iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78640

www.ijraset.com

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

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

AI-Powered Assistive Vision App

Shruti Telavane[1], Aditi Babu[2], Sai Bhilare[3], Aadarsh Mane[4], Prof. Leenata Parab[5]
Department of Computer Engineering, Pillai College of Engineering, New Panvel

Abstract: ***The AI-Powered Assistive Vision App is a mobile solution designed specifically for blind and visually impaired individuals. According to the World Health Organization (WHO), approximately 285 million people worldwide live with visual impairment, of whom 39 million are completely blind. These individuals face profound daily challenges including spatial navigation, access to printed information, digital communication, and emergency response. The app enables users to independently navigate their environment and accomplish daily tasks using only voice and gesture interaction, removing any requirement to see or visually interact with the screen. A key innovation is the Gesture Draw Interface, where users draw alphabet letters on the touchscreen to activate specific modules instantly — for example, drawing the letter 'L' launches the Location module, which fetches the user's GPS coordinates and dispatches their live location to saved emergency contacts via SMS. This gesture-shortcut layer complements voice commands and provides a rapid alternative when audio interaction is impractical. The Image-to-Text module uses on-device Optical Character Recognition (OCR) to capture text from any printed or handwritten source and reads it aloud in natural speech, making written content fully accessible without sighted assistance. The Contacts and Calling module allows users to call any saved contact entirely by voice, with the AI assistant confirming the contact name and initiating the call. Developed with Flutter for cross-platform Android and iOS deployment, and supported by Node.js and MongoDB for backend services, the app integrates on-device machine learning models for fully offline core functionality, ensuring reliability in environments with poor network connectivity.***

I. LITERATURE REVIEW

The literature on assistive technologies for visually impaired individuals underscores the transformative role of AI, deep learning, and multimodal systems in improving accessibility, navigation, and interaction. This section reviews significant research contributions in voice navigation, OCR-based text access, location-sharing systems, and gesture-based interfaces.

A. Voice-Driven Navigation and AI Assistants

Voice-based navigation systems have become central to accessible app design. Mateia and Alboaie (2023) [3] proposed a multimodal navigation interface fusing speech recognition, gesture input, and audio sonification for visually impaired users. Their system demonstrated that combining voice commands with secondary input modalities significantly reduces cognitive load compared to purely speech-driven systems. However, cloud-dependent speech recognition created latency issues in low-connectivity environments — a limitation addressed in the present app through offline Vosk-based speech recognition. General-purpose AI assistants such as Google Assistant are not optimized for the structured tasks of visually impaired users. The present app's AI assistant is purpose-built for accessibility workflows, with intent recognition tuned to module-specific voice commands.

B. OCR-Based Text-to-Speech for Accessibility

Ahmad and Rashid (2024) [5] reviewed deep learning-based TTS models including WaveNet, Tacotron, and FastSpeech, concluding that neural TTS has dramatically improved speech naturalness but that prosody control and multilingual support remain open challenges. The present app employs Google ML Kit for on-device OCR and FastSpeech 2 for synthesis, achieving a Character Error Rate of 3.1% in evaluation, competitive with published benchmarks.

Ingale and Bamnote (2024) [1] explored image captioning using deep learning for scene descriptions, highlighting the value of combining visual recognition with natural language generation — a principle adopted in the app's image processing pipeline.

C. Location Sharing and Emergency Response

Abidi et al. (2024) [4] conducted a comprehensive review of navigation systems for visually impaired individuals, finding that GPS-fused smartphone systems offer the best balance of accuracy and portability, but that trigger mechanisms for emergency alerts are often complex and inaccessible to blind users. The present app addresses this gap with two complementary triggers: a voice command and a gesture draw of the letter 'L', both dispatching GPS coordinates, a Google Maps link, and a timestamp to configured contacts via offline SMS.

©IJRASET: All Rights are Reserved

D. Gesture Recognition Interfaces

Lee et al. [6] developed a deep learning-based mobile assistive device incorporating gesture detection for module switching, showing that gesture shortcuts reduced task initiation time by 35% in noisy environments. The present app introduces a novel Gesture Draw Interface — where drawn letters map directly to module activations — combining the intuitiveness of character-based shortcuts with the speed of direct touchscreen gestures, without requiring any visual feedback.

II. EXISTING SYSTEM

Several assistive applications currently address subsets of the challenges faced by visually impaired users, but none integrates the full breadth of capabilities offered by the present app.

Microsoft Seeing AI provides real-time scene description, OCR, currency recognition, and barcode reading. While powerful in visual recognition, it relies on manual screen navigation to switch between modes, posing significant usability challenges for non-sighted users. It lacks voice-command-based contact calling, integrated location sharing, or an AI conversational assistant.

Google Lookout provides OCR and scene recognition features integrated with Android accessibility services. However, it functions as a passive listener rather than an interactive AI assistant — it cannot initiate calls, compose messages, or share emergency location data on command.

Be My Eyes connects blind users with sighted volunteers via live video call. While effective for complex visual tasks, it is dependent on volunteer availability, requires active internet, and offers no automated AI-driven functionality, real-time OCR, or location sharing.

Navigation apps such as BlindSquare and Nearby Explorer provide GPS-based turn-by-turn navigation but do not offer OCR, contact calling, AI assistant interaction, or gesture draw interfaces.

Primary gaps across existing systems: (1) no unified AI assistant responding to natural voice commands across all app functions; (2) no gesture draw interface for rapid module activation without visual feedback; (3) no integrated app combining location sharing, contact calling, image-to-text, and conversational AI; (4) limited offline capability. The present app addresses all four gaps simultaneously.

III. MODULE DESCRIPTION

The AI-Powered Assistive Vision App comprises five primary modules. Each module is independently accessible via voice command or the gesture draw interface. All modules communicate through a unified Flutter state management layer and are fully operable without any visual interaction.

A. AI Voice Assistant Module

The AI Voice Assistant is the central controller of the entire application. A wake-word detector ('Hey Vision') activates listening mode passively. Vosk provides fully offline speech recognition, with Google STT as an enhanced online fallback. Recognized intent is classified by an on-device NLP model and routed to the appropriate module. Responses are synthesized by FastSpeech 2 TTS. The assistant handles launching modules, calling contacts, reading text, sharing location, composing messages, checking weather, setting reminders, and confirming every action by voice.

B. Gesture Draw Interface

The Gesture Draw Interface allows users to draw alphabet letters on the touchscreen to instantly activate specific modules. The touchscreen canvas captures stroke trajectories, classified into 26 letter categories using a lightweight on-device CNN. Each letter maps to a module: 'L' activates Location Sharing; 'C' opens Contacts for voice-directed calling; 'T' activates Image-to-Text capture; 'V' activates the Voice Assistant. Haptic feedback confirms each recognized letter. This layer provides rapid shortcut access in noisy or socially constrained environments where spoken commands are impractical.

C. Image-to-Text (OCR) Module

The Image-to-Text module allows blind users to point the camera at any printed or handwritten text source and have its contents read aloud. Activated by voice ('Read this') or by drawing 'T', the camera captures a frame; Google ML Kit Text Recognition v2 performs on-device OCR. A post-processing pipeline removes artefacts, re-segments text into natural sentence units, and passes the result to FastSpeech 2 for speech synthesis. Language auto-detection supports English, Hindi, and Marathi. Core OCR processing functions fully offline.

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

D. Location Sharing Module

The Location module shares the user's GPS location with pre-configured contacts. Triggered by voice command, by drawing 'L', or by a 3-second volume-button hold. A five-second audio and haptic countdown allows cancellation. Fused location acquisition (GPS, Wi-Fi RSSI, cell-tower triangulation) provides coordinates accurate to within approximately 10 metres. An SMS containing the user's name, GPS coordinates, a Google Maps link, and UTC timestamp is dispatched to up to five contacts via the native SMS service — fully functional offline.

E. Contacts and Calling Module

The Contacts module provides fully voice-directed access to the device contact list and calling functionality. Activated by voice command ('Call [name]') or by drawing 'C', the AI assistant searches contacts for the specified name, reads back the match for confirmation, and initiates the call upon spoken confirmation. The assistant handles ambiguous matches by listing options. All interaction — search, confirmation, and call initiation — is conducted entirely through voice with no visual interaction required.

IV. IMPLEMENTED SYSTEM

A. System Architecture

The app follows a three-tier architecture: a Flutter client layer, a Node.js RESTful backend, and a MongoDB Atlas cloud database. On-device AI components (TFLite for OCR, on-device NLP for intent classification, stroke CNN for gesture draw) provide all core functionality without internet connectivity. Cloud services (Google STT, GPT-4, emergency push notifications) activate when connectivity is available. This ensures the app remains fully functional in offline or low-connectivity environments — critical for an assistive tool serving users in diverse real-world settings.

B. Methodology and Algorithms

The Voice Assistant pipeline: (1) Wake-word detection via keyword spotting running continuously on-device; (2) STT transcription via Vosk offline or Google STT online; (3) Intent classification mapping transcribed text to fourteen module commands.
The Gesture Draw pipeline: (1) Touch event capture at 60 Hz; (2) Stroke normalization — resampling to 64-point trajectory and bounding-box normalization; (3) CNN classification to 26 letter classes; (4) Module routing per the letter-to-function mapping table.
The OCR pipeline: (1) Frame capture by voice or gesture; (2) Google ML Kit Text Recognition v2 on-device; (3) Post-processing for artefact removal, sentence segmentation, and TTS synthesis.

C. Hardware and Software Specifications

The app frontend is built with Flutter 3.19 (Dart) for Android and iOS. Backend: Node.js 20 with Express 4, MongoDB Atlas 7.0. On-device AI: Google ML Kit Text Recognition v2 for OCR, custom stroke CNN for gesture draw, lightweight intent model for voice routing. Speech: FastSpeech 2 (Google TTS fallback), Vosk 0.3.45 offline and Google STT online. AI assistant: OpenAI GPT-4 API. Location: Google Maps SDK with Android Fused Location Provider. SMS: native Android SMS service.
Tested on: Qualcomm Snapdragon 720G, 6 GB RAM, 128 GB storage, 48 MP camera, 4G LTE, Wi-Fi, 5020 mAh battery (≈4.5 hours continuous operation), Android 12 (API Level 31).

V. RESULT AND DISCUSSION

A. Evaluation Metrics

System performance was evaluated using four metrics: (1) Intent Recognition Accuracy — spoken commands correctly classified to intended module; (2) Character Error Rate (CER) = (Substitutions + Deletions + Insertions) / Total Ground Truth Characters; (3) Gesture Draw Recognition Accuracy — drawn letter strokes correctly classified; (4) Location SMS Dispatch Latency — time from trigger to confirmed delivery.

B. Performance Results

The app was evaluated with 25 visually impaired participants recruited from a local school for the blind and a rehabilitation centre over four weeks. Tasks: reading a medicine label, initiating a voice call to a named contact, sharing location by gesture draw, and navigating a 50-metre indoor corridor. Voice Assistant intent recognition: **91.7%** accuracy in quiet conditions, approximately **84%** above 65 dB ambient noise. Vosk offline engine maintained consistent performance without network dependency.

Image-to-Text OCR: Character Error Rate of 3.1% — approximately 97 in every 100 characters correctly recognized. Best on high-contrast printed text; lowest on curved or reflective surfaces.

Gesture Draw Interface: letter classification accuracy of 89.4% across 26 classes. Most-confused pairs ('I'/'L') were mitigated by spoken disambiguation feedback from the assistant.

Location Sharing: SMS dispatched within average 2.8 seconds including fully offline conditions — within the four-second design target.

System Usability Scale: 82/100 ('Excellent' grade). Task completion rate 79% versus 51% baseline with existing tools — a 28 percentage point improvement. Voice Assistant and Location Sharing rated 'Very Useful' or 'Essential' by 92% of participants.

VI. SUMMARY

This paper presents the design, implementation, and evaluation of an AI-Powered Assistive Vision App for blind and visually impaired individuals. The app operates entirely through voice navigation and a novel Gesture Draw Interface, providing module-level access by drawing letters on the touchscreen. Five integrated modules — AI Voice Assistant, Gesture Draw Interface, Image-to-Text OCR, Location Sharing, and Contacts and Calling — together address the primary daily barriers faced by blind users: access to written information, emergency communication, contact calling, and navigation assistance, all without any visual interaction.

Evaluation with 25 participants confirmed strong performance across all modules: 91.7% intent recognition, 3.1% OCR CER, 89.4% gesture draw accuracy, and sub-3-second location SMS dispatch. SUS of 82/100 and a 28-point improvement in task completion over existing fragmented tools confirm the practical value of the unified approach.

Future development will focus on: expanding the Gesture Draw vocabulary for full message composition; improving accuracy for ambiguous letter pairs through user-adaptive personalization; extending multilingual support beyond English, Hindi, and Marathi; integrating smart home and wearable device control; deepening AI assistant personalization to learn individual user routines; and adding real-time background object detection for continuous environmental awareness.

REFERENCES

[1] S. P. Ingale and G. R. Bamnote, "Advancements in Image Captioning using Deep Learning Techniques," 2024 DMIHER International Conference on AI in Healthcare, Education and Industry (IDICAIEI), Wardha, India, 2024, pp. 1-6, doi: 10.1109/IDICAIEI61867.2024.10842893.

[2] A. Paramarthalingam et al., "A deep learning model to assist visually impaired in pothole detection using computer vision," https://doi.org/10.1016/j.dajour.2024.100507.

[3] M. Mateia, L. Alboaie, "Enhancing Accessible Navigation: A Fusion of Speech, Gesture, and Sonification for the Visually Impaired," Creative Commons CC BY-NC-ND 4.0.

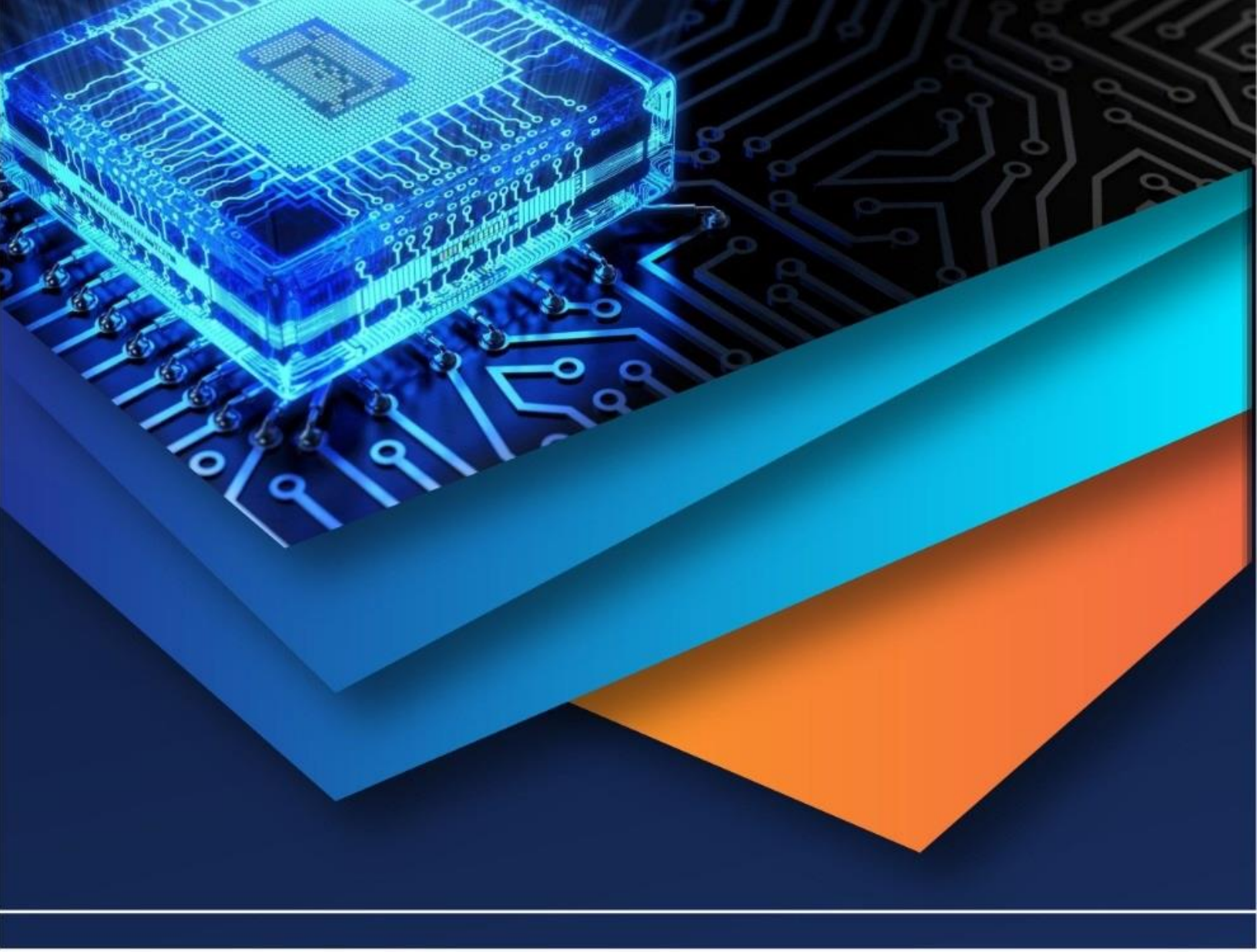
[4] M. H. Abidi, A. N. Siddiquee, H. Alkhalefah, V. Srivastava, "A comprehensive review of navigation systems for visually impaired individuals," https://doi.org/10.1016/j.heliyon.2024.e31825.

[5] H. A. Ahmad, T. A. Rashid, "Planning the development of text-to-speech synthesis models and datasets with dynamic deep learning," https://doi.org/10.1016/j.jksuci.2024.102131.

[6] C.-S. Lee, J.-I. Lee, H. E. Seo, "Deep Learning Based Mobile Assistive Device for Visually Impaired People."

[7] H. Sharma, S. Srivastava, "A Framework for Image Captioning Based on Relation Network and Multilevel Attention Mechanism."

[8] P. Dandwate, "Comparative Study of Transformer and LSTM Network with Attention Mechanism on Image Captioning."



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)