


**Please cite the Published Version**

Ahmad, Awais, Jabbar, Sohail, Akram, Sheraz, Anand, Paul, Raza, Umar  and Alshuqayran, Nuha (2024) Enhancing ChatGPT's querying capability with voice-based interaction and a CNN-based impaired vision detection model. Computers, Materials and Continua. ISSN 1546-2218

**DOI:** <https://doi.org/10.32604/cmc.2024.045385>

**Publisher:** Tech Science Press

**Version:** Published Version

**Downloaded from:** <https://e-space.mmu.ac.uk/633955/>

**Usage rights:**  [Creative Commons: Attribution 4.0](https://creativecommons.org/licenses/by/4.0/)

**Additional Information:** This is an open access article which originally appeared in Computers, Materials and Continua, published by Tech Science Press

**Data Access Statement:** The raw data supporting the conclusions of this article will be made available by the authors.

**Enquiries:**

If you have questions about this document, contact [openresearch@mmu.ac.uk](mailto:openresearch@mmu.ac.uk). Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)



ARTICLE

## Enhancing ChatGPT's Querying Capability with Voice-Based Interaction and CNN-Based Impair Vision Detection Model

Awais Ahmad<sup>1</sup>, Sohail Jabbar<sup>1,\*</sup>, Sheeraz Akram<sup>1</sup>, Anand Paul<sup>2</sup>, Umar Raza<sup>3</sup> and Nuha Mohammed Alshuqayran<sup>1</sup>

<sup>1</sup>College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, 11432, Saudia Arabia

<sup>2</sup>School of Computer Science and Engineering, Kyungpook National University, Daegu, 41566, South Korea

<sup>3</sup>Department of Engineering, Manchester Metropolitan University, Manchester, M15 6BH, UK

\*Corresponding Author: Sohail Jabbar. Email: [sjjabar@imamu.edu.sa](mailto:sjjabar@imamu.edu.sa)

Received: 25 August 2023 Accepted: 04 December 2023

### ABSTRACT

This paper presents an innovative approach to enhance the querying capability of ChatGPT, a conversational artificial intelligence model, by incorporating voice-based interaction and a convolutional neural network (CNN)-based impaired vision detection model. The proposed system aims to improve user experience and accessibility by allowing users to interact with ChatGPT using voice commands. Additionally, a CNN-based model is employed to detect impairments in user vision, enabling the system to adapt its responses and provide appropriate assistance. This research tackles head-on the challenges of user experience and inclusivity in artificial intelligence (AI). It underscores our commitment to overcoming these obstacles, making ChatGPT more accessible and valuable for a broader audience. The integration of voice-based interaction and impaired vision detection represents a novel approach to conversational AI. Notably, this innovation transcends novelty; it carries the potential to profoundly impact the lives of users, particularly those with visual impairments. The modular approach to system design ensures adaptability and scalability, critical for the practical implementation of these advancements. Crucially, the solution places the user at its core. Customizing responses for those with visual impairments demonstrates AI's potential to not only understand but also accommodate individual needs and preferences.

### KEYWORDS

Accessibility in conversational AI; CNN-based impair vision detection; ChatGPT; voice-based interaction; recommender system

## 1 Introduction

The study built on surveys [1] conducted in 39 countries that meet the criteria for inclusion. The global estimate suggests that approximately 285 million individuals of all age groups are affected by visual impairment, out of which 39 million are classified as blind, with an uncertainty range of 10–20%. The visually impaired population is primarily concentrated among those aged 50 and above, accounting for 65% and 82% of the visually impaired and blind individuals, respectively. The



leading causes of visual impairment are uncorrected refractive errors (43%) and cataracts (33%), while cataracts alone account for the majority (51%) of blindness cases.

Conversational artificial intelligence (AI) systems have witnessed remarkable advancements in recent years, revolutionizing the way humans interact with technology [2]. The system known as ChatGPT enables natural language conversation and provides valuable assistance across various domains. However, there still exists a gap to be filled with improvements in terms of user experience and accessibility. Similarly, other assistive tools like smart health wearables are crucial for continuous monitoring of vital signs and health metrics. They provide personalized insights, enable remote health monitoring, and encourage positive behaviour changes. They aid in early disease detection, integrate with telemedicine, and offer emergency assistance. Additionally, aggregated data supports research and population health analysis [3].

Traditionally, the ChatGPT system primarily relies on text-based conversations, limiting the convenience and ease of use for users [4]. Users are required to type their queries, which becomes time-consuming and cumbersome, especially when manual input is challenging and not feasible [5,6]. To highlight limitations, integrating voice-based conversation with ChatGPT can exponentially enhance the user experience by enabling them to interact through natural speech recognition [7]. Ensuring inclusivity and accessibility in conversational AI systems is of significant importance to gain better results. Especially considering users have visual impairments, making it practically impossible for them to comprehend text-based responses without any assistance. By incorporating a convolutional neural network (CNN)-based impaired vision detection model into ChatGPT, we can detect impairments in user vision and adjust the system's responses accordingly. This adaptive behaviour ensures that users with visual impairments gain appropriate assistance and information [8].

In this paper, the study a modular approach to augment ChatGPT's querying capability through the integration of voice-based interaction and a CNN-based impair vision detection model. The main objectives of this research are divided into two: (1) improve the accessibility of the system by detecting impairments in user vision and adapting the system's responses accordingly and (2) enhance the user experience by allowing voice-based interactions with ChatGPT. It aims to create a more comprehensive and user-friendly conversational AI system by removing limitations of traditional text-based AI conversation and focusing on enhanced accessibility. The proposed approach has the potential to revolutionize how users interact with ChatGPT, paving the way for a more accessible and engaging conversational AI experience.

The following advantages highlight the positive outcomes and benefits of the research conducted, specifically in the context of enhancing ChatGPT's querying capability with voice-based interaction and impaired vision detection:

- **Improved User Experience:** The integration of voice-based interaction and impaired vision detection enhances the overall user experience, making ChatGPT more intuitive and accommodating.
- **Accessibility:** This research addresses the critical issue of accessibility, particularly for users with visual impairments, by providing tailored responses and accommodations.
- **Inclusivity:** By recognizing and accommodating diverse user needs and preferences, our approach promotes inclusivity in AI interactions.
- **Efficiency:** Voice-based interaction streamlines communication, offering a more efficient and natural way for users to engage with ChatGPT.
- **Personalization:** The impaired vision detection module enables ChatGPT to offer personalized responses and support, meeting individual user requirements.

- **Scalability:** Our modular approach ensures scalability, making it feasible to implement these advancements in a range of AI systems.
- **Flexibility:** The flexibility of our system allows for future enhancements and adaptability to evolving user needs.
- **Real-World Impact:** Beyond technical innovation, our work has the potential to significantly impact the daily lives of users, particularly those with visual impairments.
- **User-Centric Development:** By focusing on user needs, our research exemplifies a user-centric approach to AI development, aligning with the broader goal of creating technology that benefits everyone.
- **Advancements in Conversational AI:** This research contributes to the ongoing advancements in the field of conversational AI, highlighting the importance of natural language understanding and accommodating diverse user needs.

The main contributions of this study are:

- (1) Blind person identification using CNN-trained modals.
- (2) An assistive system module uses the fusion of Voice-to-Text, Text-to-Voice and ChatGPT for collecting requests through voice and responding intelligently.
- (3) Assessing the current state of ChatGPT as assistive module capabilities and its gaps.

The remainder of this paper is organized as follows. [Section 2](#) provides an overview of related work in conversational AI and accessibility technologies. [Section 3](#) describes the methodology, detailing the integration of voice-based interaction and the CNN-based impaired vision detection model with ChatGPT. [Section 4](#) presents the experimental setup and evaluation metrics. The results and analysis are discussed in [Section 5](#). Finally, [Section 6](#) concludes the paper and outlines future directions for research in this domain.

## 2 Related Work

### 2.1 Conversational AI Systems

Conversational recommendation systems are AI-based systems designed to provide personalized recommendations to users through interactive conversations [9]. These systems leverage natural language processing (NLP) techniques to understand user queries, preferences, and context, and then generate relevant recommendations accordingly [10].

Conversational AI systems have gained significant attention in both academia and industry. Several approaches have been proposed to improve the natural language understanding and generation capabilities of Conversational AI systems [11,12]. In the previous era, similar systems highly relied on rule-based methods and predefined templates for generating responses. However, recent advancements in deep learning, particularly with the advent of transformer-based models like GPT, have revolutionized the field of conversational AI. Whereas models such as ChatGPT use a large-scale pre-trained model on vast amounts of text data to generate contextually relevant and coherent responses [13,14]. While these models have shown impressive performance, there is still scope for enhancing user experience and accessibility.

The major emphasis of conversational recommendation systems is to enhance the user experience by offering tailored recommendations conversationally, similar to a human conversation. Instead of presenting a list of static recommendations, these systems engage users in a back-and-forth dialogue to gather more information and refine their understanding of the user's preferences.



## ***2.2 Voice-Based Interaction***

Voice-based interaction has emerged as a popular means of interacting with AI systems, offering convenience and hands-free assistance. Several voice assistants, such as Siri, Alexa, and Google Assistant, have gained widespread adaptation among users on daily routine tasks. These systems employ automatic speech recognition (ASR) to convert spoken language into text, which is then processed by natural language understanding (NLU) components to derive user intent and generate appropriate responses [15,16]. Integrating voice-based interaction into conversational AI systems like ChatGPT can significantly improve user experience, as users can interact more naturally and effortlessly through speech.

## ***2.3 Accessibility in Conversational AI***

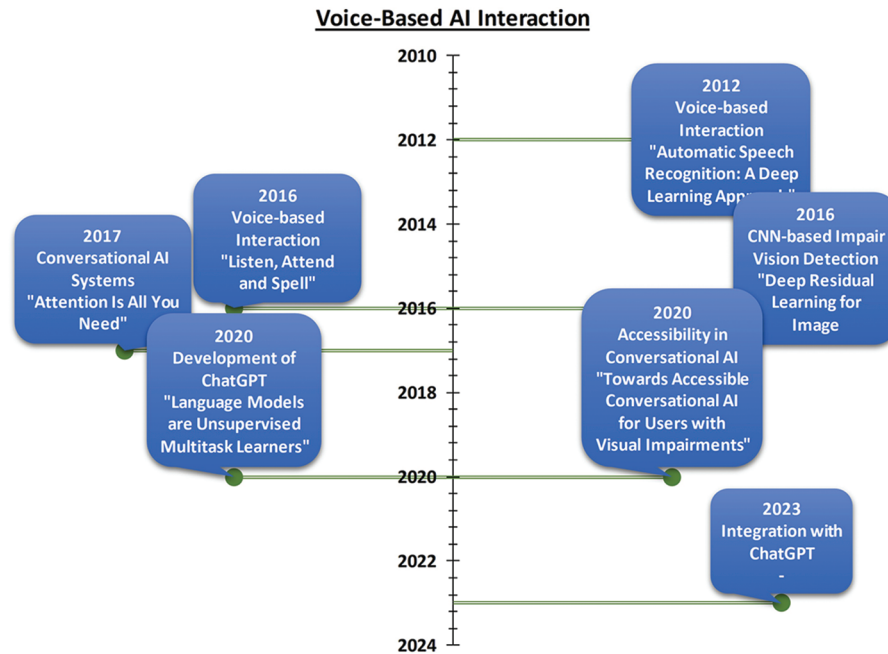
Ensuring accessibility for users with disabilities is a crucial aspect of designing inclusive conversational AI systems. Various efforts have been made to enhance accessibility, particularly for users with visual impairments. Text-to-speech (TTS) technology has been widely employed to provide auditory feedback for visually impaired users. Additionally, techniques such as screen readers and braille displays facilitate access to textual content. However, these approaches often rely on post-processing techniques and external assistive technologies. Integrating impaired vision detection directly into the conversational AI system can enable more tailored and adaptive responses, enhancing accessibility for users with visual impairments [7,17].

## ***2.4 CNN-Based Impair Vision Detection***

ChatGPT has a long list of applications, especially in the medical field. Some state-of-the-art research work, particularly in this domain are [18–22]. Convolutional neural networks (CNNs) have demonstrated exceptional performance in image classification and object detection tasks. Recent studies have explored the application of CNNs for vision impairment detection. These models analyze input images to identify specific impairments, such as blurriness, low contrast, or occlusions, which can affect visual perception [23]. By incorporating a CNN-based impaired vision detection model into ChatGPT, we can detect the presence of visual impairments in user images and modify the system's responses accordingly. This adaptive behaviour ensures that visually impaired users receive tailored and relevant information, improving their overall conversational AI experience [8].

In Fig. 1, the extended timeline highlights key advancements in voice-based interaction, CNN-based impaired vision detection, conversational AI systems, and accessibility in conversational AI. It also includes the development of ChatGPT as a significant milestone in the field, leading to the integration of voice-based interaction and impaired vision detection with ChatGPT for improved user experience and inclusivity.

In summary, previous research has focused on improving the natural language understanding and generation capabilities of conversational AI systems. Voice-based interaction has gained popularity for its convenience and hands-free operation. Furthermore, efforts have been made to enhance the accessibility of these systems, particularly for users with visual impairments. CNN-based impaired vision detection models have shown promise in identifying visual impairments from images. By integrating voice-based interaction and a CNN-based impaired vision detection model into ChatGPT, this paper aims to advance the state of the art in conversational AI systems by creating a more inclusive and accessible user experience. More work on the involvement of ChatGPT in various fields are [24,25].



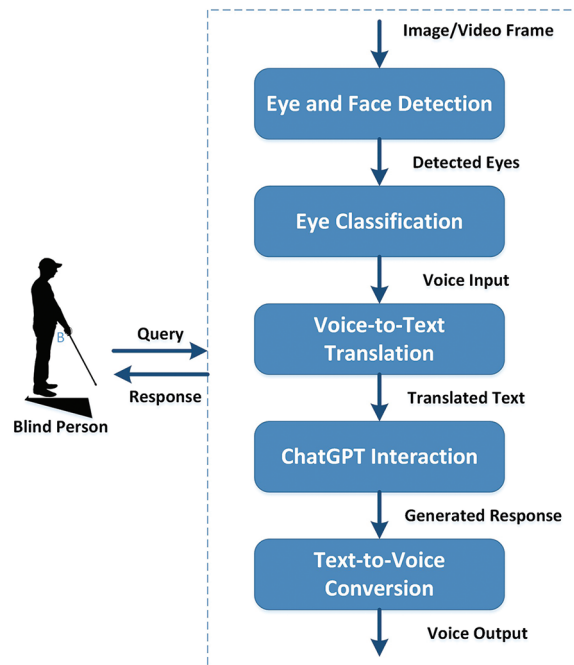
**Figure 1:** Voice-based AI evolution

### 3 Voice-Enabled ChatGPT: Improving Querying with CNN-Based Impair Vision Detection

Shafeeg et al. [26] proposed a Voice assistant integrated with ChatGPT but there is a lack of any latest AI technique and its specialized benefit for impairment. Boussougou et al. [27] proposed an attention-based 1D CNN-BiLSTM hybrid model that is enhanced with FastText word embedding for Korean voice phishing detection. In this research article, we have proposed a Voice-Enabled ChatGPT model that involves a two-step process for assisting blind individuals (as shown in Fig. 2). In the first step, the detection of the face and eyes is acquired using a trained CNN (Convolutional Neural Network) ResNet model. The ResNet 32 model is trained to accurately identify and locate the faces in an image or a video frame.

Once the eyes are extracted, a second CNN model is used for further detection by eye impairment classification. The model is trained to classify the detected eye into three categories: Blind, Normal, or Shut (for Closed). The model uses the features of the eye to determine whether the person is blind, or has a normal vision, or if the eye is closed.

In case the person is predicted as blind by the trained model, the voice-based assistance module gets activated. The assistance can take the form of a voice assistant that interacts with the blind individual. Whenever the blind individual asks a query, the system translates their voice into text format. The text gets forwarded to ChatGPT, as an AI language model capable of understanding and generating human-like text responses. ChatGPT processes the query and generates a response. The response from ChatGPT is transformed back into voice format in an attempt for the blind person to hear the generated answer. Furthermore, the system uses technologies like speech translation through synthesis to convert text into voice to enable communication as a response with the blind person. The assistance process continues in a recursive loop, where the blind person can keep asking follow-up questions or request further information. The system keeps responding with continuous assistance provided to the blind.



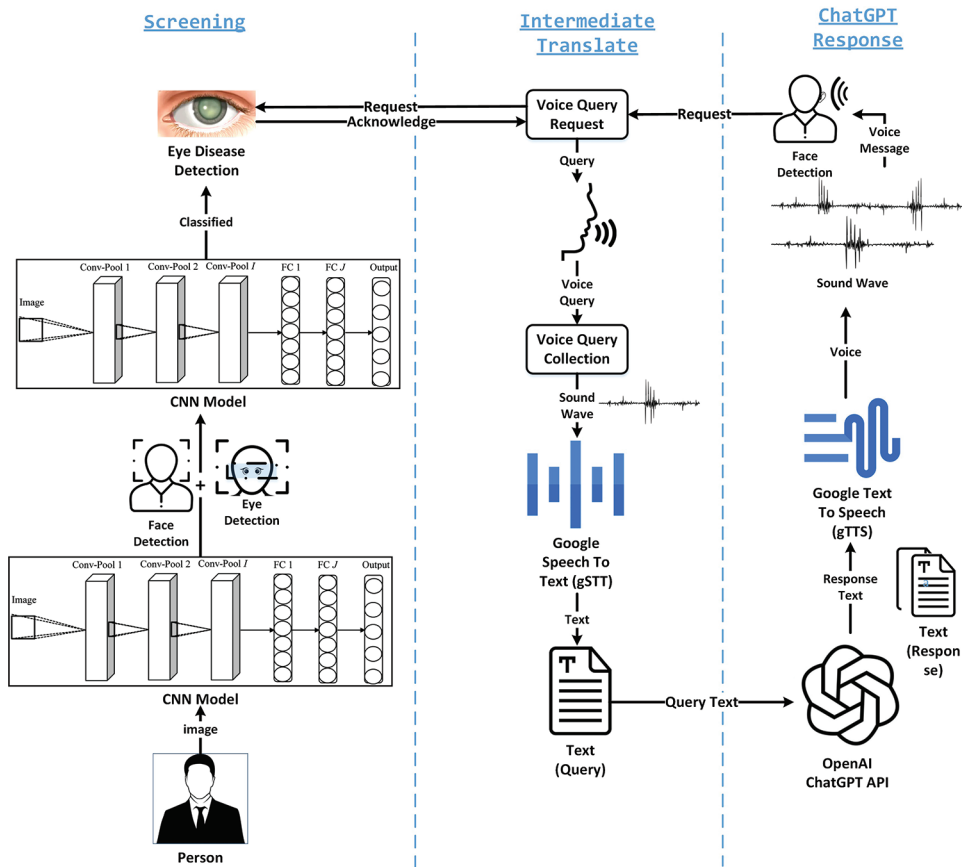
**Figure 2:** Modular model of the voice-based AI conversational system for the blind

In this study, a novel approach is used to enhance the conversational capability of ChatGPT by incorporating voice-based interaction and a CNN-based impaired vision detection model (as shown in Fig. 3). This research aims to enable users to interact with ChatGPT using voice queries by providing a more natural and efficient conversational experience. Additionally, the model includes a convolutional neural network (CNN) that can detect impairments in vision to provide tailored responses and accommodations for users. The fusion of these approaches aims to improve the accessibility and effectiveness of ChatGPT in assisting users with diverse needs.

**CNN-Based Impair Vision Detection Model:** To enhance ChatGPT’s capability while detecting impaired vision, a Convolutional Neural Network (CNN) based model plays a significant role. CNN models are suited for image-based feature detection tasks including visual impairment.

**Enhancing ChatGPT’s Querying Capability with Voice-Based Interaction:** The voice-based interaction with ChatGPT can expand its querying capability greatly along with making it more accessible and user-friendly. By allowing users to input queries through voice, ChatGPT can understand and respond to the natural language spoken by the user. In case of texting becomes difficult or impractical, the proposed approach will make the interaction more intuitive and convenient.

For voice-based interaction, a speech recognition system gets integrated with ChatGPT. The system translates spoken language into text, which can further be processed by ChatGPT to generate responses. Popular speech recognition APIs like Google Cloud Speech-to-Text are utilized for the translation from speech into text. Additionally, speech synthesis capabilities are incorporated to enable ChatGPT to respond to users. Google Text-to-Speech gets integrated to generate human-like voices for the responses.



**Figure 3:** Voice-enabled ChatGPT proposed model for visual impaired

Following is a high-level overview of the CNN-based model for impaired vision detection integrated with ChatGPT:

**Dataset collection:** Gather a diverse dataset of images representing various visual impairments, such as cataracts, glaucoma, macular degeneration, etc. Each image should be labelled with the corresponding impairment.

**Data preprocessing:** The collected dataset is augmented by resizing the images to an appropriate size, normalized pixel values, and splitting data into training and testing sets.

**Model training:** A CNN model is designed and trained using the labelled dataset. The model gets trained to classify images into different categories like Normal, Closed and Impaired.

**Integration with ChatGPT:** Finally, the trained CNN model gets integrated with the ChatGPT Query chatbot.

The image of the user gets passed through the CNN model for eye detection and another CNN Model for impairment detection. The model will classify the image by predicting the presence of the visual impairment. With the integration of voice-based interaction and a CNN-based impaired vision detection model, ChatGPT can become more adaptable by allowing users to interact through voice conversations and providing insights about query outcomes vocally.

### 3.1 Mathematical Modeling

Let us further develop the mathematical formulation of the proposed model, incorporating the CNN model for blind person detection and the querying system:

#### Phase 1: Detection of Blind Individuals Using CNN

Let:

- $N$  be the total number of individuals.
- $X(n)$  be the image of the  $n$ th individual.
- $D(n)$  be a binary variable indicating if the  $n$ th individual is blind  $D(n) = 1$  if blind, and  $D(n) = 0$  otherwise
- $C(n, c)$  be a binary variable representing the classification of the  $n$ th individual's image into category  $c$   
 $C(n, c) = 1$  if classified as  $c$ , and  $C(n, c) = 0$  otherwise
- $\theta$  be the set of parameters of the CNN model.

The mathematical model for blind person detection can be described as follows:

#### Objective:

Minimize the classification error in detecting blind individuals:

$$\text{minimize} \sum_{n=1}^N \sum_{c=1}^C (C(n, c) - D(n))^2 \quad (1)$$

#### Subject to:

CNN Model for Image Classification:

$$C(n, c) = \{1, \text{ if CNN classify } (X(n), c, \theta) > 0 \text{ otherwise}\} \quad (2)$$

for  $n = 1$  to  $N$ ,  $c = 1$  to  $C$ , where CNN classify () represents the classification function using the CNN model and its parameters  $\theta$ .

#### Phase 2: Querying System

Let:

- $P(n)$  be a binary variable indicating whether the system can be of help to the  $n$ th individual  
 $P(n) = 1$  if help can be provided, and  $P(n) = 0$  otherwise
- $Q(n, t)$  be a binary variable representing the presence of a query voice message from the  $n$ th individual at time  $t$   
 $Q(n, t) = 1$  if a message is received, and  $Q(n, t) = 0$  otherwise
- $T(n, t)$  be a string variable representing the text obtained by converting the voice message  $Q(n, t)$  into text format using the Google API.
- $R(n, t)$  be a string variable denoting the response generated by ChatGPT for the text  $T(n, t)$ .
- $V(n, t)$  be a binary variable indicating the presence of a voice message translated back into voice format for the  $n$ th individual at time  $t$   
 $V(n, t) = 1$  if message exists, and  $V(n, t) = 0$  otherwise
- $S(n, t)$  be a string variable representing the final voice message sent to the requesting individual at time  $t$ .

- $A(n, t)$  be a binary variable indicating if further assistance is needed to improve the response for the  $n$ th individual at time  $t$

$$A(n, t) = 1 \text{ if assistance needed, and } A(n, t) = 0 \text{ otherwise}$$

The mathematical model for the querying system can be described as follows:

**Objective:**

Minimize the need for further assistance to improve the response:

$$\text{minimize } \sum_n \sum_t A(n, t) \quad (3)$$

**Subject to:**

- 1) Detection of Eye Impaired Individuals:

$$N = \sum_n P(n)$$

- 2) Conversion of Voice Query to Text:

$$T(n, t) = \text{ConvertToText}(Q(n, t))$$

- 3) Generating Response from ChatGPT:

$$R(n, t) = \text{ChatGPT}(T(n, t))$$

- 4) Translation of Response to Voice Message:

$$V(n, t) = \text{TranslateToVoice}(R(n, t))$$

- 5) Delivery of Voice Message to Requesting Individual:

$$S(n, t) = V(n, t)$$

- 6) Assessment for Further Assistance:

$$A(n, t) = \text{AssessFurtherAssistance}(R(n, t))$$

The optimization problem in Phase 1 involves iteratively adjusting the parameters  $\theta$  of the Convolutional Neural Network (CNN) model. This adjustment is done in a way that aims to minimize the classification error. In this context,  $\theta$  represents the weights and biases of the CNN model's layers.

**CNN Model:** The CNN model is designed to take input eye samples and classify them as either belonging to a "normal person" or a "blind person" category. It learns to recognize patterns and features within these eye samples that distinguish between these two categories.

**Classification Error:** The classification error represents how far off the model's predictions are from the true labels of the eye samples. The goal of optimization is to reduce this error, which means improving the accuracy of the model's classification.

**Iterative Optimization:** Optimization is done iteratively, where the model's parameters  $\theta$  are adjusted in small steps to minimize the error. Techniques like gradient descent are commonly used for this purpose. The model goes through multiple iterations until it reaches a point where the error is minimized to an acceptable level.

**Second Phase Enhancement:**

In the Second Phase, the system is enhanced to reduce or eliminate the need for human assistance when interacting with blind users. Several modules are introduced for this purpose:

**Voice-to-Text Conversion:** This module is designed to convert spoken words or voice commands from the blind user into text. It utilizes speech recognition technology to accurately transcribe spoken language into written text.

**Response Generation:** Once the user’s input is converted to text, this module generates appropriate responses. These responses can include answers to user queries, providing information, or assisting with tasks. Response generation can involve natural language processing and dialogue management.

**Translation Processes:** This module may be used to translate responses into different languages or formats if needed. It ensures that the blind user can access information and assistance in their preferred language or format.

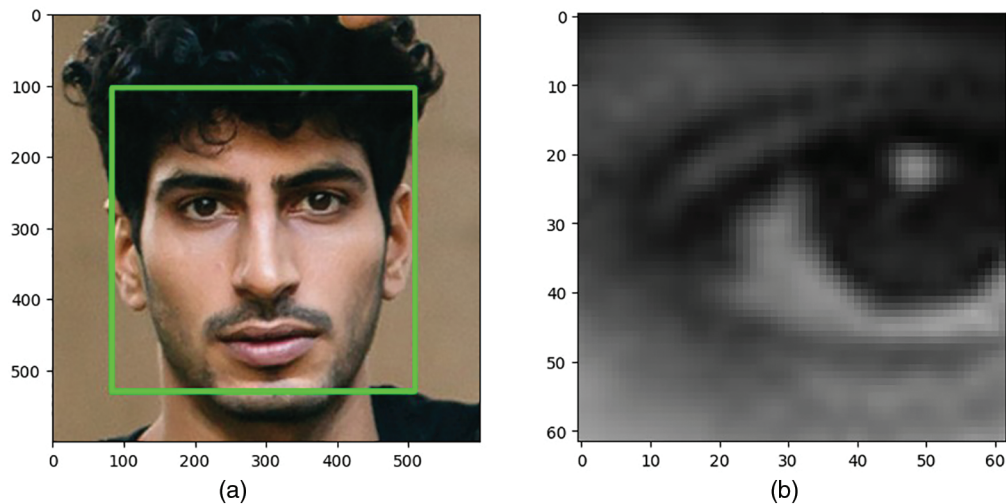
The overall objective of the Second Phase is to make the system more self-sufficient and user-friendly for blind individuals. By providing these modules, the system can understand user input, generate meaningful responses, and potentially assist with a wide range of tasks, reducing the reliance on external assistance.

The optimization problem in Phase 1 ensures that the system’s core capability, the CNN model, is as accurate as possible in distinguishing between normal and blind persons based on their eye samples. This accuracy is critical because it informs the system’s subsequent actions, such as whether to activate the assistance modules in Phase 2.

## 4 Experiments

### 4.1 Phase 1: Impaired Person Detection Using CNN

The system employs a Dlib Trained Model that has been built using the ResNet-32 architecture to accurately locate and identify the face and eyes of the person seated in front of it (as shown in Fig. 4a for face and Fig. 4b for eye). This model leverages deep learning techniques to effectively detect facial landmarks and extract eye regions with high precision.

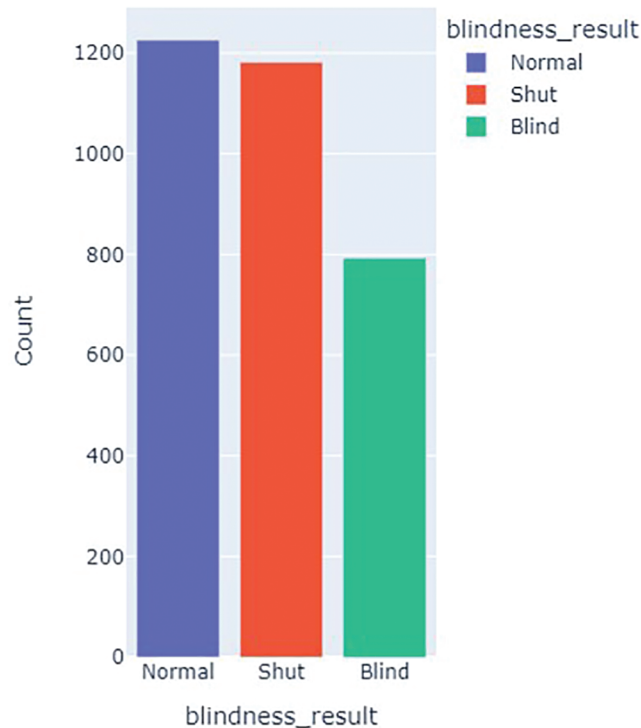


**Figure 4:** Face and eye detection using dlib ResNet-34 pre-trained model. (a) Detected face highlighted. (b) Cropped detected eye for bringing as input for categorization

Subsequently, the extracted eye samples are passed through a Convolutional Neural Network (CNN) model that has been specifically trained to differentiate between normal person eyes and blind



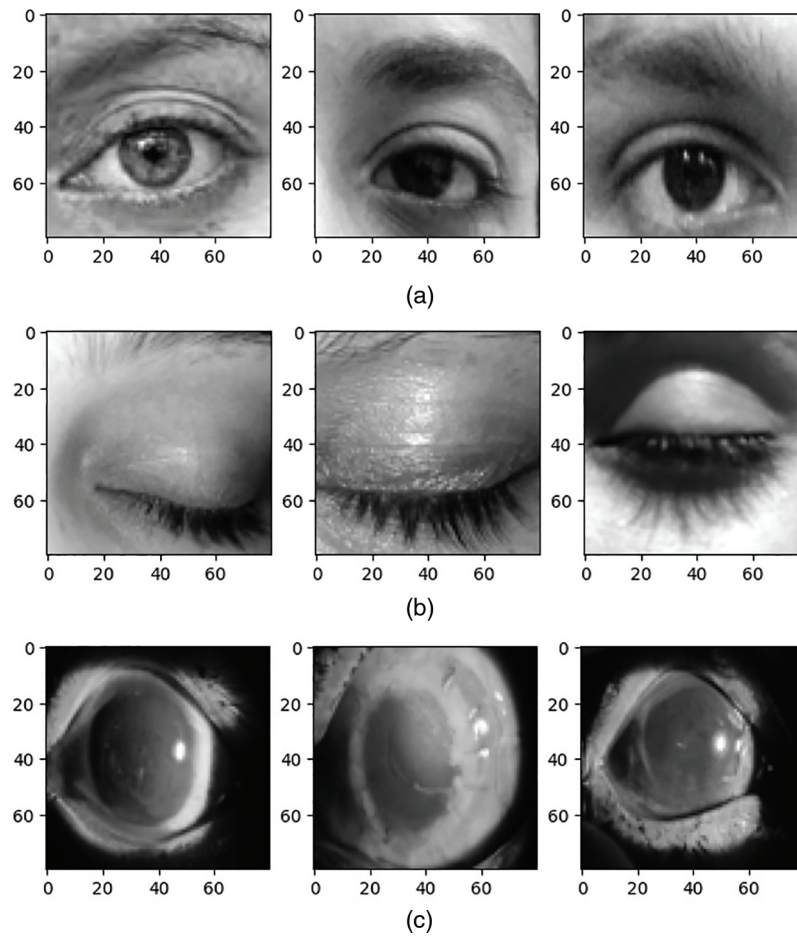
person eyes. The CNN model's training process is underpinned by a diverse dataset containing images of both normal and blind individuals' eyes. The relative strengths of each class within this dataset are visualized in Fig. 5 as a bar chart, providing a clear representation of the dataset's class distribution. This diversity in the training data is instrumental in enabling the CNN model to acquire and internalize distinctive features and patterns specific to each category. By learning from a varied range of eye images, the model becomes adept at recognizing and distinguishing between normal and blind eyes, ultimately contributing to its ability to detect visual impairments effectively.



**Figure 5:** Blind eye dataset bar chat

Once the eye sample has been processed by the CNN model, it generates a classification outcome indicating whether the person in front of the system is blind or not. If the classification indicates that the individual is blind, a subsequent phase, known as Phase 2, is initiated. Sample dataset images can be seen in Fig. 6, where Normal Eye images are shown in Fig. 6a, Shut Eye images in Fig. 6b, and Blind or Cataract Eye sample is shown in Fig. 6c.

Fig. 6 illustrates dataset samples representing three distinct eye categories: Normal, Shut, and Blind Eyes. Fig. 6a offers a glimpse of randomly selected instances from the Normal Eye category, showcasing the characteristics typical of unobstructed eyes. Fig. 6b presents three random samples from the Shut Eye category, providing a diverse representation of eyes in the shut state. Lastly, Fig. 6c reveals random selections from the Blind Eye category, offering insight into the various visual impairment characteristics captured within our dataset. These dataset samples are pivotal in training and evaluating our model's ability to detect visual impairments and deliver tailored responses, thus contributing to a more inclusive and responsive conversational AI system.



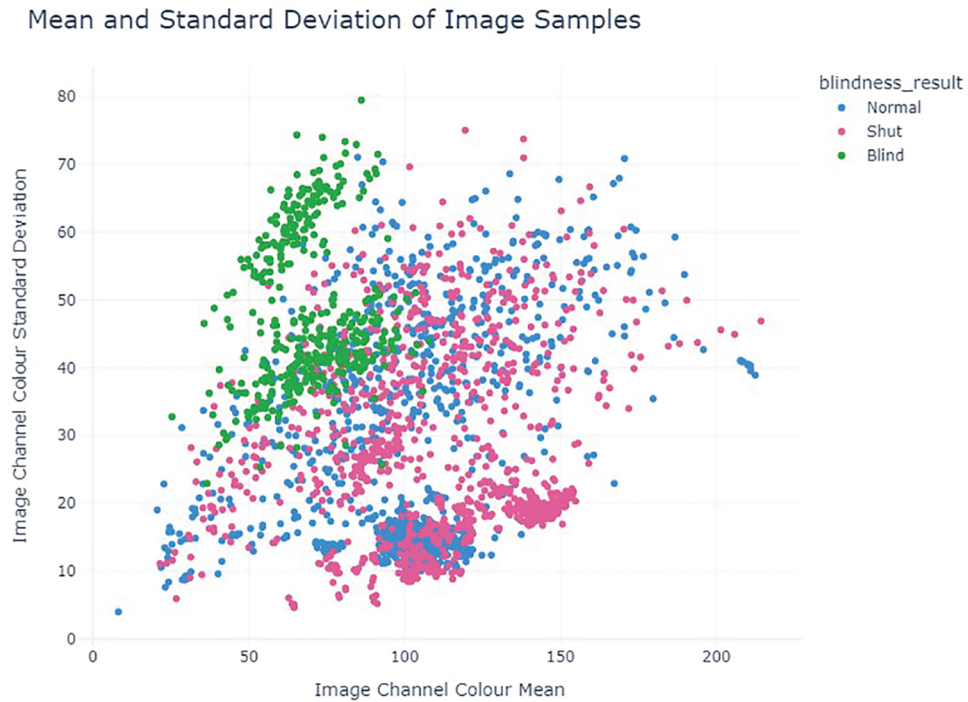
**Figure 6:** Dataset samples for normal, shut, and blind eye categories. (a) Random three normal eyes. (b) Random three shut eyes. (c) Random three blind eyes

Fig. 7 illustrates the distribution of image colour mean and standard deviation for each class in the form of scattered graphs. This figure provides a visual representation of the statistical characteristics of the dataset across different eye categories. By examining the mean and standard deviation of colour attributes, we gain insights into the variations in colour distribution within each class. This analysis is crucial for understanding the dataset's diversity and aids in the development of robust models capable of accommodating a wide range of visual characteristics.

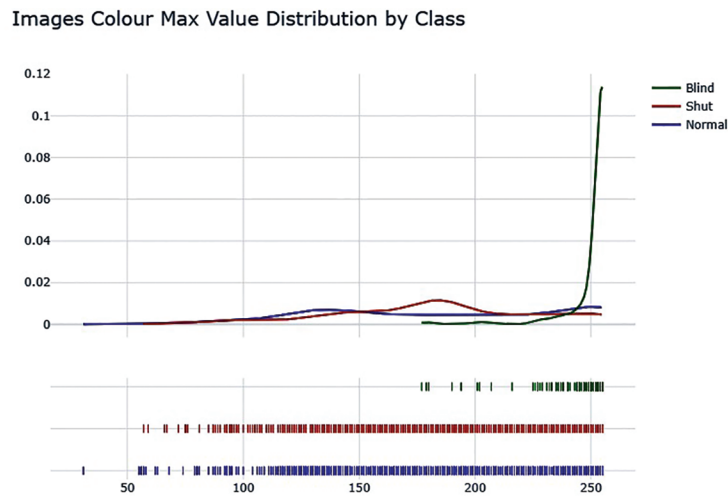
Next to gain a better understanding of class-wise distributed images detailed distribution of colour as Max, Min, and Mean Values is shown in Fig. 8.

During Phase 2, a request for assistance is automatically issued to provide the blind person with the necessary support or information. This phase serves as an opportunity for the blind user to ask specific queries or seek assistance tailored to their needs, ensuring an inclusive and accommodating experience.

By employing a combination of advanced computer vision techniques, deep learning models, and a dedicated phase for assistance, the system enhances accessibility for blind users and facilitates their interaction with the smart system.

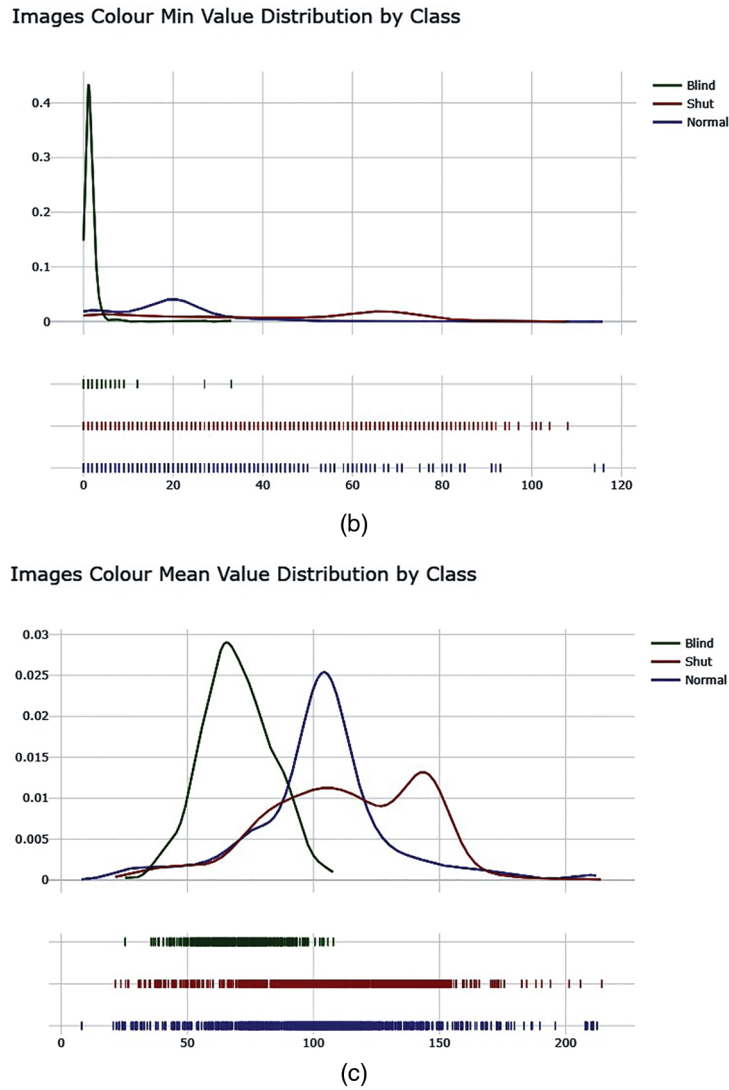


**Figure 7:** Class-wise data distribution of the images



(a)

**Figure 8:** (Continued)



**Figure 8:** Color distribution as max, min, and average among different categories of images. (a) Image color max value distribution by class. (b) Image color min value distribution by class. (c) Image color mean value distribution by class

#### 4.2 Phase 2: Voice Query & ChatGPT

Blind individuals often experience frustration when using voice-activated personal assistant (VAPA) software, such as Siri or Google Assistant, as it can be challenging to determine whether the microphone is active or inactive. This lack of clarity regarding the microphone status leads to difficulties and frustrations during interactions with these VAPA applications [28].

Once individuals with visual impairments are detected by the proposed model, it will proactively inquire if the system can provide any assistance. Upon receiving a voice query, the system utilizes the Google API to convert the message into text format. This text is then forwarded to ChatGPT for

generating concise answers. The response from ChatGPT is translated back into a voice message using another Google API and delivered to the person who made the initial request.

When blind person interacts with a smart system, their queries often revolve around entertainment, obtaining information, or experiencing visual content through descriptive explanations. Since they rely on auditory or tactile senses rather than vision, they may request assistance in exploring various forms of media, accessing news and information, or understanding visual concepts through detailed descriptions.

For entertainment, blind individuals may ask for recommendations on accessible movies, TV shows, music, audiobooks, podcasts, or games that can be enjoyed through audio or tactile interfaces. They might also inquire about live performances, theatre shows, or other events that provide audio descriptions or alternative accessibility options.

When seeking information, blind users may ask about current news, weather updates, sports scores, stock market data, or general knowledge topics. They might request summaries of articles or books, explanations of visual content like infographics or charts, or details about visual media such as movie posters or artwork.

Additionally, blind individuals might ask for descriptions of photographs, videos, or visual scenes encountered in their daily lives. They could request assistance in visualizing concepts, architecture, landscapes, or any other visual elements they are curious about.

In all these cases, the ChatGPT offer detailed verbal descriptions, summarizes information, offers recommendations, and ensures an inclusive and accessible experience for blind users.

## 5 Result and Discussion

A CNN model is used to accurately detect and locate eyes and faces within images or video frames. The of detection is crucial for further analysis and classification of the detected eyes. The subsequent eye classification module utilizes another CNN model to classify the eyes as Blindness, Normal or Closed based on their features. This classification enables the system to identify individuals with visual impairments and tailor the assistance accordingly.

The voice-based assistance provided to individuals classified as blind by the system is a crucial component of the research. By converting voice input into text, the system can effectively interact with the AI language model, ChatGPT, to generate responses and assist blind users with their queries and information needs. The text-to-voice conversion allows for seamless communication, transforming the generated responses back into voice format for blind users to hear.

CNN Model after the training and verification among 20 Epochs shows results with an accuracy of 95% over the dataset of 3200 images distributed in classes of 792 Blind Eye, 1182 Shut Eye, and 1226 Normal Eye image collection. Data is distributed as 2304 images for training, 640 images for testing, and 256 images for validation. A detailed report regarding the model is shown with their respective precision, recall, and F1-score with respect to training as in [Fig. 9a](#), validation as in [Fig. 9b](#), and testing as in [Fig. 9c](#).

Whereas, Graphs representing Accuracy and Loss are represented in [Fig. 10](#).

Based on the table represented in [Fig. 9c](#), we can observe the following improvements in the detection of blind or impaired individuals when the trained model is tested:

- Precision: The precision calculated for the class “Blind” is 0.99, representing that out of all 133 samples predicted as blind, approximately 99% of them were correctly predicted during the test.

This emphasis on the model’s ability to accurately identify blind individuals by reducing false positive predictions.

- Recall: The recall calculation for the class “Blind” is 1.00, representing the model’s ability to correctly identify approximately 100% of the actual blind individuals. Although the value represented is significantly higher compared to the other classes.
- F1-score: Similar to the recall measure the F1-score for the “Blind” class is 1.00, which represents the harmonic mean of precision and recall.

Classification Report for Train Data

	precision	recall	f1-score	support
0	0.96	0.95	0.95	864
1	0.95	0.96	0.95	855
2	1.00	1.00	1.00	585
accuracy			0.97	2304
macro avg	0.97	0.97	0.97	2304
weighted avg	0.97	0.97	0.97	2304

(a)

Classification Report for Validation Data

	precision	recall	f1-score	support
0	0.91	0.95	0.93	99
1	0.94	0.89	0.91	83
2	1.00	1.00	1.00	74
accuracy			0.95	256
macro avg	0.95	0.95	0.95	256
weighted avg	0.95	0.95	0.95	256

(b)

Classification Report for Test Data

	precision	recall	f1-score	support
0	0.95	0.95	0.95	263
1	0.95	0.94	0.95	244
2	0.99	1.00	1.00	133
accuracy			0.96	640
macro avg	0.96	0.96	0.96	640
weighted avg	0.96	0.96	0.96	640

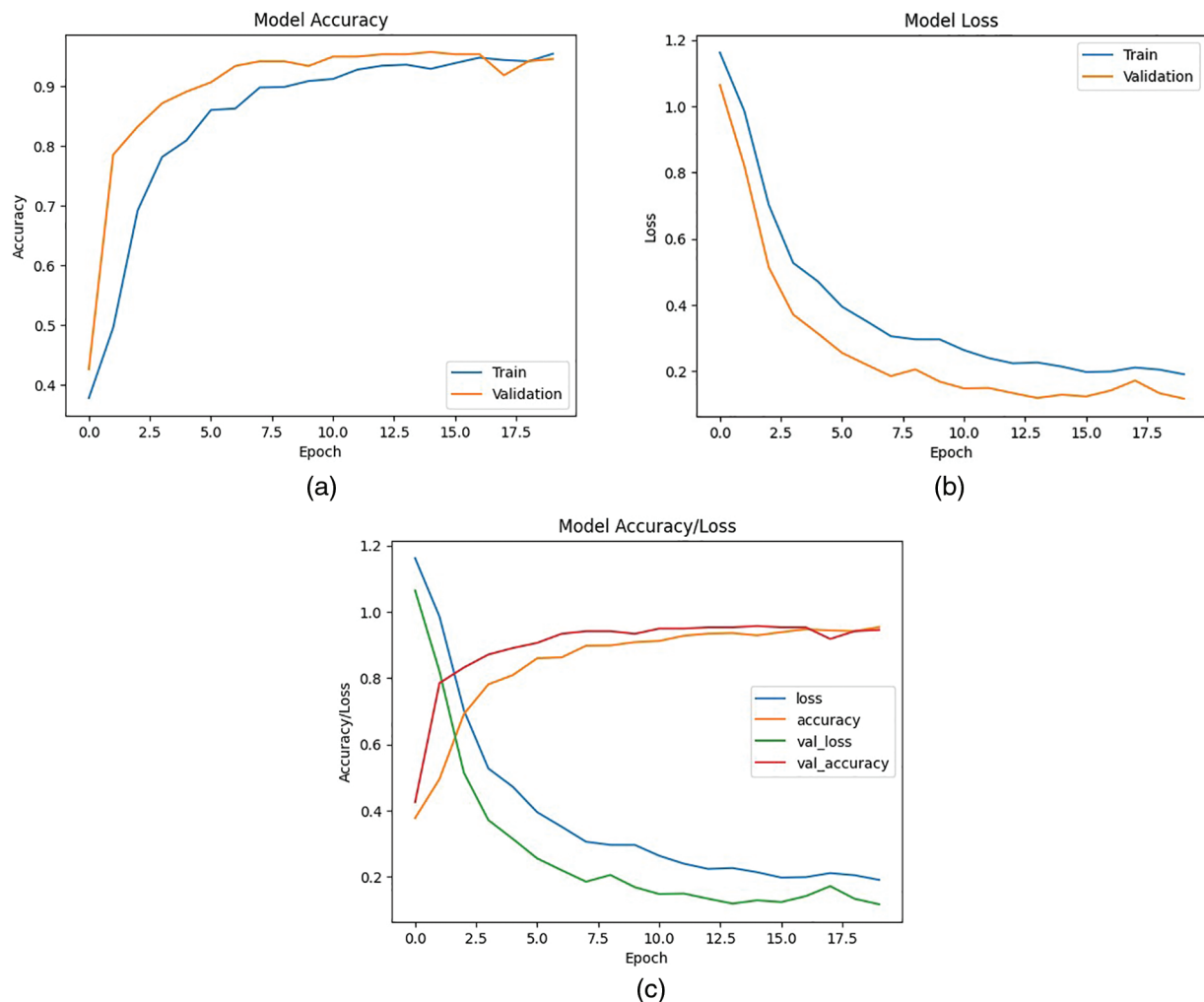
(c)

**Figure 9:** CNN model during training and validation classification reports. (a) Train classification report. (b) Validation classification report. (c) Test classification report

The above findings show that the model has a higher precision score indicating fewer false positive predictions in identifying blind or impaired individuals. This information is valuable in evaluating the effectiveness of the model for blind or impaired detection capabilities.

During multiple test cases, three individuals asked about 60 queries which were further categorized as either fact or generation enquiry by a linguistic expert. These queries were used to check vocal querying ChatGPT which are further translated into text as follows:

**Query:** “What is the time and date today?”



**Figure 10:** Model accuracy, loss, and accumulated with 20 epochs of training and validation. (a) Accuracy measurement graph. (b) Loss measurement graph. (c) Accumulated combined measurement graph

**ChatGPT Responses:** “As an AI language model, I don’t have real-time data access. However, at the beginning of our conversation, I mentioned that the current date is June 20, 2023. As for the time, I apologize, but I can’t provide you with the current time. Please refer to your device or other reliable sources for the accurate time.”

**Query:** “Can you tell me something funny?”

**ChatGPT Responses:** “Certainly! Here’s a lighthearted joke for you:

Why don’t scientists trust atoms?

Because they make up everything!

**Query:** “Where am I now?”



**ChatGPT Responses:** “As an AI language model, I don’t have access to personal information about individuals unless it has been shared with me during our conversation. I am designed to respect user privacy and confidentiality. Therefore, I don’t know where you are currently located. My primary function is to provide information and assist with answering questions to the best of my knowledge and abilities.”

**Query:** “Visualize current weather for me”

**ChatGPT Responses:** “I apologize, but as an AI text-based model, I can’t visualize or display images directly. However, I can provide you with textual information about the current weather in a specific location if you provide me with the name of the city or region you’re interested in.”

Fig. 11 shows Queries and Responses to Word Clouds, Word Frequencies, and Bigram Plots. Whereas, (a–c) show Queries Categorized as Facts Word Cloud, Word Frequency, and Bigram Plot, (d–f) show Responses Categorized as Facts Word Cloud, Word Frequency, and Bigram Plot, (g–i) show Queries Categorized as Generate Word Cloud, Word Frequency, and Bigram Plot, and (j–l) show Responses Categorized as Generate Word Cloud, Word Frequency, and Bigram Plot.

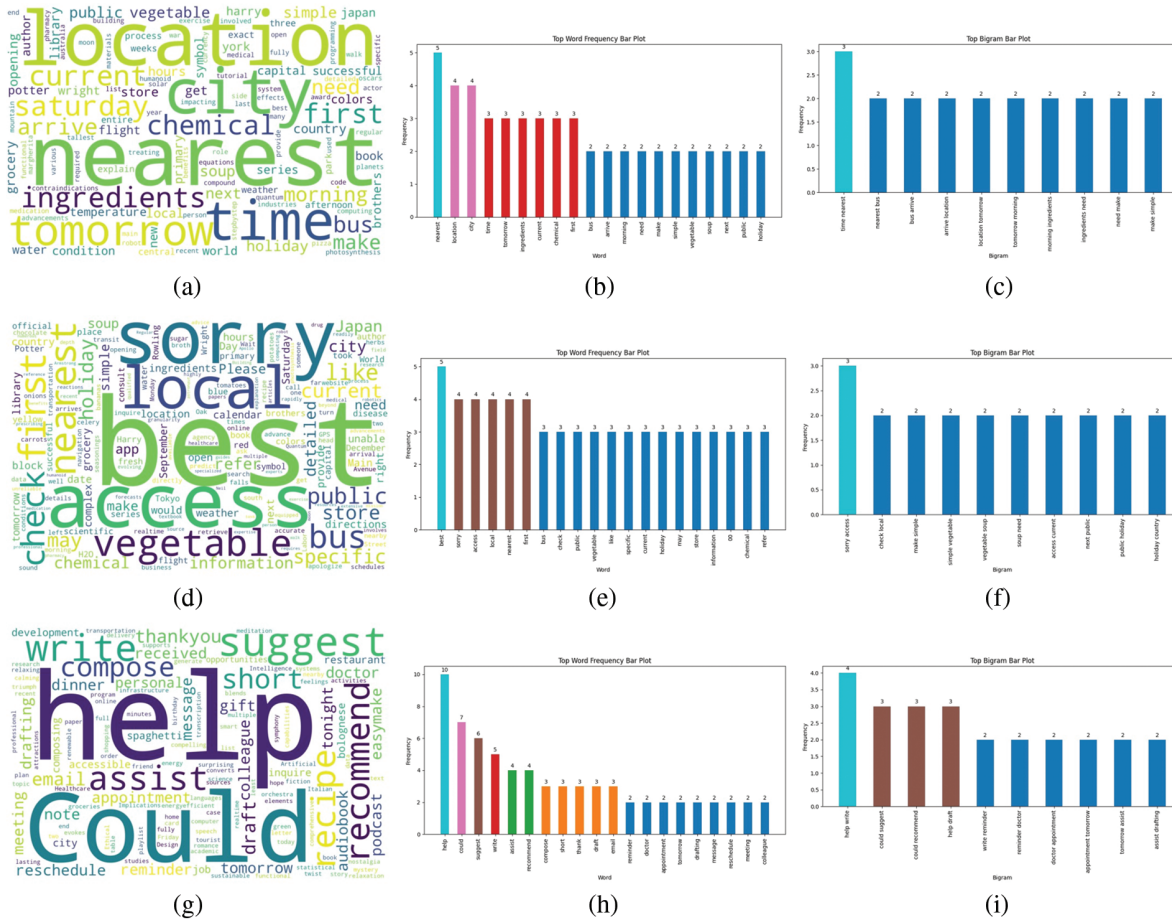


Figure 11: (Continued)



**Table 1 (continued)**

Analysis points	Summary
Challenges with personalized responses	ChatGPT faced limitations in providing personalized responses involving specific locations or personalized schedules due to the absence of real-time data and personalized context.
Limitations in medical and legal advice	The model correctly refrained from providing medical and legal advice and instead directed users to consult professionals in those fields.
Success in simple daily life queries	ChatGPT performed reasonably well in answering general daily life queries, such as assistance with writing messages and providing recipe suggestions.

## 6 Conclusion

The developed system for eye and face detection using CNN models, followed by eye classification and voice-based assistance, has shown promising potential in enhancing accessibility for individuals with visual impairments. The modular approach implemented in this research allows for a seamless flow of information and assistance between the different components of the system. This research significantly contributes to the field of accessibility technology by integrating computer vision, natural language processing, and voice technologies. By leveraging these advancements, the developed system offers a practical and efficient solution for assisting blind individuals. The modular design ensures flexibility, scalability, and ease of integration with existing assistive technologies.

However, it is important to acknowledge that further research and improvements are necessary to enhance the system's accuracy and usability. Fine-tuning the CNN models and expanding the training dataset can improve eye and face detection as well as eye classification capabilities. Additionally, incorporating real-time responses through sophisticated natural language understanding techniques and expanding the knowledge base of ChatGPT can enhance the system's ability to provide comprehensive and accurate responses. Overall, this research paves the way for the development of more advanced and robust systems for assisting visually impaired individuals. By leveraging cutting-edge technologies and integrating them into a modular framework, this research contributes to creating a more inclusive and accessible society for individuals with visual impairments. Future work can focus on refining the system, addressing its limitations, and conducting user studies to evaluate its effectiveness and impact in real-world scenarios.

**Acknowledgement:** This work was supported by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU).

**Funding Statement:** This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (Grant Number: IMSIU-RP23008).

**Author Contributions:** Conceptualization, A. Ahmad and S. Jabbar; methodology, A. Ahmad and S. Jabbar; software, U. Raza and S. Akram; validation, N. M. Alshuqayran and A. Ahmad; formal analysis, U. Raza and S. Akram; investigation, S. Akram; resources, A. Paul; data curation, N. M.

Alshuqayran; writing—original draft preparation, A. Ahmad, A. Paul and S. Jabbar; writing—review and editing, U. Raza; visualization, S. Akram; supervision, A. Ahmad and S. Jabbar; project administration, N. M. Alshuqayran; funding acquisition, A. Ahmad.

**Availability of Data and Materials:** The raw data supporting the conclusions of this article will be made available by the authors.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] D. Pascolini and S. P. Mariotti, “Global estimates of visual impairment: 2010,” *British Journal of Ophthalmology*, vol. 96, no. 5, pp. 614–618, 2012.
- [2] F. Ullah, J. Moon, H. Naeem and S. Jabbar, “Explainable artificial intelligence approach in combating real-time surveillance of COVID19 pandemic from CT scan and X-ray images using ensemble model,” *The Journal of Supercomputing*, vol. 78, no. 17, pp. 19246–19271, 2022.
- [3] M. Kashif, K. R. Malik, S. Jabbar and J. Chaudhry, “Application of machine learning and image processing for detection of breast cancer,” in *Innovation in Health Informatics*. Academic Press, Elsevier, pp. 145–162, 2020.
- [4] A. Haleem, M. Javaid and R. P. Singh, “An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges,” *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, vol. 2, no. 4, pp. 100089, 2022.
- [5] A. Rapp, L. Curti and A. Boldi, “The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots,” *International Journal of Human-Computer Studies*, vol. 151, pp. 102630, 2021.
- [6] K. R. Malik, R. R. Mir, M. Farhan, T. Rafiq and M. Aslam, “Student query trend assessment with semantical annotation and artificial intelligent multi-agents,” *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 7, pp. 3893–3917, 2017.
- [7] A. Ismail, N. S. Ghorashi and R. Javan, “New horizons: The potential role of Openai’s ChatGPT in clinical radiology,” *Journal of the American College of Radiology*, 2023.
- [8] S. Liu, A. P. Wright, B. L. Patterson, J. P. Wanderer, R. W. Turer *et al.*, “Using AI-generated suggestions from ChatGPT to optimize clinical decision support,” *Journal of the American Medical Informatics Association*, vol. 30, no. 7, pp. 1237–1245, 2023.
- [9] M. Sallam, “ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns,” *Healthcare*, vol. 11, no. 6, pp. 887, 2023.
- [10] L. Zhang, Z. Sun, J. Zhang, Y. Wu and Y. Xia, “Conversation-based adaptive relational translation method for next POI recommendation with uncertain check-ins,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 10, pp. 7810–7823, 2023.
- [11] H. Naeem, F. Ullah, M. R. Naeem, S. Khalid, D. Vasan *et al.*, “Malware detection in Industrial Internet of Things based on hybrid image visualization and deep learning model,” *Ad Hoc Networks*, vol. 105, pp. 102154, 2020.
- [12] F. Ullah, J. Wang, M. Farhan, S. Jabbar, Z. Wu *et al.*, “Plagiarism detection in students’ programming assignments based on semantics: Multimedia E-learning based smart assessment methodology,” *Multimedia Tools and Applications*, vol. 79, pp. 8581–8598, 2020.
- [13] P. P. Ray, “ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope,” *Internet of Things and Cyber-Physical Systems*, 2023.
- [14] T. H. Kung, M. Cheatham, A. Medenilla, C. Sillos, L. de Leon *et al.*, “Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models,” *PLoS Digital Health*, vol. 2, no. 2, pp. e0000198, 2023.

- [15] M. Faruqui and D. Hakkani-Tur, "Revisiting the boundary between ASR and NLU in the age of conversational dialog systems," *Computational Linguistics*, vol. 48, no. 1, pp. 221–232, 2022.
- [16] A. Carolus, Y. Augustin, A. Markus and C. Wienrich, "Digital interaction literacy model-conceptualizing competencies for literate interactions with voice-based AI systems," *Computers and Education: Artificial Intelligence*, vol. 4, pp. 100114, 2023.
- [17] X. Qu, H. Liu, Z. Sun, X. Yin, Y. S. Ong *et al.*, "Towards building voice-based conversational recommender systems: Datasets, potential solutions, and prospects," arXiv preprint arXiv:2306.08219, 2023.
- [18] A. S. George and A. H. George, "A review of chatGPT AI's impact on several business sectors," *Partners Universal International Innovation Journal*, vol. 1, no. 1, pp. 9–23, 2023.
- [19] V. Basmov, Y. Goldberg and R. Tsarfaty, "ChatGPT and simple linguistic inferences: Blind spots and blinds," arXiv preprint arXiv:2305.14785, 2023.
- [20] A. Kuzdeuov, S. Nurgaliyev and H. A. Varol, "ChatGPT for visually impaired and blind," 2023. [Online]. Available: [https://www.techrxiv.org/articles/preprint/ChatGPT\\_for\\_Visually\\_Impaired\\_and\\_Blind/22047080](https://www.techrxiv.org/articles/preprint/ChatGPT_for_Visually_Impaired_and_Blind/22047080) (accessed on 03/05/2023).
- [21] J. Witte Zimmerman, D. Hudon, K. Cramer, J. St Onge, M. Fudolig *et al.*, "A blind spot for large language models: Supradiegetic linguistic information," arXiv preprint arXiv:2306.06794, 2023.
- [22] J. Dahmen, M. E. Kayaalp, M. Ollivier, A. Pareek, M. T. Hirschmann *et al.*, "Artificial intelligence bot ChatGPT in medical research: The potential game changer as a double-edged sword," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 31, no. 4, pp. 1187–1189, 2023.
- [23] R. Sarki, K. Ahmed, H. Wang and Y. Zhang, "Automated detection of mild and multi-class diabetic eye diseases using deep learning," *Health Information Science and Systems*, vol. 8, no. 1, pp. 32, 2020.
- [24] I. S. Gabashvili, "The impact and applications of ChatGPT: A systematic review of literature reviews," arXiv preprint arXiv:2305.18086, 2023.
- [25] S. Sedaghat, "Early applications of ChatGPT in medical practice, education and research," *Clinical Medicine*, vol. 23, no. 3, pp. 278–279, 2023.
- [26] A. Shafeeg, I. Shazhaev, D. Mihaylov, A. Tularov and I. Shazhaev, "Voice assistant integrated with ChatGPT," *Indonesian Journal of Computer Science*, vol. 12, no. 1, 2023. <https://doi.org/10.33022/ijcs.v12i1.3146>
- [27] M. K. Moussavou Boussougou and D. J. Park, "Attention-based 1D CNN-BILSTM hybrid model enhanced with fasttext word embedding for Korean voice phishing detection," *Mathematics*, vol. 11, no. 14, pp. 3217, 2023.
- [28] A. Abdolrahmani, R. Kuber and S. M. Branham, "SIRI talks at you—An empirical investigation of voice-activated personal assistant (VAPA) usage by individuals who are blind," in *Proc. of the 20th Int. ACM SIGACCESS Conf. on Computers and Accessibility*, pp. 249–258, 2018. <https://doi.org/10.1145/3234695.3236344>