Voice Enabled Translation and Assistance for Rural India

1st Dogga Pavan Sekhar
Department of Computer Science and Engineering
Manav Rachna University
Faridabad, India
† doggapavansekhar@gmail.com

2nd Ketha Sathwik Reddy
Department of Computer Science and Engineering
Manav Rachna University
Faridabad, India
† sathwikreddy1120@gmail.com

3rd Kesanakurthi Naga Siddhartha
Department of Computer Science and Engineering
Manav Rachna University
Faridabad, India
† siddharthasiddu996@gmail.com

***Abstract*—In rural India, access to critical information such as government schemes, healthcare guidance, and agricultural advice is often hindered by language barriers, low literacy rates, and geographical isolation. To address these challenges, this project develops a Voice-enabled Multilingual Farmer Assistant, designed to provide vital information to farmers in their native languages using voice recognition and synthesis. The system leverages advanced Natural Language Processing (NLP), speech recognition, and generative AI models to ensure that farmers, even with minimal literacy, can interact with the technology using voice commands.**

The assistant supports multiple Indian languages, including Hindi, Marathi, Gujarati, Bengali, Tamil, Telugu, Kannada, Malayalam, Oriya, Punjabi, and Urdu. The user interacts with the system by speaking into a microphone, after which the audio is processed, cleaned, and recognized using Google Speech Recognition. The system is designed to handle noisy environments common in rural settings by employing `noisereduce` for noise cancellation. After the audio is processed, the recognized speech is translated into English for further processing, followed by a context-aware, similarity-based answer generated from a database of frequently asked questions (FAQ).

The project also integrates Word2Vec and GloVe word embeddings for semantic analysis, which allows the system to find the most relevant response to a query by comparing the similarity between the user's input and stored questions. Additionally, the assistant can generate responses using Google Gemini AI, summarizing complex agricultural information into simple, digestible paragraphs that are easy for farmers to understand.

In cases where the query is related to farming, the system utilizes generative AI to provide a tailored response, followed by the translation of the response into the user's selected language. The translated summary is then converted into speech using `gTTS` (Google Text-to-Speech) and played back to the user, ensuring seamless interaction. The system also visualizes the audio waveform and spectrograms to provide feedback on the input and output sounds.

Through this innovative solution, the Voice-enabled Multilingual Farmer Assistant aims to empower farmers with easy access to agricultural information, government schemes, and other services, overcoming literacy and language barriers. The system can be used on smartphones or simple devices, making it accessible even in remote regions of India.

keywords: Voice Assistant, Multilingual Support, Farmer Assistance, Speech Recognition, Text-to-Speech (TTS), Natural Language Processing (NLP), Streamlit Interface, Indian Agriculture, FAQ Retrieval, Human-Computer Interaction, Voice-based Query System, Vernacular Language Technology, Agricultural Information System.

I. Introduction

Accessing critical information such as government programs, healthcare guidance, and farming techniques is a significant challenge for millions of people living in rural communities. Despite advancements in technology, rural populations often face numerous barriers that hinder their access to life-changing information. In urban areas, where digital infrastructure is more developed, information is more readily available through the internet, text-based platforms, and mobile applications. However, rural communities are often left behind due to factors such as language diversity, low literacy rates, unreliable or absent internet access, and a lack of familiarity with text-based platforms. Consequently, many individuals miss out on vital resources like financial aid, medical care, and improved agricultural practices—resources that could greatly enhance their quality of life.

The need for an accessible, user-friendly system to bridge these gaps has never been more pressing. This project proposes the development of a **Voice-enabled Multilingual Farmer Assistant**, an AI-powered system that provides critical information to farmers in rural India using voice recognition and synthesis technologies. Unlike traditional systems that rely on reading and writing, this system empowers individuals with minimal literacy skills to access information by simply speaking into a microphone. This voice-based interaction opens doors for millions who may not have formal education but still need access to important resources to improve their livelihoods.

Our proposed assistant will integrate several advanced technologies, including Automatic Speech Recognition (ASR), Natural Language Understanding (NLU), Machine Translation (MT), and Text-to-Speech (TTS), to facilitate seamless communication with the user. By leveraging these technologies, the assistant will not only understand queries spoken in a variety of Indian languages but also respond with spoken answers in the same language, eliminating the need for reading and writing skills. This voice-based interface will significantly increase the accessibility of important information such as

government schemes, healthcare guidance, and farming advice for farmers.

However, creating such a system poses several challenges, particularly when considering the unique conditions and limitations of rural environments. These challenges include:

- **Understanding Speech in Noisy Environments** – Rural areas often feature high levels of background noise—such as the sounds of animals, farm equipment, marketplaces, and weather disturbances. This noise can significantly impact the accuracy of speech recognition systems, making it difficult for them to accurately capture and understand user queries.
- **Handling Dialects and Accents** – India is a linguistically diverse country with hundreds of languages and dialects. Even within the same language, different regions may have varying accents and pronunciations, which can cause difficulties for speech recognition models that are not trained to handle such diversity.
- **Delivering Clear and Relevant Information** – Simply providing answers is not enough. The information provided by the assistant must be accurate, contextually relevant, and easy to understand. Given the varying literacy levels in rural communities, it's crucial that the information is presented in a way that is straightforward and simple to comprehend, even for users with limited education.
- **Working Without the Internet** – Many rural regions suffer from unreliable or no internet connectivity. To be truly effective, the assistant must work offline, ensuring that farmers can access critical information even in areas with no internet coverage.

To overcome these challenges, we will leverage the latest advancements in deep learning, including transformer-based Natural Language Processing (NLP) models, noise-resistant speech recognition algorithms, and AI-driven contextual search frameworks. By utilizing multilingual speech datasets and fine-tuning these models to account for regional dialects, we can create a system that is both robust and adaptable to the linguistic diversity of rural India. Furthermore, to address the issue of poor internet connectivity, we will explore edge computing solutions, which allow the system to function locally on mobile devices without relying on a constant internet connection.

Through the development of this voice-enabled assistant, we aim to empower rural farmers and communities with access to life-changing information. The assistant will be able to provide users with tailored agricultural advice, updates on government schemes, and healthcare guidance, all in their native language. The integration of cutting-edge AI technologies will bridge the digital divide, ensuring that even those without access to formal education or reliable internet can benefit from the resources that can improve their lives and livelihoods.

In the long term, this system can also be adapted to other sectors beyond agriculture, creating broader opportunities for voice-based AI applications in rural India. Our goal is not only to improve the lives of farmers but also to provide a scalable solution that can be replicated in other parts of the country, and even internationally, where rural populations face similar challenges. By making information more accessible, we hope to empower individuals and communities to make informed decisions that improve their health, wealth, and overall well-being.

II. Literature Survey

Accessing critical information such as government programs, healthcare guidance, and farming techniques is a significant challenge for millions of people living in rural communities. Despite advancements in technology, rural populations often face numerous barriers that hinder their access to life-changing information. In urban areas, where digital infrastructure is more developed, information is more readily available through the internet, text-based platforms, and mobile applications. However, rural communities are often left behind due to factors such as language diversity, low literacy rates, unreliable or absent internet access, and a lack of familiarity with text-based platforms. Consequently, many individuals miss out on vital resources like financial aid, medical care, and improved agricultural practices—resources that could greatly enhance their quality of life.

Patel et al. (2009) highlight the importance of voice-based systems in bridging the information gap for rural populations in their study on Avaaj Otalo, a voice-based agricultural information system in Gujarat, India. Their research demonstrates that voice interfaces can significantly improve accessibility for low-literacy users compared to traditional text-based platforms. However, they also emphasize challenges such as the difficulty of handling multiple dialects, misinterpretation of user queries due to limited natural language processing (NLP) capabilities, and issues with speech recognition in noisy rural environments **[1]**. These challenges are precisely the ones that we aim to address in our proposed voice-enabled AI assistant for rural communities.

The challenge of multilingual speech recognition in Indian languages is another key issue that needs attention. Toshniwal et al. (2018) discuss the development of a unified speech recognition model for multiple Indian languages using a deep learning-based approach. Their work highlights the potential of a single end-to-end speech model that can handle various Indian languages, which is critical for our system, given the linguistic diversity in rural India **[2]**. However, they also identify key research gaps, such as the lack of code-switching capabilities and the inability of existing models to function optimally in noisy environments. Addressing these gaps is crucial for the success of a voice-based assistant for rural areas.

In addition to multilingual recognition, there are significant challenges associated with preserving and digitizing low-resource languages, which are common in rural regions. Mehta et al. (2021) investigate the challenges related to the Gondi language, a tribal language spoken in India, and highlight the absence of large-scale linguistic resources for low-resource languages. This issue further complicates the development of robust ASR models for rural communities **[3]**. Their study

emphasizes the need for community-driven efforts in language documentation and highlights the importance of creating standardized frameworks for language processing, which will be essential for developing an inclusive AI assistant capable of supporting diverse rural languages.

Building on the concept of accessible voice technologies, Ashwini et al. (2022) explore the use of voice-enabled chatbots for providing healthcare guidance to rural populations. Their study demonstrates that NLP-powered systems can alleviate the shortage of healthcare professionals in remote areas by providing users with healthcare insights through a voice interface **[4]**. However, the study also identifies limitations in the system's reliance on internet connectivity, which poses a challenge for rural areas with poor or no internet access. Our voice-enabled AI assistant will aim to work offline, ensuring that it can be accessed by users even in areas with limited digital infrastructure.

Further challenges arise from the noisy environments that are characteristic of rural settings. The study on speech recognition in noisy environments identifies the impact of background noise, such as farm equipment and market sounds, on ASR performance **[5]**. By incorporating noise suppression techniques and deep learning-based models, this study suggests that hybrid models can be more effective in noisy conditions. This insight is directly relevant to our system, which will need to handle noisy environments to function effectively in rural areas.

The integration of speech recognition, translation, and synthesis systems has been a key focus of research in recent years. Yuchen Liu et al. (2020) **[6]** in **"Synchronous Speech Recognition and Speech-to-Text Translation with Interactive Decoding"** explore an interactive model that simultaneously performs both speech recognition and translation tasks. Traditional systems often treat these tasks as separate processes, leading to delays and error propagation. Liu et al. propose an innovative attention mechanism that facilitates real-time sharing of information between speech recognition and translation components, improving both speed and accuracy. Their model, which uses a "wait-k" policy to enhance translation quality, outperforms conventional systems when tested on TED Talks in multiple languages. This research is aligned with the ongoing efforts to make real-time translation more efficient and accurate, a challenge also tackled by other studies.

Building on the need for efficient translation systems, the study **"ANUVAADHAK: A Two-way, Indian Language Translation System for Local Travel Information Assistance" [7]**addresses the unique challenges faced in multilingual societies like India. Unlike conventional systems, ANUVAADHAK is designed to offer two-way communication between Indian languages and English, specifically catering to local travel information needs. This system ensures context-aware, culturally relevant translations, which is vital in helping non-English speaking travelers navigate public transportation and explore local destinations. The focus on context-aware translations and user-centric design draws a direct parallel with Liu et al.'s approach of enhancing translation accuracy through real-time interaction, highlighting the importance of tailoring translation systems to specific regional needs.

Further expanding the scope of speech technology, the paper **"Comparative Study of Text-to-Speech Systems for Indian Languages"[8]** examines various text-to-speech (TTS) systems developed for Indian languages. These systems, including Dhvani, Shruti, and Vani, showcase different synthesis techniques such as unit selection and concatenative synthesis, each tailored to handle the diverse phonetic characteristics of Indian languages. The study emphasizes the importance of using the right synthesis methods for Indian languages, particularly focusing on Marathi. This is relevant to the previous studies as TTS systems are crucial in completing the speech translation pipeline—after speech recognition, machine translation, and synthesis, it is essential to ensure that the synthesized speech accurately reflects the target language's phonetic nuances. This synthesis, as demonstrated in the work of Liu et al. and ANUVAADHAK, is integral for a seamless translation experience.

On a broader scale, Vandan Mujadia et al. (2023)**[9]** in **"Towards Speech to Speech Machine Translation Focusing on Indian Languages"** take a step further by integrating speech-to-speech machine translation (SSMT) in their approach. The study presents a cascaded system combining automatic speech recognition (ASR), machine translation (MT), text-to-speech (TTS), and video synchronization. Unlike traditional systems, Mujadia et al. address challenges such as spoken disfluency and domain-specific vocabulary. Their system's ability to handle these issues and provide improved translation quality with minimal human editing directly complements the work done by Liu et al. and ANUVAADHAK, both of which emphasize real-time translation and context awareness. This study highlights how incorporating SSMT can elevate translation systems, especially in multilingual societies with complex language structures.

In a similar vein, Shivam Mhaskar et al. (2023)**10** introduce **"VAKTA-SETU: A Speech-to-Speech Machine Translation Service in Select Indic Languages"**, a system that further enhances translation services by addressing challenges such as spoken disfluency and domain-specific vocabulary. The integration of ASR, disfluency correction, machine translation, and TTS results in seamless real-time speech-to-speech translation between English, Hindi, and Marathi. Their work builds on previous studies by offering a scalable and publicly accessible service that can be applied in diverse contexts, including education, tourism, and agriculture. This aligns with the advancements in interactive decoding and context-sensitive translation seen in the previous studies and underscores the importance of developing systems that offer both high translation quality and accessibility in multilingual environments.

Sheila Cyril et al. (2015)**11**, in their systematic review titled **"Exploring the role of community engagement in improving the health of disadvantaged populations: a systematic review"**, emphasize the vital role of community engagement (CE) in improving the health of disadvantaged populations. The review of 24 studies concludes that CE

has a positive impact on various health-related outcomes, including health behaviors, public health planning, access to health services, and health literacy. Essential components that contribute to the success of CE include real power-sharing, collaborative partnerships, bidirectional learning, and active community involvement in research. However, the study also identifies gaps in evaluating the true impact of CE and stresses the need for better frameworks to measure its effectiveness. This insight into the power of community-driven approaches sets the stage for understanding the role of technology and AI in healthcare systems, especially in rural and underserved areas.

Building on the themes of healthcare equity, Kinalyne Perez et al. (2025)**[12]** investigate the application of artificial intelligence (AI) and telemedicine in rural communities in their study titled **"Investigation into Application of AI and Telemedicine in Rural Communities: A Systematic Literature Review"**. The integration of AI and telemedicine is transforming healthcare delivery, particularly in underserved regions. These technologies enable remote consultations, enhance diagnostic accuracy, and offer real-time patient monitoring, which is crucial for areas with limited access to healthcare facilities. Despite these advances, challenges such as digital literacy gaps, limited infrastructure, and privacy concerns remain. The authors argue for regulatory frameworks to ensure equitable access and safeguard data privacy. These challenges echo the need for community engagement and trust-building as highlighted by Cyril et al., as such initiatives require not only technological innovations but also strong community involvement and support for sustained impact.

While AI and telemedicine are gaining ground in improving access to healthcare, sustainable development (SD) remains a foundational principle in addressing global health disparities. Justice Mensah (2019)**[13]**, in his study **"Sustainable Development: Meaning, History, Principles, Pillars, and Implications for Human Action – Literature Review"**, delves into the concept of SD, emphasizing its importance in ensuring that development meets the needs of the present without compromising the ability of future generations to meet their own needs. The study identifies three key pillars of SD: economic, social, and environmental sustainability. These pillars are critical when considering healthcare innovations in rural and disadvantaged populations, as achieving sustainability requires integrating health equity, economic efficiency, and environmental preservation into healthcare models. The connection between community engagement, AI, telemedicine, and SD lies in the need for sustainable, inclusive solutions that prioritize the long-term well-being of all communities.

In line with this, the work by Mark Seligman et al. (2009)**[14]** on a **"Speech-Enabled Language Translation System and Method Enabling Interactive User Supervision of Translation and Speech Recognition Accuracy"** also ties into the larger conversation about healthcare equity. Seligman and colleagues describe a system that enhances cross-lingual communication by allowing real-time supervision and correction of speech recognition and translation errors. This technology has significant implications for global health, particularly in multicultural and multilingual healthcare settings, where language barriers often hinder access to healthcare services. By improving the accuracy of real-time translation and speech recognition, this technology could complement efforts to integrate telemedicine in rural areas, as discussed by Perez et al., by ensuring that language is not a barrier to receiving quality healthcare.

Finally, Nina Wallerstein and Bonnie Duran (2010)**[15]** in their paper **"Community-Based Participatory Research Contributions to Intervention Research: The Intersection of Science and Practice to Improve Health Equity"** explore Community-Based Participatory Research (CBPR) as a transformative approach to health research. CBPR fosters collaboration between researchers and communities, ensuring that health interventions are culturally relevant, locally supported, and effectively address health disparities. The study highlights how CBPR can overcome traditional barriers to research, such as power imbalances and mistrust, by promoting shared decision-making and mutual learning. This participatory approach aligns with the principles of community engagement outlined by Cyril et al., as both emphasize the importance of empowering communities to take an active role in improving their health. It also complements the AI and telemedicine innovations explored by Perez et al., as CBPR can enhance the adoption and effectiveness of technological interventions by ensuring they are tailored to the unique needs of local populations.

In conclusion, the research across these 15 studies highlights the transformative potential of voice-based and AI technologies in bridging the information and healthcare gaps in rural communities, particularly in multilingual and low-resource settings. Voice-enabled systems, such as those discussed by Patel et al. (2009) and Ashwini et al. (2022), demonstrate significant promise in improving accessibility to essential services like agriculture advice and healthcare guidance, especially for populations with low literacy and unreliable internet access. The challenges of multilingual recognition, as explored by Toshniwal et al. (2018) and Mehta et al. (2021), underscore the need for advanced speech recognition systems capable of handling diverse Indian languages and dialects in noisy environments. Moreover, the integration of speech recognition with translation and synthesis, as seen in the works of Liu et al. (2020), ANUVAADHAK (2020), and Mujadia et al. (2023), is crucial for creating a seamless user experience across language barriers. In parallel, studies like those by Sheila Cyril et al. (2015) and Kinalyne Perez et al. (2025) emphasize the role of community engagement and AI-driven telemedicine in improving healthcare access and outcomes in rural areas, which must be coupled with efforts to address digital literacy and infrastructure challenges. The incorporation of sustainable development principles, as discussed by Mensah (2019), provides a broader framework for ensuring that these technological innovations promote long-term, equitable improvements in health and well-being. Finally, the works of Seligman et al. (2009) and Wallerstein and Duran (2010)

highlight the importance of community-based participatory approaches in ensuring the relevance and success of technology-driven interventions, further emphasizing that technological advancements must be deeply integrated with the needs and values of the communities they aim to serve. Together, these studies outline a comprehensive vision for leveraging AI, speech technologies, and community-driven approaches to empower rural populations, improving their access to critical resources and services.

III. Dataset Description

This dataset was curated by visiting several official government websites, including prominent ones like **PM Kisan**, and specifically focusing on their FAQ sections. The goal was to gather questions and answers related to government schemes, primarily in agriculture and rural welfare, to provide better clarity and accessibility for users. The dataset contains questions and answers about various government schemes, particularly the **PM Kisan Scheme**, and general agricultural topics in India.

A. Dataset Structure

The dataset consists of the following columns:

- **Category**: The broader classification under which the question falls (e.g., "Government Scheme", "Agriculture").
- **Topic**: The specific program, scheme, or area related to the question (e.g., "PM Kisan Scheme", "Agriculture").
- **Question**: The specific question raised by the general public or frequently asked in the FAQ sections of the government websites.
- **Answer**: The corresponding answer provided in response to the question.

B. Example Records

Category	Topic	Question	Answer
Government Scheme	PM Kisan Scheme	What is Pradhan Mantri Kisan Samman Nidhi?	Pradhan Mantri Kisan Samman Nidhi (PM Kisan) is a new Central Sector Scheme to provide income support to all landholding farmers' families in the country to supplement their financial needs for procuring various inputs related to agriculture and
Government Scheme	PM Kisan Scheme	Whether the benefits of the PM Kisan Scheme are admissible to only Small & Marginal Farmers' (SMF) families?	No. In the beginning when the PM Kisan Scheme was launched on 24th February, 2019, its benefits were admissible only to Small & Marginal Farmers' (SMF) families, with combined landholding upto 2 hectare. The PM Kisan Scheme was later on
Government Scheme	PM Kisan Scheme	When was the PM Kisan Scheme	The PM Kisan Scheme was launched by the Hon'ble Prime
Government Scheme	PM Kisan Scheme	From which date the PM Kisan Scheme	The PM Kisan Scheme takes effect from 01.12.2018.
Government Scheme	Agriculture	What is the importance of agriculture in India?	Agriculture is the backbone of the Indian economy. It provides employment to about 58% of the population and contributes about 17% to the GDP. Agriculture is also a major source of raw
Government Scheme	Agriculture	What are the major crops grown in India?	The major crops grown in India include rice, wheat, pulses, oilseeds, sugarcane, cotton, jute, tea, coffee, and spices. India is
Government Scheme	Agriculture	What are the major challenges faced by the agriculture sector in India?	The major challenges faced by the agriculture sector in India include: (i) Climate change and its impact on crop production (ii) Water scarcity and depletion of groundwater resources (iii) Land degradation and soil erosion (iv) Lack of access to credit and financial services (v) Inadequate infrastructure and marketing
Government Scheme	Agriculture	What are the measures taken by the government to promote agriculture in India?	The government has taken several measures to promote agriculture in India, including: (i) Providing financial assistance and subsidies to farmers (ii) Implementing various schemes and programs for agricultural development (iii) Promoting research and development in agriculture (iv) Improving irrigation facilities

Fig. 1. Snapshot of the dataset records

C. Data Collection Method

The data was manually gathered by exploring various government websites, primarily through the **FAQ sections**, which frequently address common questions and concerns. These websites include official portals such as **PM Kisan**, which focuses on farmer welfare, as well as other platforms related to government schemes in rural and agricultural sectors. Some of the key websites visited include:

- **PM Kisan**
- **National Agriculture Market (eNAM)**
- **Ministry of Agriculture and Farmers Welfare**
- **Pradhan Mantri Awas Yojana (PMAY)**
- **National Rural Employment Guarantee Act (MGNREGA)**
- **Atmanirbhar Bharat Abhiyan**

Key Themes in the Dataset

- **Government Schemes**: The data set includes information on specific government schemes such as **the PM Kisan Scheme**, a central initiative that aims to provide financial support to farmers.
- **Agricultural Topics**: General agricultural information, including the importance of agriculture for the Indian economy, the major crops grown and the challenges faced by the sector.
- **Scheme Details**: Each record provides detailed answers on the specifics of various schemes, such as financial assistance, subsidies, eligibility criteria, and more, aimed at improving the livelihoods of farmers.

IV. Methodology

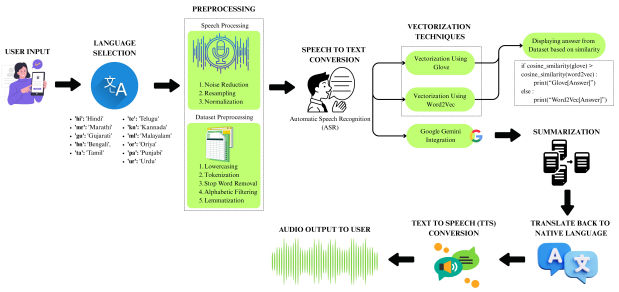

Fig. 2. Proposed Modell Architecture

The architecture of the **Multilingual Voice-Based Farmer Assistant System** enables users (especially farmers) to interact with the system using voice commands in their native language. Below is the detailed step-by-step explanation of the system components and workflow:

1) **User Input:** The user provides a voice query using a mobile device. This could be a question about weather, crop diseases, market prices, or agricultural practices.
2) **Language Selection:** The user selects their preferred language. The system supports multiple Indian languages such as:
   - Hindi (hi), Marathi (mr), Gujarati (gu), Bengali (bn), Tamil (ta),
   - Telugu (te), Kannada (kn), Malayalam (ml), Oriya (or), Punjabi (pa), Urdu (ur)

3) **Preprocessing:** This includes two major stages:
a. Speech Preprocessing:
- **Noise Reduction:** Removes background sounds to improve clarity.
- **Resampling:** Adjusts the sampling rate to a fixed value to standardize audio quality.
- **Normalization:** Balances the audio amplitude levels to ensure consistent volume.

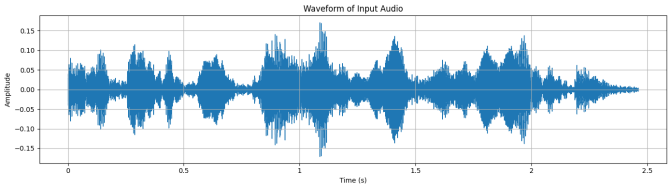

Fig. 3. Input Audio Waveform Before Preprocessing

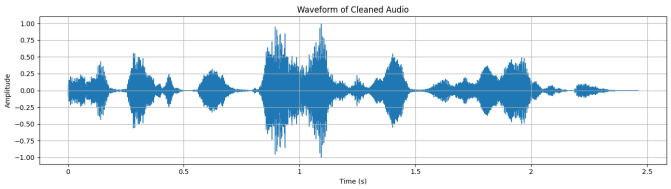

Fig. 4. Cleaned Audio Waveform After Preprocessing

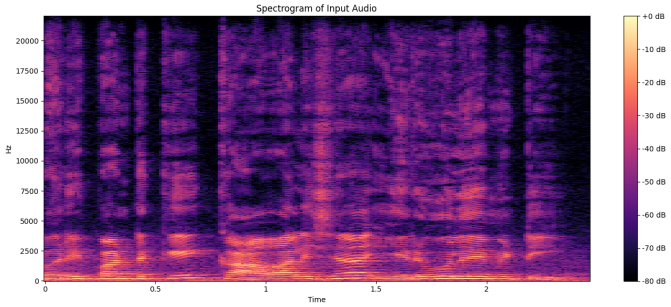

Fig. 5. Input Audio Spectogram Before Preprocessing

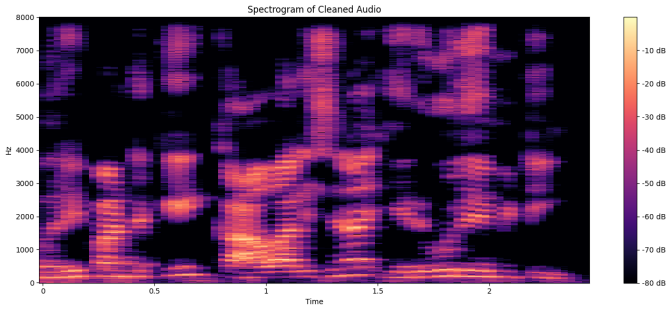

Fig. 6. Cleaned Audio Spectogram After Preprocessing

The above figures show the input audio waveform and spectrogram before and after pre-processing. The waveform shows the amplitude of the audio signal over time, while the spectrogram displays the frequency content across different time segments.

b. Dataset Text Preprocessing:
- **Lowercasing:** Converts all text to lowercase to maintain uniformity.
- **Tokenization:** Breaks the text into individual words (tokens).
- **Stop Word Removal:** Removes common words that do not contribute to the meaning (e.g., ”the”, ”and”, ”is”).
- **Alphabetic Filtering:** Filters out non-alphabetic characters to ensure meaningful tokens.
- **Lemmatization:** Reduces words to their base or root form (e.g., ”running” becomes ”run”).

4) **Speech-to-Text Conversion (ASR):** The user's voice input is converted to text using an Automatic Speech Recognition (ASR) engine. This converts the audio signals into a textual representation of the user's query.
5) **Vectorization Techniques:** The converted text is turned into a numerical vector representation using either of the two popular embedding models:
a. GloVe (Global Vectors for Word Representation): GloVe is an unsupervised learning algorithm developed by Stanford University. It learns word vectors by factoring in the global word-word co-occurrence matrix from a large corpus. The model captures semantic relationships between words. An example of this is:

$$\text{vec}(\text{“king”}) - \text{vec}(\text{“man”}) + \text{vec}(\text{“woman”}) \approx \text{vec}(\text{“queen”})$$

This property makes GloVe highly effective for capturing analogies and semantic understanding.
b. Word2Vec: Word2Vec is a predictive model developed by Google. It learns word embeddings using neural networks. It comes in two architectures:
- **CBOW (Continuous Bag of Words):** Predicts a word based on the surrounding context words.
- **Skip-Gram:** Predicts the surrounding context words given a target word.

Word2Vec excels in capturing both semantic and syntactic relationships between words, making it suitable for various NLP tasks.
6) **Answer Retrieval:** Once the user query is vectorized, the system calculates cosine similarity between the user query vector and the vectors of the dataset (i.e., FAQ dataset). Cosine similarity measures the cosine of the angle between two vectors, providing a score that indicates how similar they are. The formula for cosine similarity between two vectors $\mathbf{A}$ and $\mathbf{B}$ is:

$$\text{Cosine Similarity}(\mathbf{A}, \mathbf{B}) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$$

where $\mathbf{A} \cdot \mathbf{B}$ is the dot product of the two vectors and $\|\mathbf{A}\|$ and $\|\mathbf{B}\|$ are the magnitudes (norms) of the vectors. The system then compares the similarity scores for both GloVe and Word2Vec embeddings. The query vector is

compared with each FAQ question vector using both models, and the system selects the embedding (either GloVe or Word2Vec) that yields the higher cosine similarity score. This ensures that the system retrieves the most semantically relevant question-answer pair based on the user's query.
Optionally, a Google Gemini model can be used for deeper query understanding and better context matching.

7) **Summarization:** The retrieved answer is summarized to make it concise and easier to comprehend.
8) **Translation to Native Language:** The summarized response is translated from English back into the user's original language using a translation API or model.
9) **Text-to-Speech (TTS):** The translated text is converted to spoken language using a TTS engine, enabling audio feedback.
10) **Audio Output:** The final response is played back to the user in their chosen native language.

V. Results and Discussions

The multilingual voice-based farmer assistant system performs effectively in translating and responding to queries related to farming. When a user selects their preferred language and speaks into the microphone, the system processes the audio input and translates it into English. The translated query is then matched against a FAQ data set using Word2Vec and GloVe embeddings, which return relevant answers based on cosine similarity.

The FAQ retrieval system provides accurate responses, with the FAQ questions being processed into tokens and compared with pre-trained Word2Vec and GloVe vectors. The best match is displayed to the user, helping them get the most relevant farming-related information. Below is a representation of how the FAQ answer is fetched from an Excel file:

Fig. 7. Fetching the FAQ answer from the Excel file.

For queries related to agriculture, the system uses Google Gemini to generate a detailed summary of the query in a simple language, which is then translated back into the user's native language. The generated summary is converted into speech using Google Text-to-Speech (gTTS) and is played back to the user. The assistant also visualizes the audio input, showing both the waveform and spectrogram of the original and cleaned audio files.

In cases where the query is not directly related to farming, the system provides a fallback message asking the user to focus on farming-related questions. The input and output visualizations for a non-farming-related query are shown below:

The integration of language translation, text summarization, and speech synthesis makes the assistant a versatile tool for farmers, enabling them to access crucial agricultural information in their native languages through voice commands. The use of Word2Vec and GloVe models for semantic matching and Gemini AI for content summarization ensures that the responses are not only relevant but also comprehensible.

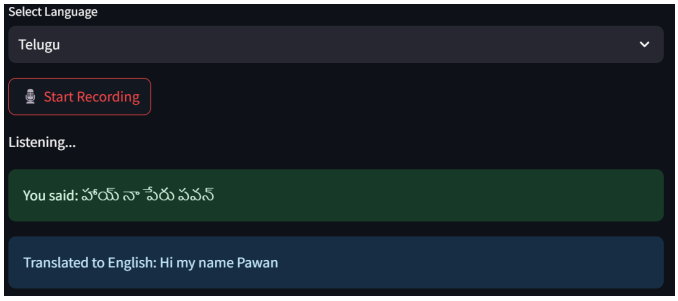

Fig. 8. Input query for a non-farming-related question.

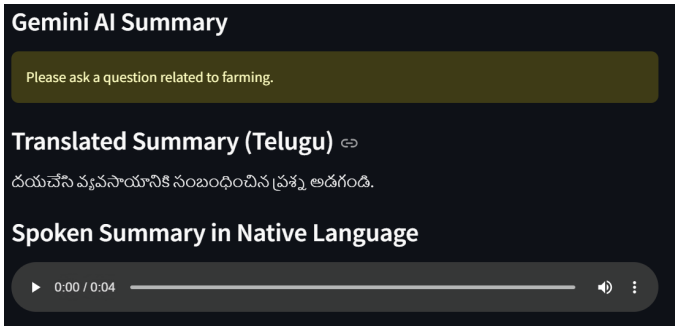

Fig. 9. Output response for a non-farming-related query.

VI. User Interface

The user interface for the multilingual voice-based farmer assistant system is designed to be intuitive and easy to use. It includes several key features to ensure smooth interaction.

To provide a visual understanding of the system, here is an image of the user interface:

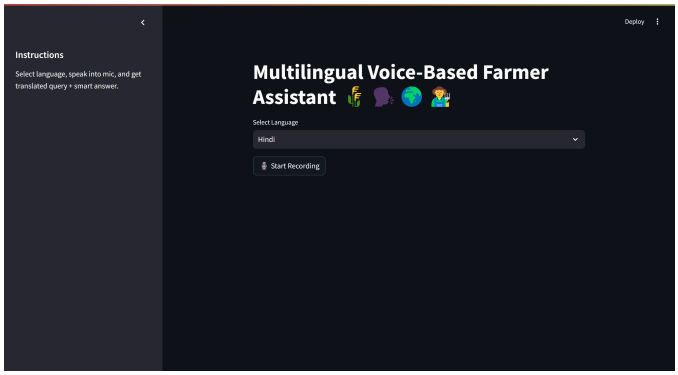

Fig. 10. User interface of the multilingual voice-based farmer assistant system.

For testing and demonstration purposes, a video showcasing the user interface in action can be found in reference [16].

VII. Libraries and Tools Used

The following libraries and tools were utilized in the development of the voice-based assistant system. Each tool plays a significant role in ensuring smooth functionality across different stages, such as speech recognition, text-to-speech conversion, and data processing.

1) **Streamlit:** Used for creating the interactive web interface for the voice-based assistant.
2) **Speech Recognition:** The `speech_recognition` library is used to recognize audio input from the user.
3) **pyttsx3:** A text-to-speech engine used for converting text to speech responses.
4) **Google Translate API:** Used for translating the input query into English and for the back translation of responses into the user's selected language.
5) **Soundfile:** A Python library for reading and writing sound files, used for handling audio data.
6) **Librosa:** A library for audio and music analysis, used for processing and cleaning audio input.
7) **NoiseReduce:** A noise reduction library used to clean the recorded audio.
8) **Matplotlib:** Used to plot and visualize audio waveforms and spectrograms.
9) **Pandas:** A data manipulation and analysis library, used for handling and processing FAQ data.
10) **NumPy:** A library for numerical operations, used for array and matrix manipulations, especially for embeddings.
11) **NLTK:** The Natural Language Toolkit is used for text processing, including tokenization, stopword removal, and lemmatization.
12) **Gensim:** Used for training Word2Vec models for generating word embeddings.
13) **Scikit-learn:** Used for computing cosine similarity to find the most similar questions from the FAQ dataset.
14) **Google Gemini API:** An API for generating AI-based content responses to the user query.
15) **gTTS (Google Text-to-Speech):** Used for converting text into speech in various languages.
16) **Pydub:** A Python library used for audio file manipulation, converting MP3 files into WAV format.
17) **Word2Vec:** A technique for generating word embeddings, used to represent text data numerically for similarity calculations.
18) **Glove:** A pre-trained word embedding model used to represent words in vector format for semantic similarity.

VIII. Project Features

The developed system incorporates several key functionalities designed to enable multilingual voice interaction, smart content generation, and audio-based input processing. These functionalities work together to provide a seamless user experience, especially for applications such as smart farming. The key functionalities are outlined below:

1) **Multilingual Voice Input:** The system allows users to interact in multiple languages by capturing voice input through the microphone. Upon capturing the audio, the system processes it using the `speech_recognition` library to convert speech into text in the selected language (e.g., Hindi, Marathi, Bengali). This ensures accessibility for users from diverse linguistic backgrounds.
2) **Noise Reduction and Audio Cleaning:** To improve the quality of the recorded voice input, the system employs noise reduction techniques. These methods eliminate unwanted background noise, ensuring clearer and more accurate transcription. The audio is then normalized to a consistent volume level, making it easier for the system to process.
3) **Speech-to-Text (STT):** The Speech-to-Text functionality utilizes Google's Speech-to-Text API to convert the captured audio into text. The transcribed text is then used for further processing, such as querying the FAQ database or generating responses. This functionality is central to enabling voice-based interaction.
4) **Text Translation:** Once the user's speech is transcribed into text, the system can translate the text into a different language (e.g., Hindi to English) using the `googletrans` library. This functionality ensures that the system can handle queries in various languages and provide responses in the preferred language of the user.
5) **FAQ-Based Response Generation:** For common queries, the system uses a predefined FAQ database. By utilizing word embeddings, such as Word2Vec or GloVe, the system compares the user's transcribed query with the FAQs using cosine similarity. The closest matching FAQ is retrieved and returned as the system's response, providing quick and relevant answers.
6) **Smart Content Generation using Gemini AI:** For domain-specific queries, such as those related to farming, the system uses Google's Gemini AI to generate intelligent, contextually appropriate responses. Gemini AI simplifies complex information, ensuring that the generated responses are both understandable and relevant to the user's inquiry.
7) **Speech Synthesis (Text-to-Speech, TTS):** The system converts the text-based responses back into speech using the `gTTS` (Google Text-to-Speech) engine. This functionality enables the system to provide spoken responses, making it fully interactive. The audio response is played back to the user, ensuring an auditory form of feedback.
8) **Visualization of Audio Signals:** For debugging or educational purposes, the system generates visual representations of the audio input. Using `librosa` and `matplotlib`, the system produces waveforms and spectrograms of the captured audio, providing insight into its structure.
9) **Recording Audio from Microphone:** The system records audio from the user's microphone using the `pyaudio` library. This functionality is initiated when the user clicks the "Record" button, capturing their speech for further processing.
10) **Audio Logging:** The system logs all interactions, in-

cluding user queries, translations, and responses, into a text file. This log is useful for tracking system performance, analyzing user interactions, and improving future responses.

11) **Smart Farming Query Detection:** The system is designed to detect whether a user's query pertains to farming. By scanning for relevant keywords, such as "crop," "fertilizer," or "field," the system activates the Gemini AI to generate a tailored response for farming-related queries.
12) **Audio Output in Multiple Languages:** After generating the response, the system converts it into speech in the selected language using the Text-to-Speech engine. This ensures that users from different linguistic backgrounds can receive spoken responses in their preferred language.
13) **System Feedback:** The system provides feedback at each stage of interaction, such as when the microphone is recording, when a response is being generated, and when the response is being played. This feedback ensures that users are aware of the system's status and actions.

REFERENCES

[1] Patel, Neil, et al. "Experiences designing a voice interface for rural India." 2008 IEEE Spoken Language Technology Workshop. IEEE, 2008.

[2] Toshniwal, Shubham, et al. "Multilingual speech recognition with a single end-to-end model." 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2018.

[3] Mehta, Devansh, et al. "Learnings from technological interventions in a low resource language: a case-study on Gondi." arXiv preprint arXiv:2004.10270 (2020).

[4] Rajalakshmi, N. R. "Dynamic NLP enabled chatbot for rural health care in India." 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA). 2022.

[5] Charan, G., et al. "Unveiling the challenges of speech recognition in noisy environments: A comprehensive review of issues and solutions." Challenges in Information, Communication and Computing Technology (2025): 407-412.

[6] Liu, Y., Zhang, J., Xiong, H., Zhou, L., He, Z., Wu, H., ... Zong, C. (2020, April). Synchronous speech recognition and speech-to-text translation with interactive decoding. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 05, pp. 8417-8424).

[7] Vemula, Venkata Vinay Babu, et al. "ANUVAADHAK: a two-way, Indian language speech-to-speech translation system for local travel information assistance." Int. J. Eng. Sci. Technol 2 (2010).

[8] Shruti Gupta and Parteek Kumar, "Comparative study of text to speech system for Indian language," *International Journal of Advances in Computing and Information Technology*, vol. 1, no. 2, pp. 199–209, Apr. 2012

[9] Mujadia, Vandan, and Dipti Misra Sharma. "Towards speech to speech machine translation focusing on indianlanguages." Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations. 2023.

[10] Mhaskar, Shivam, et al. "Vakta-setu: A speech-to-speech machine translation service in select indic languages." arXiv preprint arXiv:2305.12518 (2023).

[11] Cyril, Sheila, et al. "Exploring the role of community engagement in improving the health of disadvantaged populations: a systematic review." Global health action 8.1 (2015): 29842.

[12] Perez, Kinalyne, et al. "Investigation into Application of AI and Telemedicine in Rural Communities: A Systematic Literature Review." Healthcare. Vol. 13. No. 3. MDPI, 2025.

[13] Mensah, Justice. "Sustainable development: Meaning, history, principles, pillars, and implications for human action: Literature review." Cogent social sciences 5.1 (2019): 1653531.

[14] Seligman, Mark, et al. "Speech-enabled language translation system and method enabling interactive user supervision of translation and speech recognition accuracy." U.S. Patent No. 8,239,207. 7 Aug. 2012.

[15] Wallerstein, Nina, and Bonnie Duran. "Community-based participatory research contributions to intervention research: the intersection of science and practice to improve health equity." American journal of public health 100.S1 (2010): S40-S46.

[16] Dogga Paavn Sekhar "Multilingual Voice-Based Farmer Assistant System," *YouTube*, 2025. Available: https://youtu.be/qd2ziS6oeRA