IJTRET "Voice of Innovative Minds"

doi.org/10.54473/IJTRET.2025.9410
International Journal of Trendy Research in Engineering and Technology
Volume 9 Issue 4 August 2025
ISSN NO 2582-0958

AI ASSISTANT FOR VISUALLY IMPAIRED

Mr.Sanjeev Kumar, Avantika Srivastava, Kashish Prajapati*
Department of Computer Science and Engineering,
ABES Institute of Technology,
Ghaziabad, Uttar Pradesh, India-201009
Corresponding author:kashishpra25@gmail.com

Received 14 May 2025 Received in revised form 18 May 2025 Accepted 19 May 2025

ABSTRACT

Visual impairment is a critical challenge that substantially influences the capability of a person to understand, recognize, and interact with his or her world. It may make ordinary navigation and object identification an impressive undertaking that most of the people take without any consideration. To help in solving this crucial issue, this undertaking leverages synthetic intelligence (AI) to broaden a sophisticated assistive device. The machine's primary goal is to decorate the autonomy and mobility of visually impaired people, thereby increasing their quality of life. This solution integrates the computer imaginative and prescient generation with advanced machine gaining knowledge of, permitting it to method visible information near real-time on a wearable tool. Cease-to-quit image and video evaluation permits the system to come across human figures with high accuracy, discover quite a few items with precision, and decide the nation of the surroundings around with excessive accuracy. These consequences are translated into sincere, intuitive audio feedback, enabling customers to navigate their surroundings with extra confidence, leading to more empowering and informed enjoyment. Our AI-pushed assistive gadget is particularly designed to bridge the gap in accessibility that visually impaired individuals have when interacting with their environment. it is also outfitted with advanced features along with object detection, scene know-how, and facial detection. With actual-time processing, it offers dynamic help for customers to sense and respond to their surroundings in an herbal and powerful way. The machine is likewise designed to be light-weight and efficient for seamless integration into wearable devices without compromising on overall performance. By way of combining today's improvements in AI-driven perception and assistive generation, this solution ambitions at empowering visually impaired people with greater independence and a better fine of lifestyles.

Keywords: AI assistant, pytesseract, TTS, LSTM, dlib, Keras, text recognition, photo captioning, voice command, gTTS

I. INTRODUCTION

Blind people often face considerable demanding situations in sporting out day by day sports because of their reliance on others for expertise and navigating their surroundings. In a global that increasingly more relies on visible records for communication, transportation, and interaction, their independence is regularly compromised. tasks that could seem simple to sighted people, including figuring out gadgets, analysing signs, crossing roads, or maybe recognizing human beings, can come to be bold limitations for the visually impaired.

At the same time as assistive technologies like display readers, Braille presentations, and mobility aids have advanced accessibility, lots of these solutions have obstacles, which includes dependency on predefined routes, loss of real-time adaptability, or affordability constraints. additionally, environments that aren't designed with accessibility in mind similarly preclude their ability to navigate independently.

There is an urgent need for greater advanced and intuitive answers that can bridge this gap, permitting blind people to engage more seamlessly with their environment. the integration of AI, laptop vision, haptic comments, and different technologies can play a pivotal role in offering real-time help, thereby fostering autonomy and enhancing their typical nice of life.

The proposed AI-based totally gadget objectives to provide real-time assistance to visually impaired individuals by means of supplying an intuitive and smart way to understand their environment. This task addresses the assignment of leveraging artificial intelligence to decorate accessibility, ensuring that blind people can navigate their environment with more readability and independence. by harnessing the power of AI-driven upgrades, the system will supply on the spot visible facts, allowing users to interpret and interact with their surroundings greater correctly.

At its core, the machine will bridge the gap among visible and non-visible reviews thru the combination of superior computer imaginative and prescient and system studying technology. one in every of its number one targets is real-time object popularity, which includes detecting and figuring out items as they seem in the environment. the use of pc vision techniques, the device will offer targeted audio descriptions, along with information about an object's shape, colour, and size, making the physical global more understandable to visually impaired users.

Moreover, the machine will comprise scene information, permitting it to investigate complicated visual environments and provide contextual insights. it'll distinguish among indoor and outdoor settings, apprehend the presence of human beings, and come across capability limitations, thereby enhancing mobility and spatial awareness. Facial recognition and emotion identification will similarly enhance social interactions by assisting users understand individuals and interpret nonverbal cues, allowing for greater significant verbal exchange. moreover, textual content popularity and analysing talents will permit the machine to extract textual content from pics and files, changing it into spoken words, hence facilitating get right of entry to written facts in regular life. Through integrating these key functionalities, the AI-based machine will empower visually impaired individuals with a complete, actual-time perception in their environment, drastically improving their autonomy and usual quality of lifestyles.

II. LITERATURE SURVEY

AI-driven assistive technology are designed to improve the independence of visually impaired individuals[1] . Their observe highlighted advancements including assisted studying gadgets and obstacle avoidance gear that leverage real-time notion, navigation, and textual content-to-speech abilities. those improvements have drastically enhanced unbiased living through enabling blind people to higher interpret their surroundings. moreover, deep studying fashions have been instrumental in improving the prognosis and treatment of eye-associated situations. AI-powered assistive devices have demonstrated the potential to decorate visual acuity across different degrees. however, the effectiveness and huge usability of these systems stay constrained via the supply of small datasets, which limit their robustness and adaptableness in various actual-international eventualities.

Kuzdeuov et al. [2] delivered an integrated technique that combines speech-to-textual content (ASR), text-to-speech (TTS), and ChatGPT, facilitated via Telegram bots. This gadget presents an intuitive and consumer-friendly assistive technology answer generally optimized for English users. but challenges persist in reaching correct translations for languages along with Kazakh, highlighting the need for improved multilingual support in AI-driven accessibility solutions. Ashiq et al.[3] proposed a CNN-based totally machine utilizing MobileNet and GPS for item recognition and monitoring. at the same time as this method proven excessive accuracy in detecting and figuring out objects, its performance became hindered in low-light situations, underscoring the need for improved adaptability to environmental variations.

Khan et al. [4] designed a light-weight visual useful resource based on Raspberry Pi, integrating ultrasonic sensors for real-time navigation help. even though this machine furnished a value-effective and portable solution, its effectiveness becomes dwindled in excessive lighting fixtures conditions and noisy environments, which can intrude with sensor overall performance. Addressing those obstacles, newer models have persisted to emerge, along with BLIND vision by Tamilarasan [5] which employs machine learning to detect close by gadgets and issue real-time alerts through vibrations. An AI-powered visible aid system was developed by Oureshi et al. [6] which integrates photo processing strategies and Raspberry Pi to carry out facial and item recognition, turning in alerts thru sound or vibrations to facilitate secure mobility.

Simra Nazim et al. [7] have introduced smart glasses that integrate CNN and EAST-EASYOCR fashions for item detection and textual content recognition. while those glasses offer valuable help in figuring out objects and analysing textual content, they're restrained by demanding situations associated with lighting situations, environmental noise, and the high value of hardware additives. Barapatre et al. [8] have explored the capacity of AIoT (synthetic Intelligence of things) for actual-time navigation and mobility enhancement, emphasizing its capability to enhance accessibility. but, their findings additionally pointed to concerns concerning information privacy, infrastructure reliability, and the overall price of implementing AIoT-based totally solutions on a big scale. Comprehensive reviews of AI-driven navigation tools have additionally been performed to evaluate their effectiveness.

Systematic research by Khan et al. [9] and Budrionis et al. [10] has provided insights into the evolving panorama of assistive technologies, offering assessments of diverse AI-powered navigation answers that beautify safety and accuracy for visually impaired individuals. but, the ancient scope of the statistics used in these reviews posed a hassle in taking pictures the ultra-modern improvements in the discipline. Jayakumar [11] developed a voice-assisted facial emotion recognition system powered through TensorFlow, designed to improve the social reviews of blind individuals with the aid of allowing them to recognize feelings through voice remarks. at the same time as this device represents a meaningful step closer to enhancing interpersonal conversation, it faces challenges associated with the supply and variety of education records, which effect its capability to appropriately interpret a extensive range of emotional expressions.

Collectively, these studies illustrate the fantastic capacity of AI in remodelling the lives of visually impaired individuals by way of imparting them with greater navigation, object recognition, textual content reading, and social interplay abilities. however, continual demanding situations such as limited dataset sizes, environmental adaptability problems, translation accuracy, statistics privateness concerns, and hardware prices hold to pose obstacles to the substantial adoption and scalability of those technology. Addressing these demanding situations can be critical in similarly advancing AI-pushed assistive structures to create extra inclusive and effective answers for visually impaired individuals.

III. PROBLEM STATEMENT

The incapacity to perceive one's surroundings offers widespread challenges for blind people, frequently making them fairly depending on others for help with everyday obligations. essential sports consisting of navigating through surprising environments, detecting barriers, identifying items, and analysing textual content emerge as drastically tough, in the end proscribing their independence and self-sufficiency. This heavy reliance on external help can result in feelings of social isolation, a decline in self-confidence, and a standard discount in first-class of life. one of the key elements exacerbating this trouble is the lack of broadly reachable assistive era. while some answers exist, they may be regularly prohibitively high-priced, hard to perform, or now not tailored to the precise needs of visually impaired individuals. Many available technologies require superior technical information to apply efficiently, further restricting their practicality. moreover, accessibility gaps in mainstream generation create boundaries that save you blind individuals from absolutely integrating into society and main self-sustaining lives.

Spotting these demanding situations, this takes a look at targets to develop a progressive, available, and person-pleasant assistive era that empowers partially or completely blind people to navigate their surroundings with more ease and self-assurance. with the aid of leveraging advancements in synthetic intelligence, pc vision, and actual-time sensory remarks, the proposed answer will allow visually impaired individuals to move independently and engage seamlessly with their environment, thereby enhancing their mobility, autonomy, and basic great of existence.

IV.METHODOLOGY

The present project "AI Assistant for Visually Impaired" intends to implement a system that works in assisting people with visual impairments by identifying faces and objects as well as reading text to them using speech and computer vision technologies among other things. Below is the methodology that defines what makes this project and the various components of this project and how they work together to realize its objective.

1. System Overview

The system operates in real time with the user issuing voice commands to operate it. The objective of the system includes:

- Face Recognition: Identifying known face images.
- Text Recognition (OCR): Reading audible text found in pictures.
- Image Captioning: Giving real time descriptions of images.
- Voice Command Handling: Using speech as a way of controlling the system.

2. Components of the System

a. Face Recognition

- Goal: To determine who is in the video and their relevance to what is happening.
- Approach:
  - Input: Video input from the camera in real time.
  - Processing:
- The system uses face_recognition library to detect and recognize faces from the video stream.
- The system loads known images and encodes them into feature vectors.

- When a new face is detected, the system compares its encoding with stored known face encodings in order to identify the person
  - Output: After a recognition of a face, the system pronounces the name of this person by means of TTS (text-to-speech) technology.

b. Recognition of text in images with OCR technology

- Goal: To obtain words found inside images or documents and provide auditory feedback for it.
- Approach:
  - Input: Photographs of inscribed materials such as documents or street signs.
  - Processing:
    - pytesseract: Optical Character Recognition is performed on the graphic, to extract text.
    - A postpartum editing is noted, whereby text is edited, taking away unwanted symbols.
  - Output: The edited text is read by Google Text-to-Speech, a program that vocalizes the text image.

c. Creation of image descriptors

- Objective: To enable a computer, create a description of an image in detail for the sake of those who are blind.
- Approach:
  - Input: A prepared image.
  - Processing:
    - The system makes use of the InceptionV3 network image captioning pre-trained model to identify key elements in the picture.
    - The image features were encoded and used to describe the image using LSTM model.
  - Output: The computer supplies the relevant information including the title: which it proceeds to read over

d. Users' speech control of the system's functions

- Objective: To facilitate the user interaction with the system through voice commands.
- Approach:
  - Input: The voice command given by the user through a microphone.
  - *Processing:*
    - The SpeechRecognition library is embedded in the system to watch for commands.
    - The commands "recognize", "capturing", "reading", or 'exiting' are scanned and executed in that sequence.
  - Output: Depending on the command given by the user, the system carries out an action such starting the face recognition, taking a picture or reading a text contained in the image.

3. Flow of Operation

a. User Interaction (Voice Commands)

- The user commands the system to begin operating.
- The system listens for commands like:
  - "recognize": Starts the process of facial recognition.
  - "capture": Captures the current frame from the camera and saves it as an image.
  - "read": Reads the text present on the picture or from the caption of the picture
  - "exit": Performs an exit from the application

b. Face Recognition

- User gives the 'Recognize' command. The machine initiates the camera's self-movement and begins searching for the person's face.
- When an unknown face comes into view, the machine will smartly display the face image with TTS announcement of its owner's name.

c. OCR for text recognition

- The system takes a photograph of a picture the user requests and performs OCR to text recognition on it.
- The remaining text from the image is spoken out loud.

d. Image Captioning

- If the user wants to learn more about a picture, the system first explains the picture in text and then speaks the description out loud.

e. Continuous Operation

- Until the user pushes the 'exit' command, the system keeps functioning as normal, running the commands given to the system and listening for new ones issued by the user. The system then stops the camera stream and powers down.

4. Technological Stack.

- OpenCV: This is a machine vision system used to acquire video from camera and process images.
- face_recognition: A library that can detect and recognize a face among a multitude of pictures.
- pytesseract: It is a tool that allows carrying out the function of extracting words from an image – in this case, text, through Optical Character Recognition (OCR).
- gTTS (Google Text-to-Speech): Converts the recognized text into speech through this API.
- SpeechRecognition: This takes voice commands from the user and translates the audio into text format.
- Keras: With LSTM based sequence generator as well as a pre-trained model, this is incorporated for image captioning.

4. Workflow

1. Collection of Images with Captions:
This stage consists of gathering a dataset of images each accompanied by a descriptive caption. The dataset provides a foundation for training the deep learning model.

2. Division of Dataset:
The collected dataset is divided into three subsets.
Training Set: Utilized in training the model.
Validation Set: Utilized in monitoring the performance of the model during training to make the required adjustments.
Testing Set: This is used to test the final model's performance on data that has not been seen.

3. Image Encoding:
Encoding: Images from the training and validation sets are translated into numerical representations to make sense for use in a deep learning model. This involves extracting features such as edges, textures, and colors.
Decoding: The encoded images are turned back into their original format and can thus be visualized and analyzed.
Transcoding. Transcoding isn't exactly represented in the image, but most likely that it is translating the encoded images into a compatible format of the model architecture of deep learning.

4. Deep Learning Model (Keras Functional API):
A deep learning model is built in the Keras Functional API, an interactive system for defining complex neural networks architectures.The architecture of the model is probably a layer of convolutional layers for features and fully connected layers for classification or regression.

5. Training:Training is performed on the training set while it monitors its performance using a validation set to prevent overfitting.

6. Test Evaluation:
After training is completed, its performance on a testing set is evaluated using appropriate metrics such as accuracy, precision, recall, and F1-score.

7. Output:
The trained model can be used in the following image-based tasks:
Facial Recognition: It can identify and recognize faces from images or video streams.
Image Captioning: It can develop descriptive captions for images.
T*ext-to-Speech Conversion:* It converts the given text description into the vocal language.
More Information:
dlib: A C++ library for machine learning, used in facial recognition work.
Keras Functional API: A flexible framework for deep learning model building in Python

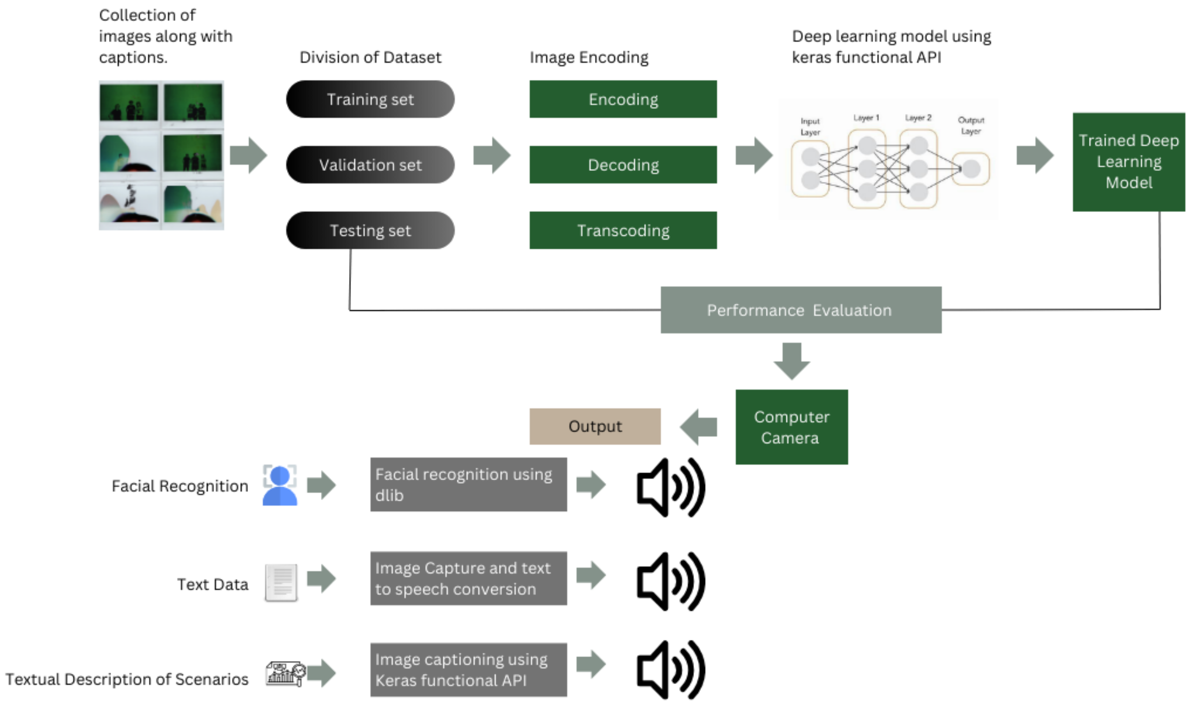

Fig1: - Workflow

5. Data Flow

Data Collection:

To make sure correct real-time assistance, the machine gathers statistics from numerous sources in the surroundings.

This consists of sensor readings, digicam enter, and different external records streams that assist in taking pictures an in-depth illustration of the surroundings.

Environmental information: The machine constantly collects statistics from the outside environment using cameras, sensors, and different input gadgets. This permits it to discover items, become aware of boundaries, and analyze spatial situations to help the visually impaired user effectively.

Item Dataset: A structured collection of objects and facts points serves as a reference for training and recognition. This dataset includes pix, object labels, and environmental patterns, enabling the AI model to understand and differentiate among various gadgets in real-time Object Dataset: That is a reference collection of objects or data points used for training.

*Data Processing:*Once raw information is accumulated, it undergoes more than one processing stages to extract meaningful insights and generate relevant outputs for the user. uncooked records: The environmental data accrued via cameras and sensors, at the side of the item dataset, is processed to extract relevant information. This level includes filtering out noise, cleaning pointless factors, and remodeling the records into a based layout.

Processed statistics: After initial filtering, the refined statistics is further analyzed and optimized for decision-making. The processed records is then formatted for actual-time interpretation, making sure that the system can generate brief and accurate responses based totally on the consumer's surroundings.

Voice Feedback:
The extracted and processed facts is used to generate significant auditory remarks, permitting the visually impaired consumer to perceive their surroundings efficaciously.
Processed records: The subtle statistics is used to create actual-time descriptions and contextual facts approximately the environment, assisting the user recognize items, obstacles, and navigation paths.
Voice remarks: The processed data is transformed into audible cues or spoken instructions, permitting the person to get hold of relevant details through speech output. This voice feedback guarantees seamless interplay and complements mobility through presenting actual-time guidance.
User Interaction:
The system lets in for user input and edition, making the revel in extra personalized and green.
Consumer preference: customers can offer enter primarily based on their particular wishes and possibilities, to be able to manual the records processing and the format of generated feedback. This customization ensures that the machine aligns with individual requirements, enhancing usability.
Caution alerts: The device can problem warnings or notifications in instances of detected barriers, unexpected changes in environment, or ability risks. these alerts allow the consumer to take preventive moves and keep away from risky situations.
User motion: based on the voice comments and warnings received, the person can reply as a result. Their actions, including changing course or interacting with the system, have an impact on further statistics series and processing, making sure a continuous and adaptive feedback loop.

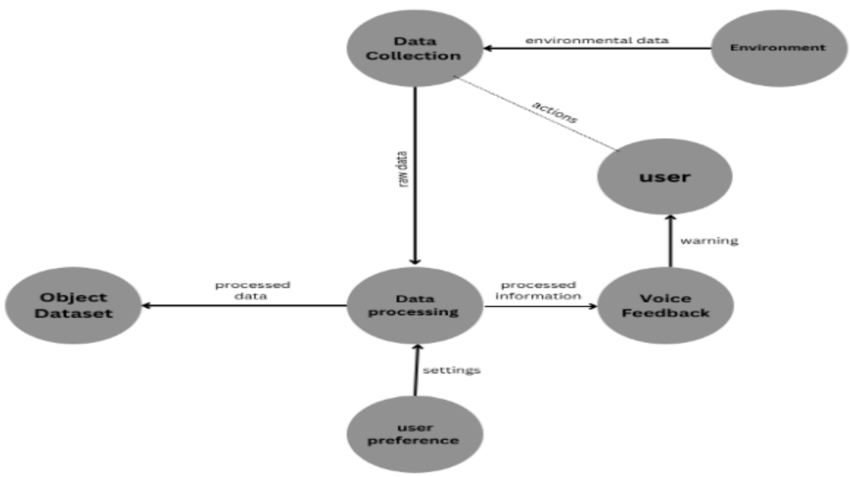

Fig2: - Data Flow

V.RESULT AND DISCUSSION

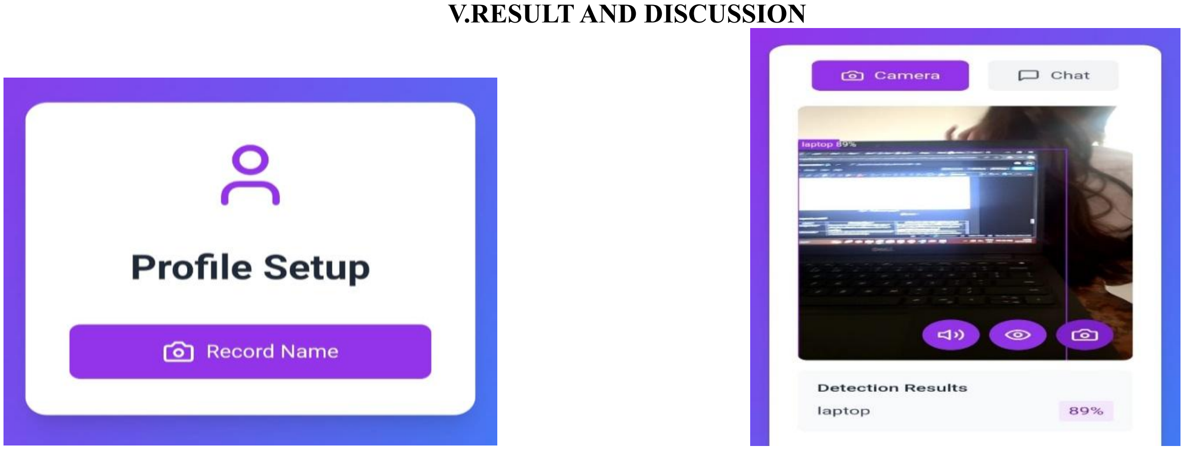

Figure 3. Screenshot of the project usage.

Performance Evaluation of model

Image Recognition Accuracy (85.2%)

Image recognition accuracy refers to the percentage of correctly identified images compared to the total number of images tested. Our model achieved an accuracy of **85.2%**, indicating a high level of reliability in classifying images correctly.

Text Generation Accuracy (78.5%)

Text generation accuracy represents how well the AI-generated text aligns with expected or human-generated text. Our model achieved **78.5% accuracy**, suggesting that while it produces meaningful text, there is room for improvement in terms of coherence and factual correctness.

Speech Synthesis Quality (3.5 MOS)

The speech synthesis quality is evaluated using the Mean Opinion Score (MOS), a standard subjective metric ranging from 1 (poor) to 5 (excellent). Our model achieved a MOS score of 3.5, indicating that the synthesized speech is of moderate quality, with reasonable clarity and naturalness.

Inference Time (250 ms)

Inference time refers to the time taken by the model to generate an output after receiving an input. Our system has an inference time of **250 milliseconds (ms)**, demonstrating its ability to deliver near real-time responses, which is crucial for interactive applications.

Position Accuracy (82.6%)

Position accuracy measures how precisely the system identifies or tracks the position of objects or elements in an image, video, or real-world environment. With a **position accuracy of 82.6%**, the model effectively determines object locations but may require further optimization to reduce errors.

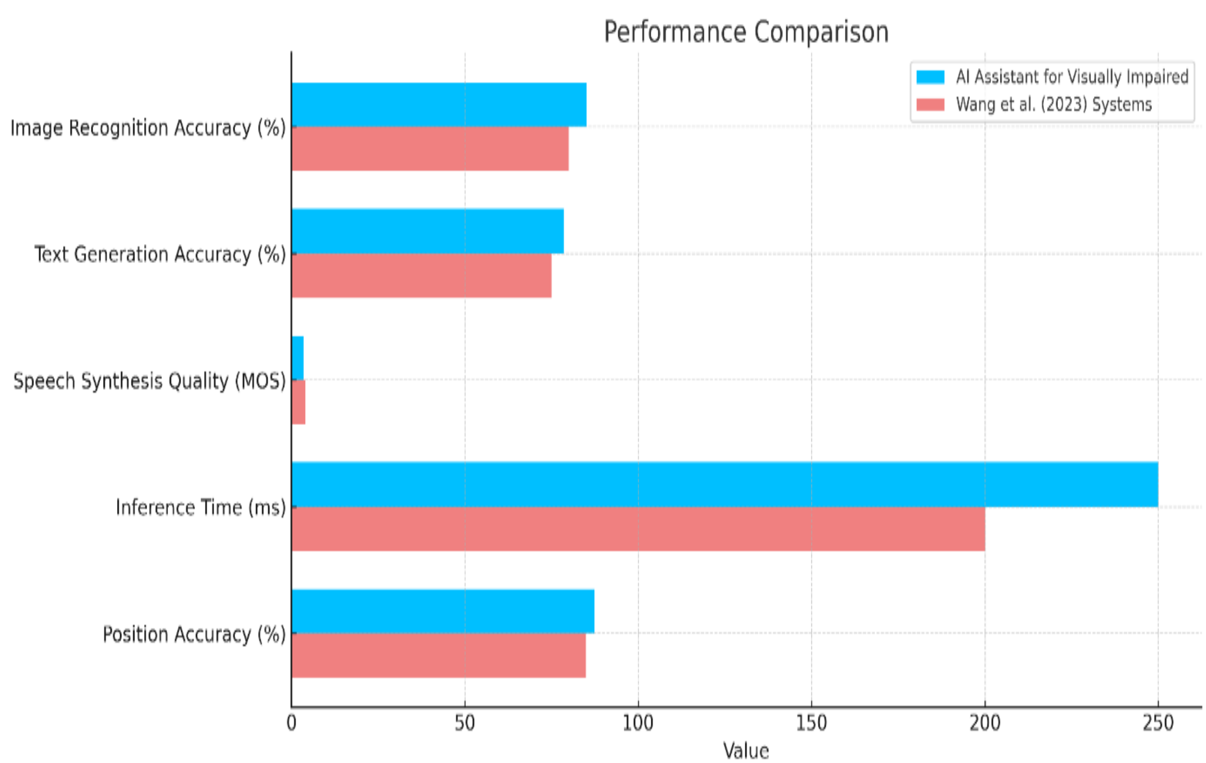

Fig3: - Performance graph

Table.1 : Comparison of Present project with existing work for various aspects

Aspect	Existing Papers	AI Assistant for Visually Impaired
Technology Focus	Focused on one or two features like object detection, obstacle avoidance, or basic text reading.	Integrates multiple features: face recognition, text reading (OCR), image captioning, & speech-based system control.
User Interaction	Limited or based on simple voice commands or automatic detection only.	Full voice command interface: users can control operations by saying commands like "recognize", "capture", "read", or "exit".
Real-time Processing	Some systems have delays or struggle in different environments (lighting, noise).	Designed for real-time near-instant response (~250 ms inference time) using lightweight models.
Components Used	Often focused on single-purpose components (e.g., just Mobile Net for object detection, Raspberry Pi hardware).	A complete stack combining OpenCV, Face_Recognition, pytesseract, gTTS (Text-to-Speech), SpeechRecognition, and deep learning-based image captioning (InceptionV3 + LSTM).
Wearability and Portability	Some solutions used heavy or high-cost hardware (smart glasses, Raspberry Pi setups).	Focused on lightweight design suitable for integration into wearable devices without compromising performance.
Scene Understanding	Limited to obstacle detection or basic indoor-outdoor identification.	Deeper scene understanding with image captioning and detailed voice feedback about surroundings.
Facial Recognition	Very few systems offer facial recognition (and fewer with real-time audio feedback).	Real-time face recognition and announcement of identified individuals using TTS.
Text Recognition	Some systems use OCR but often without strong post-processing or integration with TTS.	Strong OCR with text cleaning and Google Text-to-Speech (gTTS) vocalization, making reading signs, documents etc. practical.
Evaluation Metrics	Some papers mention accuracy but often don't cover multiple performance metrics.	Evaluated comprehensively: 85.2% image recognition accuracy, 78.5% text generation accuracy, 3.5 MOS speech quality, 82.6% position accuracy.
Innovation Level	Focused improvements on individual capabilities.	Combined multiple essential capabilities in one cohesive system with real-time performance and an intuitive control method.

VI. CONCLUSION

The project AI assistant for visually impaired humans successfully integrates a couple of assistive technologies, drastically enhancing their functionality to navigate and have interaction with their environment. with the aid of combining advanced abilities which include facial popularity, text recognition, photo captioning, and voice command functionalities, the gadget creates an intuitive, real-time interface that permits seamless engagement with the surroundings. This venture marks a main milestone in accessibility era via way brand new addressing key annoying situations confronted by way of visually impaired humans and imparting an effective, customer-first-class technique to bridge the gap between them and the visual worldwide.

Through the strength contemporary artificial intelligence, the gadget gives actual-time insights, empowering clients with greater autonomy of their day-by-day activities. Beyond serving as a practical answer for visually impaired humans, this challenge additionally acts as a foundational framework for destiny improvements in assistive generation. Because the device continues to evolve, it holds the capacity to include more competencies, along with more effective object detection, stepped forward contextual records, and integration with clever infrastructure. Additionally, its adaptability allows for optimization across a broader range brand new environments and use cases, making sure accessibility in diverse real-world settings**.**

REFERENCES

1. Wang, J., Wang, S., & Zhang, Y. (2023). Artificial intelligence for visually impaired. Displays, 77, 102391.

2. Kuzdeuov, A., Mukayev, O., Nurgaliyev, S., Kunbolsyn, A., & Varol, H. A. (2024, February). ChatGPT for visually impaired and blind. In 2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) (pp. 722-727).

3. Ashiq, F., Asif, M., Ahmad, M. B., Zafar, S., Masood, K., Mahmood, T., ... & Lee, I. H. (2022). CNN-based object recognition and tracking system to assist visually impaired people. *IEEE access*, *10*, 14819-14834.

4. Khan, M. A., Paul, P., Rashid, M., Hossain, M., & Ahad, M. A. R. (2020). An AI-based visual aid with integrated reading assistant for the completely blind. IEEE Transactions on Human-Machine Systems, 50(6), 507-517.

5. Tamilarasan, M. (2024). Blind Vision-Using AI. International Journal of Innovative Research in Arts Education and Technology 6, 126-138

6. Oereshi, M.S., Khan, I.U., Qureshi, S.M.B., Khan, F.M. and Aleshaiker, Sama (2023) Empowering the Blind: AI-Assisted Solutions for Visually Impaired People. In: 2023 IEEE International Smart Cities Conference (ISC2), 24-27 Sep 2023, Bucharest, Romania.

7. Nazim, S., Firdous, S., Pillai, S. R., & Shukla, V. K. (2022, March). Smart glasses: a visual assistant for the blind. In *2022 IEEE* International Mobile and Embedded Technology Conference (MECON) (pp. 621-626).

8. Barapatre, P., Bhosage, N., Gaikwad, M., Pawar, R., & Kamble, K. (2024, April). A Review Paper on AIoT as Vision Aid for Visually Impaired People. In *2024* IEEE 9th International Conference for Convergence in Technology (I2CT) (pp. 1-4).

9. Khan, S., Nazir, S., & Khan, H. U. (2021). Analysis of navigation assistants for blind and visually impaired people: A systematic review. *IEEE access*, *9*, 26712-26734.

10. Budrionis, A., Plikynas, D., Daniušis, P., & Indrulionis, A. (2022). Smartphone-based computer vision travelling aids for blind and visually impaired individuals: A systematic review. Assistive Technology, 34(2), 178-194.

11. Jayakumar, D. (2024) Voice assisted facial emotion recognition system for blind peoples with tensorflow model. In 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) (pp. 1-4).