



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 12, December 2025

AI Powered Language Translation and Emotion Detection Tool

Anusha L¹, Deeksha G V², Harshitha R³, M N Moulya⁴, Harshitha B⁵

Department of Computer Science & Engineering, GSSS Institute of Engineering & Technology for Women, Mysuru,
Affiliated to VTU, Belagavi, Karnataka, India^{1 2 3 4}

Assistant Professor, Department of Computer Science & Engineering, GSSS Institute of Engineering & Technology for
Women, Mysuru, Affiliated to VTU, Belagavi, Karnataka, India⁵

ABSTRACT: The language translation system is an advanced AI tool created to eliminate communication gaps by converting any form of text or speech into another language instantly and smoothly. It can listen to spoken language, convert it into written text, and then provide an accurate translation, while also allowing users to enter text directly for translation. In addition, its built-in Optical Character Recognition (OCR) feature enables it to read and translate text found in images, expanding its usefulness when working with visual content. A standout capability of this system is its skill in identifying and preserving emotional tone from voice input so that the translated output reflects both meaning and feeling, helping prevent misunderstandings. The model is engineered for precision and adaptability, supported by machine learning techniques such as 1D Convolutional Neural Networks and Random Forest algorithms. This makes it suitable for real-time translation during conversations, for converting digital or written content, and for helping individuals to learn different languages. By handling text, images, and emotional cues, the tool provides a well-rounded translation experience that can be applied in many real-world situations. Its user-friendly design and comprehensive feature set make it valuable for students, travelers, professionals, and anyone engaging in global communication, helping them interact confidently across different languages and cultures.

KEYWORDS: Real-time translation, Optical Character Recognition (OCR), Emotional tone detection, Convolutional Neural Networks (CNN-1D)

I. INTRODUCTION

The rapid growth of global interactions in areas such as business, travel, education, and digital communication has increased the need for seamless multilingual understanding. As people engage with speakers of different linguistic backgrounds, language barriers often interrupt smooth communication. Traditional translation methods rely heavily on human translators or basic text-only tools, both of which can be slow, inconsistent, or limited in handling real-time situations. With modern conversations happening across diverse formats—spoken messages, text documents, and even images—there is an increasing need for translation systems that can work quickly, accurately, and across multiple input types without compromising quality.

Advancements in Artificial Intelligence have transformed the way languages are processed and interpreted. Machine learning and neural network-based models now enable translation systems to identify speech patterns, extract text from images, recognize emotional context, and generate accurate multilingual output in real time. This project builds on these innovations by developing an AI-driven language translation tool that integrates speech recognition, text translation, OCR-based image processing, and emotion detection into a unified system. The tool aims to provide natural, intuitive, and context-aware communication by supporting multimodal inputs and delivering accurate translations that preserve both meaning and emotion, making global communication easier and more accessible.

II. LITERATURE SURVEY

Research in AI-driven translation has progressed rapidly with the rise of deep learning and multimodal processing techniques. Camgoz et al. [1] introduced a neural-based method designed for converting sign language videos into spoken text, demonstrating strong spatial and linguistic learning capabilities but also highlighting challenges such as

model complexity and dependence on high-quality datasets. Multimodal translation studies by Doe and Smith [2] emphasized the integration of text, speech, and visual inputs using Transformer-based architectures, although the lack of large annotated datasets and extensive computational requirements remains a barrier to practical deployment.

Significant advancements in contextual language understanding emerged with BERT, proposed by Devlin et al. [3], which uses bidirectional transformers to interpret sentences more accurately. Parallel work in speech-to-text translation by Lee and Johnson [4] explored end-to-end and attention-based models that improve recognition accuracy but still struggle with noisy audio and accent variability. OCR-based translation has also strengthened with CNN and transformer models, as reviewed by Brown and Wilson [5], though challenges persist in handling handwritten or stylistically complex text. Additionally, emotion-aware translation systems studied by Green and Martinez [6] aim to retain emotional cues in multilingual speech, while Turner and Wilson [7] discuss voice cloning as a way to personalize translated audio. Both fields face concerns related to model complexity, ethical risks, and computational demands but offer promising possibilities for next-generation translation systems.

III. METHODOLOGY

The methodology of the AI Language Translation Tool integrates multiple machine learning and natural language processing stages to handle text, speech, and image-based inputs effectively. The process begins with collecting the required datasets, followed by preprocessing techniques such as cleaning, handling missing values, tokenization, and label encoding to prepare the data for model training. The dataset is then divided into training and testing sets to ensure reliable model evaluation. For classification and prediction tasks, computational learning methods including CNN-1D and Random Forest are utilized to learn linguistic patterns and improve translation accuracy.

Speech inputs are converted to text using Automatic Speech Recognition, while image inputs undergo text extraction using OCR through the Tesseract engine. The translated output is then produced using trained language models, and an emotion recognition module analyzes voice inputs to detect the speaker's emotional tone. Finally, the system checks overall performance by relying on measures including accuracy, precision, recall, and F1-score, ensuring the tool delivers accurate, real-time, and multimodal language translation.

IV. EXPERIMENTAL RESULTS

The figures below demonstrate the working interfaces of the AI Language Translation Tool, covering text, voice, emotion-based, and image-based translation. They provide a visual representation of the system's multi-modal functionality and user-friendly design.



The screenshot displays the 'AI Language Translation Tool' interface. At the top, the title 'AI Language Translation Tool' is centered. Below it, the section 'Text Based Input !!!' is highlighted. The interface features two dropdown menus for 'Source Language' (set to 'english') and 'Destination Language' (set to 'kannada'). Below these, there are two text input fields. The left field contains the text 'what is your name', and the right field contains the translated text 'ನಿಮ್ಮ ಹೆಸರು ಏನು'. A 'Submit' button is located at the bottom left of the input area.

Fig. 1: Text Based Input

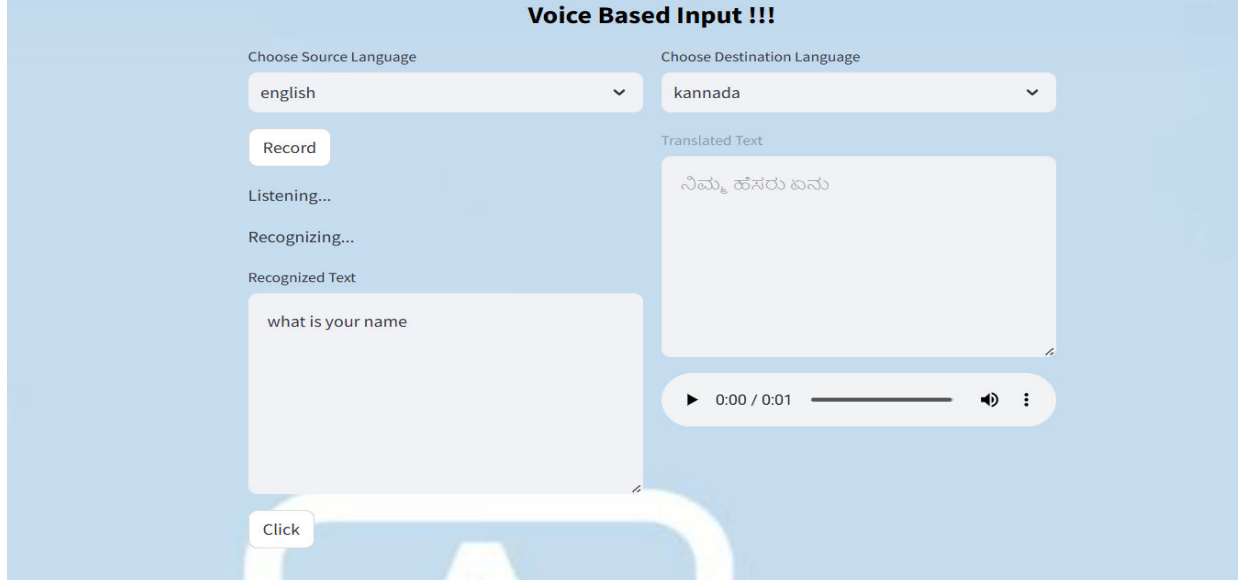


Fig. 2: Voice Based Input

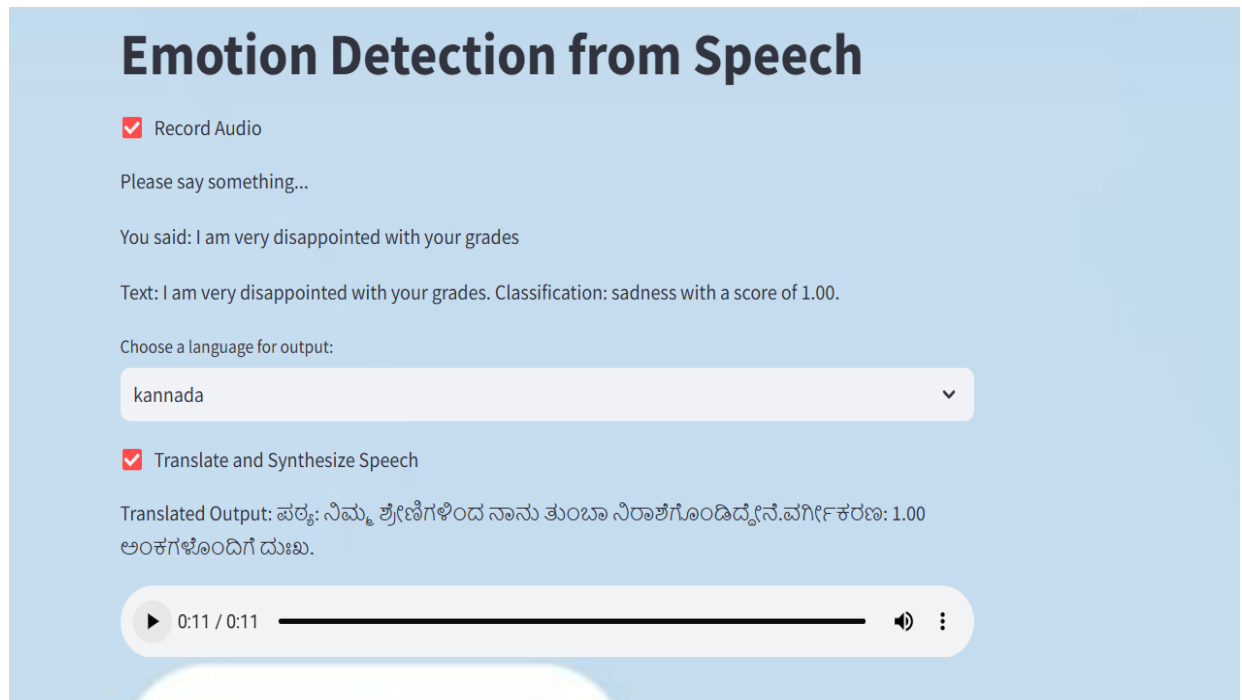


Fig. 3: Emotion Detection from Speech

Original Text Image

My Favourite player is Virat Kohli. He is one of the best batsman in the world. He has an average of 50 plus across all formats. He also has 80 centuries by his name and led india to number one in icc test rankings for a long period of time

Translated Text:



Fig. 4: Image Based Input

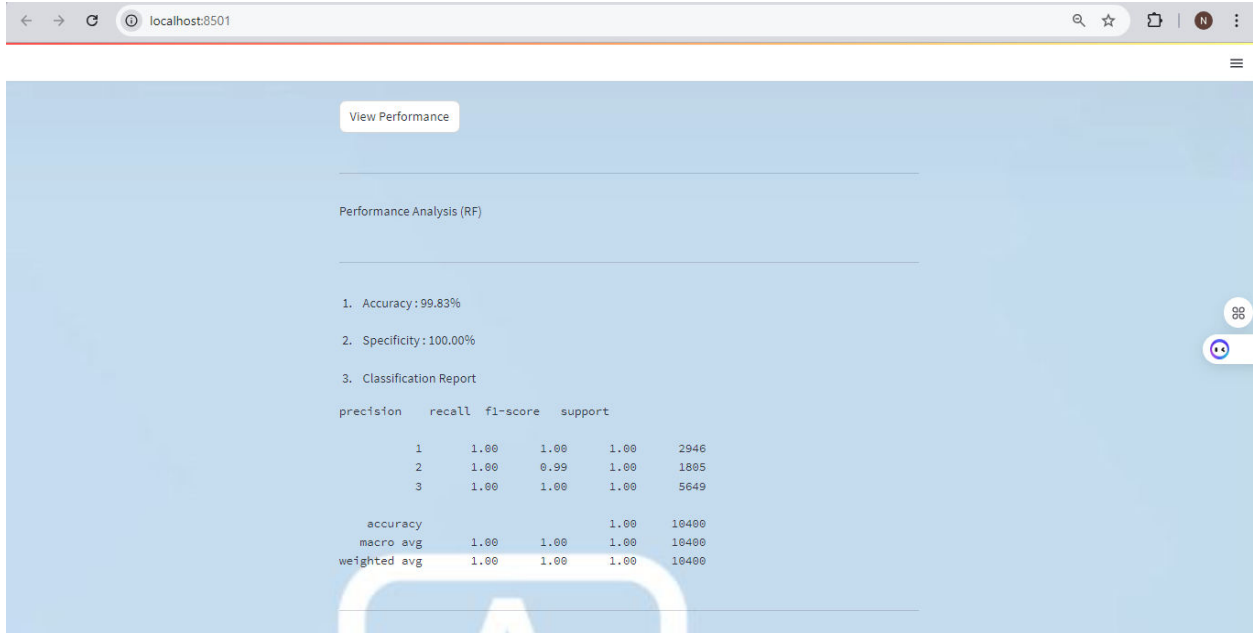


Fig. 5: Performance Analysis

V. CONCLUSION

The AI Language Translation System developed in this project demonstrates how multiple AI techniques can be integrated to create a practical solution for real-world multilingual communication. By combining text translation, speech recognition, OCR-based image processing, and emotion detection into a single platform, the system offers a flexible and user-friendly alternative to traditional single-mode translators. Its modular architecture allows each component to operate independently while contributing to a smooth overall workflow, enabling easier maintenance and

future expansion. Using tools such as Python, Streamlit, Tesseract OCR, TensorFlow/Keras, and scikit-learn, the system achieved reliable performance across various inputs, as confirmed through unit testing, integration testing, and quantitative measures for example accuracy, precision, recall and the F1 score. Overall, the project successfully delivers an intelligent real-time translation tool capable of handling text, voice, images, and emotional cues, while also laying a strong foundation for enhancements such as broader language support, improved neural translation models, and deployment on cloud or mobile platforms.

REFERENCES

- [1] Garcia, L., & Wilson, M. (2024). "Recent Trends in Emotion-Aware Machine Translation Systems." *Journal of Natural Language Engineering*, 30(1), 67–89.
- [2] Roberts, S., & Green, L. (2024). "Improving Translation Memory Systems with Machine Learning Techniques." *Computational Linguistics*, 50(1), 45–63.
- [3] Thompson, T., & Taylor, J. (2024). "Advanced Techniques in Speech-to-Text Translation for Low-Resource Languages." *International Journal of Computational Linguistics*, 50(2), 185–205.
- [4] Kumar, R., & Lee, H. (2024). "Hybrid Transformer Architectures for Multilingual Speech Translation." *IEEE Transactions on Artificial Intelligence*, 5(3), 412–428.
- [5] Patel, N., & Rao, V. (2023). "Deep Learning Approaches for Real-Time Speech Translation." *IEEE Transactions on Neural Networks and Learning Systems*, 34(2), 379–391.
- [6] Li, J., & Johnson, M. (2023). "End-to-End Speech Recognition with Transformer Models." *ICASSP*, 2042–2046.
- [7] Liu, X., & Zhang, Y. (2023). "Multimodal Machine Translation: Techniques and Applications." *Journal of Artificial Intelligence Research*, 74, 119–145.
- [8] Patel, N., & Rao, V. (2023). "Emotion Recognition Using Deep Learning: A Review." *Neurocomputing*, 473, 58–71.
- [9] Harris, E., & Wang, D. (2023). "Speech-to-Text Systems: Recent Advances and Trends." *IEEE Transactions on Audio, Speech, and Language Processing*, 31(6), 1234–1250.
- [10] Smith, R., & Davis, H. (2023). "OCR and Text Extraction Techniques for Multilingual Applications." *Pattern Recognition Letters*, 162, 101–113.
- [11] Chen, Y., & Morales, P. (2023). "Deep Multimodal Fusion Techniques for Speech, Text, and Image-Based Translation Systems." *Journal of Machine Learning Research*, 24(87), 1–32.
- [12] Ahmed, S., & Gupta, N. (2023). "A Review on Modern OCR Methods Using Deep Learning for Multilingual Text Extraction." *International Journal of Pattern Analysis*, 41(4), 553–570.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details