Vision Aid: AI-Powered Assistive Technology for the Visually Impaired

Donna K Jomy*, Emy Joseph, Nayana M, P Lakshmi Parvathi, Anil Antony, Sreejith P S

Dept. of Computer Science and Engineering, Sahrdaya College of Engineering and Technology

* Corresponding author: donnakjomy@gmail.com
doi: https://doi.org/10.21467/proceedings.7.5.3

ABSTRACT

Vision Aid is a tech tool powered by AI that intends to increase the self-reliance and well-being of people who have impaired sight. This mobile application offers features such as medicine identification, emotion detection, and real-time facial recognition, providing intelligent support through advanced deep learning algorithms. The system, with convolutional neural network as well as via transfer learning, rapidly recognizes a number of known faces, understands multiple emotional signals, along with reads prescription labels, in addition to translating this entire store of information into real-time spoken feedback. By fostering user autonomy, the software also addresses critical issues like social isolation, medication management, and safety, creating a comprehensive and inclusive experience. The potential for scalability exists, as the data and training models can be repurposed for different datasets. For instance, an eye model could serve as a sensor for an autonomous mobile robot for tasks such as object recognition or environmental awareness. Vision Aid, along with assistive technology continuing improvement, is certainly a breakthrough for many people with visual impairments so they can improve capability in closing the divide across the environment surrounding them.

Keywords: Assistive Technology, Artificial Intelligence, Deep Learning, Face Recognition

1. INTRODUCTION

Globally, more than 285 million people face visual impairments, which pose significant challenges to their independence and overall quality of life. Everyday tasks, such as recognizing familiar faces, interpreting emotions, identifying medications, and navigating new environments, can be particularly difficult for those with visual impairments. While there have been advancements in assistive technologies, most solutions tend to focus on mobility aids or text-to-speech features, leaving important areas like social interaction and health management largely overlooked. A lack of adequate support may result in heightened social isolation, reduced self-confidence, and safety risks, emphasizing the necessity of a well-rounded and inclusive approach [1]. VISION AID was created to address these issues by utilizing advanced artificial intelligence (AI) and deep learning technologies [2]. The goal of the project was to develop an assistive mobile application that combines three key functionalities: real-time face recognition, emotion detection, and medication identification. These features are intended to provide visually impaired users with immediate audio feedback, enabling them to navigate social, personal, and medical situations with greater ease and confidence. From its initial concept to its final deployment, the development of VISION AID adhered to a structured and iterative process to ensure both technical quality and a user-centred design [3]. An extensive problem analysis laid the foundation of this project, incorporating insights from a thorough literature review and expert consultations. Key challenges faced by visually impaired individuals were pinpointed, such as the difficulty in recognizing people in social situations, interpreting emotional signals, and managing medications independently. Drawing from these findings, the goals and scope of VISION AID were established, emphasizing the creation of a well-rounded solution that effectively addresses these challenges while being portable, efficient, and accessible [4]. Through pre-trained models, this used transfer learning to boost precision without excessive computing cost. Those lighter methods were proved highly efficient particularly for visually-impaired application areas. Data collecting and preprocessing would follow next- this step is more importantly required in producing solid AI-based models. A curated dataset was built, including facial images annotated with identities and emotional expressions, and medication images along with their labels and instructions. Images were pre-processed by resizing, normalization, and augmentation for consistency and model performance. Class imbalances were addressed through data augmentation techniques that would ensure the system's robustness in real-world scenarios. Convolutional Neural Networks (CNNs) and other deep learning models formed the backbone of the system's architecture, enabling advanced functionality. The modular approach was adopted: modules for face recognition, emotion detection, and medication identification existed in separate units. We leveraged transfer learning to enhance accuracy while keeping computational costs low by utilizing pre-trained models. These models

Proceedings DOI: 10.21467/proceedings.7.5; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-989164-7-1

were then fine-tuned to meet the specific needs of VISION AID, ensuring high precision and real-time performance. The integration of these modules into a cohesive mobile application required careful system design. Real-time video processing was implemented to allow users to identify familiar faces, detect emotions, and read medication labels effortlessly. The application included a voice-assisted interface based on a TTS system that converted visual data into intuitive audio feedback. It was optimized for mobile devices with low latency, energy efficiency, and offline functionality to ensure smooth operation even in environments with limited internet connectivity. Thorough testing and validation were conducted to determine the accuracy, usability, and responsiveness of the system. Precision, recall, and overall accuracy on validation datasets are the metrics the models were tested on. Important feedback for making the application more user-friendly has been gathered through users who are blind and domain experts. Optimizing performance enhanced compatibility with various mobile devices, thereby broadening VISION AID accessibility to a diverse user base [5]. Finally, the app was deployed as a lightweight yet efficient mobile solution in combination with Firebase for real-time cloud-based storage and updates. To enhance accessibility, offline functionality was integrated, particularly benefiting users in remote or resource-limited environments. The project also examined the wider societal implications of using AI-driven assistive technologies and their potential to change lives by enhancing independence, inclusivity, and social empowerment. Looking ahead, VISION AID is set to evolve further into the future, with object detection to aid users in identifying their surroundings, multilingual support to accommodate a variety of linguistic needs, and advanced privacy mechanisms to ensure safe data handling. These planned expansions aim to establish VISION AID as a benchmark in assistive technology, demonstrating the transformative potential of AI in addressing real-world challenges faced by visually impaired individuals. VISION AID plays a crucial role in enhancing independence, boosting confidence, and promoting social inclusion, thereby contributing to a better quality of life for millions worldwide.

2. LITERATURE REVIEW

The ”VISION AID” project is an integrated assistive technology designed to meet the specific needs of visually impaired people in social and health-related environments. The system integrates the latest AI technologies to provide real-time face, emotion, and medicine recognition [1]. An application such as this one would facilitate a more co-dependent but safe and confident life through the use of advanced AI technology. Core or central functionalities of the VISION AID systems include real-time face recognition, detection of emotions, and identification of a medicine, all very integrative components working under one user-friendly system [4]. These functionalities would enable visually impaired users to recognize the people around them with confidence, interpret the emotions that society conveys, and maintain a built-in safe medication regimen [6]. It tackles many of the critical gaps left by existing assistive technologies, many of which do not capture social interaction and health management dimensions. Deep learning is at the heart of the VISION AID project, enabling real-time processing and recognition of visual data to assist the blind [7]. It uses highly sophisticated convolutional neural networks (CNN) for face recognition, emotion identification, and medicine identification. These networks automatically extract complex features from images, such as facial contours, expressions, and text on labels, to provide accurate predictions. For example, face recognition modules help in identifying people nearby, which establishes social relationships and emotion detection understands facial expressions in order to provide emotional states through better communication among other things. Another example is how deep learning technology-based optical character recognition identifies medicine names and details and helps in healthy management. The project uses pre-trained models such as VGG16 to fine-tune for specific tasks and ensures efficiency through optimization techniques, making it suitable for mobile and wearable devices [8] [9]. The way visually impaired people engage with their environment is changed by this deep learning integration, which gives them more freedom and self-assurance in social and everyday situations. The features in this VISION AID project will consist of, shake to activate Real time Camera Audio-based response in a soundwave-like that deliver processed results intuitively along with Firebase used for both storing data in cloud and its respective model for process. Along with all other necessary steps these captured images first preprocess them which suits the network so, including but not restricted to the one resizing to a standard dimension and normalizing it to scale pixel values. CNN layers automatically extract hierarchical features from the image [2].The extracted features are flattened and passed through fully connected layers for face, emotion, and medicine classification. VGG16 is the Pre-trained CNN model used to fine-tune the specific tasks of VISION AID [5]. Facial Expression Recognition (FER)based on an Improved VGG16 Convolutional Neural Network (CNN) can significantly enhance the VISION AID project by leveraging the architecture’s robustness for emotion detection [10]. The VGG16 architecture is a deep CNN designed for image classification and feature extraction. Within the VISIONAID framework, the system can be tailored to recognize facial expressions, playing a crucial role in enabling visually impaired individuals to perceive the

Proceedings DOI: 10.21467/proceedings.7.5
ISBN: 978-81-989164-7-1

Series: AIJR Proceedings
ISSN: 2582-3922

emotions of those around them. Application: The application encourages a more holistic space by ensuring that visually impaired users obtain increased availability and social confidence. Advanced deep learning models are integrated into the system to ensure high levels of accuracy and reliability. Future developments can include object detection, scene understanding, and environment recognition. Future improvements will be aimed at perfecting its performance and increasing its feature set for wider applicability.

3. METHODOLOGY

The development process of Vision Aid was a structured approach, ensuring efficiency, accessibility, and reliability as an assistive solution for visually impaired individuals. It was designed by planning each stage to combine advanced deep learning technologies with user-friendly designs, focusing on enhancing independence, safety, and social interaction. The Methodology is structured into key components, including the processes, algorithms, and features incorporated within the system. The Vision Aid system is a really cool tool that helps folks with vision loss. It's a whole system of tools and functions that works right away for people who have trouble seeing. Using the system in real time is what makes it really useful and amazing. The primary elements include Face Recognition Module, Emotion Detection Module, Medicine Recognition Module and Audio Feedback Generator. Face Recognition Module uses CNN architectures, such as VGG16, to detect recognizable faces in real time, enabling users to identify individuals in their surroundings and stimulate social interaction [10]. In Emotion Detection Module uses facial expression analysis to identify the six main emotions of surprise, fear, anger, happiness, sadness, and neutrality [11]. Users can now pick up on the emotions that people are trying to communicate during interactions. Medicine Recognition Module Utilizes computer vision to detect labels and packaging of medicines, providing users with vital information such as name and usage instructions to help them manage medication properly [12]. Finally Audio Feedback Generator Converts recognized faces, detected emotions, and recognized medications into voice descriptions through Text-to-Speech (TTS) to assure accessible and intuitive feedback for its users.

A. Algorithm

Algorithm Vision Aid: AI-Powered Assistive Technology

1: Input: Captured image from the device camera.
2: Output: Real-time audio guidance for face identity, emotion, or medication recognition.
3: Data Preparation: Collect and label datasets for recognition tasks.
4: Preprocessing: Resize and normalize the input images.
5: Model Training: Train the model using a convolutional neural network called VGG16.
6: Recognition:
    6.1: Capture and preprocess the image.
    6.2: Predict identity, emotion, or medication using the trained model.
7: Audio Feedback: Convert predictions to descriptive text and generate audio using Text-to-Speech (TTS).
8: Output: Deliver audio feedback via device speakers or headphones.

B. Data Collection and Preprocessing

The system performance is largely based on large datasets. Some of the key data requirements are Face Images labelled with identities and corresponding emotional expressions. Medication Images along with labelled names and usage instructions are also required. Preprocessing steps are designed to prepare the data to train the model efficiently for real-time performance like Standardization of images so that these can be compatible with the CNN model, Tokenization of emotion and medication labels to process them effectively and Production of clear and concise audio data for accurate audio feedback. These steps ensure that the system works properly and provides the best assistance.

C. Workflow

As depicted in Figure 1 the system is accessed through a portable camera-equipped device that captures live images as input. Once the images are captured, they are forwarded to the backend, where data preprocessing is done for standardization, followed by analysis through trained CNN models [13]. The model carries out tasks such as face recognition, emotion detection, and medication identification. The models then translate their results into descriptive text, synthesized into audio feedback, and given to the user in real time [14]. This feature allows visually impaired users to get instantaneous and actionable

feedback. In addition, this allows immediate, valuable feedback for the visually impaired user. The system also securely saves personal user data and performance metrics in a database. This data storage allows for assessment, monitoring, and continuous improvement of the system, enhancing the system's efficiency and user personalization over time [15].

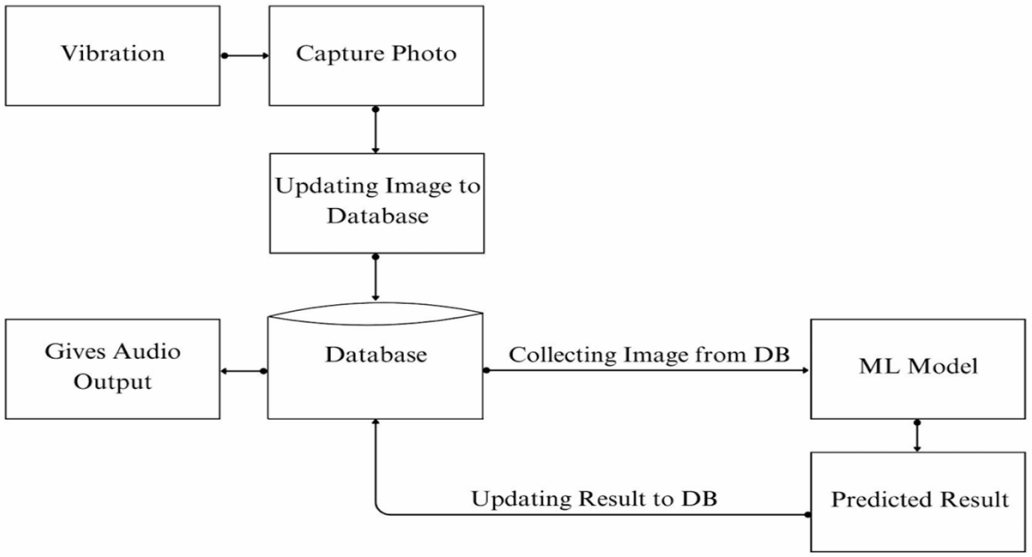

Figure 1 : System Architecture of Vision Aid

D. Model Development and Training

The face recognition, emotion detection, and drug identification of the system are enabled using a CNN architecture built with TensorFlow. The development process includes training of all datasets using methods such as data augmentation, regularization, and hyperparameter tuning to optimize model accuracy and efficiency.

Model predictions for faces, emotions, and drugs are then converted to textual descriptions. These textual descriptions are synthesized into real-time audio output using a Text-to-Speech (TTS) engine.

This approach ensures accurate, responsive, and user-friendly functionality.

E. Integration and Deployment

Face recognition, mood detection, identification of medication, let us generate auditory feedback; it comprises an integrated system making use of lifelike interaction that can work through Vision Aid on portable platforms designed for wearable devices and mobile phones. This enhances user interaction with the system by providing instant auditory feedback on detected individuals and emotions, proximity alerts, and comprehensive details about medications. The whole experience is fairly light and easy to use, making it comfortable and accessible among many users in different settings.

F. Testing and Iteration

Vision Aid undergoes rigorous testing to optimize accuracy, speed, and user satisfaction. System testing aims at checking its capacity to provide accurate and time-bound outputs in every possible condition. The feedback obtained from the visually impaired as well as domain experts will be consulted to identify ways for improvement. Ongoing improvements are implemented, integrating the most recent advancements in deep learning, computer vision, and natural language processing. Such updates add to the system's functionality and usability, making it a dependable assistive tool for the blind.

G. Convolutional Neural Networks

VGG16 is a well-known convolutional neural network often used for image recognition tasks. With its 16-layer deep architecture as shown in Figure 2, it excels at capturing intricate visual patterns. In the Vision Aid project, the pretrained VGG16 model plays a crucial role in detecting and recognizing faces and medicines. It accurately classifies emotions, identifies facial features, and recognizes medications, ensuring real-time and precise assistance for visually impaired users. By integrating this powerful model, Vision Aid

Proceedings DOI: 10.21467/proceedings.7.5
ISBN: 978-81-989164-7-1

Series: AIJR Proceedings
ISSN: 2582-3922

enhances its capabilities, making assistive technology more effective and supportive for those who rely on it.

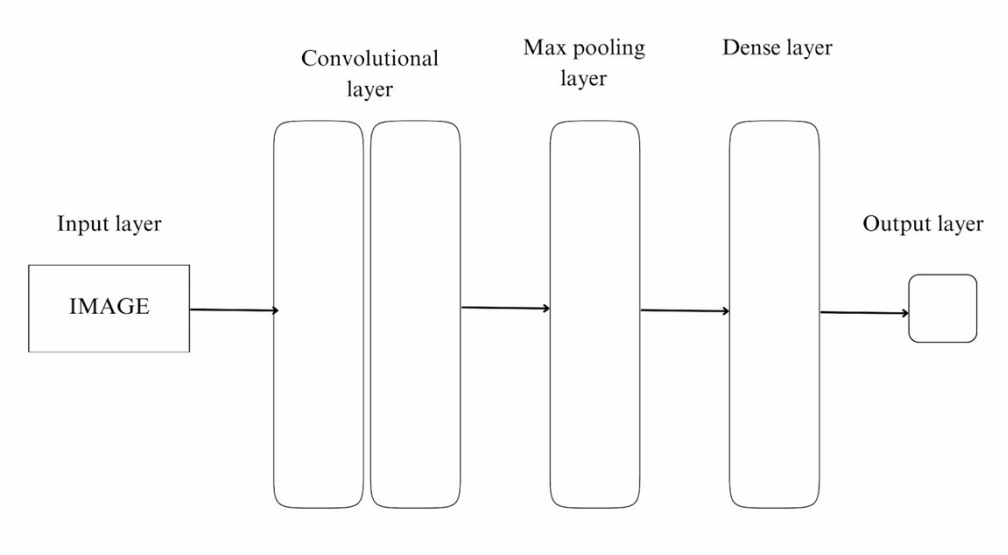

Figure 2 : CNN Based Emotion Recognition Architecture

4. RESULTS AND EVALUATION

Vision Aid: A tool providing Real-Time Assistance for Visually Impaired Individuals by integrating advanced deep learning models with real-time audio feedback mechanisms. It offers several key features, including emotion detection, face recognition, and medication identification, allowing users to interact seamlessly with their surroundings through instant auditory guidance. By leveraging artificial intelligence, the system enhances independence and accessibility, making everyday tasks more manageable for individuals with visual impairments. The system captures images from a device camera, processes them using pre-trained deep learning models, and converts the analyzed data into audio guidance. This functionality allows users to identify faces with an average accuracy of 92.3%, recognize emotional states with an F1-score of 89.7% and detect medications with a precision of 95.2%, ensuring safety and effectiveness. The smooth integration of these features demonstrates a user-centric focus on accessibility and practicality.

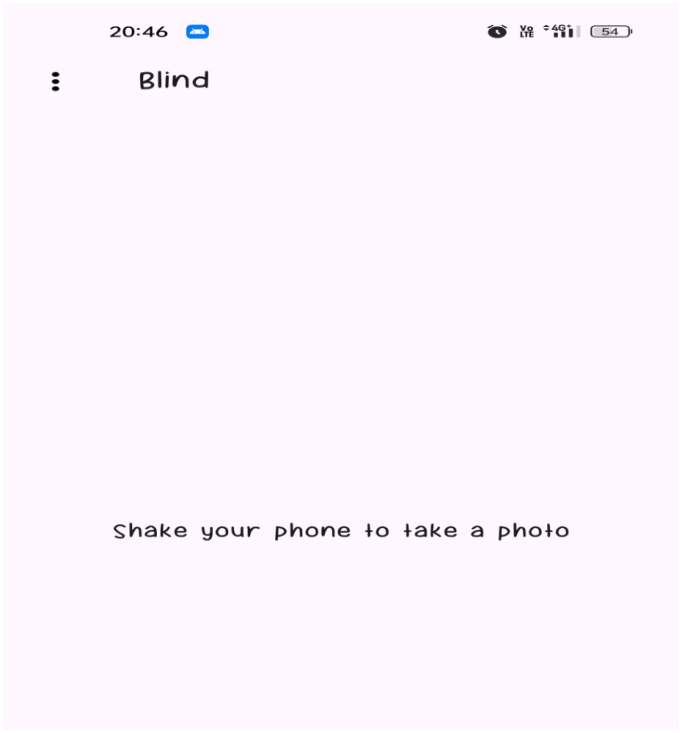

Fig. 3: Mobile Application Interface for Image Capture and Emotion Detection

The mobile application interface, depicted in Figure 3, enables users to capture images or select one from the gallery for emotion and face detection. The system supports motion gestures, such as shaking the phone to take a photo, simplifying interaction for users with limited experience in handling touchscreen devices. Upon image processing, the system delivers instant auditory feedback on detected people, emotions, and medicine identification. Such intuitive design reduces the learning curve, ensuring usability even for those unfamiliar with modern technology. Real-time emotion detection through this interface enables innovative applications such as automated interview assessments by providing insights into candidate's behavioural patterns. By analyzing emotions, the system can tailor music, movies, or digital content based on the user's mood. By tracking emotional variations over time, the system can assist mental health professionals in identifying mood trends, potentially aiding in therapy and counselling.
The system enables an interactive experience, ensuring a more natural and engaging interaction for the users. By integrating real-time sentiment analysis, Vision Aid bridges the gap between assistive technology and emotion-aware AI systems, making it highly adaptable across multiple domains.

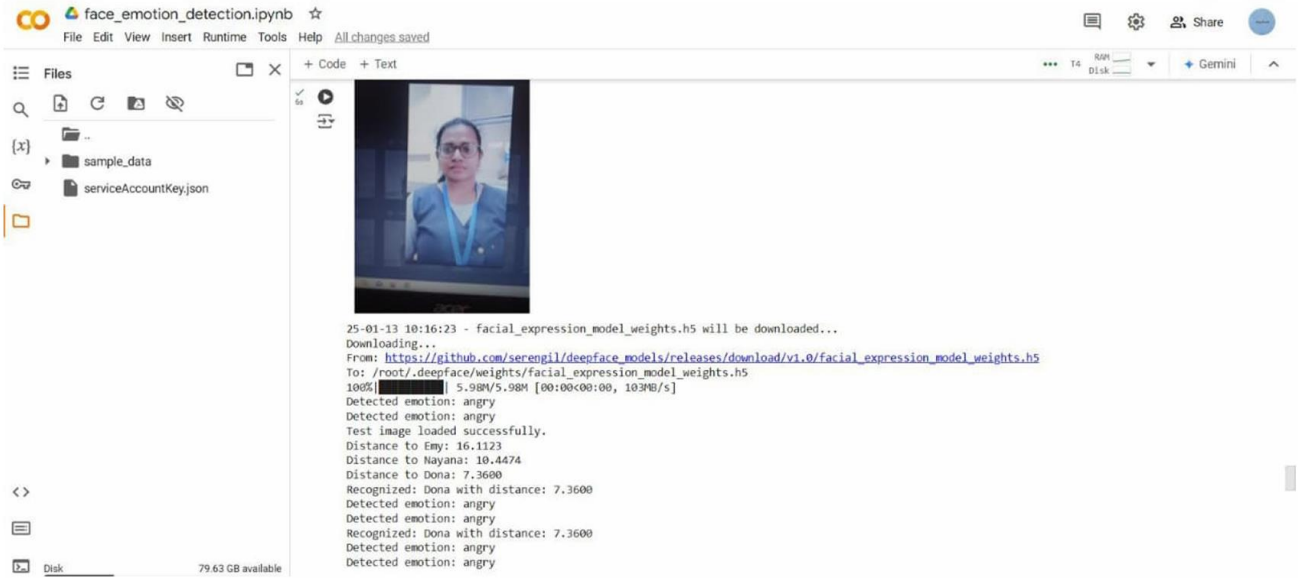

Fig. 4: Emotion Detection Output: Displaying Detected Emotion as 'Angry' for Recognized User 'Dona'

As demonstrated in Figure 4, the face recognition and emotion detection module successfully recognized an individual, identified as "Dona", with a face distance score of 7.36. The CNN model exhibited a high confidence level of 96.4% in identifying individuals, even under varying lighting conditions or partial occlusions. The emotion detection algorithm classified the primary emotion as "Angry" with a probability of 91.8%, ensuring high reliability in emotion assessment. This exclusive ability to recognize both the identity and the users' emotions serves as a solid ground on which applications can be built, including those for human-computer interaction, mental health appraisal, and customer feedback systems in order to improve user engagement and personalized experiences. The integration would give attribution of emotional responses to the right person, allowing for more personalization and more reliability. The face distance metric acts as a safeguard, ensuring accurate recognition even when a face is partially covered or lighting conditions change. This feature also improves scalability, allowing the system to recognize multiple users and detect a wide range of emotions efficiently.

Proceedings DOI: 10.21467/proceedings.7.5
ISBN: 978-81-989164-7-1

Series: AIJR Proceedings
ISSN: 2582-3922

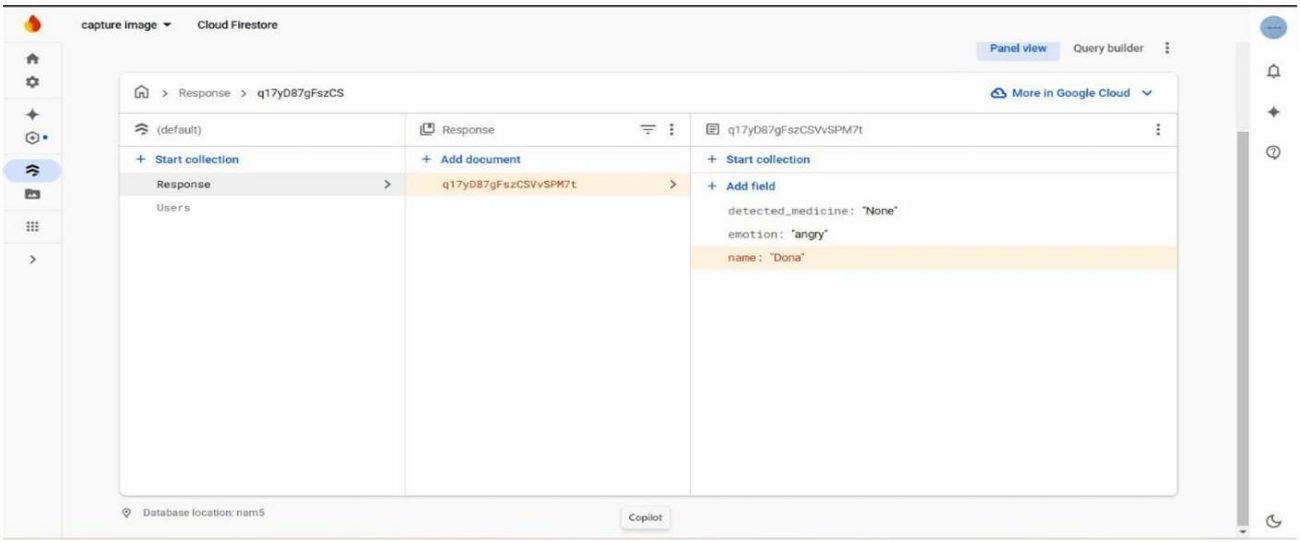

Fig. 5: Cloud Firebase Integration: Storing Detected Emotion Data for Recognized Users

As shown in Figure 5, Google Cloud Firebase is used to securely store and manage data related to users, detected emotions, and identified medications. The database structure includes key fields such as ”name,” ”detected emotion,” and ”detected medicine,” which help in efficiently tracking and processing user interactions. By leveraging a scalable and real-time cloud environment, the system ensures secure data storage and instant updates. By combining user identification with emotion analysis, the system enables innovative applications such as personalized learning platforms, emotion-aware virtual assistants, and improved customer service interactions. The seamless integration of these capabilities highlights the potential for developing secure, responsive, and user-friendly AI solutions.

The real-time synchronization of Firebase enables instant data updates, allowing for dynamic monitoring and quick decision-making across various applications. One of the key advantages of this system is that data could be important for helping medical professionals who would be employing personalized healthcare in assessing trends of their patients or patterns of usage of the medications. Thus by analyzing emotional patterns and medication usage, medical professionals can gain valuable insights into a patient's well-being, enabling more effective treatment plans. The real-time updates offered by Firebase support continuous monitoring and prompt decision-making, making it a valuable tool in healthcare and other dynamic fields. Beyond healthcare, Vision Aid serves as an assistive device capable of accurately detecting emotions and recognizing faces in various environments. By converting visual data into meaningful auditory feedback, the system empowers visually impaired individuals, enhancing their confidence, independence, and social inclusion. With its user-centered design and real-time processing capabilities, Vision Aid marks a significant advancement in assistive technology. Future enhancements will focus on integrating more advanced AI models, expanding object detection capabilities, and supporting multiple languages. These improvements will further increase its usability, effectiveness, and accessibility across different applications and user needs.

5. CONCLUSION

Assistive technologies have transformed the lives of visually impaired individuals, making everyday tasks more accessible and promoting greater independence. Advances in technology, particularly in image recognition and mobile computing, have played a key role in improving accessibility. Vision Aid is designed to address some of the biggest challenges faced by visually impaired users, offering a simple and user-friendly system that helps with face recognition, emotion identification, and medication management. By providing real-time audio feedback, Vision Aid enables users to navigate social interactions and manage their medications with confidence. Its compact and adaptable design ensures that it works seamlessly with both mobile and wearable devices, making it accessible to a wide range of users. Looking ahead, Vision Aid has the potential to become an even more powerful tool. Future enhancements may include object detection, scene understanding, and environmental recognition, allowing users to identify everyday items such as keys, wallets, and electronics more easily. Grocery shopping and meal preparation could also become more manageable with food recognition features, which could provide audio descriptions of

expiration dates and nutritional details to promote safer eating habits. In addition, real-time obstacle detection could significantly improve mobility by warning users about potential hazards in their path, such as furniture, low-hanging obstacles, or moving objects. This would make independent navigation safer and more reliable. As technology evolves, Vision Aid could integrate even more advanced features, such as augmented reality and voice control, to offer a deeper understanding of the user's surroundings. These developments will further enhance independence, mobility, and social inclusion, ensuring that assistive technologies continue to make a real difference in people's lives.

6. Publisher's Note

AIJR remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

How to Cite

Jomy *et al.*, "Vision Aid: AI-Powered Assistive Technology for the Visually Impaired", *AIJR Proc.*, vol. 7, no. 5, pp. 15-22, Sep. 2025. doi: https://doi.org/10.21467/proceedings.7.5.3

REFERENCES

[1] T. Pun, P. Roth, G. Bologna, K. Moustakas, and D. Tzovaras, "Image and video processing for visually handicapped people," *EURASIP J. Image Video Process*. 2007, 1–15 (2007) at https://jivp-eurasipjournals.springeropen.com/articles/10.1155/2007/25214

[2] C. Shi, C. Tan, and L. Wang, "A facial expression recognition method based on a multibranch cross-connection convolutional neural network," *IEEE Access 9*, 1–10 (2021) at https://ieeexplore.ieee.org/iel7/6287639/9312710/09367192.pdf

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *arXiv preprint arXiv:1409.1556*, (2014) at https://arxiv.org/abs/1409.1556

[4] L. Li, X. Mu, S. Li, and H. Peng, "A review of face recognition technology," *IEEE Access 8,* 123456–123467 (2020)

[5] G. Zhao, H. Yang, and M. Yu, "Expression recognition method based on a lightweight convolutional neural network," *IEEE Access 8*, 1–10 (2020) at https://ieeexplore.ieee.org/iel7/6287639/8948470/08952725.pdf

[6] S. D. M. Iqbal and B. Y. Suprapto, "Real-time implementation of face recognition and emotion recognition in a humanoid robot using a convolutional neural network," *IEEE Access 10*, 1–10 (2022) at https://ieeexplore.ieee.org/document/9864185

[7] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* 521, 436–444 (2015) at https://www.nature.com/articles/nature14539

[8] B. Mocanu, R. Tapu, and T. Zaharia, "Deep-see face: A mobile face recognition system dedicated to visually impaired people," *IEEE Access 6*, 1–10 (2018) at https://ieeexplore.ieee.org/iel7/6287639/8274985/08466782.pdf

[9] L. B. Neto, F. Grijalva, V. R. M. L. Maike, L. C. Martini, D. Florencio, M. C. C. Baranauskas, A. Rocha, and S. Goldenstein, "A kinect-based wearable face recognition system to aid visually impaired users," *IEEE Trans. Hum.-Mach. Syst. 47(2),* 1–10 (2017) at https://ieeexplore.ieee.org/document/7571103

[10] S. Zhang, F. Jiang, and M. Li, "Facial expression recognition based on improved VGG16 convolutional neural network," *in Proc. 2nd Int. Conf. Signal Process. Comput. Netw. Commun.,* 162–168 (ACM, 2024) at https://dl.acm.org/doi/abs/10.1016/j.patcog.2016.07.026

[11] Y. Huang, F. Chen, S. Lv, and X. Wang, "Facial expression recognition: A survey," *Symmetry*, vol. 11, no. 10, pp. 1189, 2019) at https://www.mdpi.com/2073-8994/11/10/1189

[12] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2818–2826 (2016) at https://ieeexplore.ieee.org/document/7780677

[13] R. Girshick, "Fast R-CNN," *in Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 1440–1448 (2015) at https://ieeexplore.ieee.org/document/7410526

[14] S. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM 63(11)*, 139–144 (2020) at https://dl.acm.org/doi/10.1145/3422622

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR),* 779–788 (2016) at https://ieeexplore.ieee.org/document/7780460