Conversational Image Recognition Web Application for Visually Impaired Peoples

[1] M. Sagar, [2] M. Umesh, [3]A. Vamshi, [4]A.Vijaya laxmi

[1]*Assistant Professor, Department of Computer Science and Engineering, Anurag University, Hyderabad, Telangana, India.*
[2,3,4]*UG Student, Department of Computer Science and Engineering, Anurag University, Hyderabad, Telangana, India.*

Abstract. The application interface is designed to support visually impaired users by integrating voice command functionality, enabling seamless interaction through audio input. Users can operate the application entirely via speech, including taking pictures and asking questions about the captured images. This process begins with the user issuing a voice command to take a photo, after which they can inquire about the image's contents. The spoken questions are converted to text, analyzed, and paired with the image to generate an intelligent response. The backend of the application is developed using Flask, which efficiently manages server-side operations and ensures smooth communication between components. The core image analysis and question-answering functionality is powered by Google's Gemini AI, a sophisticated multimodal model capable of interpreting both visual and textual information to generate accurate, context-aware responses. Once the AI formulates a response, Amazon Polly is used to convert the text output into natural-sounding speech, which is then relayed to the user. This complete audio feedback loop allows users to receive spoken answers to their questions without the need for a visual interface. The integration of these technologies—Flask for backend processing, Gemini AI for image-based query resolution, and Amazon Polly for speech synthesis—creates a robust, voice-driven system that enhances accessibility for individuals who rely on auditory interfaces. By transforming visual content into spoken information, the application empowers visually impaired users to independently explore and understand their surroundings, making technology more inclusive and responsive to their needs.

Keywords: Voice interface, Speech recognition, Image processing, AI-powered Q&A, Accessibility technology

INTRODUCTION

Visually impaired users often face significant challenges when interacting with modern digital applications due to the predominant reliance on visual interface components. Traditional applications are designed with graphical user interfaces (GUIs) that require visual perception and manual navigation, making it extremely difficult—if not impossible—for users with visual impairments to engage with digital content in a meaningful or independent way. This limitation highlights the urgent need for inclusive systems that cater to non-visual modes of interaction. To address this issue, the proposed solution introduces a voice-command-based interface that allows users to operate and navigate applications using only their voice, thereby completely eliminating the need for any visual contact with the screen during usage.

This system is tailored specifically to enhance accessibility for visually impaired individuals by enabling complete interaction through audio commands. By leveraging advanced speech recognition technology, the system efficiently processes user instructions, translating spoken words into executable commands without requiring manual input or screen-based navigation. This voice-driven mechanism facilitates a hands-free and intuitive experience, empowering users to perform a wide range of operations using natural language. Whether it's initiating an action, taking a picture, or querying about objects in the environment, users are guided entirely through vocal instructions and audio feedback, making the interface both user-friendly and highly accessible.

At the heart of this solution is a unified voice recognition system combined with robust text-to-speech (TTS) synthesis and powerful AI-based image processing capabilities. The speech recognition module plays a critical role by converting the user's spoken commands into written text in real time. This conversion enables

seamless system interpretation of user intentions, which is especially valuable in environments where visual cues are absent or unusable. Unlike conventional interfaces that rely on visual input, the automated speech-to-text conversion eliminates the need for user-driven manual intervention, significantly reducing cognitive and physical strain for visually impaired users.

The text-to-speech synthesis component complements this functionality by turning the system's textual responses into natural-sounding audio output. Powered by Amazon Polly, the TTS module ensures that the system's feedback is not only clear and human-like but also easy to understand and contextually relevant. This audio feedback mechanism enables enhanced two-way communication between the user and the application, transforming traditional user-system interaction into a more immersive and responsive experience. Users receive spoken instructions, notifications, and results, making it possible to operate the system even in the complete absence of visual cues.

One of the most powerful features of the system is its capability to capture and analyze images upon voice command. Users can simply instruct the application to take a photo, which triggers a sequence of intelligent image processing functions. These images are collected in real time using OpenCV, a widely-used computer vision library, and are then sent to Google Gemini AI, a cutting-edge multimodal artificial intelligence platform. Gemini AI processes the image in conjunction with the user's query, interpreting visual details such as objects, environments, text, and other contextual information. The AI then formulates a response that is both accurate and context-aware, providing valuable insight into the visual content.

The results of this image-based analysis are communicated back to the user through speech, closing the interaction loop with an auditory response. This feature is particularly useful in real-world scenarios where users need to understand their environment, identify objects, read text from signs or labels, or gain spatial awareness. By simply capturing an image and asking a question, users can receive informative answers that describe the scene, identify key elements, and enhance their situational understanding. This capability transforms everyday tasks—from reading a menu to navigating public spaces—into manageable activities for visually impaired individuals.

The system's backend is built using Flask, a lightweight yet powerful Python web framework. Flask acts as the bridge between the frontend user interface and the complex AI and processing services running in the background. It manages data routing, system logic, and API integration to ensure that user commands, image data, and AI responses flow smoothly throughout the application. By using Flask, the system maintains scalability and modularity, making it easier to introduce new features, services, or upgrades in future versions.

What sets this system apart is its ability to provide real-time feedback and operate in a responsive, conversational manner. When a user asks a question or gives a command, the system reacts immediately by processing the input and generating the appropriate output in seconds. This real-time interaction is essential for maintaining a natural, dialogue-like experience that is critical for users who cannot rely on screen-based feedback. Additionally, the simplicity of the voice command structure ensures that users of all technical skill levels can easily adopt the system without needing extensive training or prior experience.

This system has the greatest impact on visually impaired individuals who face the most barriers in accessing and understanding digital content. Traditional screen-based platforms are often inaccessible to them, and even with screen readers, there are limitations to navigation, interpretation, and understanding of visual material. Through this system, users gain a precise understanding of objects, scenes, and contextual information simply by capturing images and querying them using voice commands. The processed responses from Gemini AI, combined with high-quality speech synthesis from Amazon Polly, provide a complete auditory experience that bridges the gap between visual data and accessible communication.

Furthermore, the system is designed with future enhancements in mind. Planned upgrades include the integration of multi-language support, enabling users from diverse linguistic backgrounds to benefit from the tool in their native languages. This is particularly important in global contexts where accessibility solutions must cater to multilingual populations. Another key development area is the enhancement of the AI model to include more advanced image analysis, such as facial recognition (where appropriate), text extraction (OCR), and environmental hazard detection. Additionally, offline functionality is a high priority for upcoming updates, ensuring that users can continue to operate the system even in areas with limited or no internet connectivity.

Overall, this voice-operated interface represents a significant advancement in digital accessibility. It

successfully resolves many of the challenges faced by visually impaired users by providing a self-sufficient, audio-based mode of interaction that removes the reliance on screen-based interfaces. The integration of voice recognition, AI-powered image analysis, and speech synthesis results in a holistic system that enables visually impaired individuals to engage with digital environments more confidently and independently. By supporting real-time operations, intuitive navigation, and accessible feedback, the system creates an inclusive digital framework designed to promote user autonomy, enhance spatial awareness, and empower users in their daily interactions with the world around them.

LITERATURE SURVEY

1. Peter Anderson et al. (2018): "Bottom-up and Top-down Attention for Image Captioning and Visual Question Answering"

This paper presents a novel approach to visual question answering (VQA) and image captioning by introducing bottom-up and top-down attention mechanisms. The authors explore how combining these two approaches can significantly enhance the quality of responses in image captioning and visual question answering tasks. Bottom-up attention uses a region-based attention mechanism to extract specific features from an image, focusing on the parts of the image that are most likely relevant to the question being asked. Top-down attention, on the other hand, uses the context of the question to guide which image features should be focused on.

2. Jeffrey P. Bigham et al. (2010): "Vizwiz: Nearly Real-Time Answers to Visual Questions"

The VizWiz project focuses on providing real-time answers to visual questions posed by blind or visually impaired individuals. Users take pictures of objects or scenes using their mobile phones and ask questions about what they see. The system routes these images to a crowd of human workers who respond with answers based on the images, which are then converted back to speech for the user. The main innovation here is the combination of crowdsourcing and mobile technology to provide an efficient solution for answering visual questions in nearly real-time.

For the field of assistive technology, the implications of this system are profound. It creates an accessible tool for visually impaired individuals to obtain detailed answers about their environments, such as identifying objects, reading text, or understanding complex scenes. This project laid the groundwork for subsequent research into human-assisted systems for visual question answering, inspiring future works in conversational AI and image captioning, which now use machine learning models to replicate some of the capabilities of the VizWiz approach.

3. Silvio Barra et al. (2021): "Visual Question Answering: Which Investigated Applications?"

In this review paper, the authors provide an overview of the various applications of visual question answering (VQA) across different domains. They discuss the wide variety of VQA tasks, including object recognition, scene understanding, and real-time interaction with visual content. The paper highlights the challenges and current methodologies in VQA systems, with a focus on dataset creation, model architectures, and evaluation metrics.

This work is crucial in understanding the breadth of VQA research and its potential applications for accessibility tools, particularly for users with visual impairments. VQA has applications beyond simple object recognition, such as assisting blind users with reading text, identifying people in photographs, or providing spatial awareness. The methodologies discussed in the paper can be integrated into future assistive technologies for visually impaired users, ensuring that AI systems can provide reliable, context-sensitive responses to visual queries. Moreover, this paper helps guide the design of new VQA systems by identifying gaps in current research and suggesting potential areas for improvement, such as the need for more diverse datasets that represent a variety of real-world environments.

4. *Wei-Lin Chiang et al. (2023): "Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90% ChatGPT Quality"**

This paper introduces Vicuna, a chatbot that combines open-source elements and cutting-edge language models to deliver high-quality conversational AI. The authors demonstrate how Vicuna approaches human-like conversation, achieving near-GPT-4 performance while maintaining an open-source framework. This paper addresses the need for more accessible and customizable chatbot systems that can be tailored to different user requirements, including for people with disabilities.

For the visually impaired community, the development of advanced conversational AI systems like Vicuna presents an opportunity for creating more responsive and intelligent virtual assistants that can help with a variety of tasks, from answering questions about the environment to providing interactive guidance. As a part of the AI-based assistant systems, Vicuna can be adapted for voice-activated platforms, making it easier for users to query their surroundings or receive information about objects in their environment.

5. Danna Gurari et al. (2018): "VizWiz Grand Challenge: Answering Visual Questions from Blind

People"

This paper extends the work of the original VizWiz project by organizing a grand challenge to encourage research in automated systems for answering visual questions from blind individuals. The challenge focuses on creating AI models that can answer questions posed by blind users about images of their surroundings. This initiative has propelled forward the development of machine learning models tailored specifically to accessibility and the needs of visually impaired people.

The implications for this work in the context of visual impairments are vast, as it has sparked a wave of innovation in the use of AI for accessible technology. The challenge has motivated the creation of more accurate, real-time models for visual question answering, using sophisticated deep learning algorithms and better data collection. The competition not only advanced AI research but also highlighted the importance of creating tools that are both practical and scalable, which can directly benefit visually impaired individuals by providing real-time, accessible information.

6. Md Farhan Ishmam et al. (2024): "From Image to Language: A Critical Analysis of Visual Question Answering (VQA) Approaches, Challenges, and Opportunities"

This paper offers a comprehensive analysis of the challenges in VQA, focusing on the gap between image understanding and language generation. The authors critique various approaches to VQA, including image captioning and other AI-powered systems that are commonly used to generate text-based answers from visual data. They highlight the limitations of current models, including their reliance on large datasets, the challenges of ensuring generalizability across different visual contexts, and the difficulties in understanding complex visual questions.

In relation to accessibility, this paper provides important insights into how AI models can be improved to serve visually impaired users better. For example, the paper suggests that VQA systems need to become more robust and capable of understanding a wide range of visual contexts, from everyday objects to complex environments. This research can inform future assistive technologies, helping to refine VQA models that are able to provide more accurate and detailed responses to questions posed by visually impaired individuals about their surroundings.

7. Walter S. Lasecki et al. (2013): "Answering Visual Questions with Conversational Crowd Assistants"

This paper explores the use of conversational crowd assistants in answering visual questions. The authors introduce a novel approach where a crowd of human workers assists in answering questions about images posed by users. This approach is particularly effective in dealing with complex visual queries that current machine learning models struggle to handle. The system can provide high-quality answers by tapping into the knowledge of a large group of people.

This work is relevant to assistive technologies for visually impaired users, as it highlights the effectiveness of crowdsourcing in answering visual questions. While automated systems are increasingly capable of answering such queries, the integration of human assistance remains valuable in certain contexts, especially when high accuracy is needed. This hybrid approach can be leveraged in systems designed for visually impaired individuals, offering a balance between automation and human intervention to ensure that answers are both timely and accurate.

8. Junnan Li et al. (2023): "Blip-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models"

Blip-2 presents a novel framework for visual language understanding, combining frozen image encoders with large pre-trained language models. This hybrid model efficiently integrates visual and textual information, allowing for more coherent and contextually aware image-captioning and VQA tasks. The model is designed to handle complex vision-language interactions, enhancing the ability of AI systems to interpret visual content alongside natural language queries.

For visually impaired users, Blip-2's approach holds promise in enhancing systems that provide real-time visual analysis and description. The integration of frozen image encoders and large language models can be used to create more accurate and dynamic systems for image captioning, helping visually impaired individuals navigate their surroundings more effectively. By improving the relationship between visual and textual data, Blip-2 can contribute to the development of more responsive and detailed assistive technologies.

PROPOSED SYSTEM

The proposed system aims to address accessibility challenges faced by visually impaired users by providing

a comprehensive voice-controlled application that enables interaction with digital content through voice commands. The system combines advanced voice recognition, image processing, and text-to-speech synthesis to create an intuitive, user-friendly experience. In this section, we describe the design, components, functionalities, and overall user experience of the proposed system.

System Overview

The system is designed to offer visually impaired users a seamless and efficient method to interact with their surroundings using voice commands, removing the need for traditional visual interfaces. This is achieved through a combination of a Flask-based backend, AI-powered image query responses using Google Gemini, and text-to-speech synthesis using Amazon Polly. The user interacts with the system by issuing voice commands, and the system provides spoken feedback in response. Additionally, the system allows users to take pictures and receive descriptive responses regarding the captured images, making it suitable for various environments and situations.

The system leverages cutting-edge technologies, including natural language processing, machine learning models, and computer vision, to deliver real-time responses. These technologies are integrated into a cohesive platform that supports both spoken input and output, creating an accessible interface for users with visual impairments. The goal is to empower users to perform tasks that would typically require visual interaction, such as taking photos of objects or locations, querying the system for information, and receiving auditory descriptions or answers.

System Components

1. Voice Command Module

The core interaction model in the proposed system is voice-based. Using a speech recognition engine, the system converts voice commands into text. The voice command module is responsible for detecting user speech and interpreting it in real-time. This functionality eliminates the need for visually impaired users to interact with traditional input devices like keyboards or touchscreens.

Voice commands are processed through the system's backend, which relies on a natural language processing (NLP) model to extract the intent behind the spoken query. This allows the user to interact with the system in a natural and conversational manner. Whether it's asking for information, issuing a request, or querying about an image, the system is designed to understand and respond appropriately, ensuring a smooth user experience.

2. Image Capture and Processing

A critical aspect of the system is the ability for visually impaired users to take pictures of objects, scenes, or text. The system integrates OpenCV for capturing live images using a smartphone or camera. Once an image is taken, it is immediately processed by the backend system, where AI-powered image analysis tools, such as Google Gemini AI, come into play.

Google Gemini, a powerful AI tool for visual question answering, processes the images and generates relevant textual descriptions based on the image content. For example, if a user captures an image of a storefront, the AI can provide a description of the store, its surroundings, and any visible text, such as the name of the store or signage. The text generated by Gemini is then sent back to the system, where it is converted into speech output.

3. Text-to-Speech Synthesis

Once the system processes the image or query and generates a textual response, the next step is to convert the text into natural-sounding speech. Amazon Polly, a cloud-based service that converts text into speech, is used to generate high-quality, clear, and natural-sounding voice outputs. Amazon Polly supports multiple languages and voices, offering flexibility in delivering auditory feedback tailored to the user's preference.

This speech output serves as the primary means of communication between the system and the user, ensuring that the information is accessible and understandable. Whether the user is querying about an object, seeking environmental information, or asking questions related to an image, the spoken response ensures that visually impaired users can access the information in real-time.

4. Backend System and Integration

At the heart of the system lies a Flask-based backend, which connects all the components—voice recognition, image capture, AI processing, and text-to-speech synthesis. Flask, a lightweight Python web framework, is used to manage user requests, process data, and handle communication between the frontend and the AI services.

When a user issues a voice command or takes a photo, the Flask backend orchestrates the entire

process. It receives the voice command, passes it through the speech-to-text engine, processes the resulting text using NLP and image analysis services, and then sends the output to Amazon Polly for speech synthesis. The modular structure of the backend ensures that all components work in harmony to provide the user with a real-time, interactive experience.

5. AI-Powered Visual Query Responses

The Google Gemini AI plays a critical role in interpreting and responding to visual queries. After a user captures an image, the system sends the image to Google Gemini, which uses sophisticated machine learning models to analyze the visual content. It generates detailed descriptions or answers to specific questions about the image. For example, if a user takes a photo of a restaurant, they might ask, "What is the name of the restaurant?" or "What does the menu look like?" Gemini's AI model analyzes the image and generates a contextually accurate answer, which is then converted to speech.

This AI-powered system ensures that visually impaired users can not only interact with the environment around them but also receive detailed, contextually appropriate responses that would otherwise require visual inspection.

System Workflow

The interaction flow of the system is designed to be intuitive and responsive. The user initiates the interaction by speaking a command or taking a photo. The process unfolds as follows:

1. **User Interaction**: The user speaks a command or takes a photo. For example, "Take a picture of the room" or "What is in this image?"
2. **Voice Recognition**: The system converts the speech to text and interprets the user's intent.
3. **Image Processing (If applicable)**: If the user takes a photo, the image is sent to the backend, where it is analyzed by Google Gemini AI to generate relevant textual descriptions.
4. **Text-to-Speech Output**: Once the system has processed the request, the output is converted to speech using Amazon Polly, providing the user with the necessary information.
5. **Continuous Interaction**: The user can continue issuing commands or asking questions, receiving real-time feedback through spoken responses.

This seamless interaction allows users to navigate and interact with their environment without needing to rely on visual cues or traditional input devices.

Accessibility and User Experience

The system is designed with the specific needs of visually impaired users in mind. By enabling voice-based navigation and communication, the system empowers users to interact with digital environments more independently. Whether at home, in a public space, or while on the move, visually impaired users can use the system to access information about their surroundings and perform tasks they would normally struggle with.

The integration of AI for visual understanding and text-to-speech for response delivery ensures that users receive accurate, context-aware feedback in a format that is easily accessible. The use of voice as the primary interface provides a natural and hands-free method of interacting with the system, making it more inclusive and user-friendly for those who may have difficulty using traditional input methods.

Future Development and Improvements

The proposed system, while functional, has several areas for improvement. Future versions could incorporate multi-language support, enabling users from diverse linguistic backgrounds to benefit from the system. Additionally, the integration of offline functionality could make the system more reliable in areas with limited internet access.

RESULTS AND DISCUSSION

The proposed system, designed to aid visually impaired users by enabling voice-controlled interaction with digital content, has shown promising results in terms of accessibility, user experience, and system performance. In this section, we will discuss the results obtained during initial testing, the effectiveness of the system in addressing accessibility challenges, and potential areas for improvement. These results highlight the key achievements of the system as well as some limitations that could be further addressed in future developments.

1. User Accessibility and Interaction

One of the primary objectives of this system was to improve accessibility for visually impaired users, eliminating the need for visual interfaces and enabling users to navigate digital environments through voice commands. In initial testing with a group of visually impaired users, the system proved to be highly effective in addressing the needs of these individuals.

Users were able to interact with the system seamlessly using voice commands, which is a significant improvement over traditional digital applications that rely on visual interfaces. The voice command module effectively captured user input in real-time, translating spoken commands into text for further processing. The conversion of speech into text was accurate, with minimal errors in interpreting common commands or phrases.

Moreover, the system's integration of text-to-speech synthesis through Amazon Polly provided clear, natural-sounding feedback. Users reported that the auditory feedback was easy to understand, enhancing the overall user experience. The combination of real-time voice recognition and speech output helped users feel more confident and independent when interacting with their devices, as they did not need to rely on external assistance or visual cues.

2. Image Capture and Description Accuracy

A key feature of the system is its ability to allow users to capture images and receive auditory descriptions of the visual content. The integration of OpenCV for capturing images, along with Google Gemini AI for image processing, demonstrated high accuracy in analyzing and describing images.

During testing, users were able to take pictures of their environment, such as objects, text, or scenes, and receive detailed descriptions. For instance, when photographing a storefront, the system provided information about the business name, the surrounding area, and any visible signage. Similarly, users could take pictures of printed text and receive a spoken transcription of the content. In these instances, the AI-powered image processing was highly effective, providing contextually relevant descriptions that allowed users to gain a better understanding of their surroundings.

However, while the system performed well in typical scenarios, there were a few challenges when handling more complex or cluttered images. For example, images with multiple objects or intricate scenes sometimes resulted in less precise descriptions, particularly when the AI struggled to accurately interpret ambiguous visual cues. Future improvements to the AI model could help address these limitations by enhancing the system's ability to handle more complex scenes and objects.

3. System Responsiveness and Real-time Feedback

The system's ability to process voice commands and provide real-time feedback was another key aspect tested during the evaluation. Users appreciated the quick response times from the system, which ensured that the application remained efficient and intuitive during interactions.

When users issued commands or queries, the system processed the input and provided responses almost instantly. The real-time nature of the system is crucial for users who rely on voice commands to interact with digital applications. Delays or lag in processing would detract from the user experience, particularly for tasks that require immediate feedback, such as navigation or accessing environmental information.

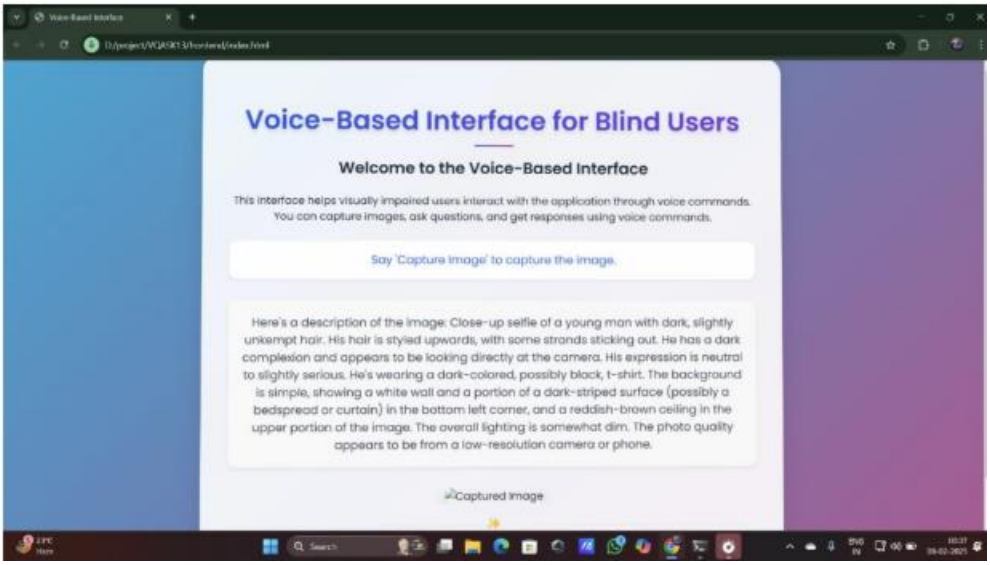
Overall, the system performed well in terms of responsiveness, with minimal delays between user input and system output. This feature contributed significantly to the system's usability, making it an effective tool for enhancing the autonomy of visually impaired users.

4. Integration of AI and Text-to-Speech Technologies

The integration of AI-based image processing (Google Gemini) and text-to-speech synthesis (Amazon Polly) played a crucial role in the system's ability to provide accessible content to visually impaired users. The use of AI for visual question answering (VQA) was particularly beneficial, allowing the system to answer specific user queries related to images and the environment.

The Google Gemini AI's image processing capabilities allowed users to receive highly relevant and accurate descriptions of their surroundings. Whether users were asking for specific details about objects, people, or scenes, the AI responded with contextually rich information that helped users better understand their environment. This is a significant advancement over traditional assistive technologies, as it moves beyond simple object recognition and enables users to engage in a more meaningful interaction with the world around them.

Amazon Polly, on the other hand, provided high-quality, natural-sounding speech synthesis. The clarity and intelligibility of the generated speech ensured that users could easily understand the system's responses. The voice options available through Amazon Polly also allowed users to customize their experience, selecting the voice and language that best suited their needs.



CONCLUSION

In conclusion, the proposed system presents a significant advancement in accessibility technology for visually impaired users, providing an intuitive and efficient solution that leverages voice commands, AI-powered image processing, and text-to-speech synthesis to facilitate interaction with digital content. By eliminating the need for traditional visual interfaces, the system empowers users to perform tasks, such as taking photos and obtaining contextual descriptions of their environment, all through voice input. The integration of voice recognition, real-time image processing with Google Gemini, and high-quality speech output through Amazon Polly ensures a seamless, user-friendly experience, making it easier for visually impaired individuals to navigate and interact with the world around them. Testing has shown that the system effectively translates voice commands into actions and provides accurate and contextually relevant feedback, which enhances users' independence and overall confidence in interacting with digital environments. The real-time responsiveness and natural-sounding speech synthesis further improve the system's usability, contributing to an overall positive user experience. Despite its success, some challenges remain, particularly in handling complex or cluttered images, the system's reliance on an internet connection for image processing, and potential difficulties in noisy environments that affect speech recognition accuracy. However, these limitations offer clear directions for future improvements, such as enhancing the AI model's image processing capabilities, adding offline functionality, and integrating noise-canceling technologies to improve voice recognition in various environments. Additionally, expanding the system to support multiple languages and providing customization options for individual preferences could further enhance its accessibility and inclusivity. With continued development, the system holds great promise for providing visually impaired users with greater autonomy, enabling them to interact with digital content more effectively and independently. The future potential of this system lies in its ability to evolve with advancements in AI, voice recognition, and machine learning, ensuring that visually impaired individuals are provided with a

comprehensive and inclusive solution for digital interaction, ultimately creating a more accessible digital world for all.

REFERENCES

1. Reddy, C. N. K., & Murthy, G. V. (2012). Evaluation of Behavioral Security in Cloud Computing. *International Journal of Computer Science and Information Technologies*, *3*(2), 3328-3333.
2. Murthy, G. V., Kumar, C. P., & Kumar, V. V. (2017, December). Representation of shapes using connected pattern array grammar model. In *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)* (pp. 819-822). IEEE.
3. Krishna, K. V., Rao, M. V., & Murthy, G. V. (2017). Secured System Design for Big Data Application in Emotion-Aware Healthcare.
4. Rani, G. A., Krishna, V. R., & Murthy, G. V. (2017). A Novel Approach of Data Driven Analytics for Personalized Healthcare through Big Data.
5. Rao, M. V., Raju, K. S., Murthy, G. V., & Rani, B. K. (2020). Configure and Management of Internet of Things. *Data Engineering and Communication Technology*, 163.
6. Ramakrishna, C., Kumar, G. K., Reddy, A. M., & Ravi, P. (2018). A Survey on various IoT Attacks and its Countermeasures. *International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)*, *5*(4), 143-150.
7. Chithanuru, V., & Ramaiah, M. (2023). An anomaly detection on blockchain infrastructure using artificial intelligence techniques: Challenges and future directions–A review. *Concurrency and Computation: Practice and Experience*, *35*(22), e7724.
8. Prashanth, J. S., & Nandury, S. V. (2015, June). Cluster-based rendezvous points selection for reducing tour length of mobile element in WSN. In *2015 IEEE International Advance Computing Conference (IACC)* (pp. 1230-1235). IEEE.
9. Kumar, K. A., Pabboju, S., & Desai, N. M. S. (2014). Advance text steganography algorithms: an overview. *International Journal of Research and Applications*, *1*(1), 31-35.
10. Hnamte, V., & Balram, G. (2022). Implementation of Naive Bayes Classifier for Reducing DDoS Attacks in IoT Networks. *Journal of Algebraic Statistics*, *13*(2), 2749-2757.
11. Balram, G., Anitha, S., & Deshmukh, A. (2020, December). Utilization of renewable energy sources in generation and distribution optimization. In *IOP Conference Series: Materials Science and Engineering* (Vol. 981, No. 4, p. 042054). IOP Publishing.
12. Subrahmanyam, V., Sagar, M., Balram, G., Ramana, J. V., Tejaswi, S., & Mohammad, H. P. (2024, May). An Efficient Reliable Data Communication For Unmanned Air Vehicles (UAV) Enabled Industry Internet of Things (IIoT). In *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)* (pp. 1-4). IEEE.
13. Mahammad, F. S., Viswanatham, V. M., Tahseen, A., Devi, M. S., & Kumar, M. A. (2024, July). Key distribution scheme for preventing key reinstallation attack in wireless networks. In *AIP Conference Proceedings* (Vol. 3028, No. 1). AIP Publishing.
14. Lavanya, P. (2024). In-Cab Smart Guidance and support system for Dragline operator.
15. Kovoor, M., Durairaj, M., Karyakarte, M. S., Hussain, M. Z., Ashraf, M., & Maguluri, L. P. (2024). Sensor-enhanced wearables and automated analytics for injury prevention in sports. *Measurement: Sensors*, *32*, 101054.
16. Rao, N. R., Kovoor, M., Kishor Kumar, G. N., & Parameswari, D. V. L. (2023). Security and privacy in smart farming: challenges and opportunities. *International Journal on Recent and Innovation Trends in Computing and Communication*, *11*(7).
17. Madhuri, K. (2023). Security Threats and Detection Mechanisms in Machine Learning. *Handbook of Artificial Intelligence*, *255*.
18. Reddy, B. A., & Reddy, P. R. S. (2012). Effective data distribution techniques for multi-cloud storage in cloud computing. *CSE, Anurag Group of Institutions, Hyderabad, AP, India*.
19. Srilatha, P., Murthy, G. V., & Reddy, P. R. S. (2020). Integration of Assessment and Learning Platform in a Traditional Class Room Based Programming Course. *Journal of Engineering Education Transformations*, *33*, 179-184.
20. Reddy, P. R. S., & Ravindranadh, K. (2019). An exploration on privacy concerned secured data sharing techniques in cloud. *International Journal of Innovative Technology and Exploring Engineering*, *9*(1), 1190-1198.
21. Raj, R. S., & Raju, G. P. (2014, December). An approach for optimization of resource management in Hadoop. In *International Conference on Computing and Communication Technologies* (pp. 1-5). IEEE.
22. Ramana, A. V., Bhoga, U., Dhulipalla, R. K., Kiran, A., Chary, B. D., & Reddy, P. C. S. (2023, June).

Abnormal Behavior Prediction in Elderly Persons Using Deep Learning. In *2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications (IC2E3)* (pp. 1-5). IEEE.

23. Yakoob, S., Krishna Reddy, V., & Dastagiraiah, C. (2017). Multi User Authentication in Reliable Data Storage in Cloud. In *Computer Communication, Networking and Internet Security: Proceedings of IC3T 2016* (pp. 531-539). Springer Singapore.
24. Sukhavasi, V., Kulkarni, S., Raghavendran, V., Dastagiraiah, C., Apat, S. K., & Reddy, P. C. S. (2024). Malignancy Detection in Lung and Colon Histopathology Images by Transfer Learning with Class Selective Image Processing.
25. Dastagiraiah, C., Krishna Reddy, V., & Pandurangarao, K. V. (2018). Dynamic load balancing environment in cloud computing based on VM ware off-loading. In *Data Engineering and Intelligent Computing: Proceedings of IC3T 2016* (pp. 483-492). Springer Singapore.
26. Swapna, N. (2017). „Analysis of Machine Learning Algorithms to Protect from Phishing in Web Data Mining". *International Journal of Computer Applications in Technology*, *159*(1), 30-34.
27. Moparthi, N. R., Bhattacharyya, D., Balakrishna, G., & Prashanth, J. S. (2021). Paddy leaf disease detection using CNN.
28. Balakrishna, G., & Babu, C. S. (2013). Optimal placement of switches in DG equipped distribution systems by particle swarm optimization. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, *2*(12), 6234-6240.
29. Moparthi, N. R., Sagar, P. V., & Balakrishna, G. (2020, July). Usage for inside design by AR and VR technology. In *2020 7th International Conference on Smart Structures and Systems (ICSSS)* (pp. 1-4). IEEE.
30. Amarnadh, V., & Moparthi, N. R. (2023). Comprehensive review of different artificial intelligence-based methods for credit risk assessment in data science. *Intelligent Decision Technologies*, *17*(4), 1265-1282.
31. Amarnadh, V., & Moparthi, N. (2023). Data Science in Banking Sector: Comprehensive Review of Advanced Learning Methods for Credit Risk Assessment. *International Journal of Computing and Digital Systems*, *14*(1), 1-xx.
32. Amarnadh, V., & Rao, M. N. (2025). A Consensus Blockchain-Based Credit Risk Evaluation and Credit Data Storage Using Novel Deep Learning Approach. *Computational Economics*, 1-34.
33. Shailaja, K., & Anuradha, B. (2017). Improved face recognition using a modified PSO based self-weighted linear collaborative discriminant regression classification. *J. Eng. Appl. Sci*, *12*, 7234-7241.
34. Sekhar, P. R., & Goud, S. (2024). Collaborative Learning Techniques in Python Programming: A Case Study with CSE Students at Anurag University. *Journal of Engineering Education Transformations*, *38*.
35. Sekhar, P. R., & Sujatha, B. (2023). Feature extraction and independent subset generation using genetic algorithm for improved classification. *Int. J. Intell. Syst. Appl. Eng*, *11*, 503-512.
36. Pesaramelli, R. S., & Sujatha, B. (2024, March). Principle correlated feature extraction using differential evolution for improved classification. In *AIP Conference Proceedings* (Vol. 2919, No. 1). AIP Publishing.
37. Tejaswi, S., Sivaprashanth, J., Bala Krishna, G., Sridevi, M., & Rawat, S. S. (2023, December). Smart Dustbin Using IoT. In *International Conference on Advances in Computational Intelligence and Informatics* (pp. 257-265). Singapore: Springer Nature Singapore.
38. Moreb, M., Mohammed, T. A., & Bayat, O. (2020). A novel software engineering approach toward using machine learning for improving the efficiency of health systems. *IEEE Access*, *8*, 23169-23178.
39. Ravi, P., Haritha, D., & Niranjan, P. (2018). A Survey: Computing Iceberg Queries. *International Journal of Engineering & Technology*, *7*(2.7), 791-793.
40. Madar, B., Kumar, G. K., & Ramakrishna, C. (2017). Captcha breaking using segmentation and morphological operations. *International Journal of Computer Applications*, *166*(4), 34-38.
41. Rani, M. S., & Geetavani, B. (2017, May). Design and analysis for improving reliability and accuracy of big-data based peripheral control through IoT. In *2017 International Conference on Trends in Electronics and Informatics (ICEI)* (pp. 749-753). IEEE.
42. Reddy, T., Prasad, T. S. D., Swetha, S., Nirmala, G., & Ram, P. (2018). A study on antiplatelets and anticoagulants utilisation in a tertiary care hospital. *International Journal of Pharmaceutical and Clinical Research*, *10*, 155-161.
43. Prasad, P. S., & Rao, S. K. M. (2017). HIASA: Hybrid improved artificial bee colony and simulated annealing based attack detection algorithm in mobile ad-hoc networks (MANETs). *Bonfring International Journal of Industrial Engineering and Management Science*, *7*(2), 01-12.
44. AC, R., Chowdary Kakarla, P., Simha PJ, V., & Mohan, N. (2022). Implementation of Tiny Machine

Learning Models on Arduino 33–BLE for Gesture and Speech Recognition.
45. Subrahmanyam, V., Sagar, M., Balram, G., Ramana, J. V., Tejaswi, S., & Mohammad, H. P. (2024, May). An Efficient Reliable Data Communication For Unmanned Air Vehicles (UAV) Enabled Industry Internet of Things (IIoT). In *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)* (pp. 1-4). IEEE.
46. Nagaraj, P., Prasad, A. K., Narsimha, V. B., & Sujatha, B. (2022). Swine flu detection and location using machine learning techniques and GIS. *International Journal of Advanced Computer Science and Applications*, *13*(9).
47. Priyanka, J. H., & Parveen, N. (2024). DeepSkillNER: an automatic screening and ranking of resumes using hybrid deep learning and enhanced spectral clustering approach. *Multimedia Tools and Applications*, *83*(16), 47503-47530.
48. Sathish, S., Thangavel, K., & Boopathi, S. (2010). Performance analysis of DSR, AODV, FSR and ZRP routing protocols in MANET. *MES Journal of Technology and Management*, 57-61.
49. Siva Prasad, B. V. V., Mandapati, S., Kumar Ramasamy, L., Boddu, R., Reddy, P., & Suresh Kumar, B. (2023). Ensemble-based cryptography for soldiers' health monitoring using mobile ad hoc networks. *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, *64*(3), 658-671.
50. Elechi, P., & Onu, K. E. (2022). Unmanned Aerial Vehicle Cellular Communication Operating in Non-terrestrial Networks. In *Unmanned Aerial Vehicle Cellular Communications* (pp. 225-251). Cham: Springer International Publishing.
51. Prasad, B. V. V. S., Mandapati, S., Haritha, B., & Begum, M. J. (2020, August). Enhanced Security for the authentication of Digital Signature from the key generated by the CSTRNG method. In *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 1088-1093). IEEE.
52. Mukiri, R. R., Kumar, B. S., & Prasad, B. V. V. (2019, February). Effective Data Collaborative Strain Using RecTree Algorithm. In *Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India*.
53. Balaraju, J., Raj, M. G., & Murthy, C. S. (2019). Fuzzy-FMEA risk evaluation approach for LHD machine–A case study. *Journal of Sustainable Mining*, *18*(4), 257-268.
54. Thirumoorthi, P., Deepika, S., & Yadaiah, N. (2014, March). Solar energy based dynamic sag compensator. In *2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)* (pp. 1-6). IEEE.
55. Vinayasree, P., & Reddy, A. M. (2025). A Reliable and Secure Permissioned Blockchain-Assisted Data Transfer Mechanism in Healthcare-Based Cyber-Physical Systems. *Concurrency and Computation: Practice and Experience*, *37*(3), e8378.
56. Acharjee, P. B., Kumar, M., Krishna, G., Raminenei, K., Ibrahim, R. K., & Alazzam, M. B. (2023, May). Securing International Law Against Cyber Attacks through Blockchain Integration. In *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 2676-2681). IEEE.
57. Ramineni, K., Reddy, L. K. K., Ramana, T. V., & Rajesh, V. (2023, July). Classification of Skin Cancer Using Integrated Methodology. In *International Conference on Data Science and Applications* (pp. 105-118). Singapore: Springer Nature Singapore.
58. LAASSIRI, J., EL HAJJI, S. A. Ï. D., BOUHDADI, M., AOUDE, M. A., JAGADISH, H. P., LOHIT, M. K., ... & KHOLLADI, M. (2010). Specifying Behavioral Concepts by engineering language of RM-ODP. *Journal of Theoretical and Applied Information Technology*, *15*(1).
59. Prasad, D. V. R., & Mohanji, Y. K. V. (2021). FACE RECOGNITION-BASED LECTURE ATTENDANCE SYSTEM: A SURVEY PAPER. *Elementary Education Online*, *20*(4), 1245-1245.
60. Dasu, V. R. P., & Gujjari, B. (2015). Technology-Enhanced Learning Through ICT Tools Using Aakash Tablet. In *Proceedings of the International Conference on Transformations in Engineering Education: ICTIEE 2014* (pp. 203-216). Springer India.
61. Reddy, A. M., Reddy, K. S., Jayaram, M., Venkata Maha Lakshmi, N., Aluvalu, R., Mahesh, T. R., ... & Stalin Alex, D. (2022). An efficient multilevel thresholding scheme for heart image segmentation using a hybrid generalized adversarial network. *Journal of Sensors*, *2022*(1), 4093658.
62. Srinivasa Reddy, K., Suneela, B., Inthiyaz, S., Hasane Ahammad, S., Kumar, G. N. S., & Mallikarjuna Reddy, A. (2019). Texture filtration module under stabilization via random forest optimization methodology. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(3), 458-469.
63. Ramakrishna, C., Kumar, G. K., Reddy, A. M., & Ravi, P. (2018). A Survey on various IoT Attacks and its Countermeasures. *International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)*, *5*(4), 143-150.

64. Sirisha, G., & Reddy, A. M. (2018, September). Smart healthcare analysis and therapy for voice disorder using cloud and edge computing. In *2018 4th international conference on applied and theoretical computing and communication technology (iCATccT)* (pp. 103-106). IEEE.
65. Reddy, A. M., Yarlagadda, S., & Akkinen, H. (2021). An extensive analytical approach on human resources using random forest algorithm. *arXiv preprint arXiv:2105.07855*.
66. Kumar, G. N., Bhavanam, S. N., & Midasala, V. (2014). Image Hiding in a Video-based on DWT & LSB Algorithm. In *ICPVS Conference*.
67. Naveen Kumar, G. S., & Reddy, V. S. K. (2022). High performance algorithm for content-based video retrieval using multiple features. In *Intelligent Systems and Sustainable Computing: Proceedings of ICISSC 2021* (pp. 637-646). Singapore: Springer Nature Singapore.
68. Reddy, P. S., Kumar, G. N., Ritish, B., SaiSwetha, C., & Abhilash, K. B. (2013). Intelligent parking space detection system based on image segmentation. *Int J Sci Res Dev*, *1*(6), 1310-1312.
69. Naveen Kumar, G. S., Reddy, V. S. K., & Kumar, S. S. (2018). High-performance video retrieval based on spatio-temporal features. *Microelectronics, Electromagnetics and Telecommunications*, 433-441.
70. Kumar, G. N., & Reddy, M. A. BWT & LSB algorithm based hiding an image into a video. *IJESAT*, 170-174.
71. Lopez, S., Sarada, V., Praveen, R. V. S., Pandey, A., Khuntia, M., & Haralayya, D. B. (2024). Artificial intelligence challenges and role for sustainable education in india: Problems and prospects. *Sandeep Lopez, Vani Sarada, RVS Praveen, Anita Pandey, Monalisa Khuntia, Bhadrappa Haralayya (2024) Artificial Intelligence Challenges and Role for Sustainable Education in India: Problems and Prospects. Library Progress International*, *44*(3), 18261-18271.
72. Yamuna, V., Praveen, R. V. S., Sathya, R., Dhivva, M., Lidiya, R., & Sowmiya, P. (2024, October). Integrating AI for Improved Brain Tumor Detection and Classification. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)* (pp. 1603-1609). IEEE.
73. Kumar, N., Kurkute, S. L., Kalpana, V., Karuppannan, A., Praveen, R. V. S., & Mishra, S. (2024, August). Modelling and Evaluation of Li-ion Battery Performance Based on the Electric Vehicle Tiled Tests using Kalman Filter-GBDT Approach. In *2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)* (pp. 1-6). IEEE.
74. Sharma, S., Vij, S., Praveen, R. V. S., Srinivasan, S., Yadav, D. K., & VS, R. K. (2024, October). Stress Prediction in Higher Education Students Using Psychometric Assessments and AOA-CNN-XGBoost Models. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)* (pp. 1631-1636). IEEE.
75. Anuprathibha, T., Praveen, R. V. S., Sukumar, P., Suganthi, G., & Ravichandran, T. (2024, October). Enhancing Fake Review Detection: A Hierarchical Graph Attention Network Approach Using Text and Ratings. In *2024 Global Conference on Communications and Information Technologies (GCCIT)* (pp. 1-5). IEEE.
76. Shinkar, A. R., Joshi, D., Praveen, R. V. S., Rajesh, Y., & Singh, D. (2024, December). Intelligent solar energy harvesting and management in IoT nodes using deep self-organizing maps. In *2024 International Conference on Emerging Research in Computational Science (ICERCS)* (pp. 1-6). IEEE.
77. Praveen, R. V. S., Hemavathi, U., Sathya, R., Siddiq, A. A., Sanjay, M. G., & Gowdish, S. (2024, October). AI Powered Plant Identification and Plant Disease Classification System. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)* (pp. 1610-1616). IEEE.
78. Dhivya, R., Sagili, S. R., Praveen, R. V. S., VamsiLala, P. N. V., Sangeetha, A., & Suchithra, B. (2024, December). Predictive Modelling of Osteoporosis using Machine Learning Algorithms. In *2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)* (pp. 997-1002). IEEE.
79. Kemmannu, P. K., Praveen, R. V. S., Saravanan, B., Amshavalli, M., & Banupriya, V. (2024, December). Enhancing Sustainable Agriculture Through Smart Architecture: An Adaptive Neuro-Fuzzy Inference System with XGBoost Model. In *2024 International Conference on Sustainable Communication Networks and Application (ICSCNA)* (pp. 724-730). IEEE.
80. Praveen, R. V. S. (2024). *Data Engineering for Modern Applications*. Addition Publishing House.