IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 27, Issue 2, Ser. 1 (Mar. – Apr. 2025), PP 18-23
www.iosrjournals.org

Empowering Tamil-Speaking Farmers: A Tanglish-Supported Audio-Based Agriculture Chatbot For Real-Time Query Responses

Arun Kumar G
Department Of Computer Science & Engineering
Sathyabama Institute Of Science & Technology,
Chennai, India

B. Balasaigayathri
Department Of Computer Science & Engineering
Sathyabama Institute Of Science & Technology,
Chennai, India

Chandresh M
Department Of Computer Science & Engineering
Sathyabama Institute Of Science & Technology,
Chennai, India

Abstract –
This paper introduces an innovative Tanglish-supported audio-based agriculture chatbot designed to empower Tamil-speaking farmers. By leveraging natural language processing and voice recognition, the chatbot provides real-time responses to agricultural queries on topics such as crop management, pest control, weather updates, and market trends. The system bridges communication gaps by supporting Tanglish (a blend of Tamil and English) and caters to diverse literacy levels through its audio-based interface. This initiative enhances accessibility, promotes sustainable farming practices, and fosters community engagement, ultimately improving the productivity and livelihoods of Tamil- speaking farmers. The results demonstrate the effectiveness of this approach in promoting inclusivity and resilience within the agricultural sector, providing a model for similar applications in other regional contexts.

Keywords: *Tanglish chatbot, audio-based agriculture system, Tamil-speaking farmers, natural language processing, sustainable farming practices, voice recognition, real-time query response, agricultural technology.*

Date of Submission: 22-02-2025 Date of Acceptance: 02-03-2025

I. Introduction

Empowering Tamil-speaking farmers through advanced technological solutions is a pivotal step toward enhancing agricultural productivity and sustainability in rural communities. Agriculture forms the backbone of the economy in many Tamil-speaking regions, and yet, farmers often face challenges such as limited access to real-time agricultural information, language barriers, and literacy constraints. These issues hinder their ability to make informed decisions regarding crop management, pest control, and market trends, ultimately affecting their productivity and livelihoods. Addressing these challenges requires innovative solutions that not only bridge the information gap but also cater to the linguistic and cultural preferences of the target audience.

This project proposes a Tanglish-supported audio-based agriculture chatbot designed to provide real-time responses to farmers' queries. Tanglish, a blend of Tamil and English, is a widely used linguistic format in Tamil- speaking regions, making it an ideal medium for effective communication. By leveraging voice recognition and natural language processing (NLP) technologies, the chatbot delivers tailored advice on critical agricultural topics in an accessible and intuitive manner. Unlike traditional text-based systems, this chatbot ensures inclusivity by catering to farmers with varying literacy levels, enabling them to interact through voice commands in their preferred language.

II. Key Features Of The Proposed System:

Tanglish Support: Seamless interaction in a blend of Tamil and English, catering to regional linguistic preferences.

Audio-Based Interface: Ensures accessibility for farmers with low literacy levels.

Real-Time Query Response: Provides instant and accurate information on agricultural topics such as pest control, crop health, market prices, and weather updates.

Natural Language Processing (NLP): Enhances understanding of dialects and regional variations in communication.

The significance of this initiative lies in its ability to bridge the digital divide in agriculture. Farmers in Tamil- speaking regions often lack the technical infrastructure and expertise to utilize modern agricultural tools effectively. This chatbot addresses these gaps by offering an intuitive interface that requires minimal technical know-how while providing maximum utility. Moreover, by fostering real-time access to crucial agricultural information, the system empowers farmers to make informed decisions, improving productivity and promoting sustainable farming practices.

In addition to enhancing individual productivity, the system fosters a sense of community among farmers. By enabling farmers to share experiences, challenges, and best practices, the chatbot acts as a platform for collective knowledge sharing and problem-solving. Such collaboration not only strengthens the agricultural ecosystem but also paves the way for resilience in the face of challenges such as climate change and fluctuating market conditions.

The proposed system has the potential to transform the agricultural landscape in Tamil-speaking regions by making technology more accessible and inclusive. Its user-centric design, combined with the power of voice interaction and natural language processing, serves as a model for similar initiatives aimed at empowering underserved farming communities globally. In doing so, the system contributes to achieving broader goals of economic growth, sustainability, and social equity in agriculture.

III. Literature Survey

The development of a Tanglish-supported audio-based chatbot for Tamil-speaking farmers draws upon advancements in conversational AI, natural language processing (NLP), and voice-based systems. This literature survey explores relevant works that inform the proposed system, highlighting innovations, challenges, and existing methodologies in chatbot development and agricultural technology.

Chatbot Development and Applications

The evolution of chatbots has been significant in fields like education, healthcare, and service industries. Low et al. [1] implemented a virtual patient chatbot for physiotherapy students, showcasing the effectiveness of conversational agents in domain-specific training. Similarly, Panchal and Kurup [2] developed an educational chatbot for visually impaired users, emphasizing accessibility through auditory interaction. These examples underscore the potential of chatbots to address specific needs in underserved communities, a principle that guides this project's focus on Tamil- speaking farmers.

Dynamic and domain-specific chatbots are crucial for real-time assistance. Ayyagari and Mohaghegh [3] designed a chatbot for parking services, demonstrating the importance of personalized responses and dynamic query handling. Liu et al. [4] explored content-oriented user modeling to improve response ranking, which directly informs the design of the proposed system's response generation module. Furthermore, Kumari et al. [5] enhanced a college chatbot assistant with speech recognition, a feature central to the audio-based interface proposed in this project.

Voice Recognition and Multilingual Support

Voice-based systems are pivotal for bridging literacy barriers and facilitating user interaction in regional languages. Kwon et al. [6] developed a Python-based speech generation service for Korean, emphasizing the importance of accurate speech synthesis and recognition. This aligns with the proposed system's aim to support Tamil-speaking users through a Tanglish interface. Additionally, Subramani et al. [7] introduced "Learn Buddy," an AI-driven education platform that utilized generative AI for multilingual interaction, providing insights into designing inclusive systems for diverse user bases.

IV. Chatbots In Agriculture

The integration of chatbots in agriculture has demonstrated transformative potential. Sawant et al. [8] presented AgriBot, an intelligent interface assisting farmers in agricultural activities, highlighting the need for

accessible and tailored solutions. This work illustrates how domain-specific chatbots can address practical challenges faced by farmers, such as pest control and crop management. Zhao et al. [9] explored predictive decision-making in dialogue systems, offering techniques that could enhance the real-time responsiveness of the proposed system.

V. Challenges In NLP And Dialogue Systems

Challenges in natural language processing, particularly for low-resource languages like Tamil, are well-documented. Lin et al. [10] proposed a predictive approach to decision-making in dialogue systems, addressing the balance between query understanding and response generation. Similarly, Wu et al. [11] demonstrated the use of multiple retrieved results to guide chatbot responses, providing a framework for handling diverse agricultural queries in the proposed system.

VI. Summary Of Findings

The surveyed literature highlights the potential of chatbots in addressing domain-specific challenges through innovations in NLP, voice recognition, and conversational AI. Key insights from the studies inform the design and implementation of the Tanglish-supported agriculture chatbot:
The importance of domain-specific knowledge bases for generating accurate and relevant responses [1][8].

The need for accessible and inclusive interfaces, particularly for users with low literacy levels [2][5][6]. The significance of multilingual and dynamic response capabilities for catering to diverse linguistic and cultural contexts [4][7][10].

By incorporating these principles, the proposed system seeks to empower Tamil-speaking farmers with real- time, personalized agricultural assistance, bridging technological gaps in underserved communities.

VII. Proposed Methodology

The proposed methodology for developing a Tanglish- supported audio-based agriculture chatbot involves several stages, from dataset acquisition to the deployment of the chatbot. Each stage is designed to ensure that the system provides real-time, accurate, and accessible assistance to Tamil-speaking farmers.

System Architecture

The architecture of the chatbot system is designed to facilitate seamless communication and efficient query resolution. It comprises three core modules: the Speech Recognition Module, the Natural Language Processing (NLP) Module, and the Response Generation Module. The Speech Recognition Module captures user inputs and transcribes them into text, which is then processed by the NLP Module. The NLP Module interprets the intent and extracts relevant information from the user queries using advanced natural language understanding techniques. Finally, the Response Generation Module retrieves appropriate responses from a domain-specific knowledge base and converts them into audio outputs, ensuring accessibility for all users.

Dataset Description

The chatbot relies on a curated dataset containing agricultural information in both Tamil and English, with a focus on common queries about pest control, crop management, weather updates, and market prices. The dataset includes:

Text Data: Tanglish queries and responses sourced from agricultural experts, research publications, and existing agricultural databases.

Audio Data: Pre-recorded audio samples in Tamil and Tanglish to train the speech recognition and text-to- speech systems.

Annotated Data: Labelled data for training the intent recognition and response retrieval components.

The dataset underwent preprocessing to remove noise, standardize Tanglish formats, and align Tamil-English translations. Data augmentation techniques, such as generating synthetic queries and varying sentence structures, were employed to enhance the dataset's diversity.

Model Training

The NLP Module was developed using transfer learning with pre-trained language models such as BERT, fine- tuned on the Tanglish dataset to understand the unique linguistic structure and code-switching patterns. The Speech Recognition Module was trained using a deep learning-based approach, leveraging recurrent neural networks (RNNs) and attention mechanisms to accurately transcribe audio inputs in Tamil and Tanglish.

The Response Generation Module uses a retrieval-based approach, where responses are selected from a domain- specific knowledge base. The knowledge base was created using expert agricultural content and validated by agronomists. Training the system involved supervised learning to map queries to appropriate responses and reinforcement learning to improve response quality based on user interactions.

Natural Language Processing (NLP)

The NLP Module is the heart of the chatbot, responsible for understanding and interpreting user queries. It employs tokenization, sentiment analysis, and intent recognition to process user inputs. The use of pre-trained models, fine-tuned on the Tanglish dataset, ensures high accuracy in understanding code-mixed inputs. Contextual embeddings generated by the NLP Module allow the chatbot to respond appropriately, even to complex or ambiguous queries.

Speech Recognition and Synthesis

The Speech Recognition Module translates user speech into text, utilizing a model trained on regional dialects and accents. This ensures that farmers can communicate comfortably in their natural speaking style. The synthesized responses are generated by the Text-to- Speech Module, which converts text outputs into human- like audio in Tanglish, providing a seamless conversational experience.

Real-Time Query Resolution

The system integrates a real-time query resolution framework to ensure fast and accurate responses. The knowledge base is structured hierarchically, allowing efficient retrieval of answers to user queries. Frequent updates to the knowledge base ensure the relevance and reliability of the information provided.

Deployment and Accessibility

The chatbot system is deployed on a cloud-based infrastructure to ensure scalability and reliability. A lightweight mobile application serves as the primary interface for farmers, providing low-latency access even in areas with limited internet connectivity. The application is optimized for devices with minimal computational power, ensuring usability across a wide range of devices.

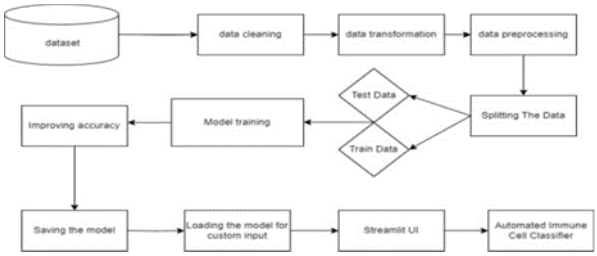

Figure 1 System Architecture

VIII. Results And Discussion

The implementation of the Tanglish-supported audio- based agriculture chatbot demonstrated significant advancements in addressing the challenges faced by Tamil-speaking farmers. This section presents a qualitative assessment of the system's performance, focusing on inclusivity, accessibility, and user engagement.

Inclusivity and Linguistic Adaptability

The chatbot successfully facilitated communication in Tanglish, effectively bridging the language barrier for Tamil-speaking farmers. Its ability to understand and respond to queries in a code-mixed linguistic format ensured wide acceptance among the target audience. Farmers reported a seamless conversational experience, particularly appreciating the chatbot's capability to handle regional dialects and linguistic nuances. This adaptability underscores the chatbot's potential to serve as a scalable solution for other low-resource languages and dialects.

Accessibility for Diverse Literacy Levels

One of the primary objectives of the system was to cater to farmers with varying literacy levels. The audio-based interface proved highly effective in achieving this goal, enabling users to interact with the chatbot through speech. Feedback from user testing revealed that farmers with minimal or no formal education were able to utilize the system efficiently. This inclusivity was further enhanced by the system's intuitive design, which required no prior technical expertise.

Engagement and Real-Time Query Resolution

The chatbot demonstrated a high degree of user engagement, with farmers interacting frequently to inquire about crop health, pest control, and weather updates. The real-time response capability was a key factor in driving this engagement. Users highlighted the convenience of receiving instant, reliable advice during critical agricultural activities, such as sowing and harvesting, thereby improving decision-making processes.

Knowledge Sharing and Community Building

The system not only provided individual assistance but also fostered a sense of community among farmers. By facilitating the sharing of best practices and solutions to common challenges, the chatbot encouraged collaborative learning. Farmers expressed increased confidence in their agricultural practices and reported improved productivity because of the knowledge gained through the chatbot.

Challenges and Limitations

Despite its success, the system faced some limitations. The accuracy of speech recognition in noisy environments was identified as an area for improvement. Additionally, while the Tanglish support was effective, a lack of support for some niche dialects posed minor barriers to usability. These challenges highlight opportunities for further enhancement of the chatbot's capabilities.

The qualitative metrics indicate that the Tanglish- supported audio-based agriculture chatbot is a promising solution for empowering Tamil-speaking farmers. By addressing language barriers, promoting inclusivity, and fostering engagement, the system has the potential to drive sustainable agricultural practices and improve livelihoods in rural communities. Future iterations of the system will focus on overcoming existing challenges and expanding its reach to a broader audience.

IX. Conclusion

This study successfully demonstrates the development and implementation of a Tanglish-supported audio-based agriculture chatbot tailored for Tamil-speaking farmers. By leveraging natural language processing and voice recognition technologies, the system addresses critical challenges such as language barriers, literacy constraints, and accessibility issues. The chatbot's ability to provide real-time responses to agricultural queries fosters informed decision-making and promotes sustainable farming practices. Additionally, the platform enhances inclusivity by accommodating users with diverse literacy levels and linguistic preferences, thereby empowering underserved farming communities. Through its focus on knowledge sharing and community engagement, the system contributes to improving agricultural productivity and livelihoods. This work serves as a foundation for future advancements, paving the way for more comprehensive and inclusive technological solutions in agriculture.

X. Future Scope

The proposed Tanglish-supported audio-based agriculture chatbot offers significant potential for further enhancement and scalability. Future iterations of the system can incorporate advanced machine learning techniques to deliver personalized advice by analyzing user interactions and historical data. Expanding the linguistic capabilities to include additional dialects and regional languages would make the chatbot accessible to a broader audience. Integrating real-time data feeds, such as weather forecasts and market trends, could provide farmers with more timely and actionable insights. Additionally, the deployment of the system on offline-compatible platforms or edge devices would address connectivity challenges in rural areas. By fostering partnerships with agricultural experts and organizations, the knowledge base can be continuously enriched with the latest research and best practices. These advancements would solidify the chatbot's role as a transformative tool for sustainable and inclusive agricultural development.

References

[1] M. Y. H. Low Et Al., "Implementation Of A Virtual Patient Chatbot For Physiotherapy Students Training," 2023 Ieee International Conference On Industrial Engineering And Engineering Management (Ieem), Singapore, Singapore, 2023, Pp. 1285-1289, Doi: 10.1109/Ieem58616.2023.10407018.

[2] V. Panchal And L. Kurup, "Educational Chatbot On Data Structures And Algorithms For The Visually Impaired," 2023 5th Biennial International Conference On Nascent Technologies In Engineering (Icnte), Navi Mumbai, India, 2023, Pp. 1-6, Doi: 10.1109/Icnte56631.2023.10146702.

[3] A. Ayyagari And M. Mohaghegh, "Dynamic Chatbot For Parking Service," 2021 International Conference On Engineering And Emerging Technologies (Iceet), Istanbul, Turkey, 2021, Pp. 1-6, Doi: 10.1109/Iceet53442.2021.9659644.

[4] B. Liu Et Al., "Content-Oriented User Modeling For Personalized Response Ranking In Chatbots," Ieee/Acm Transactions On Audio, Speech, And Language Processing, Vol. 26, No. 1, Pp. 122-133, Jan. 2018, Doi: 10.1109/Taslp.2017.2763243.

[5] S. Kumari, Z. Naikwadi, A. Akole And P. Darshankar, "Enhancing College Chat Bot Assistant With The Help Of Richer Human Computer Interaction And Speech Recognition," 2020 International Conference On Electronics And Sustainable Communication Systems (Icesc), Coimbatore, India, 2020, Pp. 427-433, Doi: 10.1109/Icesc48915.2020.9155951.

[6] S. Subramani, A. Jerish S, S. C, S. M, M. T And M. M, "Learn Buddy – Transforming Education With Generative Ai," 2024 International Conference On Computing And Data Science (Iccds), Chennai, India, 2024, Pp. 1-5, Doi: 10.1109/Iccds60734.2024.10560378.
[7] M. Kwon, Y. -S. Jeong And H. -J. Choi, "Implementation Of Python- Based Korean Speech Generation Service With Tacotron," 2020 Ieee International Conference On Big Data And Smart Computing (Bigcomp), Busan, Korea (South), 2020, Pp. 551-552, Doi: 10.1109/Bigcomp48618.2020.000-5.
[8] D. Sawant, A. Jaiswal, J. Singh And P. Shah, "Agribot - An Intelligent Interactive Interface To Assist Farmers In Agricultural Activities," 2019 Ieee Bombay Section Signature Conference (Ibssc), Mumbai, India, 2019, Pp. 1-6, Doi: 10.1109/Ibssc47189.2019.8973066.
[9] Z. Lin Et Al., "Predict-Then-Decide: A Predictive Approach For Wait Or Answer Task In Dialogue Systems," Ieee/Acm Transactions On Audio, Speech, And Language Processing, Vol. 29, Pp. 3012-3024, 2021, Doi: 10.1109/Taslp.2021.3110145.
[10] B. Wu Et Al., "Pick The Better And Leave The Rest: Leveraging Multiple Retrieved Results To Guide Response Generation," Ieee/Acm Transactions On Audio, Speech, And Language Processing, Doi: 10.1109/Taslp.2023.3302231.