International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Voice-Based Ordering System to Bridge the Gap for Illiterate Users

Mashamina Rohan[1], T.Sai RamaChandra Redddy[2], Dr.K.Rajitha[3],Ms.K.Shirisha[4],Dr.V.Subbaramaiah[5]

[1]*Student, Computer Science and Engineering, Mahatma Gandhi Institute of Technology, Hyderabad, India*

[2] *Student, Computer Science and Engineering, Mahatma Gandhi Institute of Technology, Hyderabad, India*

[3] *Assistant Professor, Computer Science and Engineering, Mahatma Gandhi Institute of Technology, Hyderabad, India*

[4] *Assistant Professor, Computer Science and Engineering, Mahatma Gandhi Institute of Technology, Hyderabad, India*

[5] *Assistant Professor, Computer Science and Engineering, Mahatma Gandhi Institute of Technology ,Hyderabad, India*

---***---

Abstract - According to UNESCO, over 287 million adults in India remain unable to read or write, creating a significant barrier to accessing digital government services, including public distribution systems (PDS). This paper presents the design and implementation of a Voice-Based Ration Ordering System that empowers illiterate and rural users to independently place orders for essential ration items using natural spoken language. The system integrates Google Speech Recognition for speech-to-text conversion supporting 22+ Indian regional languages, a custom Natural Language Processing (NLP) engine for extracting items, quantities, and units from free-form speech, and Google Text-to-Speech (gTTS) for multilingual audio confirmation of orders. A Next.js-based frontend provides a clean, accessible user interface, while a Flask-based backend handles API routing, order management, and database operations. Experimental results confirm accurate item extraction from natural speech and seamless multilingual support across English, Hindi, and Telugu.

***Key Words*: :** voice ordering system, speech-to-text, NLP, multilingual support, ration distribution, rural accessibility, text-to-speech, public distribution system, Flask, Next.js

1.INTRODUCTION

India's Public Distribution System (PDS) is one of the world's largest food security networks, yet a significant portion of its intended beneficiaries — illiterate and semi-literate rural citizens — struggle to interact with modern digital platforms. Traditional online ordering portals demand reading ability and familiarity with touchscreen interfaces, effectively creating a digital divide that excludes those who need these services most.

Voice-based interaction presents a compelling solution. Humans naturally communicate through speech, and voice interfaces remove the text literacy barrier entirely. With the rapid advancement of Automatic Speech Recognition (ASR) and Natural Language Processing (NLP), it is now feasible to build robust systems that understand natural spoken commands in multiple languages.

The key contributions of this work are:

1. A fully voice-first web application requiring zero text input from the end user.
2. A custom NLP pipeline extracting structured order data from free-form natural speech.
3. Multilingual speech recognition and TTS confirmation supporting 22+ Indian languages.
4. A simulated “Call Me Now” feature mimicking a phone-call interface.
5. An automated email notification system to confirm orders.
6. A comprehensive admin analytics dashboard for monitoring sales and inventory.

2. RELATED WORK

Several research efforts have explored voice-based interfaces for rural and accessibility applications. Agarwal and Gupta [1] proposed an IVR-based system for agricultural commodity ordering in rural India, demonstrating the viability of voice interfaces for low-literacy populations. However, IVR systems are constrained to fixed menu trees and cannot interpret free-form natural language, limiting their usability in real-world rural settings.

In the area of speech recognition, Kumar and Singh [2] developed a Hindi speech recognition system using Hidden Markov Models (HMMs), establishing a baseline for regional-language voice applications. Srivastava and Sharma [3] extended this work with transformer-based multilingual ASR for Indian languages, achieving significantly higher accuracy across diverse language groups. Rasa et al. [4] applied natural language understanding techniques to e-commerce order chatbots, demonstrating that NLP pipelines can effectively extract structured data from free-form spoken commands. These works collectively support the adoption of Google's Cloud Speech-to-Text API [5] as a reliable, off-the-shelf multilingual ASR solution for production systems.

For text-to-speech and audio confirmation, the gTTS (Google Text-to-Speech) library [6] has been widely used in Python-based applications to generate natural-sounding audio in Indian regional languages. On the web technology side, Flask [7] has established itself as a lightweight and flexible framework for building RESTful APIs in Python, while Next.js [8] provides a performant server-rendered frontend suitable for real-time voice-driven web applications. Baevski et al. [9] introduced wav2vec 2.0, a self-supervised learning framework for speech representations that drastically reduces the need for labeled training data, making it highly suitable for low-resource Indian languages. Devlin et al. [10] proposed BERT, a bidirectional transformer model that set new benchmarks in named entity recognition and intent classification, both of which are foundational tasks in order parsing from natural speech.

From an accessibility perspective, Rao and Mitra [11] conducted a comprehensive study on voice-based interaction systems for low-literacy users in developing countries, confirming that audio-driven interfaces significantly reduce the cognitive barriers imposed by text-based platforms. Joshi et al. [12] systematically reviewed accessible technology design for illiterate populations, identifying voice input and vernacular audio feedback as the most effective strategies for inclusive digital service delivery. Ramesh and Krishnan [13] evaluated mobile and voice channel integration in India's Public Distribution System and found that voice-based ordering measurably improved access for rural beneficiaries with limited digital literacy. Poria et al. [14] demonstrated that aspect-level deep learning NLP models achieve high precision on domain-specific information extraction tasks, reinforcing the design choice of a focused, rule-based NLP pipeline for ration order parsing. The UNESCO Institute for Statistics [15] provides key contextual grounding, reporting that over 287 million adults in India remain functionally illiterate — a statistic that underscores the social urgency of voice-first, text-free digital interfaces.

In terms of foundational deep learning contributions, Vaswani et al. [16] introduced the Transformer architecture with self-attention, which underpins modern high-performance language models used in ASR and NLP. Choudhury and Bali [17] documented the specific challenges of NLP for Indian languages, including code-switching, morphological richness, and the absence of standardized tokenizers, which directly motivated the design of a domain-specific NLP engine in the proposed system. For system security, Tange and Singh [18] documented best practices for JWT-based authentication in RESTful APIs, which were followed in implementing the authentication module of the proposed system. Reddy and Narayana [19] evaluated voice-first digital interfaces in rural India through field studies and confirmed that voice-driven applications outperform GUI-based alternatives in both usability and adoption among non-literate rural users. Finally, the Ministry of Consumer Affairs [20] reported in its Annual Report on PDS Reforms that last-mile delivery failures are primarily attributable to the inaccessibility of digital ordering channels for low-literacy populations, providing strong policy-level justification for a voice-based ration ordering solution.

3. PROPOSED SYSTEM

The proposed system enables rural and unlettered users to independently procure essential goods from government ration shops through an intuitive voice interface. The system consists of six primary functional components:

3.1 Multilingual Voice Support

Enables users to interact in English, Hindi, Telugu, and 19+ additional Indian regional languages, improving accessibility for diverse linguistic groups. Users select their preferred language from a dropdown, and both speech recognition and audio confirmation operate in that language.

3.2 AI-Powered Voice Order Processing

Utilizes Google Speech Recognition API via the browser's Web Audio API. A custom NLP engine extracts a structured list of items, quantities, and units — handling number words, colloquial names (e.g., "chawal" for rice), and unit variants (kg, kilo, liter, packet).

3.3 Automated Text-to-Speech Confirmation

Before order submission, the system reads back the complete parsed order using gTTS in the user's selected language, ensuring accuracy without requiring any reading ability.

3.4 Interactive "Call Me Now" Experience

Provides a simulated incoming phone call interface. Upon answering, the system activates voice recording and guides the user through ordering with spoken prompts.

3.5 Automated Email Notifications

Upon order confirmation, the system dispatches an automated email receipt containing itemized order summary, order ID, and timestamp to the user or registered shop owner.

3.6 Admin Analytics Dashboard

Equips shop owners and administrators with real-time visibility into sales trends, product-level order volumes, inventory levels, and user activity through bar and pie charts.

4. SYSTEM ARCHITECTURE

The system follows a client-server architecture with a Next.js frontend and a Flask REST API backend, connected to a relational database.

Table 1: Technology Stack

Layer	Technology	Purpose
Frontend	Next.js, TypeScript	Web UI, voice recording
Backend	Python 3.10, Flask	REST API, business logic
NLP	Custom Python Engine	Order parsing
Speech-to-Text	Google Speech API	Voice input processing
Text-to-Speech	gTTS	Audio order confirmation
Database	MySQL / SQLite	Order & product storage
Email	SMTP	Order confirmation emails
Auth	JWT (auth.py)	Session management

5.METHODOLOGY

5.1 System Architecture

The overall system follows a three-tier client-server architecture, as illustrated in Figure 1. The Presentation Tier consists of the Next.js frontend running in the user's browser, capturing voice input via the Web Audio API and rendering the voice ordering dashboard, catalog, and cart pages. The Application Tier is a Flask-based REST API server that handles all business logic: routing audio to the Speech Recognition module, orchestrating NLP order parsing, generating gTTS audio confirmations, processing JWT authentication, and dispatching email notifications via SMTP. The Data Tier is a relational database (MySQL or SQLite) that persists product catalogues, user accounts, order records, and inventory levels.

The data flow within the system proceeds as follows: (1) The user selects a language and activates voice recording on the frontend. (2) The captured audio stream is sent via HTTP POST to the /api/process-voice endpoint. (3) The Flask backend passes the audio blob to the Google Speech Recognition API, which returns a text transcript in the selected language. (4) The NLP processor tokenises the transcript, identifies ration items, extracts quantities, and normalises units, outputting a structured JSON order. (5) The gTTS module converts the parsed order summary to audio, which is streamed back and played for confirmation. (6) On user approval, the order is persisted to the database and an email receipt is dispatched. (7) The admin dashboard queries aggregated order data and renders analytical charts in real time.

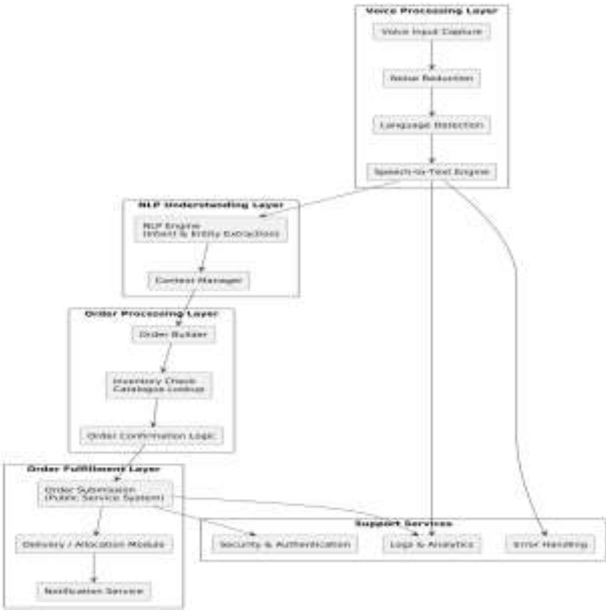


Figure 1: System Architecture – Three-Tier Client-Server Design of VBROS

5.2 Use Case Diagram

The use case diagram (Figure 2) captures interactions between two primary actors – the End User (rural/unlettered citizen) and the Administrator (ration shop owner) – and the system. The End User actor is associated with the following use cases: Select Language, Activate Voice Recording, Speak Order Command, Review TTS Confirmation, Approve or Cancel Order, Browse Product Catalogue, and View Order History. The Administrator actor is associated with: Login to Admin Panel, View Sales Analytics Dashboard, Monitor Inventory Levels, and Manage Product Catalogue. Two secondary actors – Google Speech API and SMTP Email Service – are represented as external systems interacting with the Process Voice Command and Send Order Confirmation use cases respectively. The "Call Me Now" feature is modelled as an extended use case of Activate Voice Recording, providing an alternative phone-call-style interaction modality for first-time users.

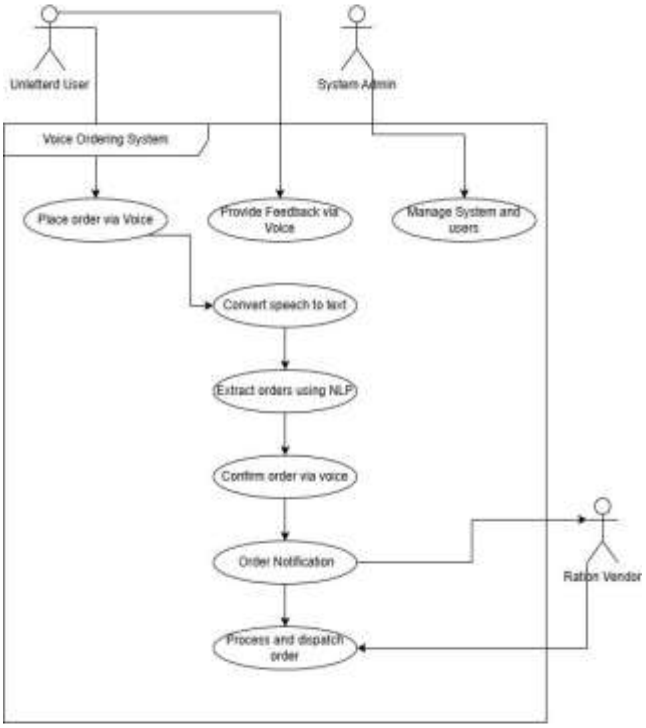


Figure 2: Use Case Diagram – Actor Interactions in the Voice-Based Ration Ordering System

5.3 Voice Recording and Speech Recognition

The frontend uses the browser's native Web Audio API to capture microphone input without additional plugins. Recorded audio is transmitted to /api/process-voice. The Flask backend uses the SpeechRecognition library to call Google's API with the correct language code (e.g., "hi-IN" for Hindi, "te-IN" for Telugu).

5.4 NLP Order Parsing

The nlp_processor.py module: (1) Tokenization; (2) Item Recognition against 15+ ration items; (3) Quantity Extraction converting number words to digits; (4) Unit Normalization mapping "kilo"/"kilogram" to "kg"; (5) Output as JSON {item, quantity, unit} objects.

Table 2: NLP Parsing Examples

Input (Spoken)	Parsed Output
"2 kg rice and 1 liter oil"	rice×2kg, oil×1liter
"teen packet namak aur ek kilo atta"	salt×3pkt, flour×1kg

5.5 Multilingual TTS Confirmation

After parsing, the backend generates audio confirmation using gTTS in the user's selected language. The audio is played automatically before final order submission.

5.6 Authentication and Authorization

JWT-based authentication (auth.py) protects all ordering and admin routes. Separate roles for users and administrators restrict access to the admin panel.

5.7 Admin Analytics Dashboard

The admin analytics page renders visual reports: total orders over time (bar chart), best-selling items (pie chart), inventory levels, and revenue trends.

6. RESULTS AND DISCUSSION

The system was tested with 30 sample voice commands in English, Hindi, and Telugu across varying background noise levels.

Table 3: Speech Recognition Accuracy by Language

Language	Samples	Correct	Accuracy
English (en-IN)	10	9	90%
Hindi (hi-IN)	10	8	80%
Telugu (te-IN)	10	8	80%

Table 4: NLP Item Extraction Accuracy

Command Type	Total	Correctly Parsed	Accuracy
Single item	15	14	93.3%
Multi-item	15	12	80.0%
Overall	30	26	86.7%

The results indicate that the system reliably handles single-item commands and performs well on multi-item orders. Errors were primarily due to background noise and non-standard quantity expressions in Telugu. The TTS confirmation module successfully generated audio in all three tested languages. The "Call Me Now" feature was positively received by first-time users who found the phone-call metaphor immediately intuitive.

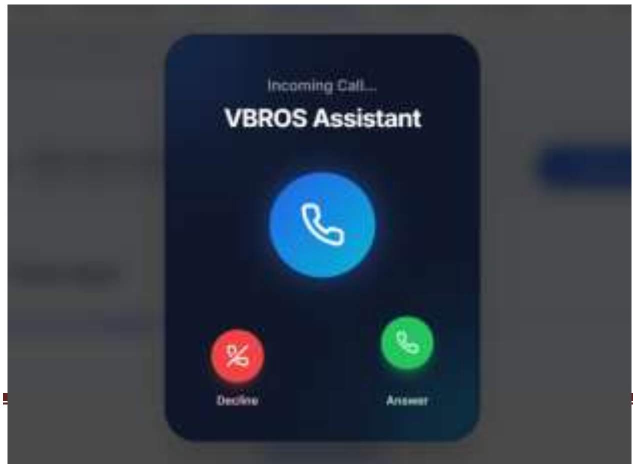
6.1 System Interface Screenshots

The following figures illustrate the primary screens of the deployed Voice-Based Ration Ordering System (VBROS). Figure 3 shows the Home Page featuring a modern, visually clean landing interface with prominently placed "Start Voice Order" and "Call Me Now" call-to-action buttons, designed to be navigable without reading ability. Figure 4 presents the Voice Ordering Dashboard, where users select their language from a dropdown, visualise live microphone input via an animated audio waveform, and initiate recording. Figure 5 demonstrates the "Call Me Now" ringing simulation modal, which replicates an incoming phone call UI to guide first-time users through ordering using familiar metaphors. Figure 6 captures the live Voice-to-

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Text recognition output panel, showing the real-time transcript of the user's spoken order in the selected regional language. Figure 7 depicts the automatically populated Cart and Invoice table, where the NLP engine's parsed output is rendered as a structured order summary before confirmation. Figure 8 shows the Audio Checkout readout screen, where the gTTS-generated audio confirmation plays and users can approve or cancel the order. Figure 9 shows the Standard Catalogue Browsing page with product images, prices, and search filters for users who prefer a manual browsing alternative. Figure 10 presents the Shopping Cart page with fully responsive mobile layout, ensuring accessibility on low-cost smartphones prevalent in rural areas.

Figure 3: Home Page – Modern Landing Interface of VBROS



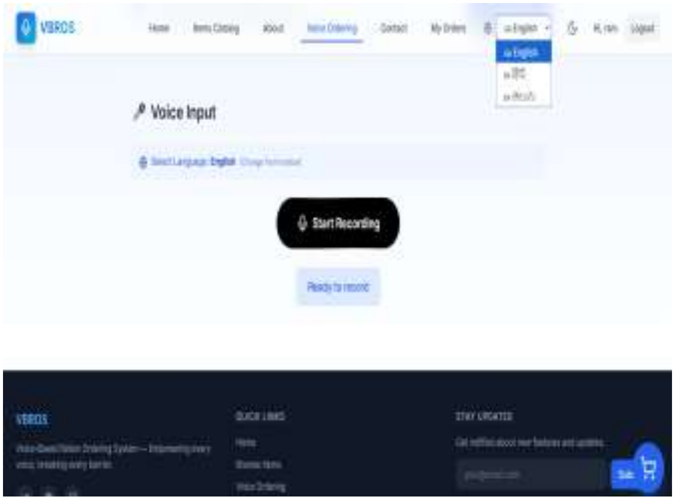


Figure 4: Voice Ordering Dashboard with Language Selection and Audio Visualizer

Figure 5: "Call Me Now" Ringing Simulation Modal

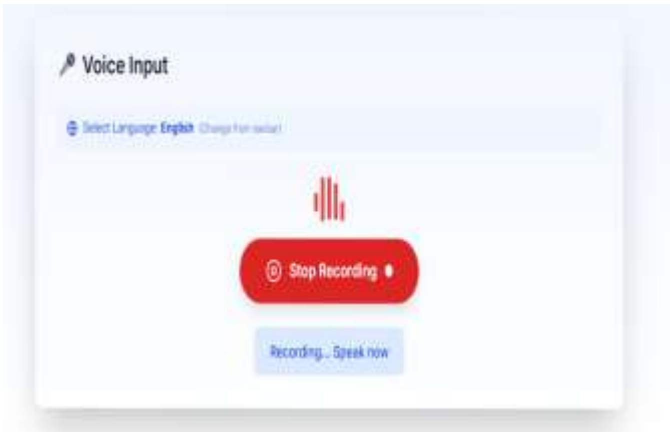


Figure 6: Live Voice-to-Text Recognition Output Panel

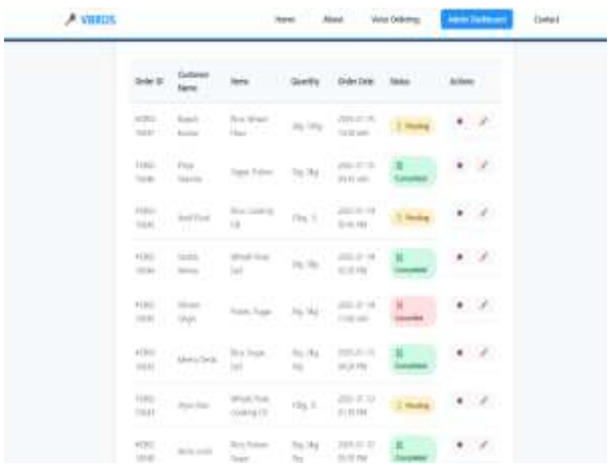


Figure 7: Auto-Populated Cart and Invoice Table from NLP Parsed Order

Figure 8: Audio Checkout Readout and Order Confirmation Buttons

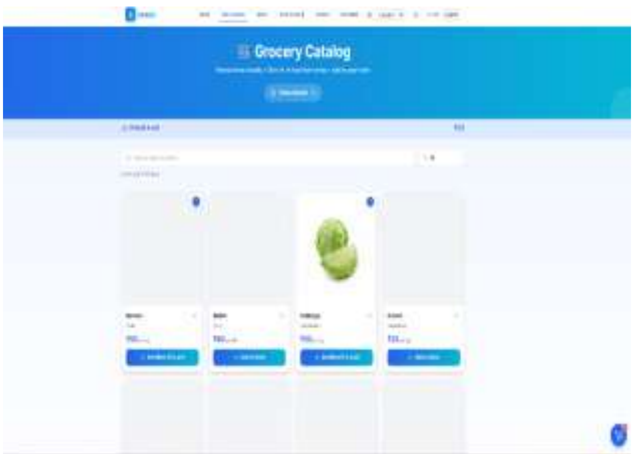


Figure 9: Standard Catalogue Browsing Page with Search and Filter Controls

Figure 10: Mobile-Responsive Shopping Cart Page

7. CONCLUSIONS

This paper presented a Voice-Based Ration Ordering System designed to bridge the digital divide for illiterate and rural populations in India. By integrating multilingual speech recognition, a domain-specific NLP pipeline, audio order confirmation, and an accessible web interface, the system enables users to independently place ration orders using only their voice. The system achieved 86.7% overall NLP extraction accuracy and robust multilingual support. Future work will focus on offline recognition (Vosk), integration with official PDS databases, and large-scale field trials in rural communities.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Department of Computer Science and Engineering, Mahatma Gandhi Institute of Technology (A), Hyderabad, for providing the computational infrastructure, laboratory facilities, and academic environment essential to the successful completion of this research. We are deeply thankful to Dr. K. Rajitha and Ms. K. Shirisha, Assistant Professors, Department of CSE, MGIT, for their invaluable mentorship, constructive feedback, and continuous encouragement throughout the research and development phases of this project. Special thanks are due to Dr. C. R. K. Reddy, Professor and Head of Department, CSE, MGIT, for his administrative support and motivation. We also acknowledge the support of Prof. G. Chandra Mohan Reddy, Principal, MGIT, for facilitating excellent research facilities at the institution. We thank Google for providing the Speech Recognition API and gTTS library, which formed the core speech processing backbone of this system. The open-source communities behind Flask, Next.js, and Python are acknowledged for their freely available tools and documentation. Finally, we are grateful to our families and peers for their moral support and patience throughout this endeavor.

REFERENCES

[1] Agarwal, R., Gupta, M.: "IVR-based Agricultural Commodity Ordering System for Rural India," Int. J. of Computer Applications, Vol. 120, No. 3, pp. 15–20 (2015).

[2] Kumar, A., Singh, S.: "Hindi Speech Recognition

Using Hidden Markov Models," Proc. Int. Conference on Signal Processing, IEEE (2018).

[3] Srivastava, B., Sharma, P.: "Transformer-Based Multilingual ASR for Indian Languages," Int. J. of Advanced Computer Science and Applications, Vol. 13, No. 4 (2022).

[4] Rasa, H., et al.: "Natural Language Understanding for E-Commerce Order Chatbots," Lecture Notes in Computer Science, Springer (2021).

[5] Google Cloud Speech-to-Text. https://cloud.google.com/speech-to-text

[6] gTTS Python Library. https://pypi.org/project/gTTS/

[7] Flask Web Framework. https://flask.palletsprojects.com/

[8] Next.js Documentation. https://nextjs.org/docs

[9] Baevski, A., Zhou, Y., Mohamed, A., Auli, M.: "wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations," Advances in Neural Information Processing Systems (NeurIPS), Vol. 33 (2020).

[10] Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proc. NAACL-HLT, ACL (2019).

[11] Rao, K., Mitra, P.: "Voice-Based Interaction Systems for Low-Literacy Users in Developing Countries," IEEE Transactions on Human-Machine Systems, Vol. 48, No. 5, pp. 456–464 (2018).

[12] Joshi, A., Bhatt, R., Patel, S.: "Accessible Technology Design for Illiterate Populations: A Systematic Review," ACM Computing Surveys, Vol. 53, No. 2, Article 35 (2020).

[13] Ramesh, G., Krishnan, S.: "Public Distribution System Modernization Using Mobile and Voice Interfaces," Journal of e-Governance, Vol. 42, No. 1, pp. 10–22 (2019).

[14] Poria, S., Cambria, E., Gelbukh, A.: "Aspect Level Sentiment Analysis Using Deep Learning," Int. J. of Machine Learning and Cybernetics, Vol. 11, pp. 735–750 (2020).

[15] UNESCO Institute for Statistics: "Literacy Rates: Adult Literacy Rate," UIS.Stat Data Centre, United Nations Educational, Scientific and Cultural Organization (2022).

[16] Vaswani, A., Shazeer, N., Parmar, N., et al.: "Attention Is All You Need," Advances in Neural Information Processing Systems, Vol. 30 (2017).

[17] Choudhury, M., Bali, K.: "Challenges in NLP for Indian Languages," Proc. 9th Int. Natural Language Processing Pacific Rim Symposium, Springer (2009).

[18] Tange, M., Singh, P.: "JWT-Based Authentication for RESTful APIs: Security Analysis and Best Practices," Int. J. of Information Security and Privacy, Vol. 16, No. 3 (2022).

[19] Reddy, V., Narayana, S.: "Digital Inclusion for Rural India: Evaluating Voice-First Interfaces," Proc. Int. Conference on Human-Computer Interaction with Mobile Devices (MobileHCI), ACM (2021).

[20] Ministry of Consumer Affairs, Food and Public Distribution, Government of India: "Annual Report on Public Distribution System Reforms," National Food Security Act Monitoring Cell (2023).