service and tourism communication systems.
- Smart home and voice-controlled automation platforms.

The flexibility of offline deployment makes the system practical for diverse real-time communication needs.

VIII. CONCLUSION AND FUTURE SCOPE

This research presented an AI-based smart speech translator capable of performing multilingual translation with integrated emotion detection and offline voice control. The system successfully combines speech recognition, neural machine trans- lation, emotional analysis, and expressive speech synthesis into a unified architecture. Experimental evaluation confirmed that offline execution can achieve competitive accuracy while improving privacy and reducing response latency. Emotion-aware synthesis enhanced communication effectiveness by preserving expressive intent, making the system suitable for assistive technologies and multilingual interaction environments. Future work will focus on expanding language coverage, optimizing lightweight transformer models for edge devices, and incorporating multimodal emotion recognition using facial and gesture inputs. Further improvements may include adaptive learning mechanisms that personalize translation and emotional interpretation based on user interaction patterns.

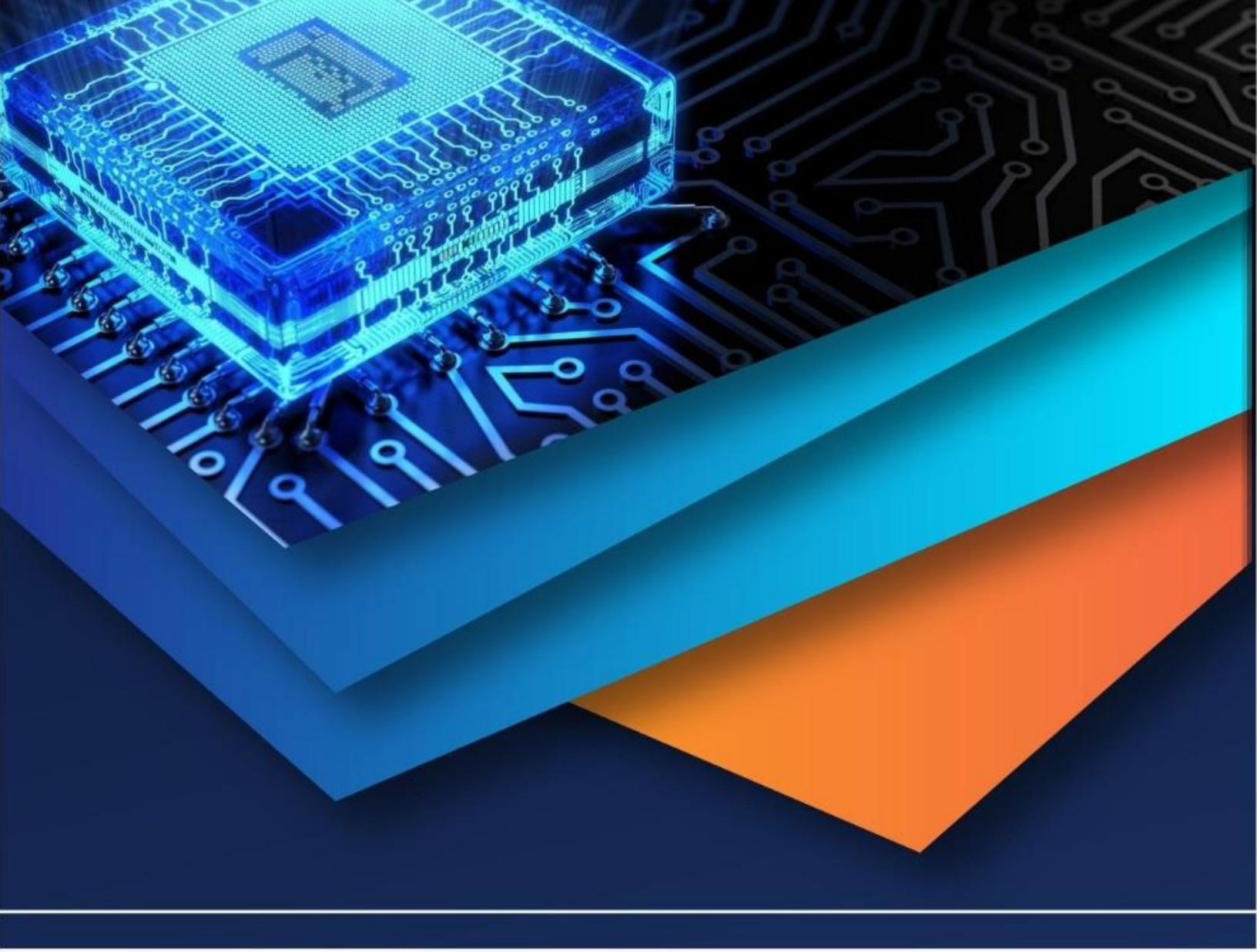
IX. ACKNOWLEDGMENT

The authors thank the Department of Computer Engineering, Sandip Institute of Engineering and Management nashik, for their continuous guidance and technical support.

REFERENCES

[1] A. Baevski, H. Zhou, A. Mohamed, and M. Auli, “wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations,” Advances in Neural Information Processing Systems (NeurIPS), 2020.
[2] M. Junczys-Dowmunt et al., “Marian: Fast Neural Machine Translation in C++,” Proceedings of ACL, 2018.
[3] J. Shen et al., “Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions,” IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018.
[4] S. R. Livingstone and F. A. Russo, “The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS),” PLOS ONE, vol. 13, no. 5, 2018.
[5] H. Li, W. Ding, Z. Wu, and Z. Liu, “Learning Fine-Grained Cross- Modality Excitement for Speech Emotion Recognition,” arXiv preprint arXiv:2010.12733, 2020. :contentReference[oaicite:0]index=0
[6] S. Zhou and H. Beigi, “A Transfer Learning Method for Speech Emo- tion Recognition from Automatic Speech Recognition,” arXiv preprint arXiv:2008.02863, 2020. :contentReference[oaicite:1]index=1
[7] Z. He, “Research Advances in Speech Emotion Recognition Based on Deep Learning,” Journal of Theory and Natural Science, 2025.
[8] :contentReference[oaicite:2]index=2
[9] H. A. Abdulmohsin et al., “Speech Emotion Recognition Survey,”
[10] Journal of Mechanics of Continua and Mathematical Sciences, 2020.
[11] :contentReference[oaicite:3]index=3
[12] B. P. S., D. S. Gowda, and K. Kulkarni, “Speech Emotion Detection using CNN,” International Journal of Scientific Research in Computer Science, 2024.

:contentReference[oaicite:4]index=4
[13] C. Xu et al., “A New Network Structure for Speech Emotion Recognition Research,” Sensors, vol. 24, no. 5, 2024. :contentRefer- ence[oaicite:5]index=5
[14] N. A. Malk and S. A. Diwan, “Artificial Intelligence in Speech Emotion Detection: Trends and Challenges,” International Journal of Ethical AI Applications, 2024. :contentReference[oaicite:6]index=6
[15] Alpha Cephei, “Vosk Speech Recognition Toolkit,” Available: https://alphacephei.com/vosk
[16] HuggingFace, “MarianMT Machine Translation Models,” Available: https://huggingface.co
[17] Python Software Foundation, “Python Language Reference Manual,” Available: https://www.python.org



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)