

Research Article

Smartphone-Based Cognitive Assistance of Blind People in Room Recognition and Awareness

Muhib Ullah,¹ Shah Khusro ,¹ Mumtaz Khan ,¹ Iftikhar Alam ,² Inayat Khan ,³ and Badam Niazi ⁴

¹Department of Computer Science, University of Peshawar, Peshawar 25000, Pakistan

²Department of Computer Science, City University of Science and Information Technology, Peshawar, 25010, Pakistan

³Department of Computer Science, University of Buner, Buner 19290, Pakistan

⁴Department of Computer Science, University of Nangarhar, Jalalabad, Nangarhar 2600, Afghanistan

Correspondence should be addressed to Badam Niazi; niazi5.48@gmail.com

Received 5 February 2022; Revised 18 April 2022; Accepted 24 May 2022; Published 13 June 2022

Academic Editor: Mukhtaj Khan

Copyright © 2022 Muhib Ullah et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Blind people face multiple challenges in performing their daily activities, such as localization, navigation, and communication with other people following the social status of the ambient environment. The outdoor navigation and localization are mitigated through assistive technologies, such as white cane, smartphones, and Global Positioning System (GPS). However, little attention has been paid to assisting a blind person in judging the room environment, occupants in a room, communicating with an intended person, social status, and so on. This study attempts to provide cognitive assistance to blind people in predicting room types and finding the intended person to communicate their message by understanding their ambient environment and social status in terms of their age, gender, and the number of people inside a room. The proposed solution uses the microphone and speaker to recognize room types and a camera for understanding the ambient environment and social status. The information is conveyed to the blind person through haptic feedback. We analyzed different evaluation metrics, including the movement of people, ambient sounds, orientation, and position. We conducted an extensive user study to validate the proposed solution in real-world scenarios and achieved 87.64% accuracy in room type recognition in 11 rooms, 64.57% gender recognition, 61.73% age group, and 61.71% correctly identified the number of people.

1. Introduction

Global societies are established on the ability to see [1]. Vision is the most important five senses and an integral part of the human body [2]. It plays a vital role in every aspect of our life. Every year, World Sight Day is observed to spread awareness about blindness and visual impairment and get the world's attention for its prevention and care [2].

Usually, a human perceives up to 80% of all information through sight (<https://medicalfuturist.com/future-of-vision-and-eye-care/>). World Health Organization (WHO), in its 2019 world report on vision, states that approximately 2.2 billion people around the globe are visually impaired or blind, of whom at least 1 billion have a vision loss that could have been avoided or is yet to be tackled. Vision is essential

to interpersonal and social interactions in face-to-face communication, conveying information through nonverbal cues, such as facial expressions and body gestures. Interpersonal communication is either verbal or nonverbal. Almost 65% of communication is nonverbal [3]. Nonverbal communication [4] (facial expression, eye contact, and bodily gestures) plays a vital role in social interaction. Blind people cannot access nonverbal communication, leading to social segregation and discrimination.

Sighted people usually use their empirical senses to navigate from one point to another, get knowledge of the surrounding environment, or communicate their message to the intended person per their social status. According to Mandru et al. [5], 83% of the information regarding the surrounding environment is received through vision, 11% by

hearing, and 3.5% by smell. Therefore, vision contributes more to everyday activities and enables people to grow in every phase of living. In contrast, blind people are deprived of vision-based information. The situation adds fuel to the fire that 90% of the world's visually impaired people live in developing and low-income countries and most blind people are in the age group from 50 to 82 [6]. They are facing many challenges in performing their daily life activities.

Some of the essential activities where blind and visually impaired people need assistance include navigation [7–9] (indoor and outdoor), localization (indoor and outdoor), and interaction with a concerned person while taking care of the social status of the surrounding environment [10, 11]. Among these, navigation and localization can be mitigated by different types of assistive aids, which can be simple such as guide dogs, human guides, and white cane, and complicated high-tech devices [6] such as virtual white cane [12], ioCane [13], and smart robot [14]. These technological devices can be categorized into three main groups based on their application: electronic orientation aids, position locator devices, and electronics travel aids [15]. However, these solutions are either expensive, difficult to use, or do not accurately meet the intended purpose [16]. For example, using these aids, a blind user may reach a target room inside an academic building, but understanding the room, its nature, layout, and social status is a daunting task. Traditional solutions cannot address it.

Summarizing the existing techniques, we can say that these solutions cannot recognize the room type where the blind person enters and find the room's social status by identifying the number of occupants, gender, and age group. Besides, these solutions are limited due to additional hardware, customized infrastructure, or socially unacceptable. Therefore, it is an immense need to develop a solution that can understand the room type and human occupants, improve a simple blueprint of the room, and efficiently communicate the blind with the intended person in an academic building. Such a solution should be cheap and socially acceptable so that a blind user, while standing at the door, easily understands the room type and recognizes its social status by finding the number, age, and gender so that the blind user can adapt their behavior accordingly. The solution should avoid additional hardware or customized infrastructure. Therefore, we proposed a solution that uses the ubiquitous smartphone, which is cheap and socially acceptable.

Although both indoor and outdoor environments are important for blind users regarding navigation, localization, and understanding, the former becomes more important when a blind has to reach the desired room and understand its social status for communication [17]. By social status, we mean the number, age, gender, and seating/standing arrangement so that a blind user understands the occasion (e.g., meeting, seminar, and teaching) finding the intended candidate person to talk per the social status or find an empty seat to sit, and so on.

Besides the availability of technology-assisted solutions for navigation and localization, it is of immense importance to find the number of people, understand their social status, age, and gender, and navigate to and find the intended

audience/person to communicate their message. For this, we use a smartphone because it is a better off-the-shelf gadget available with a large number of sensors [18–21]. These sensors are responsible for sensing the ambient environment and positioning of the smartphone [22, 23]. Besides robust sensors, the smartphone comes with high processing CPUs and GPUs that can process sophisticated graphics and 3D animated data with high speed and accuracy. In a nutshell, the smartphone comes with various robust sensors liable to be used in different situations without involving cognitive overload and extra hardware.

Therefore, we can use a smartphone that will ease the cognitive overload of blind people by providing information about room type, internal environment, number of occupants, and their social status. We use the microphone and speaker to recognize the room type and a camera to recognize the age, gender, and the number of people inside a room. This information is vital for blind people to know; otherwise, it is challenging for the blind to perceive the intended person to communicate their message. We use the information regarding room type and occupants inside a room and convey it to a blind person so that blind user can communicate their message to the intended person without cognitive overload. The aim of the research work is the cognitive assistance of blind people regarding the ambient environment so that the blind can behave gracefully and communicate their message to the intended person. The objectives of this proposed work are as follows:

- (i) To assist blind users in identifying the type of room using smartphone-based acoustic fingerprinting
- (ii) To identify the layout, arrangement of objects, and social status (in terms of age, gender, and the number of people) using the smartphone
- (iii) To infer the appropriate person for communication per their social status through the classification method

The proposed work offers cognitive assistance to blind people in room type recognition and awareness regarding its occupants to enable blind users to navigate independently and carry the message appropriately. The available solutions are limited due to either additional hardware, customized infrastructure, or both. Unlike these solutions, it is required to develop a solution that can understand the type, nature, and layout of a room in an academic building by taking the door as a frame of reference and identifying the room type and human occupants to develop a simple blueprint of the room by understanding the social status of the room and communicate it to the blind. Such a solution should be cheap and socially acceptable so that a blind user, while standing at the door, easily understands the room type and identify its social status by finding the number, age, and gender and adapting their behavior accordingly. The solution should also assist the user in identifying the candidate person (e.g., the teacher in the classroom, chairperson in a meeting, and officer in an office) to talk to or convey their message and navigate properly without a human guide.

The rest of the paper is organized into five sections. Section 2 discusses literature regarding the challenges faced by a blind person in everyday life and solutions both manual and technology-based to cope with them. Section 3 consists of a methodology and different steps involved in the proposed solution. Section 4 presents the results and discussion and analysis of the proposed solution. Section 5 concludes the paper along with a discussion on future research work.

2. Literature Review

Blind people are facing multiple challenges in performing day-to-day activities. These challenges range from localization to navigation and visual information to social interaction [17]. While some challenges are not well understood, a study is conducted to find the pictorial questions to which a blind person would want answers [24, 25]. The sighted people avoid hazards and curiosities of the world through vision, spend leisure time playing video games, climbing into the rock, social gathering, and chatting with friends; exclusion of blind people as a participant leads them to discrimination and social avoidance. The situation aggravates when blind people are cut off from commerce and business opportunities, leading to incompetence and inability². Social interaction is mostly based on vision, like exchanging services and goods, and information is primarily available to the eye, which excludes the blind from the social interaction. Moreover, blindness, in any sense, causes significant complications for blind people in performing daily activities, such as mobility, awareness of surroundings, and interpersonal communication. The prominent challenges blind people face in daily life activities can be broadly categorized as follows.

The arrangements and settings of objects, navigating to a specific point in indoor as well as in outdoor environments is one of the prominent challenges a blind person faces. Blind people memorize the arrangement of objects, settings, and properties in the indoor environment [26]. Blind people are unable to navigate independently outdoor and in unknown places. There are obstacles, moveable and unmovable, like ditches, maintenance holes, drains, vehicles, roads, and so on, which may cause severe injury or even death. Restricting the movement of blind people can cause social isolation, unemployment, and psychological alienation (<http://wafthb.net/blindness-challenge-and-achievement>).

2.1. Human and Animal Guides. Although a human guide is a very simple technique, it is expensive assistive aid used by blind people for guidance in mobility. A human guide or sighted guide is a person who helps a blind partner in navigation [27]. The dog guide technique is less popular among blind people, and only 2% of blind people use this technique. The dog needs to be trained carefully for this purpose [28]. The blind user commands, and the dog obeys accordingly. The guide dog assists the blind individual in obstacle detection and wayfinding.

2.2. Long White Cane. The long white cane is mostly used as a blind assistive technique for mobility and obstacle detection. The blind individuals move long white cane left and

right and back and forth around their bodies, synchronized steps to detect obstacles in front and surface change. Further, a white cane enables the blind user to detect elevation (https://www.apsguide.org/chapter2_travel.cfm) and stairs. However, white cane fails to detect obstacles at head level, for example, hanging objects and open windows [29]. The Virtual White Cane (VWC) [12] uses the camera to capture the reflected beam to compute the distance from the obstacle and communicate it to the blind user through vibration patterns. The magnitude of vibration gives a rough measure of the distance. However, the solution could not give information about the geometry of the obstacle.

The ioCane [13] is a sensory device mounted in the middle of the white cane to send contextual data wirelessly to the smartphone. The ultrasonic sensor is used for distance estimation, and obstacle detection is used as a processing and output device [30, 31]. However, it uses an extra ultrasonic sensor for obstacle detection and cannot recognize the obstacle in front.

2.3. Obstacle Detection. Rahman et al. [32] use the already stored exclusive ground images for reference in the indoor environment. The image taken is compared with prestored pictures via an algorithm, which calculates the Mean Square Error (MSE) and then compares it with a threshold value α . If the value is less than α , there is no obstacle; otherwise, there are two possibilities: either floor is changed, or an obstacle is detected. However, the value of α is taken heuristically. It provides no method to handle the obstacle detected.

Yi et al. [33] used a network of cameras for object detection in the indoor environment. The cameras installed at important points of a building identify the nearest objects of daily use. When a blind user requests an item, the cameras capture and trigger image recognition. The Speeded-Up Robust Feature (SURF) and the Scale Invariant Feature Transformation (SIFT) [33] techniques extract features and detect the object. Each camera recognizes the object and sends the report to the host. The host collects and analyzes the data and reports the nearest location to the item. Moreover, it needs networks of cameras, which is expensive.

2.4. Navigation. Ding et al. [34] proposed a solution using RFID technology with wireless mobile communication to assist blind people in walking along the road and finding the best route and bus station. The RFID reader and antenna mounted in a walking cane, which uses the built-in RFID labels beneath the road, provide a virtual blind road. The GPS provides localization with high precision and is a universally accepted system for localization and navigation [35]. The integration of real-time GPS with a customized tactile-foot unit fixed in the shoe assists the blind user in safe navigation in the outdoor environment [36]. The experimental results show that the blind user recognizes the tactile-foot feedback with high precision. Khusro et al. [37] proposed a real-time feedback system specifically designed for blind people to convey information to them based on vibration patterns. The proposed solution has been evaluated

through an empirical study by collecting data from 24 blind people through a mixed-mode survey using a questionnaire. Results show the average recognition accuracy for 10 different vibration patterns as 90%, 82%, 75%, 87%, 65%, and 70%.

ARIANNA [38] is an indoor assistant in navigation system based on a customized infrastructure like specialized painting on the floor, stickers, or barcodes being recognized through the smartphone application. However, it uses customized infrastructure.

Khan et al. [39] proposed a smartphone-based navigation solution for blind users, operating in sighted mode and blind mode. The sighted user enters references about infrastructure in a web server. The blind user uses this information for navigation and object recognition. It uses a tagged door, color, pattern, or building geometry for indoor navigation. Although it contributes to navigation in an indoor environment, it requires a customized infrastructure and a web server, making it an expensive option [40].

2.5. Face Recognition. The proposed solution [41] uses a neural network-based (CNN) technique to detect straight and front views of the human face in grayscale pictures. The portion of the input image is processed through one or more CNN-based algorithms directly, and then the results of these CNN-based algorithms are arbitrated. Each CNN-based network is trained to generate the input image's result (presence or absence of a human face). The system arbitrates in the networks to improve performance and accuracy. Similarly, V. Vpnik and his team (AT&T Bell Labs) developed a deep learning technique called Support Vector Machine (SVMs) [42]. The proposed system [42] detects human faces by thoroughly scanning an image for human face-like patterns at many scales, dividing the original image into partly covered images, and categorizing them through a Support Vector Machine to find a specific class. Moreover, this solution investigates the application of SVMs in computer vision, like a training dataset of 50,000 points to discriminate the face and non-face patterns. The SVM can train deep learning networks (i.e., neural networks, Radial Basis Function classifiers, etc.).

The Local Binary Pattern (LBP) is one of the simplest appearance-based feature extraction methods [43]. This method uses neighboring pixels to carry out computation on them. The proposed solution [44] converts the image into a greyscale, detects the face, and extracts the facial points through the Local Binary Pattern (LBP) and Active Shape Model (ASM). The LBP and LDP reduce the dimensionality, and at the final stage, classification is performed through the Support Vector Machine. The accuracy of the face detection improved from 92.8% to 94.5%.

2.6. Gender and Age Recognition. The research work [45] proposed video-based age and gender estimation by combining different classifiers. The results obtained from the different classifier is combined and generated single output. The solution used the Age net and Gender net models and a deep VGG-16 neural network trained for age/gender

prediction [45]. The gait-based gender recognition system used the gait energy image (GEI) feature as the main feature for gender recognition. The proposed system [46] captures human images (still and moving) and extracts the moving human silhouettes by thresholding and subtracting the background. Then, the human information is extracted and used to improve appearance-based gender recognition. The gender recognition system [47] fused different face recognition features such as facial texture, hair geometry, and mustache feature. In the first phase, the facial and hair features are extracted through the AdaBoost algorithm, which roughly classifies the image into males and females.

The age estimation is more complicated than gender classification because different people age in different ways [48]. This solution provides an automatic age guesstimate technique called AGES (Aging Pattern Subspace). The main idea is to model the pattern for age estimation, which is restricted to classifying a specific person's face images organized in time order by creating a demonstrative subspace.

2.7. Counting People in an Indoor Environment. Although counting the number of people in a crowd saw many advances, the solutions have some limitations: the background must not be complicated, the crowd in the crowd must not be stationary, and the image resolution must not be below. The proposed solution [49] claims to develop a method for guessing the total number of persons in a low-resolution image with complex background. The solution is based on three steps: The first background is subtracted in a complicated scene (moving people). Second, a method has been developed (Expectation-Maximization, EM) to locate individuals in low-quality scenes. The method used a new cluster model to identify each individual in a complex scene and does not require precise front contour. The method is validated using a 4-hour video with 10% errors.

2.8. Understanding Room Layouts. Jeon et al. [50] developed a prototype model for understanding the unknown room layout; objects of interest like chairs, desks, tables, computers, microwave, switch, door, trash can, and window have been identified to understand the unknown room layout. Moreover, the sounds of the refrigerator, computer, and other home appliances are considered. The Quick Response (QR) code was assigned to every item in the room. The proposed solution has two modes, namely, overview and detail. The overview mode scans the room from left to right and saves the scanned data with a specific voice command, date, and time. In the detailed mode, objects are recognized by keeping the camera in an interesting direction. The identified objects resulted in voice messages.

2.9. Room Type Recognition. Rossi et al. [51] use smartphone sensors (i.e., microphone and speaker) to sense the room type in the indoor environment. This solution is based on active sound probing. A human chirp is generated from the speaker, and impulse response from the room is recorded through the microphone. The captured patterns are used for

generating acoustic fingerprinting of the rooms. Acoustic fingerprinting is further used for training the Support Vector Machine (SVM) [51]. This system used audible sound, which is susceptible to ambient sound. The same approach was adopted by Song et al. [52] with cloud service and RESTful Application Program Interface (API). The proposed solution uses a speaker and microphone of a smartphone for room recognition. Moreover, it creates an inaudible chirp and records the acoustic fingerprinting for minimal time to avoid privacy issues in the room. However, the performance decreases in a crowd environment.

2.10. Social Interaction. Social interaction is mostly based on nonverbal communication, including facial expression, head nodding, eye contact, and body language [53]. These nonverbal cues can only be visible to sighted people, while blind people are unable to catch nonverbal cues, leading blind people to discrimination and social segregation. Interpersonal communication is verbal or nonverbal; almost 65% of communication is nonverbal [3]. Nonverbal communication (facial expression, eye contact, and bodily gestures) plays a vital role in social interaction. Blind people are unable to access nonverbal communication, leading to social segregation and discrimination.

VizWiz Social is an iPhone application used to interact with the crowd. The blind person takes the picture and asks about the picture and gets back the answer from the crowd. Brady et al. [24] categorize the 4000 questions asked through VizWiz Social into primary and subcategories. The primary category consists of animals/persons, settings, and objects.

Philips and Proulx [53] designed a prototype of a multimodal assistive system that contains two functions, namely, the ID and GAZE. However, the prototype uses customized hardware for social interaction, which is difficult to move with blind users everywhere. Similarly, Panchanathan et al. [54] proposed a Social Interaction Assistant (SIA) to enrich the nonverbal communication of visually impaired people, which is based on a pair of glasses with a camera embedded. The camera records the video stream, analyzes it through machine learning and computer vision, and extracts nonverbal cues like facial expression, head nodding, and body gestures. However, the system uses customized hardware, which is less efficient and uneconomical.

For improved social interaction, Sarfraz et al. [55] proposed a prototype multimodal assistive system consisting of a PC, camera, headphone, and vibrotactile built; all these components are wirelessly connected to the PC. The proposed system performs two important tasks: (1) Detecting who is in the camera frame, the number of people, their names, and relative position to the camera. (2) The system informs the user who is looking towards you. And his name and position through audio-haptic feedback. Moreover, the system performs two functions: ID and the GAZE: the ID function tells the user about the number of people in the room while the GAZE function automatically activates when someone looks towards the user via audio feedback. However, the system uses customized hardware for improved social interaction. Most blind assistive technologies focus on

localization, navigation, and obstacle detection and recognition. The solutions described above are limited due to the use of additional hardware and customized building. However, these solutions are either expensive, difficult to use, need customized infrastructure, are socially unacceptable, or do not accurately meet the intended purpose.

3. Proposed Methodology

The proposed solution integrates two approaches. The first approach used the audience dataset and CAFFE model to recognize faces and estimate age group, gender, and the number of people, and approach two used acoustic/reverberation of room to guess the room type [52]. However, we created our dataset, trained the model on the dataset, and developed a classifier for classification. After that, we drew inferences and communicated to the blind. The architecture diagram consists of three layers, namely, the data acquisition layer, application layer, and inferencing layer, as shown in Figure 1. In the data acquisition layer, there are smartphone sensors (i.e., speaker, microphone, and camera capture sensory data). The sensory data from the camera is required for module-1 and data from the microphone and speaker is required for module-2.

In the data processing layer, data acquired from smartphone sensors (i.e., camera and microphone) and processed in two modules; module-1 needs visual data for age, gender, and the number of people recognition, and module-2 needs audio/impulse response for room type identification.

3.1. Recognition of Age, Gender, and Number of People.

Module-1 is responsible for scanning the room from left to right and recording the video stream. The recorded video stream is further processed by filtering. The required frames are separated from extra frames. The required frames (images) are further passed to the classifier. The classifier converts the image into an appropriate format and further passes it to the trained model. After processing, the trained model returns probabilistic results in terms of age, gender, and the total number of people inside the room.

For face detection and recognition, we use the OpenCV library. The OpenCV detects faces using facial landmarks. Facial landmarks are used to identify and signify the most important sections of the human face in the image. The facial landmarks used by OpenCV consist of eyes, eyebrows, nose, mouth, jawline, and so on, as shown in Figure 2.

This module scans the occupants in the room via a smartphone camera. The recorded video is passed to a deep learning framework for feature extraction and classification. For this purpose, we use a deep learning framework called CAFFE (<https://caffe.berkeleyvision.org/>). The CAFFE model has two associated files: “.pretext” and “.caffemodel”. The pretext file contains CNN definition in terms of layers, input, output, and functionality, while the Caffe model file consists of information on the trained model. In our case, we use two trained Caffe models one for age estimation (age_net.caffe model) and one for gender

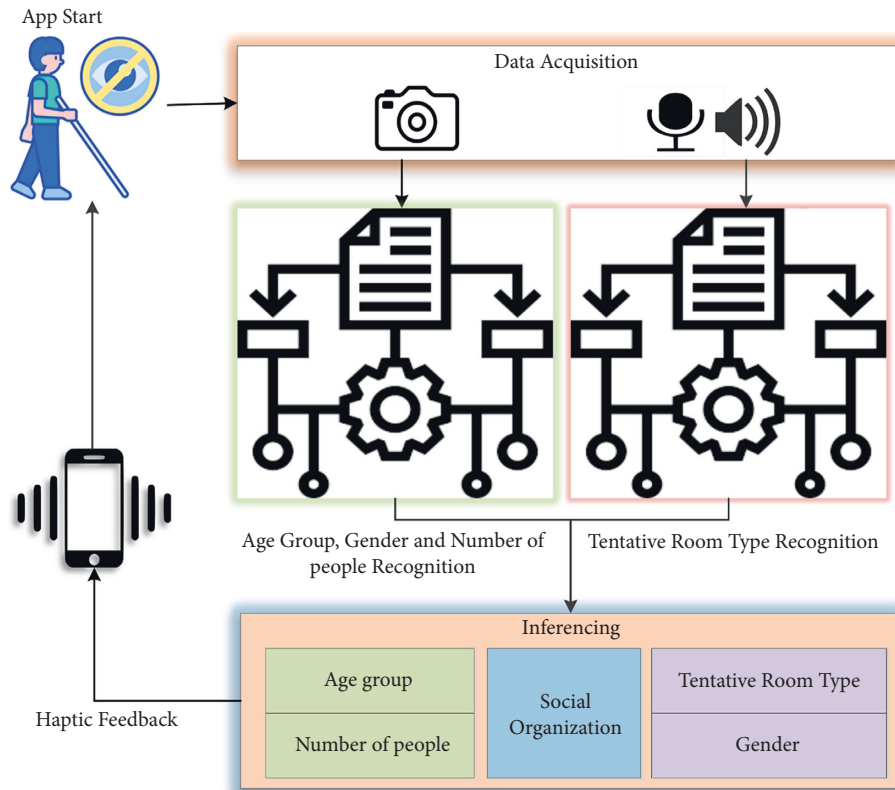


FIGURE 1: A schematic diagram of the proposed solution.

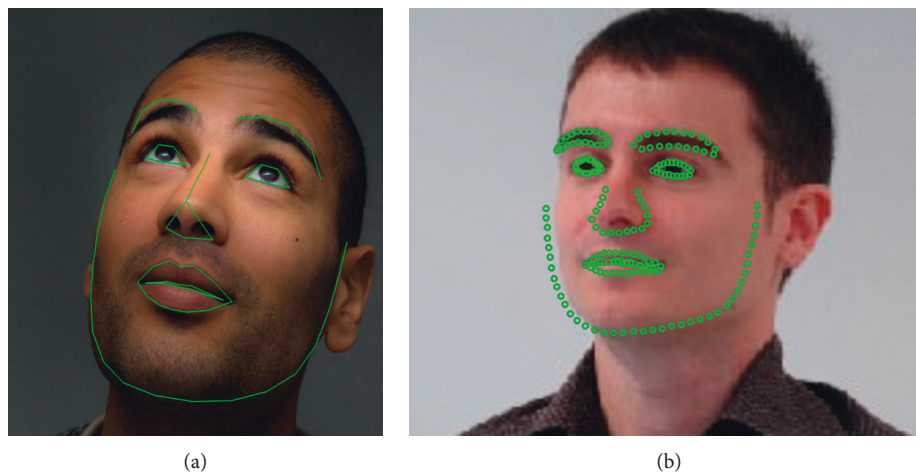


FIGURE 2: Facial landmarks, important facial features in an image (source (<https://blog.dlib.net/2014/08/real-time-face-pose-estimation.html>)).

recognition(gender_net.caffemodel). These models are trained with a dataset (Audience (<https://www.kaggle.com/ttungl/adience-benchmark-gender-and-age-classification>) Images Dataset) available online. The output from the feature extractor is then passed to OpenCV (<https://opencv.org/>) for analyzing and considering different features, and CNN predicts the possible output. This real-time output is then checked against the pretrained model for appropriate prediction. The predicted age return for further processing. The gender and age recognition are almost similar except that the matching pretext file and the Caffe model files.

Moreover, the CNN's prediction layer (output layer) consists of 8 categories of age groups. The classes for age groups are as follows: ("0-2", "4-6", "8-13", "15-20", "25-32", "38-43", "48-53" and "60-"). For example, the input of age range 1-12 is labeled with child, 12-18 is labeled with teenager, and 18- n is labeled with an adult.

3.2. Room Type Recognition. When a blind user enters a room, module-2 recognizes the type of room through echolocation. This module uses smartphone-based sensory

data (i.e., microphone and speaker). A chirp of a human is generated through the speaker and recorded back through a microphone in a specific format (wav) for further processing. The recording of the impulse response is kept for one second, as the echo/impulse response vanishes after 100 milliseconds. So, we record both the chirp generated from the speaker and the impulse response of the room and then covert room acoustics into a spectrogram. We convert audio/wav to spectrogram because it gives the best result while training a model through CNN [52].

The spectrogram is an image-like visual representation of audio. It is a 2-dimensional graph with one additional dimension, represented through color, which is the intensity (amplitudes) of the signal over time. The x -axis is the time while the y -axis represents frequency. The red color represents the intensity of the sound in terms of frequency, as shown in Figure 3. The audio/echo is converted to spectrogram using an online service (<https://academo.org/articles/spectrogram/>).

For training the model on the dataset, we used an online model training service provided by a teachable machine (<https://teachablemachine.withgoogle.com/train/image>) the dataset contains a spectrogram of different room types fed into the Convolutional Neural Network (CNN), train the model, and export the trained models TensorFlow lite. After downloading the model, we load the trained model in android. For coding, we use Android Studio 3.5.3 and Java 8 (64 bits) as editor and version, respectively. For classification, the image classifier is used to feed the spectrogram to the trained model and the model returns the probability of different room types. Figure 4 demonstrates the whole process of classification.

3.3. Inferencing. In module-3, inferencing regarding the social status of the room based on the output of module-1 and module-2 is drawn. Module-1 gives information about the total number of people, their age group, and gender, while module-2 gives a rough estimation of room type. Although it is hard to predict the exact age of a person by looking at their face, we divide age into different age ranges such as age range 1–12 is labeled as a child, 12–18 teenager, 18–33 adult, and 33– n old. We will create different scenarios in terms of age and gender and draw inferences. These inferences are based on the results of module-1 and module-2.

- (i) *Scenario-1.* If the age group detected is 1–12, the number of faces is one or greater, and the gender is male or female, then most probably, the setting is a residential room and family environment. Blind people may behave per the situation.
- (ii) *Scenario-2.* If faces detected are more than 10 and the age groups detected are 18–33 and 33– n and gender is either male or female, then inferences may be made as classroom environment and blind people may convey a message or talk to a person with age group 33– n because, in the classroom environment, an older person is more appropriate to talk or convey a message.

- (iii) *Scenario-3.* When faces detected are between 1 and 5, and the age group detected are 18–33 and 33– n , then most probably the environment is an office and probably, the right person is the elder among the detected faces.

Similarly, inferences may be made regarding the social interaction of people. For example, meeting with an elder person or teacher may need to behave more politely than meeting with friends, and communication with a child needs more caring behavior than an adult. Moreover, gender variance may be considered vital because of communication with a female desire more graciousness than male in some societies.

4. Results and Evaluations

We have evaluated our app using different evaluation metrics through physical experiments. These experiments are performed in different rooms (rooms selected as per availability) of the Department of Computer Science, University of Peshawar. We have also conducted an extensive user study to further test the performance of our app. To assist blind users in identifying the social status of the people nearby them accurately and help them in their challenging life by overcoming the cognitive overload, we evaluated our smartphone app by comparing the results from our app with the dataset using different evaluation metrics and keep care of the accuracy of the system, to check whether or not we get the desired result. We also identified some important factors for room type recognition that affect performance. We also conducted extensive user studies to further understand the utility of the proposed solution in reducing the cognitive overload regarding understanding the type, nature, and social status of the room for ease of navigation and communication.

4.1. Vulnerability to Ambient Sounds. We have evaluated our app and observed room type recognition vulnerability to ambient sounds. This experiment was conducted in 11 rooms. We choose three faculty offices, three laboratories, four classrooms, and one library as listed in Table 1.

During sample collection and training data, we keep the rooms quiet. When we test our app recognition precision in both cases: absence of ambient sounds or quiet, and presence of ambient sounds or play music via a laptop in the tested rooms. The respective average room type recognition accuracy is 87.47% and 44.74%. Table 1 indicates that the average accuracy falls (almost 50%) in the presence of ambient sound.

4.2. Impact of Still and Movement. Our app's performance is evaluated through an experiment conducted in the laboratory; the description of the room is shown in Table 1 to analyze the impact of the movement of people on the proposed solution. We collect datasets for training and testing from the same spot in the laboratory. During the collection of data, we keep the room quiet and empty. We

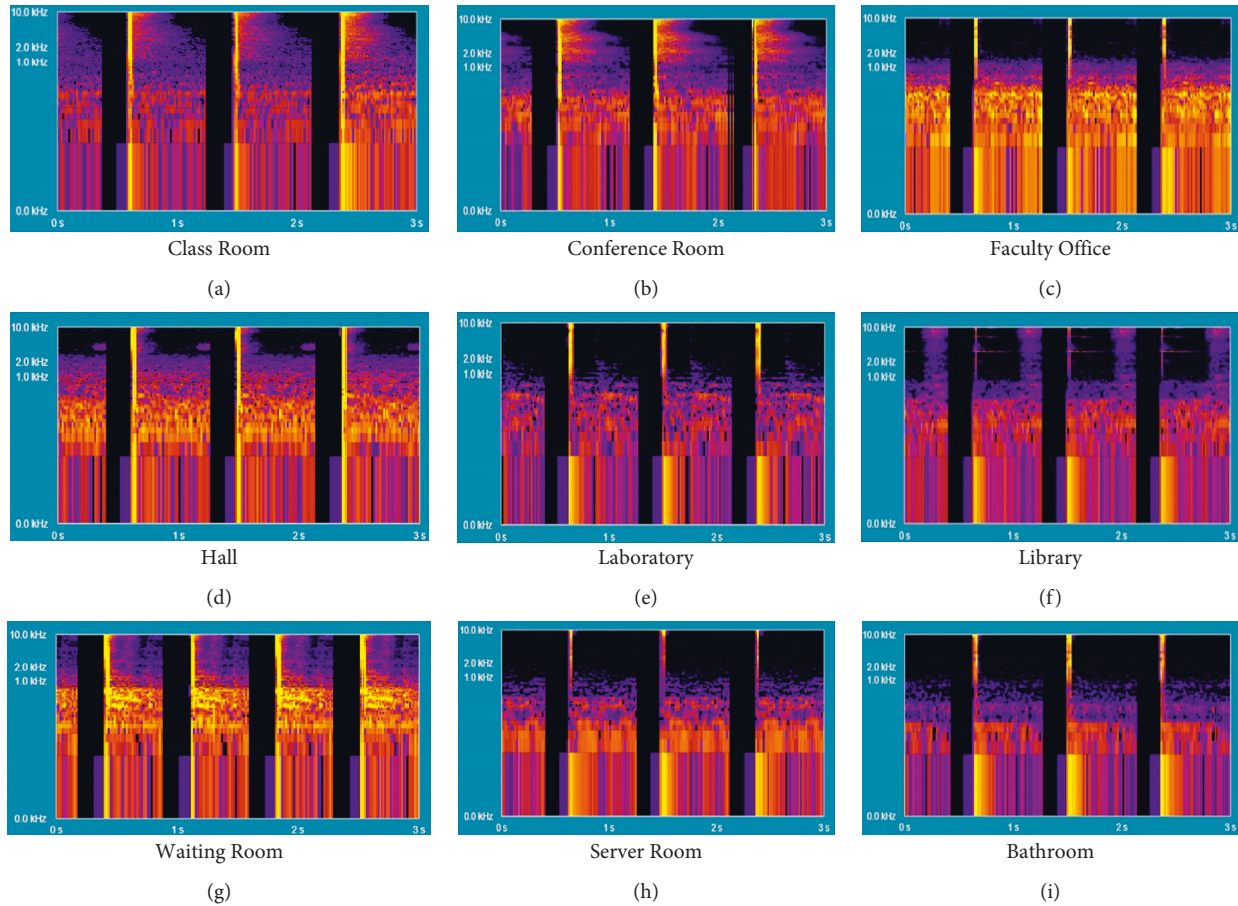


FIGURE 3: Spectrograms of room type in the academic building of C. S (UoP). (a) Classroom, (b) conference room, (c) faculty office, (d) hall, (e) laboratory, (f) library, (g) waiting room, (h) server room, and (i) bathroom.

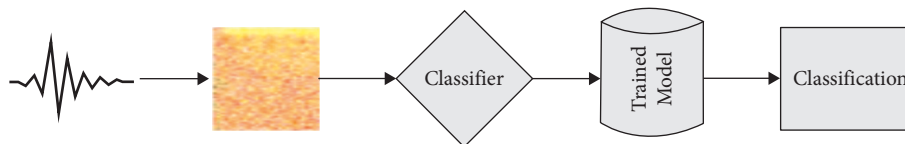


FIGURE 4: Wav to spectrogram and feed to trained model for classification.

performed our test when a number of our fellows were invited to walk in the laboratory. Figure 5 illustrates that the accuracy rate (%) against the number of walking individuals decreases with the increase in the number of walking people in the room. The room's response to the chirp changes because of the reflection of the acoustic signals from walking people.

4.3. Effect of Changes in Furniture Layouts. The changing position of furniture impacts the performance of room type recognition. For this purpose, we analyze a room having movable furniture. Figures 6(a)–6(d) represent the original layout, removed chairs, randomly arranged chairs, and moved the table and chairs position, respectively. We collected the training data set for this in the original setting, as shown in Figure 6(a). In the testing phase, we changed the

setting as shown in Figures 6(b)–6(d) and observed the probabilistic results as 84.21%, 83.98%, and 81.23%, respectively. The percentage performance recorded in each furniture set is graphically represented in Figure 7. It is observed that changes in the position of furniture in a room may reduce the correctness of room type recognition. It is because of the changes of reflection of acoustic signal of room, to the response of the chirp.

4.4. Assessment in Similar Rooms. We conducted another experiment to test the performance of room type recognition in almost similar rooms. For this purpose, we select four classrooms in the Department of the Computer Science University of Peshawar. The selected rooms comprised almost similar furniture, layouts, and sizes. The settings of 4 similar rooms are shown in Figures 8(a)–8(d). We collect

TABLE 1: Description of absence and presence of ambient sound.

Room type	Number of times	Absence of ambient sound (%)	Presence of ambient sound (%)
Faculty Office 1	127	90.55	68.50
Faculty Office 2	147	84.35	54.42
Faculty Office 3	137	79.56	52.55
Laboratory 1	143	83.92	37.06
Laboratory 2	105	78.1	45.71
Laboratory 3	133	89.47	42.11
Classroom 1	111	90.99	45.05
Classroom 2	134	91.04	29.10
Classroom 3	128	85.16	37.50
Classroom 4	145	95.86	42.76
Laboratory	150	93.12	37.33
Total	1460	87.47	44.74

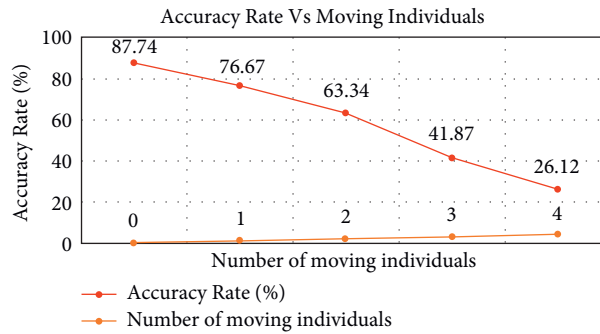


FIGURE 5: Accuracy rate (%) versus the number of walking people.

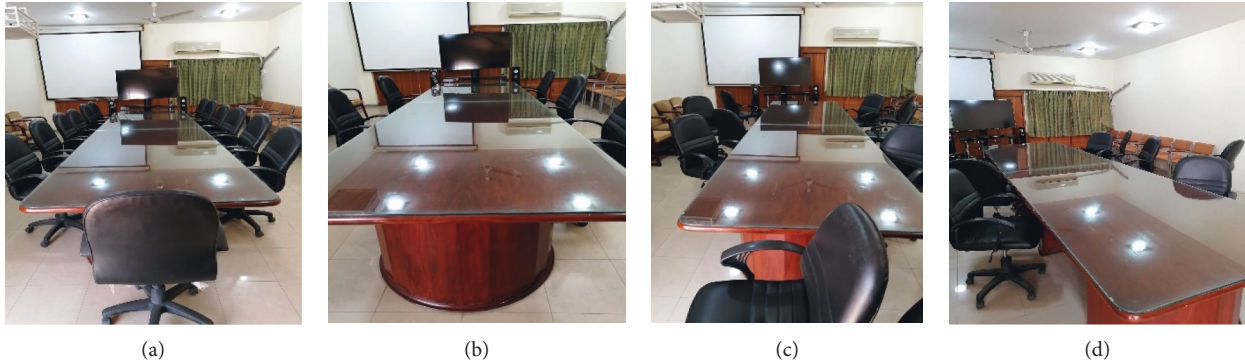


FIGURE 6: Furniture changes in conference room. (a) Original layout, (b) removed chairs, (c) randomly arranged chairs, (d) moved table and chairs.

500 samples, one spot in each of the four rooms, for training and testing. The average accuracy of 4 similar rooms was observed as 86.32%. The performance in each room is graphically represented in Figures 9(a) and 9(b). It is observed that the higher similarity in rooms may drop the average precision rate.

4.5. *Empirical Study.* This section conducted an extensive study of the users to test the App. The app is tested 3524 times by five blind users in 11 different rooms. The details of blind users in terms of their age group, gender, educational background, and smartphone usage experience are shown in Table 2. The blind users choose on their willingness and

availability. Each of the blind users is briefed about the experiment and given essential training about the usage of the app. During the experiment, the blind user having smartphone usage experiment and educated noted a higher precision rate (%) compared with illiterate and having lesser smartphone usage experience.

Each of the blind users tested the app between 50 and 100 times randomly in each room. The average percentage accuracy of the proposed system is noted as 87.64%.

Similarly, the app screenshots shown in Figure 7 are tested in different academic building rooms. These rooms are chosen as per the availability. The app is tested 100 times in each room. The percentage average accuracy for age group, gender, and the number of people were noted as

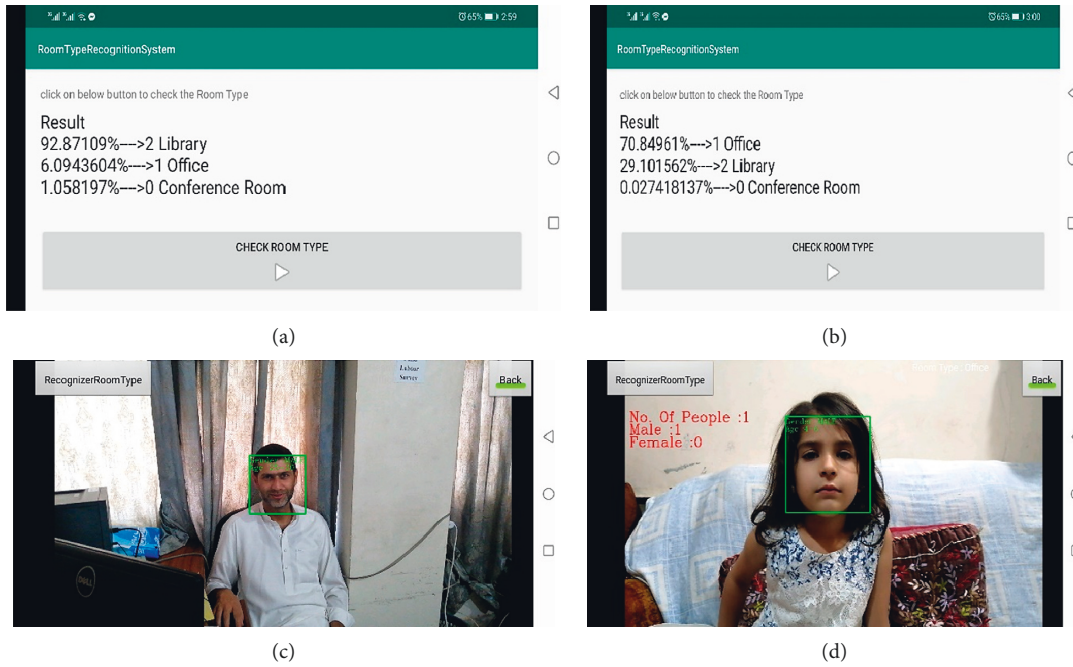


FIGURE 7: Different screenshots of the app. (a) App result in library, (b) app result in office, (c) male aged 25–30 detected, and (d) female of age 4–6 detected.

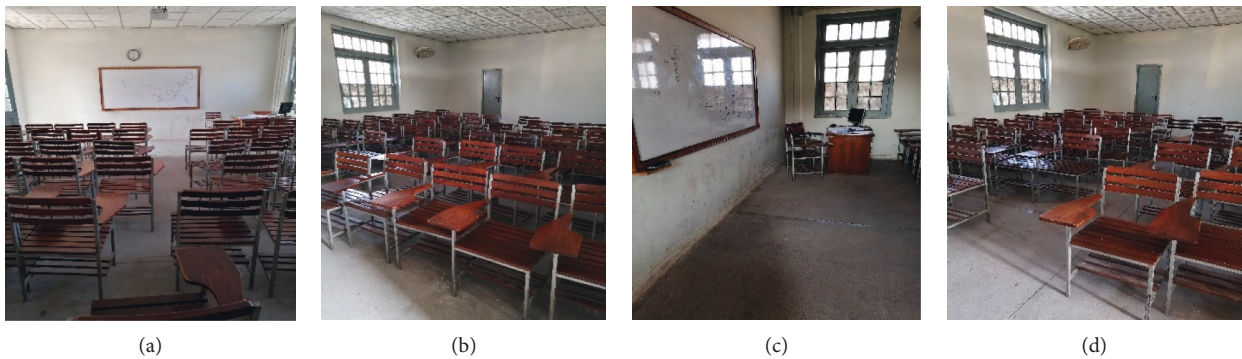


FIGURE 8: Similar classrooms. (a) Classroom1, (b) classroom2, (c) classroom3, and (d) classroom4.

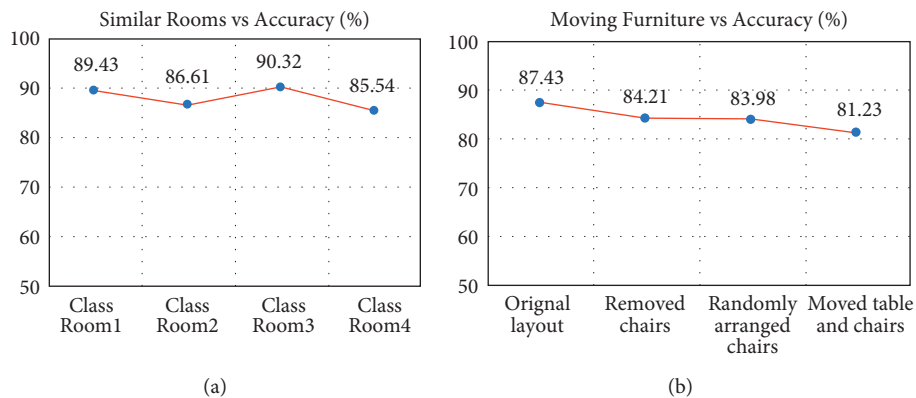


FIGURE 9: (a) Four almost similar rooms. (b) Different furniture settings.

TABLE 2: Description of a blind user.

Blind user	Age	Gender	Background	Smartphone usage experience
User 1	22–33	Male	Educated	3 months
User 2	34–40	Male	Educated	6 months
User 3	41–45	Male	Literate	2 months
User 4	46–50	Male	Illiterate	1 month
User 5	22–33	Male	Literate	3 months

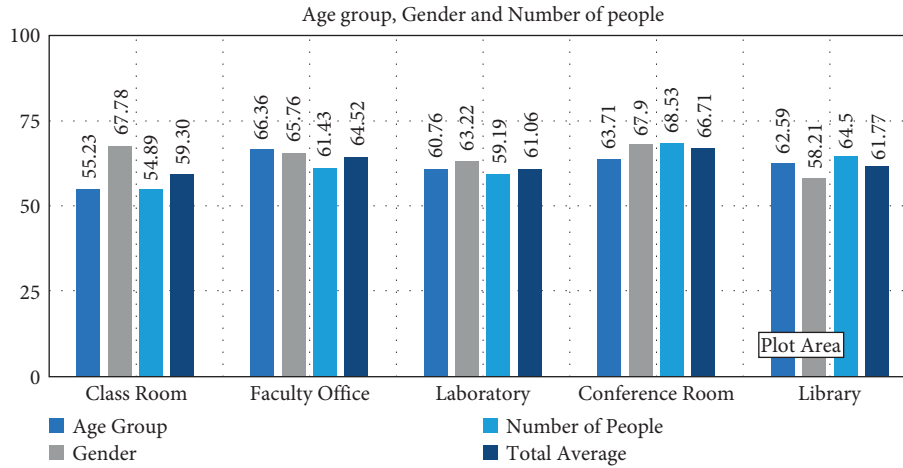


FIGURE 10: Accuracy (%) in terms of age, gender, and number of people.

61.73%, 64.57%, and 61.71%, respectively. The total average accuracy recorded is 62.67%.

Figure 10 graphically represents the accuracy rate (%) in six rooms (chosen as per availability), such as the classroom, faculty office, laboratory, conference room, and library. Each room is scanned 100 times. The percentage average accuracy was recorded as 62.67%.

To summarize this section, we conducted different experiments in different rooms of academic buildings and evaluated our app from a different perspective using different evaluation metrics. Furthermore, to make it more accurate and efficient for blind users, we also conducted a general user study and found the results satisfactory. The next section will conclude this research and provide guidelines for future work.

5. Discussion

The deficiency of vision is a misfortune and imposes several constraints upon specific movements in every single part of living. Therefore, it is vital to explore techniques and approaches for supporting visually impaired people to give them appropriate assistance and a sufficient chance to live. They are facing enormous challenges, but the most prominent challenges are localization, navigation, and communication with other humans per the social status of the ambient environment. Some of these challenges like navigation and localization are mitigated through assistive technology such as white cane, technology-assisted white cane, Global Positioning System (GPS) for outdoor navigation, and so on while assisting blind people in room

recognition and communication to the intended person is not addressed by the traditional solution. Therefore, to assist the blind user in room type recognition, we proposed a solution that provides cognitive assistance to blind people in room type recognition and awareness of the surrounding in terms of the number of people, gender, and age group in an indoor environment. We focus on the ubiquitous smartphone because the widespread current research uses smartphones as a research platform for various surveys and evaluations. The powerful sensors make it the best choice for the assistance of blind users. To achieve the aims and objectives, we divide the proposed system into three modules: Module-1 uses a camera to understand the social status and ambient environment by exploiting the deep learning and coffee model for face detection, age group estimation, and gender recognition. Module-2 uses a microphone and speaker for room type recognition by generating a chirp of a specific frequency from the speaker and recording the echoes of rooms. To achieve the purpose, we create a dataset of different rooms' acoustics and further convert it into a spectrogram. The model is trained on a spectrogram dataset of different rooms. In module-3, we draw inferences based on module-1 and module-2. The blind user is not only conveyed information regarding the room type but is also aware of the social status of the room through haptic feedback. Moreover, we evaluate the proposed solution using different evaluation metrics and conduct an extensive user study. The experimental result achieved from the study is satisfactory and may further be investigated. However, an image-based approach may lead to privacy issues. The alternative approaches may be used for estimating occupancy

in a room. Similar is the case for gender and age-group recognition. The major limitation of the acoustic method is the susceptibility of the acoustic signal that may be easily disturbed by several outside factors. Scalability is another challenge that may be solved by presenting different (potentially radio frequency-based) approaches. The experimental results indicate that the movement of people, changes in furniture setting, and ambient sounds significantly decrease the performance of the solution. The proposed solution is tested indoors and does not work outdoor. A solution can be proposed that enables blind people to be aware of ambient surroundings both indoors and outdoors.

6. Conclusion and Future Work

To provide cognitive assistance to blind people in predicting room types and finding the intended person to communicate their message by understanding the ambient environment and social status of the people in terms of their age, gender, and the number of people inside a room, this study proposed a solution that used the microphone and speaker for recognizing room types and the camera for understanding the social status and ambient environment. We analyzed different evaluation metrics, including the movement of people, ambient sounds, orientation, and position, and conducted an extensive user study to validate the proposed solution in real-world scenarios and achieved very good accuracies. The proposed solution considers the people inside a room. The furniture and objects, such as computer, table, chair, printer, whiteboard, desk, inside a room, are also vital for further improvement in the proposed system. Another important thing is the standing and sitting position of people inside a room; for example, a person standing in a classroom is probably the right person to whom a blind may convey his message. The proposed solution uses image processing for counting the number of people, age group, and gender. However, scanning a room may lead to privacy issues. The non-image-based solutions may solve the privacy issue of counting the number of people in the indoor environment [56]. In the future, we aim to extend this work to controlled outdoor environments, such as natural scenes.

Data Availability

The data that support the findings of this study are available upon request from the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

References

- [1] A. Khan, S. Khusro, B. Niazi, J. Ahmad, I. Alam, and I. Khan, "TetraMail: a useable email client for blind people," *Universal Access in the Information Society*, vol. 19, no. 1, pp. 113–132, 2020.
- [2] Who, "World Sight Day," *South African Ophthalmology Journal*, vol. 14, pp. 56–58, 2019.
- [3] N. Ambady and R. Rosenthal, "Thin slices of expressive behavior as predictors of interpersonal consequences: a meta-analysis," *Psychological Bulletin*, vol. 111, no. 2, pp. 256–274, 1992.
- [4] S. Qiu, J. Hu, and M. Rauterberg, "Nonverbal signals for face-to-face communication between the blind and the sighted," in *Proceedings of the International Conference on Enabling Access for Persons with Visual Impairment*, pp. 157–165, Athens, Greece, February 2015.
- [5] D. Mandru, I. Lungu, A. Mociran, and O. Tătar, "Development of a mechatronic blind stick," *Fascicle of Management and Technological Engineering*, vol. 6, pp. 797–802, 2007.
- [6] M. A. Hersh and M. A. Johnson, "Disability and assistive technology systems," in *Assistive Technology for Visually Impaired and Blind People*, pp. 1–50, Springer, Berlin, 2008.
- [7] N. A. Giudice and G. E. Legge, "Blind navigation and the role of technology," *The Engineering Handbook of Smart Technology for Aging, Disability, and Independence*, vol. 8, pp. 479–500, 2008.
- [8] C. M. Covarrubias, *Exploration of Quality of Life with the Use of Assistive Technology for Individuals that Are Visually Impaired or Blind*, The Chicago School of Professional Psychology, Chicago, IL, USA, 2022.
- [9] J. B. Lanbon, K. C. S. Lee, and S. Yan-Li, "Enhancing effective teaching and learning of ICT in the schools for the blind in Ghana," *International Journal of Web-Based Learning and Teaching Technologies*, vol. 17, no. 3, pp. 1–12, 2022.
- [10] I. Khan, S. Khusro, and I. Ullah, "Technology-assisted white cane: evaluation and future directions," *PeerJ*, vol. 6, Article ID e6058, 2018.
- [11] N. Tyagi, D. Sharma, J. Singh, B. Sharma, and S. Narang, "Assistive navigation system for visually impaired and blind people: a review," in *Proceedings of the 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV)*, pp. 1–5, Gujarat, India, September 2021.
- [12] P. Vera, D. Zenteno, and J. Salas, "A smartphone-based virtual white cane," *Pattern Analysis & Applications*, vol. 17, no. 3, pp. 623–632, 2014.
- [13] B. Leduc-Mills, H. Profita, S. Bharadwaj, and P. Cromer, *ioCane: A Smart-Phone and Sensor-Augmented Mobility Aid for the Blind*, 2013.
- [14] K. Yelamarthi, D. Haas, D. Nielsen, and S. Mothersell, "RFID and GPS integrated navigation system for the visually impaired," in *Proceedings of the 2010 53rd IEEE International Midwest Symposium on Circuits and Systems*, pp. 1149–1152, Washington, DC, USA, August 2010.
- [15] L. Nieto, C. Padilla, and M. Barrios, "Design and implementation of electronic aid to blind's cane," in *Proceedings of the 2014 III International congress of Engineering Mechatronics and Automation (CIIMA)*, pp. 1–4, Colombia, October 2014.
- [16] W. Jeamwattananachai, M. Wald, and G. Wills, "Indoor navigation by blind people: behaviors and challenges in unfamiliar spaces and buildings," *British Journal of Visual Impairment*, vol. 37, no. 2, pp. 140–153, 2019.
- [17] B. Kuriakose, R. Shrestha, and F. E. Sandnes, "Tools and technologies for blind and visually impaired navigation support: a review," *IETE Technical Review*, vol. 13, pp. 1–16, 2020.
- [18] S. Jan, S. Musa, T. Ali et al., "Integrity verification and behavioral classification of a large dataset applications pertaining smart OS via blockchain and generative models," *Expert Systems*, vol. 38, no. 4, Article ID e12611, 2021.
- [19] I. Khan and S. Khusro, "Towards the design of context-aware adaptive user interfaces to minimize drivers' distractions,"

- Mobile Information Systems*, vol. 2020, pp. 1–23, Article ID 8858886, 2020.
- [20] I. Khan and S. Khusro, “ConTEXT: context-aware adaptive SMS client for drivers to reduce risky driving behaviors,” *Soft Computing*, vol. 2022, pp. 1–18, 2022.
- [21] I. Khan, S. S. Rizvi, S. Khusro, S. Ali, and T.-S. Chung, “Analyzing drivers’ distractions due to smartphone usage: evidence from AutoLog dataset,” *Mobile Information Systems*, vol. 2021, pp. 1–14, Article ID 5802658, 2021.
- [22] I. Khan, S. Khusro, S. Ali, and J. Ahmad, “Sensors are power hungry: An investigation of smartphone sensors impact on battery power from lifelogging perspective,” *Bahria University Journal of Information & Communication Technologies (BUJICT)*, vol. 9, 2016.
- [23] I. Khan, S. Ali, and S. Khusro, “Smartphone-based lifelogging: an investigation of data volume generation strength of smartphone sensors,” *Simulation Tools and Techniques*, in *Proceedings of the International Conference on Simulation Tools and Techniques*, pp. 63–73, University of Colorado, Chengdu, China, July 2019.
- [24] E. Brady, M. R. Morris, Y. Zhong, S. White, and J. P. Bigham, “Visual challenges in the everyday lives of blind people,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2117–2126, Paris France, April 2013.
- [25] K. Anwar, T. Rahman, A. Zeb et al., “Improving the convergence period of adaptive data rate in a long range wide area network for the internet of things devices,” *Energies*, vol. 14, no. 18, 5614 pages, 2021.
- [26] L. Ran, S. Helal, and S. Moore, “Drishti: an integrated indoor/outdoor blind navigation system and service,” in *Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications*, pp. 23–30, Orlando, Florida, March 2004.
- [27] G. P. Soong, J. E. Lovie-kitchin, and B. Brown, “Preferred walking speed for assessment of mobility performance: sighted guide versus non-sighted guide techniques,” *Clinical and Experimental Optometry*, vol. 83, no. 5, pp. 279–282, 2000.
- [28] M. Bousbia-Salah, M. Bettayeb, and A. Larbi, “A navigation aid for blind people,” *Journal of Intelligent and Robotic Systems*, vol. 64, no. 3-4, pp. 387–400, 2011.
- [29] S. Cardin, F. Vexo, and D. Thalmann, “Wearable obstacle detection system for visually impaired people,” in *VR Workshop on Haptic and Tactile Perception of Deformable Objects*, pp. 50–55, Inforscience, Tokyo, Japan, 2005.
- [30] S. I. Akkalkot, G. H. Nandini, Y. Krishna, L. Sowmya, and K. G. Ravikumar, “Ultrasonic blind stick using ATmega328P,” in *ICDSMLA 2020*, pp. 1313–1319, Springer, Berlin, 2022.
- [31] I. Khan, S. Khusro, N. Ullah, and S. Ali, “AutoLog: toward the design of a vehicular lifelogging framework for capturing, storing, and visualizing LifeBits,” *IEEE Access*, vol. 8, Article ID 136546, 2020.
- [32] S. Rahman, S. Ullah, and S. Ullah, “Obstacle detection in indoor environment for visually impaired using mobile camera,” *Journal of Physics: Conference Series in Proceedings of the Journal of Physics: Conference Series*, vol. 960, Article ID 012046, Ancona, Italy, June 2018.
- [33] C. Yi, R. W. Flores, R. Chinchá, and Y. Tian, “Finding objects for assisting blind people,” *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 2, no. 2, pp. 71–79, 2013.
- [34] B. Ding, H. Yuan, L. Jiang, and X. Zang, “The research on blind navigation system based on RFID,” in *Proceedings of the 2007 International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 2058–2061, Shanghai, China, September 2007.
- [35] M. A. Rahman, S. Siddika, M. A. Al-Baky, and M. J. Mia, “An automated navigation system for blind people,” *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 1, pp. 201–212, 2022.
- [36] R. Velázquez, E. Pissaloux, P. Rodrigo, M. Carrasco, N. Giannoccaro, and A. Lay-Ekuakille, “An outdoor navigation system for blind pedestrians using GPS and tactile-foot feedback,” *Applied Sciences*, vol. 8, no. 4, 578 pages, 2018.
- [37] S. Khusro, B. Shah, I. Khan, and S. Rahman, “Haptic feedback to assist blind people in indoor environment using vibration patterns,” *Sensors*, vol. 22, no. 1, 361 pages, 2022.
- [38] D. Croce, P. Gallo, D. Garlisi, L. Giarré, S. Mangione, and I. Tinnirello, “ARIANNA: a smartphone-based navigation system with human in the loop,” in *Proceedings of the 22nd Mediterranean Conference on Control and Automation*, pp. 8–13, Palermo, Italy, June 2014.
- [39] I. Khan, S. Khusro, I. Ullah, and S. Mahfooz, “Towards smartphone-based navigation for visually impaired people,” in *Proceedings of the Computational Methods in Systems and Software*, pp. 366–373, Springer, Cham, Switzerland, 2018.
- [40] L. Oksana, T. Ihor, and L. Pavlo, “Navigation assistive application for the visually impaired people,” in *Proceedings of the 2020 IEEE 11th International Conference on Dependable Systems, Services and Technologies (DESSERT)*, pp. 320–325, Kyiv, Ukraine, May 2020.
- [41] H. A. Rowley, S. Baluja, and T. Kanade, “Neural network-based face detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23–38, 1998.
- [42] E. Osuna, R. Freund, and F. Girosit, “Training support vector machines: an application to face detection,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 130–136, Puerto Rico, June 1997.
- [43] Y.-M. Cheung and J. Deng, “Ultra local binary pattern for image texture analysis,” in *Proceedings of the 2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, pp. 290–293, Wuhan, Hubei, China, October 2014.
- [44] M. H. Bayana, S. Viriri, and R. Angulu, “Gender classification based on facial Shape and texture features,” in *International Symposium on Visual Computing*, pp. 157–166, IEEE Xplorer, Durban, South Africa, 2018.
- [45] A. Kharchevnikova and A. V. Savchenko, “Video-based age and gender recognition in mobile applications,” in *Proceedings of the International Conference on Information Technology and Nanotechnology*, vol. 4, no. 9, Russia, April 2018.
- [46] S. Shiqi Yu, T. Tieniu Tan, K. Kaiqi Huang, K. Kui Jia, and X. Xinyu Wu, “A study on gait-based gender classification,” *IEEE Transactions on Image Processing*, vol. 18, no. 8, pp. 1905–1910, 2009.
- [47] C.-C. Lee and C.-S. Wei, “Gender recognition based on combining facial and hair features,” in *Proceedings of the International Conference on Advances in Mobile Computing & Multimedia*, Vienna Austria, December 2013.
- [48] X. Geng, Z.-H. Zhou, and K. Smith-Miles, “Automatic age estimation based on facial aging patterns,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2234–2240, 2007.
- [49] Y.-L. Hou and G. K. H. Pang, “People counting and human detection in a challenging situation,” *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 41, no. 1, pp. 24–33, 2011.

- [50] M. Jeon, N. Nazneen, O. Akanser, A. Ayala-Acevedo, and B. Walker, "Listen2dRoom: helping blind individuals understand room layouts," in *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pp. 1577–1582, ACM Digital Library, Austin, TX, USA, 2012.
- [51] M. Rossi, J. Seiter, O. Amft, S. Buchmeier, and G. Tröster, "RoomSense: an indoor positioning system for smartphones using active sound probing," in *Proceedings of the 4th Augmented Human International Conference*, pp. 89–95, Stuttgart, Germany, March 2013.
- [52] H.-R. Zhang, F. Min, Z.-H. Zhang, and S. Wang, "Efficient collaborative filtering recommendations with multi-channel feature vectors," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 5, pp. 1165–1172, 2018.
- [53] M. Phillips and M. J. Proulx, "Social interaction without vision: an assessment of assistive technology for the visually impaired," *Technology & Innovation*, vol. 20, no. 1, pp. 85–93, 2018.
- [54] S. Panchanathan, S. Chakraborty, and T. McDaniel, "Social interaction assistant: a person-centered approach to enrich social interactions for individuals with visual impairments," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 5, pp. 942–951, 2016.
- [55] M. S. Sarfraz, A. Constantinescu, M. Zuzej, and R. Stiefelhagen, "A multimodal assistive system for helping visually impaired in social interactions," *Informatik-Spektrum*, vol. 40, no. 6, pp. 540–545, 2017.
- [56] S. T. Kouyoumdjieva, P. Danielis, and G. Karlsson, *Survey of Non-image Based Approaches for Counting People*, IEEE Communications Surveys & Tutorials, 2019.