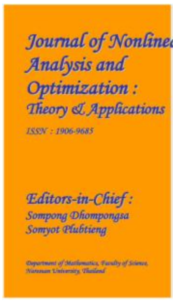
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1 : 2024
ISSN : **1906-9685**



BLIND ASSISTANCE SYSTEM - USING DEEP LEARNING

Mrs. Yogitha Pushpa , Associate Professor of Computer Science Department (CSE),Vignan's Institute of Information Technology, Vishakapatnam, India, pushpa.viyyapu@gmail.com
Pureddy Lavanya , Computer Science Department (CSE), Vignan's Institute of Information Technology, Vishakapatnam, India, pureddylavanya2002@gmail.com
Siraparapu Venkata Sampath, Computer Science Department (CSE), Vignan's Institute of Information Technology, Vishakapatnam, India, ucanreachsampath@gmail.com
Sirapu Murali, Computer Science Department (CSE), Vignan's Institute ofInformation Technology, Vishakapatnam, India, muralibeast143@gmail.com
SanapalaYaswanth , Computer Science Department (CSE), Vignan's Institute ofInformation Technology, Vishakapatnam, India, sanapalayaswanth49@gmail.com
Adarsh Kottapu, Computer Science Department (CSE), Vignan's Institute ofInformation Technology, Vishakapatnam, India itsmezed09@gmail.com

ABSTRACT:
Real – time object detection posses challenges due to theneed for significant computing powerto recognize objects quickly. An additional challenge arises from the lack of labeled data generatedby real- time systems, making effective training difficult withoutaccess to large quantities of Data with assigned labels. Single Shot Multi-Box Detection (SSD) this approach utilizes a convolutional neural network model for swift and real-time object detection.The SSD model streamlines the process by eliminating the feature
re-sampling stage and consolidating all environments with constrained computational resources like mobile devices (e.g., laptops, mobilephones), a lightweight network model is essential. In this context,the study suggests adopting a light weight network model called MobileNet,which utilizes depth-wise seperable.

Keywords:
SSD,MoblieNet,Convolutional

1. INTRODUCTION

Innovative technologies hold significant promise in assisting visually impaired individuals to achieve greater self- sufficiency and overcome social restrictions. One such technology involves the use of systems capable of recognizing items in the environment through voice commands and performing text analysis to interpret text from hard copy documents.This approach enables visually impaired individuals to interact more effectively with their surroundings and enhances their independence. As per the World Health Organization (WHO), vision impairment affects a considerable portion of the global population, with approximately 285 million individuals worldwide are facing some level of vision impairment, with around 39 million Moreover,around 3% of the total population across all age groups is visually impaired. People who have trouble seeing encounter many obstacles in their routines, such as problems with getting around unfamiliar environments, findingdirections, and accessing places they do not frequent. By leveraging innovative technologies like those described, it becomes possible to every individual.

2. REVIEW OF LITERATURE

Sight, the primary sense we rely on, holds immense importance in all aspects and phases of our existence. Despite often overlooking its significance, the absence of vision poses challenges in learning, mobility, literacy, engagement in education, and employment[1].Blindness denotes a substantial impairment of vision that cannot be rectified through conventional means such as glasses, contact lenses, medication, or surgical procedures. It hampers an individual's capacity to carry out daily tasks. A thorough examination of existing literature and meta-analyses (where applicable) was conducted, employing The GRADE approach is utilizedto update the prior American Academy of Sleep Medicine Practice Parameters concerning the management of intrinsic circadian rhythm sleep-wake disorders. Older individuals experiencing visual impairment face anincreased risk of falls, yet there is a lack of comprehensive evaluations of fall prevention strategies tailored for community-dwelling older adults with vision impairment.[2-4]. The prevalence of non-fatal consequences arising from diseases and injuries is progressively compromising the overall health and well- being of the global population. Vision impairment (VI) is commonly evaluated by assessing the best- corrected visual acuity (BCVA) of the eye with superior vision [5-6].The goal of Warrington VIP (Visually Impaired People) is to support individuals, as well as their families and friends, in coping with sight loss, improving quality of life, fostering ongoing independence, and increasing awareness of the needs of visuallyimpaired individuals within our communities [8].During the summer, I instructed a cohort of Japanese university students in an English and Culture summer program at Kent University. One of my students, ‘C’, was visually impaired. It presented a significan curve for me as I’d never worked with a VIP before. I gainedvaluable that I compelled to share drawing from my modest four weeks’ experience in the hope that it wouldbenefit others as detailed in journal[9].

EXISTING SYSTEM

The below explains the existing system work with a flow:

In existing system (Fig.1.), The system uses a webcam to capture surroundings, processes the images, identifies objects, and outputs the results in text.Transfer Learning:Leverage pre- trained models on largedatasets to enhance the detection capabilities.

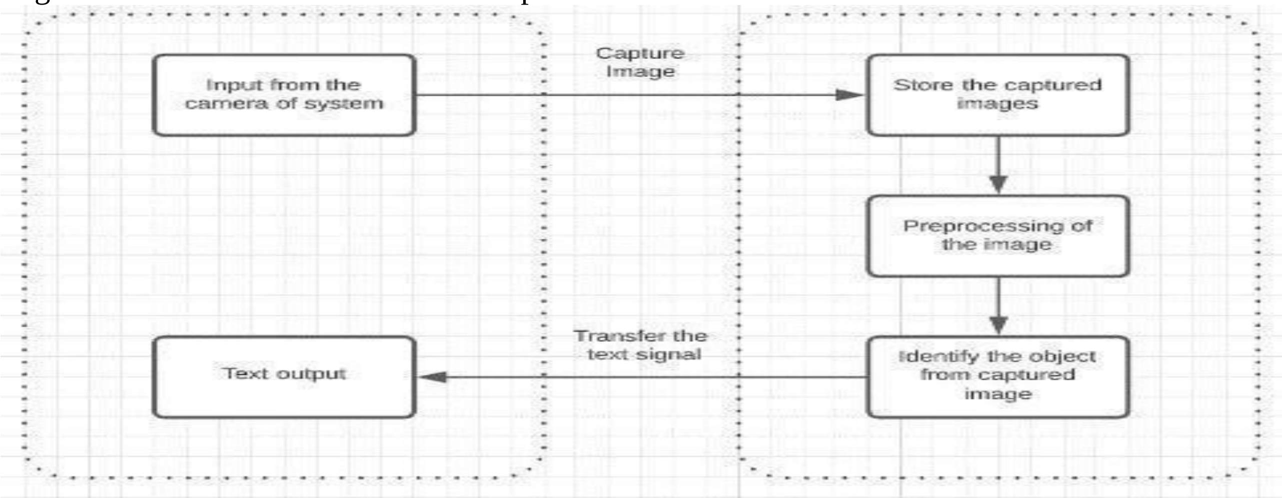


Fig 1 :- Flow Chart Of Existing Work

4. PROPOSED SYSEM

The proposed system for assisting visually impaired individuals represents a significant advancement in addressing the challenges they face when navigating outdoor environments independently. Built upon object/obstacle detection technology, this system operatesthrough a series of steps, beginning with frame extraction and culminating in output recognition. By comparing eachframe with a database of known objects,the system can accurately detect items in the user's surroundings. Once identified, an audio file is activated

to relay information about each detected object, simultaneously addressing both object detection and identification needs.

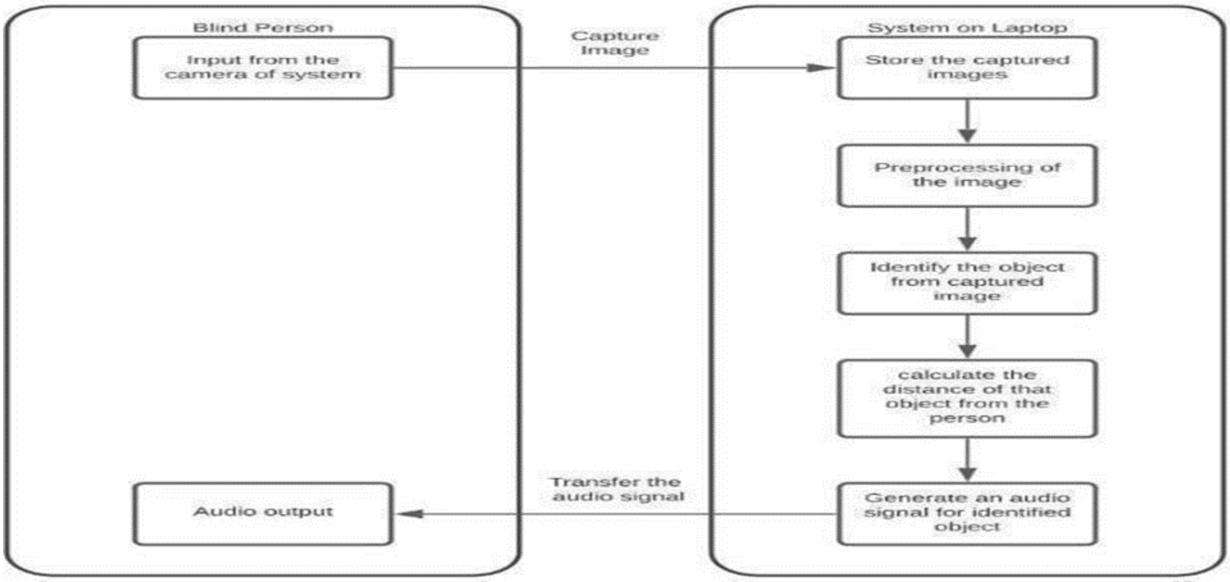


Fig 2 :- Flow Chart Of Proposed Work

Furthermore, the system may incorporate features such as real-time processing, voice interaction, and obstacle detection to enhance usability and safety. Customization options and integration with navigation systems further tailor the experience to individual preferences and requirements.With mechanisms for continuous improvement and feedback, The system pledges to provide visually impaired individuals with the assurance and autonomy to navigate outdoor environments with greater ease.In this system, each frame is analyzed by comparing it with a database of known objects. This comparison helps detect items present inthe environment. Once objects are identified,an audio file is activated to provide information about each detected object. This integrated approach tackles both object detection and identification simultaneously, aiming to improve the freedom and movement of people who have difficulty seeing when navigating outdoor environments.

5. METHODOLOGY

1. Real-time frames are captured by the systemtoprovide live visual input.
2. A Laptop Based Server utilizes a pre-trained SSD detection model trained on the COCO dataset toanalyze the captured frames. The model identifies objects present in the scene and evaluates its detection accuracy against predefined metrics.
3. Upon detection, the system employs voice modules toconvert the object's class into default voice notes. These voice notes are then dispatched to visually impaired individuals to assist them in understanding their surroundings.
4. In addition to object detection, an alert system is integrated to calculate the proximity of the detected objects to the visually impaired individual. Voice- based outputs, accompanied by distance units, are generatedto indicate whether the individual is in close proximity to the object or situated at a safer distance.

6. TENSOR FLOW

The implementation of the system utilizes TensorFlow APIs, providing a set of common operations for efficient development. By leveraging these APIs, developers can streamline the process of building deep learning networks without starting from scratch, saving time and effort.Specifically, the TensorFlow objectdetection API serves asapowerful tool for addressing object detection challenges. It offers a framework thatincludes pre-trained models, knownas the Model Zoo, which encompasses datasets such as COCO, KITTI, and Open Images. The primary focus lies on the COCO dataset, which facilitates robust object detection capabilities.

7. SDD

The Single Shot MultiBox Detector (SSD) architecture comprises two main parts: the backbone model andthe SSD head.

The main model acts as a tool to pull out important features in object detection setups such as the

Single Shot MultiBox Detector (SSD). Typically, the backbone is typically a pre-trained image classification network taught using extensive collections of data like ImageNet. Examples of backbone networks includeResNet, VGG, and MobileNet

The SSD head is a crucial component of the Single Shot MultiBox Detector (SSD) architecture, responsible for generating predictions of bounding boxes and object classifications based on the features extracted by the backbone model. SSD creates a complex computer system that understands the meaning of images whilekeeping their layout, even though it might be less detailed.

ResNet34 generates 256 feature maps of size 7x7 for an input image. SSD partitions the picture into gridsquares, with each square is tasked with spotting objects within its boundaries. Object detection involvespredicting an object's class and position within the grid cell.

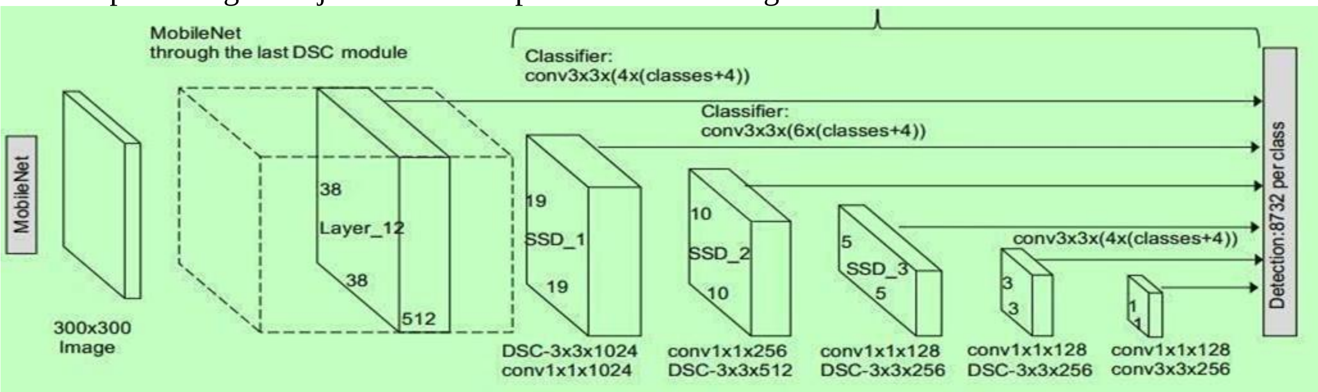


Fig 3 :- Architecture Of SSD

8. MOBILE NET

The model is inspired by MobileNet's concept of depthwise separable convolutions, employing factorized convolutions. Depthwise convolutions, derived from basic convolutions, are used to separate spatial and channel-wise information. Additionally, point wise convolutions, also known as 1x1 convolutions, further enhance the model's capabilities. The depth wise convolutions in MobileNets apply applying one filter to every input channel, whereas point wise convolutions combine their outputs using 1x1 convolutions. Thesefilters act similarly to traditional convolutions, combining inputs into new outputs in one level. Depth-wiseseparable convolutions divide this process into filtering and mixing layers, significantly reducing computation time and model size.

9. VOICE GENERATION MODULE

After detecting an object, it's essential to alert the individual about its presence. PYTTSX3 plays a vital role in the voice generation module, converting text to speech. PYTTSX3 is a Python library compatible with both Python2 and 3, simplifying the process of text-to- speech conversion. It serves as a straightforward toolfor generating speech output from text inputs, aiding in providing auditory alerts to users.

The technique operates as follows: upon identifying an object,We estimate the distance and display relevanttext on the screen using the cv2 library and its cv2.putText() function. Additionally, we employ Python- tesseract., an Optical Character Recognition (OCR) tool, is then employed to acknowledge and extract text incorporated within the image. This process facilitates the retrieval of textual information from images, enhancing the system's ability to provide comprehensive object identification and contextual information to users.OCR (Optical Character Recognition) is a technology that scans and analyzes images to recognize and encode text .content in a computer-readable format. Python- tesseract is a tool used for OCR, capable of detecting and "reading" text embedded within images. Once the text is recognized, it can be further processed and linked to a text-to-speech engine, such as pyttsx, enabling the conversion of the detected textinto audible speech. This integration allows for the automatic vocalization of text content extracted from images, enhancing accessibility and usability for users. The system produces audio alerts based on the proximity of detected objects. Using libraries like PyTorch for object detection, pyttsx3 for text-to- speech, and pytesseract for OCR, it warns users if objects are too close with messages like "Warning: Theobject is too close to you." Conversely, it signals objects at a safe distance with messages like "The object is at a safedistance"

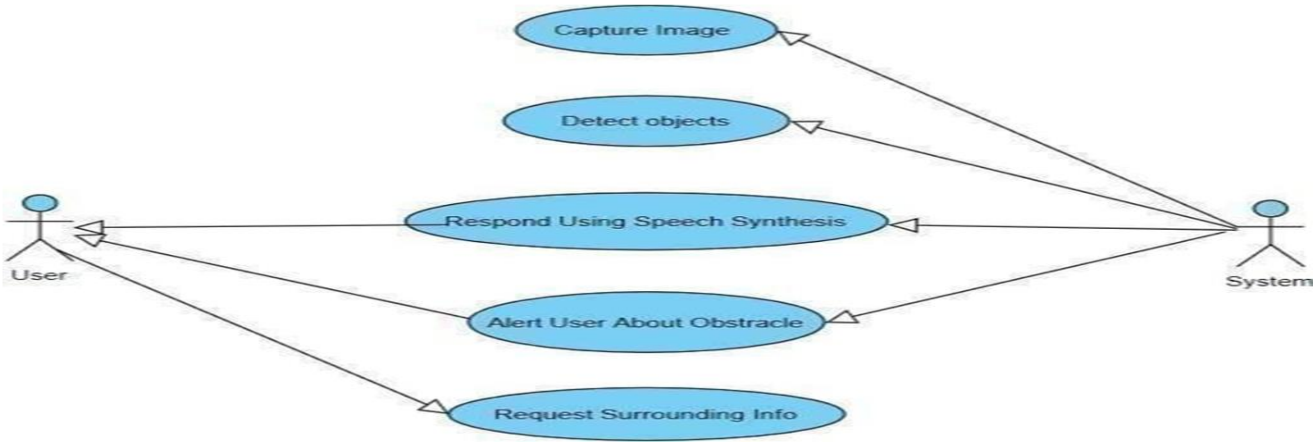


Fig 4 :- Use Case Diagram Of System

In above Fig.4. as you can see, In this interactive process, the user provides surrounding information to the system. Upon receiving this input, the system captures an image of the surroundings, detects objects within the image using computer vision techniques, and generates a response using speech synthesis technology. Finally, the system alerts the user about specific obstacles detected in the environment, providing crucial information for navigation and safety. This iterative exchange allows for real-time assistance and awarenessof obstacles in the user's surroundings.

10. IMAGE PROCESSING

Image processing involves applying various operations on an image to enhance its quality or extract relevant information from it. These tasks may involve filtering, enhancement, segmentation, and feature extraction, among others. The goal aims to enhance the visual presentation of the image or extract specific features or patterns that are relevant to a particular task or application. Image processing techniques are widely used in fields like computer vision, medical imaging satellite imaging, and digital photography, among others, to analyze and manipulate images for various purposes. The fundamental definition of image processing states that it involves the examination and manipulation of digital images, primarily aimed at improving their quality. This encompasses various methods and algorithms employed to improve, analyze, and extract information from digital images, thereby contributing to various fields such as computer vision, medical imaging, and remote sensing.

11. OPEN CV

OpenCV is a freely available, open-source library renowned for its comprehensive functionality in the field of computer vision. It offers a broad spectrum of functions and algorithms for various tasks related to computer vision, including motion-tracking, facial-identification, object-detection, segmentation, and recognition, among others. With OpenCV, users have the ability to manipulate images and process real-timevideo streams. for customization them to their specific requirements. Its versatility and extensive capabilitiesmake it a popular choice for researchers, developers, and enthusiasts working in the realm of computer vision.

12. WHAT IS COCO

The COCO dataset, which is abbreviated as (Common Objects in Context),It was made to improve tasks related to recognizing images. It offers top-notch image datasets for computer vision uses, often casted for training state-of-the-art neural networks. The dataset consists of a wide range of images annotated with various objects, facilitating the development and evaluation of image recognition algorithms. In total, the COCO dataset contains annotations for 90 predefined objects,enabling researchers and developers to tacklechallenging computer vision tasks effectively.

13. RESULTS

Object detection using deep learning has yielded remarkable results, achieving unprecedented degrees of precision and effectiveness in identifying objects within images and video streams. Deep learning models, like SSD, YOLO, and Faster R-CNN, have demonstrated the capability to identify objects with a high degree of accuracy across various domains and datasets. These models not only accurately localize objects but also classify them into predefined categories, enabling sophisticated applications in domains like autonomous-vehicles, surveillance, medical-imaging, and robotics. With ongoing advancements in deep learning methodologies and model architectures, object detection systems continue to improve, pushing the boundaries of what's possible in computer vision and artificial intelligence.

Fig 5:- Ouput 1

Fig 6 :- Ouput 2

Fig 7 :- Ouput 3

Fig 8 :- Ouput 4

14. CONCLUSION

In our research, we aimed to recognize objects displayed in front of a webcam. To accomplish this, we utilized the TensorFlow Object Detection API frameworks for testing and training our model. However, reading frames from a webcam posed challenges, necessitating a 1solution to optimize

frames per second and minimize Input/Output issues. We focused on threading methodology, significantly improving frames per second and reducing processing time for each object. Despite accurately identifying objects in front of the webcam, there remains a delay of approximately 3-5 seconds for the object detection box to transition tothe next object.
Utilizing the findings from this study, we can apply object recognition and tracking techniques to sports fields, enabling computers to engage in deep learning applications. By implementing deep learning models trained on sports-related datasets, such as footage from training sessions, computers can learn to recognize and track various objects such as players,balls, and equipment in real-time. This application has significant potential for enhancing sports analytics, player performance analysis, referee assistance, and fan engagement.Utilizing object detection to identify ambulance vehicles in traffic through public surveillance cameras can facilitate the management of traffic signals, allowing for the prioritization of emergency vehicles. This technique holds promise for improving emergency response times and enhancing road safety. Moreover, the application of object detection technology extends beyond traffic management, with potential applications in various fields. For instance, in literature, it can aid in text recognition and analysis. In finance, it can assist in currency authentication and fraud detection. Additionally, in language translation, it can support real-time translation services. Overall, the versatility of object detection technology opens doors to numerous innovative applications across different domains, promising advancements in efficiency, safety, and convenience.
Indeed, the project's outcomes hold the potential to significantly benefit society by enabling the detection and tracking of various items. This capability can streamline processes, enhance efficiency, and ultimately make life easier for individuals across different contexts.Whether it's in traffic management, surveillance, logistics, or other areas, the ability to detect and track items effectively can lead to improved safety, resourceallocation, and decision-making. Overall, the project's contributions have the promise to positively impact various aspects of daily life, offering convenience, reliability, and peace of mind to users.

15. FUTURE SCOPE

Beyond its primary application for assisting the visually impaired, the project's algorithms find diverse applications in finding pictures or photos, keeping things safe and protected, watching and monitoring closely, and advanced tools to help drivers with tasks like parking or staying in line, underscoring its potential impact across various domains. The innovative methodology ensures not only enhanced safety andindependence for the visually impaired but also contributes to broader advancements in technology with far-reaching societal implications.

16. REFERENCES

1. "Visual impairment and blindness Fact Sheet N°282". August 2014. Archived from the original on12 May 2015. Retrieved 23 May 2015.
2. ^ Jump up to:a b c d e f "Blindness and Vision Impairment". February 8, 2011. Archived from theoriginal on 29 April 2015. Retrieved 23 May 2015.
3. ^ Auger RR, Burgess HJ, Emens JS, Deriy LV, Thomas SM, Sharkey KM (October 2015). "Clinical Practice Guideline for the Treatment of Intrinsic Circadian Rhythm Sleep-Wake Disorders: AdvancedSleep-Wake Phase Disorder (ASWPD), Delayed Sleep-Wake Phase Disorder (DSWPD), Non-24- Hour Sleep-Wake Rhythm Disorder (N24SWD), and Irregular Sleep- Wake Rhythm Disorder (ISWRD). An Update for 2015: An American Academy of Sleep Medicine Clinical Practice Guideline". Journal of Clinical Sleep Medicine. 11 (10): 1199– 236. doi:10.5664/jcsm.5100. PMC 4582061. PMID 26414986.
4. ^ Blaylock, SE; Vogtle, LK (June 2017). "Falls prevention interventions for older adults with low vision: A scoping review". Canadian Journal of Occupational Therapy. 84 (3): 139147. doi:10.1177/0008417417711460. PMID 28730900. S2 CID 7567143.
5. ^ Jump up to:a b Vos T, Allen C, Arora M, Barber RM, Bhutta ZA, Brown A, et al. (GBD 2015 Disease and Injury Incidence and Prevalence Collaborators) (October 2016). "Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015". Lancet. 388 (10053): 1545–

1602. doi:10.1016/S0140- 6736(16)31678-6. PMC 5055577. PMID 27733282
6. "Vision Impairment and Blindness |Examination-Based Studies | Information on Data Sources | Vision and Eye Health Surveillance System | Vision Health Initiative (VHI)| CDC". www.cdc.gov. 2021-08-10. Retrieved 2022-03-03.
7. ^ "Visually Impaired People (VIP) - Education & Engagement". Bat ConservationTrust. Retrieved 2023-01- 16.
8. ^ "Home page". Warrington VIP Centre. Retrieved 2023- 01-16.
9. ^ Rountree, Helen (Autumn 2018). "Working with a VIP (Visually Impaired Person)". IH World. 45.Retrieved 2023-01-16.
10. ^ Jump up to:a b c GLOBAL DATA ON VISUALIMPAIRMENTS 2010 (PDF). WHO. 2012.p. 6.Archived (PDF) from the original on 2015-03-31