Vol. 07, No. 1 (2025) 405-416, doi: 10.24874/PES07.01C.007

Proceedings on Engineering Sciences

www.pesjournal.net

USER PROFILE MANAGEMENT USING DEEP LEARNING FOR ASSISTIVE HEADGEAR

Siddharth Dhondiyal[1]
Harshit Goyal
Shubham Ahuja
Nishtha Jatana

Received 13.02.2024.
Revised 20.03.2024.
Accepted 22.04.2024.

Keywords:

Blinds, User Profiling, MTCNN, FaceNet, assistive Technology, Headgear

A B S T R A C T

Individuals with visual impairments face significant challenges in interacting with their environment, particularly in identifying and recognizing people. The learning disability associated with visual impairment compounds these difficulties, hindering their ability to navigate and engage with their surroundings effectively. This study proposes an innovative AI-powered user profiling solution using a multi-task cascaded convolutional neural networks (MTCNN) algorithm for rapid face detection, FaceNet for accurate face recognition, region-based and within-region priority setting, and a tailored feedback loop with a voice-assisted contact-saving feature. The research also aims to integrate the user profiling solution into an assistive headgear prototype called AXIE. Networking involves server and client-side communication via Message Queuing Telemetry Transport(MQTT) for maintaining low network latency. Experimental results confirm the successful integration of the user profiling system with the headgear. Our profiling model, trained using Labeled Faces in the Wild(LFW) and VGGFace2 datasets, demonstrated a detection accuracy(F_D) of 99.23% and a face recognition accuracy of 95.3% (F_R). Additionally, the network latency (N_L) was measured at 60ms. Developing a user profiling system and integrating it with an assistive headgear marks a significant step towards empowering the visually impaired

1. INTRODUCTION

The population of visually impaired individuals has increased in recent decades, with an estimated 285 million people globally, as reported by the World Health Organization (WHO) ("Who—visual impairment and blindness," 1948). However, schools and workplaces often struggle to accommodate them due to a lack of assistive technologies and economic barriers (Velázquez, 2010), leaving 90% of visually impaired individuals in low-income situations.

User profiling in the context of this research involves the capability to detect and identify individuals based on pre-saved information within the headgear. When the blind or visually impaired person wearing the assistive headgear encounters another person, the device employs advanced facial detection and recognition AI algorithms (Yang, Yu & Gong, 2009) to discern the presence of known contacts. This process utilizes the named contacts previously stored in the headgear, allowing the device to create a personalized profile for each

[1] Corresponding author: Siddharth Dhondiyal
Email: siddharth.dhondiyal2512@gmail.com

individual in the user's social or professional network. User profile management system interoperability issues often arise, limiting seamless integration with other applications or platforms. This lack of compatibility can result in disconnected user experiences(Ma, Silver & Shashuki, 2010).

Our research introduces a novel user profiling technology through multi-task cascaded convolutional neural networks (MTCNN). This includes face detection, followed by the application of FaceNet for highly accurate recognition. Furthermore, there is a tailored feedback loop based on user preference with region-based(using latitude and longitude) and within-region priority settings for facial recognition built into the user profiling module. We also propose a design for an assistive headgear called AXIE(Assistive Extension with Interaction Enabler) to assist the user profile management system in various tasks which incorporates the Raspberry Pi 4 single-board computer with Debian Linux, a camera, and wired earphones for information conveyance and helping with voice assistance, and an Android app interface. To optimize performance, the design incorporates low network latency (Avasalcai, Tsigkanos & Dustdar, 2021) by processing information and running the face recognition model on the server side rather than the client side (i.e. the Raspberry Pi module) and using MQTT for connection between multiple clients and servers, also facilitating real-time proximity feature and voice-assisted contact- saving capabilities for successful integration of user profile technology into the headgear. By utilizing a combination of cameras, ultrasonic sensors, and an emergency response(ER) service and alert system the headgear can identify people close to the blind user and avoid nearby objects enabling users to navigate with greater confidence and safety (Kulkarni, Potdar, Hegde & Balidar, 2019).

The study addresses the challenge of optimizing communication and interaction technologies for people with visual impairments. The algorithm has been trained using the LFW and VGGFace2 datasets. Implementing the presented user profiling technology demonstrates promising results of detection and recognition accuracy (F_D and F_R) of faces through the camera, and low network latency(N_L) calculated using Round Trip Time. This paper makes two core contributions to assistive technology for the blind and the visually impaired. First, we are introducing a novel user profile management system using the deep learning algorithm of MTCNN and FaceNet with a tailored feedback loop based on the visually impaired preference, a three-stage contact-saving feature with assistive voice technology, and dynamic face recognition priority-based on location(latitude and longitude) and also the people within a particular region or location. Secondly, we are presenting and designing a prototype assistive headgear named AXIE aiding the user profile technology by providing a plethora of features and server-side facial recognition processing rather than on the client-side using an MQTT connection for better networking and maintaining low network latency.

The remaining paper is organized as follows. Section 2 highlights the related work. In Section 3, we explain the implementation of the user profiling management system and the design and architecture of the AXIE headgear. Section 4 further evaluates the experimental results of the user profiling technology and the design of the AXIE headgear. The threats to validity and the future scope of the proposed work are presented in Section 5. Lastly, section 6 gives the conclusion of this research.

2. LITERATURE REVIEW

The application of Deep learning algorithms like Multi-Task Cascaded Convolutional Neural Network (MTCNN) and FaceNet for facial profiling in images and video is significantly relevant. The MTCNN stands out as a robust solution for face detection and alignment (Zhang, Luo & Gao, 2020). MTCNN's ability to perform these tasks concurrently contributes to its efficiency in handling real-world scenarios, making it a popular choice in facial analysis applications. Complementing MTCNN, FaceNet, another prominent deep learning algorithm, has gained significant attention for its prowess in generating facial embeddings with high accuracy. FaceNet employs a unique triplet loss function, optimizing the neural network to minimize the distance between embeddings of the same individual while maximizing the distance between those of different individuals.(Ming, Chazalon, Luqman, Visani & Burie, 2017). This results in highly discriminative facial representations, forming a foundation for precise face verification and recognition (William, Rachmawanto, Santosa & Sari, 2019). Assistive glasses were simple and had basic features like a visual interface for remote systems in the beginning but nowadays assistive glasses have become more advanced and provide a plethora of features (Prabha, Saraswathi, Haily, Sindhuja & Udhaya, 2023). Assistive gadgets represent cutting-edge technology (Ali Hassan & Tang, 2016) . For instance, OrCam, a commercially available pair of glasses, incorporates an embedded computer with a gesture recognition system (activated by pointing a finger) to execute various tasks such as reading, and conveying the information to the user in an audio format (Amore, Silvestri, Guidobaldi, Sulfaro, Piscopo, Turco & Markowitz ,2023). Another example is eSight, designed for individuals with low vision. It captures and processes real-time scenes, presenting them on a specialized screen positioned in front of the user's eyes (Viraganoor & Nadu, 2020). Table 1. gives an overview of previous work on more of these assistive headgear designs.

Table 1. Previous work of assistive headgear

Product title	**Eyesynth-Assistive glasses**	**Google Glasses**	**Smart Glasses for Blind people**	**The eSight 3 (Kalaganis, Migkotzidis, Georgiadis, Chatzilari, Nikolopoulos, & Kompatsiaris, 2021)**	**Oton Glass**
Developed by	Eyesynth Company and designed	Google X	Hawra, Lina, Shoaa and Aqeela (CCES university)	Jason Kuhn (Optical sciences pro-	Keisuke Shimakage (Institute of Advanced Media Arts and Sciences)
Features	1. Eyesynth glasses are equipped with 3D cameras 2. These glasses give a nuanced understanding of the position and size of the surroundings using 3d-volumetric analysis	Google Glasses uses voice assistance to let user use the internet and show information hands-free.	Aids the blinds in read and translate the typed text which is written in the English language (Said, Alkhatib, Aloraidh, & Alhaidar, 2018).	It helps individuals with low vision using a camera with highresolution	These glasses help dyslexic individuals to read and scans the text and conveys it to the user via earpiece
Uses	Enable visu- ally impaired individuals	Allows the user to record video, take pictures, send messages, translation using google lens (Kumar & Sharma, 2014).	1. Uses ultrasonic sensors for object detection and notifies them when object is near. 2. Scans the image using the camera and provides information about the image.	Help users with low vision.	These glasses convert symbols into audio cues.
Drawbacks	1. It is very costly (about 570$) 2. It solely identifies objects and provides directions.	Not tailored to be used heavily by the blinds and visually impaired and are costly	1. Only has assistance for the English language. 2. Display of pictures, video, or data in front of the eyes makes it unsuitable while driving as it can distract the user.	Doesn’t improve the sight of users.	Only aids the people who have difficulty reading and not much use for blind people.

Wu and Zhang (2021) proposed an enhanced FaceNet network, integrating MTCNN for swift detection and FaceNet with an improved loss function for highly accurate verification and recognition. The experimental results demonstrated superior real-time recognition performance, surpassing traditional and deep learning algorithms, making it applicable to Access Control Systems.

Daescu, Huang, and Weinzierl (2019) addressed prosopagnosia challenges by proposing a wearable glasses-based face recognition system. Unlike local systems, they introduced a client-server architecture, achieving 98.18% accuracy using Deep CNN for backend recognition, designed to handle new identities without model reconstruction.

Chaudhry and Chandra (2017) gave a visual assistive system employing mobile face detection and recognition via convolutional neural networks, addressing challenges like camera shakes. Dataset creation and performance evaluation against cascade classifiers were conducted, emphasizing robust implementation for visually impaired guidance.

Ruffieux, Hwang, Junod, Caldara, Lalanne, and Ruffieux (2023) explored the correlation between visual impairment, quality of life, and expectations for smart glasses as multipurpose assistive tools. Utilizing NEI-VFQ-25 and two newly developed questionnaires, they assessed specific challenges, strategies, and assistive technology expectations. By combining these instruments, the researchers were able to gain a comprehensive understanding of the diverse needs and preferences within the visually impaired population, informing the development of tailored interventions and assistive technologies to enhance their quality of life.

Rajaraman and Shivapriya (2023) stated the challenges faced by India's visually impaired population through an innovative solution, A-Eye. This smart glasses system integrates AI features seamlessly, leveraging Mobile Edge Computing (MEC) for efficient data processing on mobile phones, ensuring low latency, scalability, and enhanced assistance for the visually impaired.

Jafri and Ali (2014) highlight the transformative potential of natural eyewear-based display devices like Google Glass for visually impaired users. Emphasizing discreetness and versatile functionality, they explore existing features' utility, suggest accessibility enhancements, and propose novel applications. Insights from an interview with a blind individual underscore the practicality and benefits of these devices in enhancing autonomy and self-reliance.

Suh, Kang, and Woo (2005) proposed a framework featuring wearable personal stations (WPS) and decentralized user and application services, facilitating context exchange only when needed. This approach, avoiding a centralized server, efficiently addresses individual user preferences, desires, and needs, ensuring privacy by allowing users to share their profiles selectively.

The combination of MTCNN and FaceNet is particularly notable in facial profiling, where comprehensive user analysis is achieved through the union of rapid face detection and the generation of discriminative facial embeddings (Jin, Li, Pan, Ma, & Lin, 2021) . Such an integrated approach enhances the overall performance of facial profiling systems. Moreover, headgear-based facial profiling has found applications across different domains, like retail analytics, and healthcare. Studies have highlighted the potential of these devices in enhancing security measures, improving customer experiences, and facilitating personalized healthcare interventions.

3. IMPLEMENTATION OF PROPOSED WORK

This work introduces an integrated user profiling technology within an assistive headgear. The functioning of the entire system is illustrated in Figure 2. The proposed system comprises of multitude of interconnected devices and components, each providing a functionality in the assistive headgear. These components are mentioned as follows:

- The AXIE device in the architecture is the Raspberry Pi 4, acting as the main processing unit for the headgear and also transmitting audio and visual data(via camera and earphones) of people the visually impaired individual encounters, to the server. It also helps in establishing connections with other nearby headgear using Discover ID.
- An Android application, designed specifically for visually impaired users, is equipped with location services, facilitating real-time tracking of the user's position in terms of latitude and longitude. In the event of gear failure, detected through the absence of communication pings, the application triggers an Emergency Response (ER) service and sends an alert, ensuring the blind user's safety and well-being. Furthermore, it acts as an internet provider for the Raspberry Pi.
- For non-blind users, an Android application is incorporated into the system, offering additional features. This device facilitates the storage of contacts based on actual images, utilizing a global database connection rather than relying on the local Raspberry Pi storage. It also provides the gear's location and issues alerts in case of gear failure or prolonged communication disruptions.
- The server side is equipped with a GPU, for the computation of deep-learning models of MTCNN and FaceNet. The MQTT protocol acts as a gateway, facilitating communication between devices on the client side and server.MQTT efficiently transmits video bit streams captured by AXIE's camera—to the server-side for tasks like facial recognition. This real-time data transfer is optimized for low latency, ensuring prompt updates. MQTT's streamlined mechanism compresses and encodes the video bit streams, maximizing bandwidth efficiency. This integration establishes a robust link, facilitating bidirectional communication for timely and accurate information exchange. Additionally, it establishes a connection with Amazon S3 for image storage, enabling efficient data retrieval for non-blind Android devices in the case of failed voice recognition.

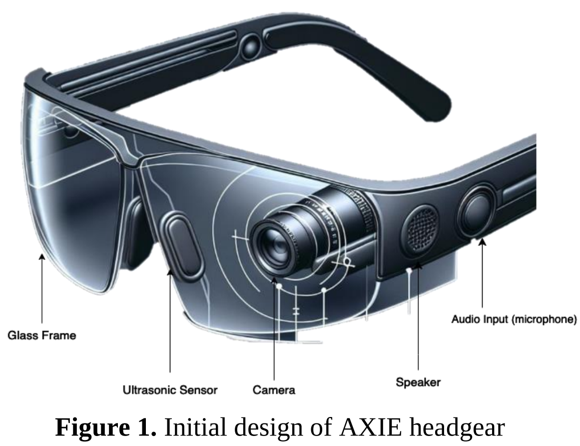

Figure 1. Initial design of AXIE headgear

Further research plans for encryption updates are envisioned to enhance the security and privacy aspects of the system.

These components form a robust and interconnected ecosystem designed to address the specific needs of visually impaired individuals.

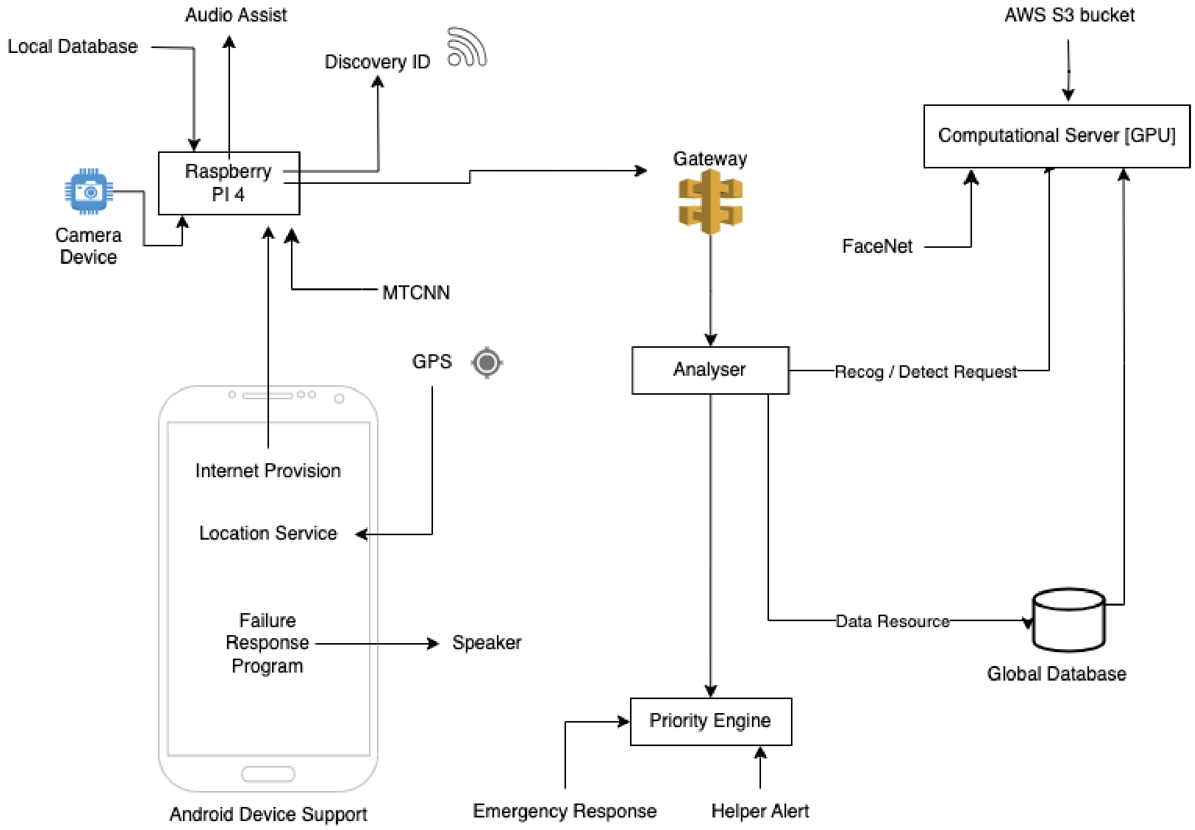

Figure 2. Operational structure

3.1 Overview of AXIE headgear's hardware design

At the center of the assistive headgear is the Raspberry Pi 4, a single-board computer with 8GB of RAM and Debian Linux as an operating system. The emulation environment mirrors th production setup, simulating Debian 11.X with 4GB of RAM and 2 cores. The Raspberry Pi, functioning as a vital computational component, handles various client-side operations and facilitates the headgear's interfacing with its essential components, including the audio and camera modules.

The camera and wired earphones contribute significantly to the headgear's functionality. The camera captures video at 60 frames per second, providing high-quality recordings parsed in 720P resolution. Meanwhile, the earphones provide the headgear with voice assistance that is enhanced through text-to-speech processing on the Raspberry Pi. Another part of the headgear is the Ultrasonic module (as shown in Figure 3.) that offers non-contact measurement functionality ranging from 40 cm to 150 cm ((Rahman, Chakma, Raza, Akter, & Sattar, 2021; Ghosh, 2023). Comprising ultrasonic transmitters, receivers, and control circuitry, the module autonomously emits eight high-frequency ultrasonic pulses and detects the presence of a returning pulse signal. Upon signal reception, it generates a high-level output, with the duration of the high-level IO corresponding to the time taken for the ultrasonic signal to travel and return.

The AXIE hardware architecture also consists of two buttons that are connected to the Raspberry Pi module and pressing one of them starts the headgear which in turn activates the camera/audio devices while the other button provides contact-saving functionality to the blind user.

Moreover, it has a glass frame to which the sensor, camera, and earpiece are attached and the Raspberry Pi module would be embedded into the frame, creating a wearable configuration for the user as demonstrated in Figure 1.

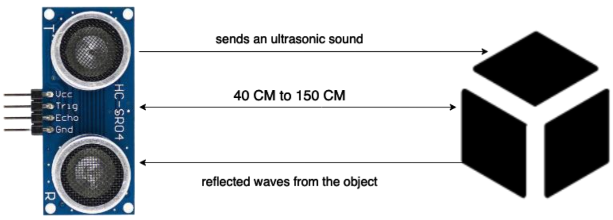

Figure 3. Ultrasonic sensor mounted on the AXIE headgear

3.2 User Profiling system

User profiling in the context of this work means the ability to identify and recognize the people interacting with blind individuals throughout their daily endeavors. This involves saving the identity of unknown faces in the database according to the blind individual's preference and notifying them when a known face comes near them with the help of the AXIE headgear. Both the client-side headgear and the server-side GPU with the database are (Sanchez-Gordon, Aguilar-of AXIE and User Profiling system Mayanquer, & Calle-Jimenez, 2021) working in tandem to deploy the user profiling technology.

One of the primary responsibilities of the AXIE headgear on the client end is managing a local database and providing local storage for essential data to enhance the headgear's functionality. Establishing connections with both the central server, equipped with a GPU for rapid computational tasks, and the Android device designed for visually impaired users, the Raspberry Pi helps in performing swift communication among these components.

Moreover, on the client side, an Android application was designed for the assistive headgear. The user profiling process of client-side application goes as follows. When a user opens the application and toggles the on/off button of the AXIE, the phone application will ask the user to make a Bluetooth connection with the glasses. This activates the camera and audio input and starts detecting the people near the blind user. The device continuously scans the images and renders audio of the people the AXIE device user is interacting with and the Android application sends the image captured to the server via an HTTP post request for the subsequent face recognition process. The face detection process happens on the client side as it doesn't require a lot of computational power and can be handled easily by the Raspberry Pi processor. Detection happens using the MTCNN algorithm which captures the frames of people around the blind individual and if the blind user is interacting with someone they are familiar with but the AXIE's database has no record of their identity then the user profiling algorithm starts calculating the number of detected frames of that person. The prototype's camera captures video at 60 frames per second and the algorithm has a threshold of 10 seconds which means that if the number of frames of that person detected by the headgear exceeds 600 frames then the user profiling algorithm will store the image data of that person in the local database of the headgear(refer to Algorithm 1).

The unknown user's image data is stored along with a timestamp, capturing the precise date, time, and location of the blind user's encounter with that individual. This timestamp provides a unique record of unfamiliar people to the AXIE headgear and also aids the blind user in saving their contact later based on the time, date, and location of interaction. Contact-saving is a server-side feature designed to empower blind users to recognize and remember individuals within their social interactions. When the AXIE device encounters an unknown face, the system initially stores the person's data in the AXIE's local database.

When the blind user is in their home or any other safe space they can use the button present on AXIE to initiate the contact-saving process or the AXIE device would do it automatically when it is idle and hasn't detected any faces for more than 30 minutes. This feature allows the blind to store the locally saved contact details in the global database using voice assistance as illustrated in Table 2. The information stored in the global database is the username of the unknown contact which is assigned by the blind user along with the image data and audio recording. The AXIE headgear will navigate them through this process using the earphones, iterating through the unknown contact list asking the user for an input username for each contact, and then storing it in the global database.

Algorithm 1 User Profiling

```
for all face hash in faces do
    if face hash in known database then
        if alerted then
            do nothing if already alerted
        else
            notify user about the recognized face
            speak(face hash.username)
            set alerted = true
        end if
    else
    if face hash in unrecognised database then
            detected frames += 1
            if detected frames > 600 then
  // record sound if the unknown user is detected
  microphone) for more than 600 frames using
  timestamp
record sound(timestamp.wav, input database
            end if
  else
  // Add the unknown face to the unrecognised
unrecognised database.append(face hash)
        end if
    end if
end for
```

When multiple faces are detected simultaneously, it leads to images with identical timestamps stored in the AXIE local database. This problem is countered by storing a small voice recording of each person along with their image data which would be played during the contact-saving process when the blind user is in a safe space. It aids the blind user to recognize the voice along with the timestamp and location to get a comprehensive understanding of whom they were talking to while saving their contact. Additionally, in cases where the detected person already has a record in the global database then they will be recognized by the AXIE device with the help of the FaceNet algorithm running on the server-side and the AXIE alerts the visually impaired user about the contact details of that person i.e. their username as saved by the blind user. Another feature that aids the user profiling strategy is region-based and within-region priority settings. The region-based priority entails using latitude and longitude to categorize encountered contacts according to their geographical location.

This approach ensures that contacts made at frequently visited places, such as home or office, receive higher priority in the user profiling algorithm. Also, within each region, a fine-grained priority system is established based on the frequency of recognition. For instance, individuals recognized more frequently in a specific location, like a family member at home, are accorded higher priority, optimizing the sorting of contacts within the database. These prioritization mechanisms facilitate quicker and more efficient access to important contacts.

Table 2. Dialogue between AXIE headgear and blind user during contact-saving process

1: AXIE: *you have a total of 1 contacts to save ! say yes to confirm*
2: Process: speak
3: User: *proceed*
4: AXIE : *alright let's start*
5: AXIE: *you met this person at 12 today in your office*
6: AXIE: *want me to save it as ?*
7: Process: speak
8: User: *xyz*
9: AXIE: *confirm name by saying confirm, selected name XYZ*
10: Process: speak
11: User: *confirm*
12: AXIE: *saving this person as xyz*
13: Process: ('1700897329guy': 'xyz

3.3 Face detection and recognition model

Dataset The user profiling system uses two datasets for training the face recognition model. The Labeled Faces in the Wild (LFW) dataset serves as the training ground for FaceNet's verification prowess. It provides numerous labeled image pairs, indicating whether two faces belong to the same individual. To enhance FaceNet's recognition capabilities, the VGGFace2 dataset offers an abundance of images per person, captured under varied poses and lighting conditions. We used 8600 identities with 2721600 total images from the VGGFace2 dataset to train the face recognition model. Complementing these datasets is application-specific user data, sourced from real-world images captured through the camera. Including all the datasets, we used 5646 identities with 12582 total images for testing.

Preprocessing For MTCNN preprocessing, images were first resized to a resolution of 240x240 pixels for consistent processing, and pixel intensities were then normalized, ensuring a zero mean and unit variance across all images, and facilitating numerical stability during neural network training. FaceNet preprocessing leverages MTCNN's landmark locations, extracting the facial regions from captured images and then these extracted regions are aligned to a standardized size and cropped to focus on the key facial features, eliminating distracting backgrounds. The identification of facial regions in images is achieved through the use of the open-source library Dlib5. Dlib face detector was employed to crop faces in both the VGGFace2 and LFW datasets. The Dlib face detector exhibits a high accuracy of 99.23% when employed for detecting facial regions on the LFW dataset.

Training The neural network for the face recognition model was trained using the stochastic gradient method and the VGGFace2 dataset. Hyperparameters are the same as the m recommended values in (Schroff, Kalenichenko, & Philbin, 2015). Sampling was done from the images and identities of the dataset in producing triplets, minimizing the distance between the anchor and positive images (faces belonging to the same person), while maximizing the distance between the anchor and negative image (different person). Since the loss function that guides the network to learn a face embedding space where distances reflect facial similarity is defined as:

$$\mathbf{L_{triplet}} = \max(\|f_a - f_p\|_2 - \|f_a - f_n\|_2 + \text{margin}, 0). \quad (1)$$

where,
$\mathbf{L_{triplet}}$ = triplet loss
$\mathbf{f_a}$ = embedding of the anchor image
$\mathbf{f_p}$ = embedding of the positive image (same person as anchor)
$\mathbf{f_n}$ = embedding of the negative image (different person)
margin = hyperparameter controlling the minimum distance between positive and negative pairs

For every batch, we sampled 700 people and 5 images per person. The margin value is set to 0.3. The images for each triplet (anchor, positive, negative) were each fed through the model to produce an output vector. We start with a learning rate of 0.06. Our model was trained for 230 hours and stopped when it appeared to stagnate in performance improvements.

4. RESULT

During testing of the headgear prototype the camera and earphones were connected to the Raspberry Pi 4 module and the ultrasonic sensor was connected to the breadboard via jumper wires. The breadboard was attached to the tester's upper arm and the camera and ultrasonic sensor themselves were mounted on the glass frame which was worn by the tester as illustrated in Figure 4.

Figure 4. Final AXIE prototype

Testing of the user profiling technology was done by measuring key parameters, including the algorithm's accuracy in face recognition and detection, along with the network latency for bidirectional data transmission between the client-side and server- side.

Network latency (N_L) is the aggregate sum of all the possible delays the packet faces during transmission(refer to Equation (3)) and is expressed using Round Trip Time (RTT). Possible delays include processing (D_P), queuing, transmission, and propagation delays. Other factors may contribute to the delay, but this research specifically addresses the typical causes of delay within its scope, and since queuing delay is not calculable as it varies this work only considers transmission and processing delays. Transmission delay (D_L) is the time to push all the available data in a transmission medium or wire which is calculated by dividing the number of bits (N_B) by the transmission rate (T_R) as shown in Equation (2).

$$D_T = N_B / T_R \tag{2}$$

Network latency is calculated as:

$$N_L = D_P + D_T \tag{3}$$

Table 3. Experimental values of parameters for calculating network latency

Parameters	Value
Transmission speed	200 Mbps
Data size	500 KB
Processing speed on Raspberry pi	15 ms
Processing speed on server GPU	25 ms

The transmission speed through MQTT connection is measured at 200 Mbps. The data size, totaling 500 KB, encompasses both the facial numeric data generated by the MTCNN algorithm on the client side and the audio data from user interactions. On the client side, the Raspberry Pi module, equipped with a 1.8 GHz processor, efficiently executes the MTCNN algorithm in approximately 15 ms. Meanwhile, the server side, featuring a 3.8 GHz processor, processes images and audio data through the FaceNet algorithm for recognition within 25 ms. These factors that contribute to network latency are described collectively in Table 3.

Therefore, network latency (N_L)

$$= N_B / T_R + D_P = [(1000 * 8 * 500)/ 200000000] \text{ sec} + 40 \text{ ms} = 60 \text{ ms} \tag{4}$$

Table 4. Confusion matrix for face detection using wider face dataset

True Positive 393703	False Positive 2940
False Negative 115	**True Negative** 0

Figure 5. Detection of faces using AXIE headgear

The MTCNN detection algorithm's performance was assessed using a confusion matrix as illustrated in Table 4. Results demonstrated an impressive accuracy of 99.23% on the wider face dataset (Yang, Luo, Loy, & Tang, 2016), also indicating the user profiling algorithm's robustness in effectively detecting and precisely localizing faces within the visual data stream acquired from the AXIE camera live feed (refer to Figure 5). This underscores the algorithm's proficiency in handling various facial orientations, sizes, and poses. Within the AXIE device's operational framework, the MTCNN algorithm continuously processes the real-time camera feed on the Raspberry Pi. Leveraging its hierarchical structure comprising three stages—face detection, facial landmark localization, and bounding box refinement—the algorithm quickly and accurately identifies individuals within the device's vicinity (Gu, Liu, & Feng, 2022). The convolutional neural networks in each stage allows for efficient feature extraction and robust face representation, contributing to the algorithm's exceptional performance. Additionally, the AXIE device alerts a user upon continuously detecting more than 300 frames of an individual using a 60 fps(frames per second) camera, to enhance situational awareness for the blind user.

Accuracy for MTCNN using confusion matrix (F_D) =

$$\frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

where ,
TP = True Positive (correctly detected or recognized faces)
TN = True Negative (correctly identified absence of faces)
FP = False Positive (incorrectly detected or recognized faces)
FN = False Negative (faces that are present but not detected or recognized)

Putting the values of TP, TN, FP, FN in the Equation (4), the accuracy was calculated to be 99.23% for user profiling algorithm's detection.

The face recognition system running on the server side GPU, an integral component of the AXIE device, employs convolutional neural networks (CNNs) trained on extensive datasets of VGGFace2 and LFW datasets. VGGFace2 comprises a vast collection of facial images captured under diverse conditions, offering a rich source of data for training deep learning models. The FaceNet architecture utilizes a triplet loss function to learn discriminative feature embeddings, where face images of the same individual are encouraged to be closer in the embedding space compared to those of different individuals. This methodology enables the network to generate compact yet highly discriminative representations of faces, facilitating accurate recognition. Evaluation of the FaceNet algorithm is also done using confusion matrix analysis as demonstrated in Table 5., the FaceNet model achieved an impressive accuracy rate of 95.3% (by putting the values of TN, TP, FN, and FP from Figure 7 in Equation (4)) on the VGGFace2 dataset, underscoring its effectiveness in robust face recognition tasks.

Table 5. Confusion matrix for face recognition using VGGFace2 dataset

True Positive 110578	False Positive 2387
False Negative 6136	**True Negative** 62201

Accuracy for FaceNet using confusion matrix (F_P) =

$$\frac{TP + TN}{TP + TN + FP + FN} = 95.3\%. \tag{6}$$

Furthermore, the user profiling algorithm implemented on the AXIE device demonstrated significant accuracy in recognizing stored contacts and it was evaluated by calculating the ratio of recognized frames of an individual to the total frames detected—set at 600 frames for this algorithm—accurate contact identification was achieved as illustrated in Figure 6. Table 6. presents the approximated accuracy of known contacts that are already stored in the global database the AXIE headgear sends the image data to the server side and it is processed by the FaceNet algorithm and sent back to the client-side, if the face recognition accuracy is more than 90% it the blind user is notified with the username of the individuals.

Figure 6. Recognition accuracy of known faces already stored in global database

Table 6. Testing user profiling on already saved contact

S.No	Contact saved in database	Recognized		Accuracy (approximate values)
		Number of frames recognized	Total number of frames detected	
1	Sparsh	576	600	0.96
2	Kartikay	577	600	0.96
3	Siddharth	575	600	0.96
4	Mayank	564	600	0.94
5	Zubin	570	600	0.95
6	Aashwin	576	600	0.96
7	Maanvika	571	600	0.95
8	Ananya	565	600	0.94
9	Aayushi	564	600	0.94
10	Saanvi	568	600	0.95

5. THREATS TO VALIDITY AND FUTURE SCOPE

In this section, we have assessed the possible threats that might impede the advancement of this research, including issues related to high-latency networking and hardware limitations. The discussion also briefly touches on our future implementation, aiming to make AXIE more scalable, customizable, inclusive, and enhancing the AXIE headgear's performance in low-network areas.

5.1 High latency in low network environment

The primary concern lies in the comparatively high latency networking primarily stemming from issues related to low network connectivity in rural areas (Villapol, Liu, Gutierrez, Qadir, Gordon, Tan, & Zhang, 2018) and places with no network. This hinders the ability of the user profile system integrated with the head-gear to identify people around the blind and visually impaired as network-dependent processes, such as facial recognition and video bit packets streaming from the client side to the server side, are particularly susceptible to disruptions caused by inadequate network conditions, leading to compromised accuracy in identifying people. This concern necessitates the exploration of strategies to optimize the system's performance under challenging network scenarios, ensuring robust functionality regardless of connectivity limitations.

5.2 Hardware constraints in headgear optimization

The use of the Raspberry Pi 4 introduces inherent limitations, given its processing power and memory constraints (Motti, 2019). The absence of bone conduction hardware makes the system incompatible with deaf-blind individuals. Additionally, hardware degradation over time also presents a concern and, potential issues affecting the system's long-term performance.

5.3 Enhancing User Profiling Technology and AXIE

With an emphasis on User Profiling, the gears' functionality can be expanded and improved upon to accommodate various tasks by integrating additional models into the core program, subject to the limitations of the high network latency issues along with the extensibility and inclusivity of AXIE:

- **Improving Networking using Metaheuristics:** Dijkstra's Algorithm Integration. The optimization of networking within the assistive headgear system is a paramount objective to minimize latency and enhance real-time responsiveness. We propose the integration of metaheuristics, specifically a dynamic Dijkstra algorithm, to streamline the networking processes. By employing Dijkstra's algorithm, we aim to optimize the routing of data, facilitating the simultaneous execution of features with minimal latency. This approach not only ensures efficient communication between server and client components but also contributes to the overall responsiveness of the system. Further improvements of this work strive to create a seamless and instantaneous user experience, reinforcing our commitment to providing cutting-edge assistive technology solutions for individuals with visual impairments.

- **Making AXIE more Inclusive and into an Expandable Platform:** The evolution of AXIE extends beyond its initial design, aiming to extend the headgear's hardware design to foster inclusivity for individuals and accommodate users with diverse visual disabilities. The vision is to optimize networking and faster processing for transforming AXIE into a scalable and expandable platform, inviting contributions from the community to integrate new features into AXIE seamlessly.

6. CONCLUSION

This research implements a user profiling solution for individuals struggling with visual impairments to identify people around them, using deep learning algorithms of MTCNN and FaceNet for face detection and priority-based recognition system along with a three-stage contact-saving feature integrated with low network latent prototype of an assistive headgear leveraging the Raspberry Pi 4 single-board computer, a camera, an audio device for voice assistance, and an Android app that provides an interface for both the blind and non-blind to interact with the headgear. Delving into the system's design, operational mechanisms, and principles, supplemented by experimental results this work envisions to enhance the lives of visually impaired students and aims to surmount economic constraints. Testing confirms the efficacy of user profile technology, demonstrating a remarkable detection accuracy of 99.23% and recognition accuracy of 95.3%. Furthermore, network latency stands at 60ms. These findings underscore the practical utility and effectiveness of the User profile management, showcasing its potential to make a meaningful impact on the lives of the visually impaired. We further discuss the limitations pertaining to the hardware and high latency in network deficient areas and propose ideas for improvement of the AXIE headgear and fostering inclusion of a broader spectrum of users.

References:

Ali Hassan, E., & Tang, T. B. (2016). Smart glasses for the visually impaired people. In *Computers Helping People with Special Needs: 15th International Conference, ICCHP 2016, Linz, Austria, July 13-15, 2016, Proceedings, Part II 15* (pp. 579-582). Springer International Publishing.

Amore, F., Silvestri, V., Guidobaldi, M., Sulfaro, M., Piscopo, P., Turco, S., ... & Markowitz, S. N. (2023). Efficacy and Patients' Satisfaction with the ORCAM MyEye Device Among Visually Impaired People: A Multicenter Study. *Journal of Medical Systems*, *47*(1), 11.

Avasalcai, C., Tsigkanos, C., & Dustdar, S. (2021). Resource management for latency-sensitive IoT applications with satisfiability. *IEEE Transactions on Services Computing*, *15*(5), 2982-2993.

Chaudhry, S., & Chandra, R. (2017). Face detection and recognition in an unconstrained environment for mobile visual assistive system. *Applied Soft Computing*, *53*, 168-180.

Daescu, O., Huang, H., & Weinzierl, M. (2019, June). Deep learning based face recognition system with smart glasses. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments* (pp. 218-226).

Ghosh, J.. S. (2023: Flex sensor and machine learning-based mute communication system. Indian Scientific Journal of Research in Engineering and Management. https://doi.org/10.55041/ijsrem24537

Gu, M., Liu, X., & Feng, J. (2022). Classroom face detection algorithm based on improved MTCNN. *Signal, Image and Video Processing*, *16*(5), 1355-1362.

Jafri, R., & Ali, S. A. (2014, September). Exploring the potential of eyewear-based wearable display devices for use by the visually impaired. In *2014 3rd International Conference on User Science and Engineering (i-USEr)* (pp. 119-124). IEEE.

Jin, R., Li, H., Pan, J., Ma, W., & Lin, J. (2021). Face recognition based on MTCNN and Facenet.

Kalaganis, F. P., Migkotzidis, P., Georgiadis, K., Chatzilari, E., Nikolopoulos, S., & Kompatsiaris, I. (2021, July). Lending an artificial eye: beyond evaluation of CV-based assistive systems for visually impaired people. In *International Conference on Human-Computer Interaction* (pp. 385-399). Cham: Springer International Publishing.

Kulkarni, A. U., Potdar, A. M., Hegde, S., & Baligar, V. P. (2019, July). Radar based object detector using ultrasonic sensor. In *2019 1st International Conference on Advances in Information Technology (ICAIT)* (pp. 204-209). IEEE.

Kumar, G., & Sharma, P. (2014). Google Glasses Impediments. *medicine*, *1*(2).

Ma, Z., Silver, D. L., & Shakshuki, E. M. (2010). User profile management: reference model and web services implementation. *International Journal of Web and Grid Services*, *6*(1), 1-34.

Ming, Z., Chazalon, J., Luqman, M. M., Visani, M., & Burie, J. C. (2017, October). Simple triplet loss based on intra/inter-class metric learning for face verification. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)* (pp. 1656-1664). IEEE.

Motti, V. G. (2019, September). Assistive wearables: opportunities and challenges. In *Adjunct proceedings of the 2019 acm international joint conference on pervasive and ubiquitous computing and proceedings of the 2019 acm international symposium on wearable computers* (pp. 1040-1043).

Prabha, M., Saraswathi, P., Hailly, J., Sindhuja, C., & Udhaya, P. (2023, April). Smart Glasses: A Visual Assistant for Visually Impaired. In *2023 11th International Conference on Emerging Trends in Engineering & Technology-Signal and Information Processing (ICETET-SIP)* (pp. 1-6). IEEE.

Rahman, M. M., Chakma, S., Raza, D. M., Akter, S., & Sattar, A. (2021, July). Real-Time Object Detection using Machine Learning. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-5). IEEE.

Rajaraman, S., & Shivapriya, P. (2023, March). A-Eye: Computer vision and Deep learning based Smart-Glasses with Edge computing for Visually Impaired. In *2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP)* (pp. 1-5). IEEE.

Ruffieux, S., Hwang, C., Junod, V., Caldara, R., Lalanne, D., & Ruffieux, N. (2023). Tailoring assistive smart glasses according to pathologies of visually impaired individuals: an exploratory investigation on social needs and difficulties experienced by visually impaired individuals. *Universal Access in the Information Society*, *22*(2), 463-475.

Said, H., Alkhatib, L., Aloraidh, A., & Alhaidar, S. Smart Glasses for Blind People. (2018). Retrieved from https://www.pmu.edu.sa/attachments/academics/p- -pt/udp/cces/smart-glasses-for-blind-people-report.pdf

Sanchez-Gordon, S., Aguilar-Mayanquer, C., & Calle-Jimenez, T. (2021). Model for profiling users with disabilities on e-learning platforms. *IEEE Access*, *9*, 74258-74274.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 815-823).

Suh, Y., Kang, D., & Woo, W. (2005, September). Context-based User Profile Management for Personalized Services. In *ubiPCMM* (pp. 64-73).

Velázquez, R. (2010). Wearable assistive devices for the blind. *Wearable and Autonomous Biomedical Devices and Systems for Smart Environment: Issues and Characterization*, 331-349.

Villapol, M. E., Liu, W., Gutierrez, J., Qadir, J., Gordon, S., Tan, J., ... & Zhang, W. (2018). A sustainable connectivity model of the internet access technologies in rural and low-income areas. In *Smart Grid and Innovative Frontiers in Telecommunications: Third International Conference, SmartGIFT 2018, Auckland, New Zealand, April 23-24, 2018, Proceedings* (pp. 93-102). Springer International Publishing.

Viraganoor, M., & Nadu, T. (2020). For (E) Sight: A Perceptive Device to Assist Blind People.

Who—visual impairment and blindness. who. (1948). Retrieved from http://www.who.int/mediacentre/factsheets/fs282/en/.

William, I., Rachmawanto, E. H., Santoso, H. A., & Sari, C. A. (2019, October). Face recognition using facenet (survey, performance test, and comparison). In *2019 fourth international conference on informatics and computing (ICIC)* (pp. 1-6). IEEE.

Wu, C., & Zhang, Y. (2021). MTCNN and FACENET based access control system for face detection and recognition. *Automatic Control and Computer Sciences*, *55*, 102-112.

Yang, J., Yu, K., Gong, Y., & Huang, T. (2009, June). Linear spatial pyramid matching using sparse coding for image classification. In *2009 IEEE Conference on computer vision and pattern recognition* (pp. 1794-1801). IEEE.

Yang, S., Luo, P., Loy, C. C., & Tang, X. (2016). Wider face: A face detection benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5525-5533).

Zhang, N., Luo, J., & Gao, W. (2020, September). Research on face detection technology based on MTCNN. In *2020 international conference on computer network, electronic and automation (ICCNEA)* (pp. 154-158). IEEE.

Siddharth Dhondiyal
Maharaja Surajmal Institute of Technology,
New Delhi,
India
siddharth.dhondiyal2512@gmail.com
ORCID 0009-0004-7905-1464

Harshit Goyal
Maharaja Surajmal Institute of Technology,
New Delhi,
India
hgoyal2099@gmail.com
ORCID 0009-0003-9265-1766

Shubham Ahuja
Maharaja Surajmal Institute of Technology,
New Delhi,
India
shubhamahuja482@gmail.com
ORCID 0009-0006-1870-6273

Nishtha Jatana
Maharaja Surajmal Institute of Technology,
New Delhi,
India
nishtha.jatana@gmail.com
ORCID 0000-0002-6863-6756