TRINETRA - THE THIRD EYE

Cyril Thomas Easow
Computer Science and Engineering
Amal Jyothi College of Engineering
Kottayam, India
cyrilthomaseasow2023@cs.ajce.in

Diya Mereen Thomas
Computer Science and Engineering
Amal Jyothi College of Engineering
Kottayam, India
itidiya@gmail.com

Felix Kurian
Computer Science and Engineering
Amal Jyothi College of Engineering
Kottayam, India
felixkurian2023@cs.ajce.in

Gowripriya Dileep
Computer Science and Engineering
Amal Jyothi College of Engineering
Kottayam, India
gowripriyadileep2000@gmail.com

Anu Rose Joy
Asst. Prof. Computer Science and Engineering
Amal Jyothi College of Engineering
Kottayam, India
anurosejoy@amaljyothi.ac.in

***Abstract*—Technological advancements have been aiding visually impaired individuals in many aspects of their daily lives. According to the World Health Organization, there are 285 million people globally who are blind or visually impaired, with 39 million of them being completely blind. This makes it challenging for them to understand their surroundings and carry out tasks like navigation, object recognition, and reading. Hence, it's crucial for them to have access to technology that can assist them. This can be in the form of devices like screen readers, text-to-speech software, smart glasses, wearable devices with sensors, Braille displays, and smart canes equipped with obstacle detection sensors.The objective of the proposed project is to create a mobile application for blind or visually impaired individuals using the Android platform. The purpose of the system is to identify objects and Indian currencies in real-time using a camera on a mobile device and then provide audio or vocal notifications to the user to inform them of the objects detected.**

I. INTRODUCTION

Sight is a crucial sense that allows us to perceive our environment. Individuals who lack this sense are known as visually impaired. People who are blind or visually impaired face a variety of challenges in their daily lives that sighted individuals may take for granted. Software applications with speech synthesizers that enable internet browsing and access to text documents are particularly helpful. Another significant challenge for the visually impaired is mobility and safety while walking. Since the 1970s, researchers have conducted various forms of research to address this issue. They have used different types of devices, such as sonar and cameras, to record environmental information and translate it into perceptible information, aiming to help visually impaired individuals walk independently with reduced safety concerns.

Blindness and visual impairment can significantly impact a person's ability to navigate their surroundings, recognize objects, and perform everyday tasks like reading. To address these challenges, this paper proposes an Android application that detects objects and Indian currency in real-time using a mobile camera. The system provides audio notifications to the user to inform them of the detected items. The objective of this project is to provide visually impaired individuals with an accessible and convenient tool to help them interact with their environment. This paper presents a comprehensive analysis of the system design, implementation, and testing, as well as potential areas for future development and enhancement.

The proposed system represents a significant advancement in assistive technology for visually impaired individuals. The integration of a mobile camera and real-time object and currency detection capabilities provides a practical and effective solution to many of the challenges faced by the blind community. The system is designed to be user-friendly, accessible, and cost-effective, making it a valuable tool for a large portion of the population. This paper outlines the development of the Android application, including the object and currency detection algorithms' design and implementation. The testing and evaluation results are presented, demonstrating the system's accuracy and reliability in real-world situations. Additionally, we discuss the challenges encountered during the system's development and potential solutions to overcome them.

It is worth emphasizing that the current system's scope is limited to object and Indian currency detection, but it lays the groundwork for the advancement and expansion of assistive technology for visually impaired individuals. Future developments may involve integrating additional sensors and features such as obstacle avoidance and navigation to offer a comprehensive solution for people with visual impairments.

The generation of parking slot maps using Python involves creating a graphical representation of a parking lot that displays the size and location of individual parking spaces. This map assists drivers in locating their designated parking spot promptly. To generate this map, Python libraries and tools, such as OpenCV and matplotlib, are utilized. The resulting map can be used for a variety of purposes, including parking management, transportation planning, and navigation.

II. Literature Review

Pias Paul et al. [1] proposed a model that serves as a blind person's visual help. The functionality of the suggested prototype has been shown using a Raspberry Pi 3 Model B+ due to its inexpensive price, small size, and simplicity of integration. For object detection and obstacle avoidance, the design includes a camera, sensors, and powerful image processing algorithms. The camera and ultrasonic sensors calculate the separation between the user and the obstruction. The system has an integrated image-to-text converter and audio feedback after the integrated reading assistance. The complete system is inexpensive, simple to use, and can be installed onto a conventional pair of eyeglasses. It is also lightweight and portable. The research article proposes a revolutionary pair of eyeglasses for completely blind people as a sort of visual help. The key features are object detection, distance measurement, reading assistant and its more of an ergonomic in nature which again adds more comfort level.

Kanchi Kedar Sai Nadh Reddy et al. [2] proposed a model that Utilizes image recognition from an object database, the model is created. The system's aural feedback will inform the user of the image recognition results. The suggested system has been successful in determining the worth of Indian currency notes, accurately identifying a few nearby items, such as a bottle or a person, and measuring the user's distance from the target and its direction. The pre-processing stage of the proposed model uses the Mobilenet pre-trained model and entails depth-wise separable convolution, depth-wise and point-wise layers, batch normalisation, and ReLU. The Single Shot Detector (SSD) framework and the MobileNet architecture have both been used in this model to develop a quick, efficient deep-learning approach for item recognition that aids in achieving many detections in a single scene. The brute force descriptor matcher of the OpenCV library was used to recognise dollar bills using the KNN technique, and OpenCV and a custom mathematical formula were utilised to determine the item's angle.

Dhruv Dahiya et al. [3] proposed a framework, the proposed framework's key addition is its use of deep learning techniques, which make it extremely robust due to the algorithm's high accuracy. TensorFlow has been used to train the framework on a dataset with more than 450 photos. Based on this, the framework classifies scene photos as belonging to ATMs, pharmacies, restrooms, and metro stations, if any of these objects are visible in the scene images. Due to its quick processing, this framework supports real-time applications and is extremely accurate. This study suggests a precise and efficient real-time system built on deep learning. The algorithm was trained using a total of more than 450 shots, more than 100 of which were taken at each public place that was taken into account. 140 photographs from each of the four classes, which were different from the training images, were used to test this system, and the results showed that ATMs, pharmacies, bathrooms, and metro stations could be accurately and quickly detected. Through the use of this method, blind people can gain sight while requiring less assistance from others.

Jawaid Nasreen et al. [4] from the analysis of medical imaging to driverless cars, machine learning is increasingly used in a variety of industries. Finding objects in images is one of the most important research areas today, and computers are becoming more and more adept at not just finding objects but also drawing bounding boxes on them.The suggested system includes a website that provides an interface for a camera to capture an image from the phone camera and send it to a web server where a machine learning model (YOLO) detects the image and finds the objects in it and passes back the results to the client where a browser-based voice library (google voice) converts those text to voice and narrates it to the blind or visually impaired person. The system is only intended for use by those with vision impairments, and it may be installed on any camera-equipped device. People with visual impairments can use this software on their smartphones. The system will identify the objects in the photograph that the user captures of their surroundings and narrate it.This study describes a system that can be used to help the blind person understand their surroundings by describing the nearby things. The developed solution is based on using a website that, when it loads, grabs an image from the phone's rear camera and sends it to a server. To identify the objects in that image, a trained machine learning model known as YOLO (You only look once) is loaded on the server.

Lina Li et al. [5] proposed a system where the two components of the suggested system for identifying various student actions in a classroom video are ESRGAN and YOLOv5s with a small object detection module. Low-quality video frames are transformed using ESRGAN to a higher resolution for sharper visuals, which are subsequently utilised as input for YOLOv5s. YOLOv5s divides high-resolution photos into eleven categories using its tiny object identification module. The system uses a particular variant of YOLOv5, YOLOv5s, for its minimal memory and energy requirements, making it suited for real-time classroom behaviour identification. The system is intended to be quick and accurate. Three sections make up the YOLOv5s architecture: the Backbone network, Neck network, and Head network. Focus structure, CSPNet, and SPP are components of the backbone network that execute feature extraction, feature fusion, and feature reduction.The Neck network uses PAN and FPN to combine features from many levels and increase network robustness. The Head network finds objects of various sizes and produces the expected bounding box and categories. The system is intended to track student conduct in the classroom. To increase precision for small objects, a module for tiny object detection has been included. The small object detection module is made up of FPN and PAN in the Neck structure, which produces a feature map in the Head network that is 160x160 in size. The upgraded YOLOv5s can learn more precise image features and recognise targets as small as 4x4 pixels or more when the input size of the image is increased and this module is integrated.

Sejal Gianani et al. [6] proposed a model where finding lost or dropped personal things is one of the many difficulties that visually impaired people encounter when navigating their surroundings. A solution is suggested to help people handle everyday objects for those with partial or complete visual loss since they require better positioning and control of their personal items. They will be able to pinpoint and pick up objects in their immediate area with the aid of this software. In this research work, an object detection and positioning system is discussed. While object positioning establishes where an object is in relation to the user's hand, object detection identifies objects using computer vision and image processing techniques.The system works as follows first It receives real-time video input from a camera then It recognises items in the video stream and enables voice input from the user to designate a desired object ,the system then scales, maps, and determines if the user's hand can reach the selected object in the video and If the user can reach the object, the system will speak instructions to direct them there. The only means of communication between the user and the machine is voice instructions.

Milad Ghantous et al. [7] proposed a model named iSee.iSee is an Android-based application that uses commercially available technologies to enhance the daily routines of persons who are blind. In iSee, a single screen tap can act as a virtual eye by giving the visually impaired individual a sense of seeing by loudly transmitting the object(s) names and descriptions. The suggested method makes use of the object detection algorithms SURF and ORB. ORB has performance that is comparable to SIFT and SURF while being faster, whereas SURF is a scale and rotation invariant detector and descriptor. Because of its benefit in terms of computation, ORB is chosen as the underlying algorithm. The usercaptured scene and a training object are the algorithm's two inputs.A nearest neighbour matching procedure is used to weed out the good matches after interest point identification and descriptor computation. The remaining "poor" matches are then all eliminated using a thresholding strategy based on the number of matches. The remaining matches are then contrasted in order to select the most. At this point, the user is informed of the object's name and type.Using assistant software, the app is opened by tapping, and the built-in back camera starts recording immediately. The user can point the phone in any direction and tap the screen to take a photo.

Fahad Ashiq et al. [8] proposed a model where in people who are visually impaired make up a sizeable section of the population, and they may be found everywhere in the world. Technology has recently demonstrated its presence in every field, and cutting-edge gadgets help people in their daily lives. In order to help visually impaired people navigate their environment and maintain their safety, this article suggests a clever and sophisticated solution.The suggested system consists of a camera, headphones, and a Raspberry Pi digital signal processing (DSP) board with GSM and GPS modules. The object detection and recognition module (CNN model) receives a live feed from the video camera that is captured by the DSP and passed on to it. The text-to-speech converter (SAPI) module reads the names of the objects the model predicts will appear in the current video frame while wearing headphones. Additionally, the tagged photo is JPEG encoded by the Joint Photographic Experts Group. The server's database stores the encoded image along with the user's precise location. Additionally, VIPs have access to a user-driven function that allows family members to track their whereabouts (receive the most recent labelled frame and position) while unwinding at home.

Wan-Jung Chang et al. [9] proposed a model In order to achieve the objectives of aerial obstacle avoidance and fall detection, this study suggests an intelligent assistive system based on wearable smart glasses and an intelligent walking stick for visually impaired people. An intelligent walking stick, wearable smart glasses, a smartphone app, and a cloud-based information management platform make up the suggested assistive system.The recommended assistive system consists of a pair of wearable smart glasses, a smart walking stick, the V-Protector software for mobile devices, and a cloud-based information management platform. smart glasses that can be worn. The wearable smart glasses are equipped with a Bluetooth low energy (BLE) wireless communication module, a microcontroller unit (MCU), a 6-axis (gyro + accelerometer) MEMS motion tracking sensor module, an IR transceiver sensor module, a 6-axis (gyro + accelerometer) MEMS motion tracking sensor module, a battery charging module, and a charging module. To combine the two fall recognition techniques, BLE wireless communication is used to connect and pair the suggested intelligent walking stick and wearable smart glasses. The proposed wearable smart glasses are used to detect impending airborne obstacles and falls.

S. A. Jakhete et al. [15] proposed an Android app to aid blind persons in using a mobile phone or other handheld device to see. It combines different techniques to create a sophisticated Android application that will help visually impaired people as rapidly as possible by recognising objects around them in real time and providing an audio output. For both object detection and recognition, the SSD (Single Shot Detector) algorithm is employed. The programme also makes use of the Android Text To Speech and TensorFlow APIs to provide audio output.In this study, we used the COCO (Common Objects in Context) 2014 Database, which contains 83K training photos and 41K test images for 80 different object classes. The labelled dataset, which is helpful to train the model, is the dataset that was employed. The labelled dataset, which is helpful to train the model, is the dataset that was employed.This research created an Android application to assist persons who are blind or visually impaired. It helps users identify the objects they encounter and sends them an audio label based on the predicted object's confidence score in the frame. The SSD approach is faster than other comparable algorithms and provides results for real-time object detection that are nearly correct. The suggested system accurately recognises a variety of things, including a bottle, a person nearby, a book, etc., and also produces vocal output.

III. METHODOLOGY

A. Dataset :

Our android application uses artificial neural networks for object recognition, which requires a dataset of objects and currency to train the classifier. The COCO 2017 dataset is used for object detection, which contains over 330,000 labeled images with more than 2.5 million object instances across 80 different categories. The dataset is divided into three subsets: train, validation, and test. The train subset has 118,000 images, the val subset has 5,000 images, and the test subset has 41,000 images. This labeled dataset is useful for training the model to recognize objects, with examples of objects including those from 80 different categories.The 80 different object categories in the COCO dataset include animals, vehicles, household items, food, furniture, people, and more. To detect currency, a dataset was developed in-house with 7 different classes of currency denominations. The dataset contains a total of approximately 584 images and is divided into three subsets: train, validation, and test. The train subset consists of 510 images, the val subset consists of 43 images, and the test subset consists of 31 images.

B. Data Preparation:

The COCO dataset for object detection was obtained from Kaggle, which is an online platform for data science competitions, datasets, and notebooks. To enable currency detection in the Android application, a dataset was developed in-house. This involved taking photographs of different denominations of currency in various angles and under different lighting conditions. The resulting dataset was then labeled and annotated to indicate the denomination of each currency note in the image. The dataset was then split into three subsets: the train, test and validation sets.

C. Data Labeling :

For object detection, we used images that were already labeled and annotated, and which were provided with the annotation file along with the dataset. To label the images in the currency detection dataset, the organization used Roboflow, which is an online platform for creating, managing, and improving image datasets.The annotation file contains parameters such as object class, unique object ID, x-coordinate for the center, y-coordinate for the center, width, and height for each image. These parameters are used to train the object detection model, enabling it to recognize and classify objects in new images and to indicate the denomination of each currency note in the image.

D. Train-Test Split:

Once the dataset for both object and currency detection was collected and annotated, we randomly shuffled the data and selected 80% of the dataset for training the model. The remaining 20% of the data, which was not seen by the model during the training phase, was used for testing the performance of the model.

E. Model training:

The main concept behind developing the object detection and classification models for the Android application is to utilize Transfer Learning, which involves utilizing a pre-trained model that has already been trained on a large dataset, making it efficient in recognizing various objects. This approach is aimed at reducing the time and computational resources required for training the models from scratch. By using a pre-trained model, the organization can build and fine-tune models for object and currency detection, and improve their accuracy and performance on new data. To achieve this, we used Yolov5s, a pre-trained model that has been shown to be effective in object detection. Two separate models were created, one for currency detection and one for object detection. The object and currency detection datasets were tested using both models, and the results were analyzed to evaluate the effectiveness of each model.

F. Real Time Video Processing:

The real-time video processing rate in YOLOv5s depends on various factors such as the hardware specifications, image resolution, and number of objects in the video. Typically, YOLOv5s can process up to 40 frames per second on modern GPUs such as Nvidia RTX 2080Ti, and up to 15 frames per second on a CPU like Intel Core i9-9900K.

The stable output is generated for the real-time input.

G. Object Detection and Currency Detection:

Pretrained models for object detection use anchor-based object detection to form boundary boxes. This technique divides the input image into a grid of cells and assigns anchor boxes to each cell. During inference, the model predicts the probability of an object being present in each cell and adjusts the coordinates of the anchor boxes to align with the actual object location. The model also predicts the class of the object within each boundary box.

When generating bounding boxes, a confidence score is also generated which indicates the likelihood of the presence of an object within the box. Additionally, each bounding box is associated with a predicted class, represented as a probability distribution across all classes in the model. By combining the confidence score and class probability, the final score is calculated to indicate the probability of the bounding box containing a specific object.

In YOLOv5s, which stands for "small" variant of the YOLOv5 family, the default minimum detection confidence threshold is set to 0.25. This means that any detected object with a confidence score below 0.25 will be discarded and not considered as a valid detection.

Bounding boxes whose score is more than the threshold is given as an output along with the respective class name.

H. Text-to-Speech Converter:

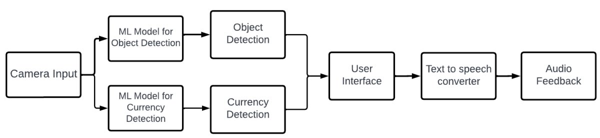
The Google Text-to-Speech API can be used in conjunction with an object and currency detection model to generate an audio output that speaks the labels of the objects and currencies detected by the model. Using the Google Text-to-Speech API can improve the accessibility of our application for visually impaired users, making it easier for them to interact with the system and understand the information it presents.

I. Android Application:

The model is integrated into Android Application. The app uses a rear camera of the smartphone for real time processing.

IV. Workflow

The Android application is designed to offer object and currency detection functionalities to the user. To use the app, the user simply presses a button on their phone to open the application. Upon opening the app, the user is prompted with an audio note that instructs them to tap on the top of the screen for object detection or the bottom for currency detection. Once the user interacts with the app, the camera starts capturing the live feed from the surroundings. An audio note is then played, notifying the user that a snapshot has been captured and is ready to be analyzed. If the result of the snapshot is shaky or unclear, the user is prompted to retake the snapshot.



Workflow of the System

Once the snapshot is captured, it is passed to the object detection and recognition module. The model predicts the objects in the current video frame, and if a match is found, the name of the object is communicated to the user via a voice message. In case no match is found, the user is notified with a “not found” message.

Overall, the Android application offers an intuitive and user-friendly interface for object and currency detection, enabling users to quickly and easily identify objects and currency within their surroundings.

V. Results and Evaluation

The field of Computer Vision, specifically Object Detection and Recognition, has undergone extensive research and experimentation resulting in a variety of algorithms and models that aim to achieve highly accurate results. Among these models are Single Shot Multibox Detector (SSD), Faster R-CNN, and You Only Look Once (YOLO). Our chosen model for this project is YOLOV5s due to its advantages and ability to process real-time data efficiently.The experimental results show that the improved model’s mAP (mean average precision) is 94.9%, an improvement of 7.0% over the original model. The size of the improved model is 4.49 MB, and the mean FPS (Frames Per Second) is 62. The convergence curve of the box loss, the objectness loss, and the classification loss, and the performance curve of the precision, the recall, and the mean average precision (mAP) over the training epochs for the training and validation set of currency detection model are shown together in Figure 1. The box loss is speculated as to the GIoU loss function, which is used as the loss of the bounding box. The objectness loss reflects the loss whether the bounding box is the object or the background. The classification loss is speculated as to the mean of classification loss. The confusion matrix of the YOLOv5s model on the currency detection validation set is shown in Figure 2. The number in the grid on the diagonal of the confusion matrix represents the recall rate of each category.well. It can be seen from the confusion matrix that the overall recognition rate of the method proposed by us on the validation data set is high.

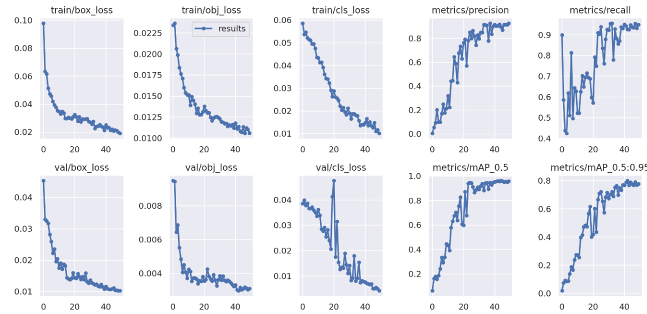


Fig. 1. YOLOV5s training loss functions and performance curves

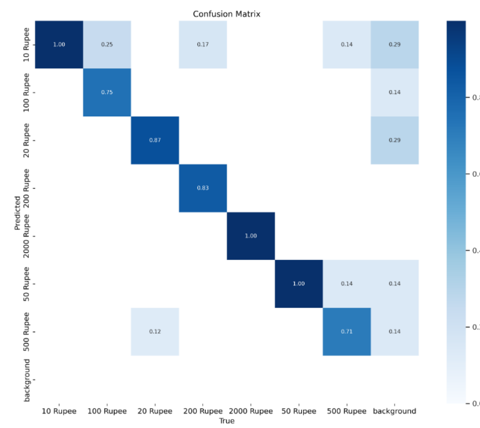


Fig. 2. Confusion matrix with proposed method on currency detection validation set

We achieved exceptional accuracy with our object and currency detection model using YOLOV5. Our model was able to accurately identify various types of objects and currency denominations. We conducted experiments using different objects found in our surroundings and various types of currency, and our model was able to detect them accurately.The results for object detection are shown in Figure 3 and the results for currency detection for different denominations are shown in Figure 4.

VI. Conclusion

Visually impaired individuals face a number of challenges in their daily lives, one of which is the difficulty in identifying

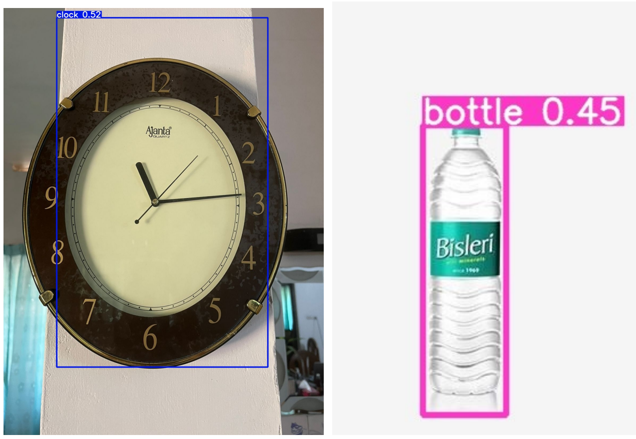


Fig. 3. Output: Clock and Bottle Detection

Fig. 4. Output: Currency Detection

objects and currency. For instance, visually impaired individuals may struggle with identifying everyday objects such as food items, clothing, and household appliances. This can make tasks such as grocery shopping, cooking, and cleaning more difficult and time-consuming. Similarly, visually impaired individuals may have difficulty identifying different denominations of currency, which can make transactions and financial management more challenging.

This project aims to develop an Android application designed to assist visually impaired individuals in recognizing the objects and currencies they encounter. The application uses object recognition technology to analyze the camera feed and generate an audio label for the object and currency based on its predicted confidence score. This provides the user with a way to identify and interact with the world around them, improving their independence and quality of life.

References

[1] 1. M. A. Khan, P. Paul, M. Rashid, M. Hossain and M. A. R. Ahad, "An AI-Based Visual Aid With Integrated Reading Assistant for the Completely Blind," in IEEE Transactions on Human-Machine Systems, vol. 50, no. 6, pp. 507-517, Dec. 2020.

[2] 2. K. K. S. N. Reddy, C. Yashwanth, S. KVS, P. A. T. V. Sai and S. Khetarpaul, "Object and Currency Detection with Audio Feedback for Visually Impaired," 2020 IEEE Region 10 Symposium (TENSYMP), Dhaka, Bangladesh, 2020, pp. 1152-1155.

[3] 3. D. Dahiya, H. Gupta and M. K. Dutta, "A Deep Learning based Real Time Assistive Framework for Visually Impaired," 2020 International Conference on Contemporary Computing and Applications (IC3A), Lucknow, India, 2020, pp. 106-109.

[4] 4. J. Nasreen, W. Arif, A. A. Shaikh, Y. Muhammad and M. Abdullah, "Object Detection and Narrator for Visually Impaired People," 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS), Kuala Lumpur, Malaysia, 2019, pp. 1-4.

[5] 5. L. Li, M. Liu, L. Sun, Y. Li and N. Li, "ET-YOLOv5s: Toward Deep Identification of Students' in-Class Behaviors," in IEEE Access, vol. 10, pp. 44200-44211, 2022.

[6] 6. S. Gianani, A. Mehta, T. Motwani and R. Shende, "JUVO - An Aid for the Visually Impaired," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, India, 2018, pp. 1-4.

[7] 7. Ghantous, M., Nahas, M., Ghamloush, M., Rida, M. (2014). iSee: An Android Application for the Assistance of the Visually Impaired. In: Hassanien, A.E., Tolba, M.F., Taher Azar, A. (eds) Advanced Machine Learning Technologies and Applications. AMLTA 2014. Communications in Computer and Information Science, vol 488. Springer, Cham.

[8] 8. F. Ashiq et al., "CNN-Based Object Recognition and Tracking System to Assist Visually Impaired People," in IEEE Access, vol. 10, pp. 14819-14834, 2022.

[9] 9. W. -J. Chang, L. -B. Chen, M. -C. Chen, J. -P. Su, C. -Y. Sie and C. -H. Yang, "Design and Implementation of an Intelligent Assistive System for Visually Impaired People for Aerial Obstacle Avoidance and Fall Detection," in IEEE Sensors Journal, vol. 20, no. 17, pp. 10199-10210, 1 Sept.1, 2020.

[10] 10. S. A. Jakhete, P. Bagmar, A. Dorle, A. Rajurkar and P. Pimplikar, "Object Recognition App for Visually Impaired," 2019 IEEE Pune Section International Conference (PuneCon), Pune, India, 2019, pp. 1-4.

[11] 11. B. Nayak, N. P. Thota, J. D. Kumar, D. R and B. S, "Currency Detector for Visually Impaired using Machine Learning," 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), Puducherry, India, 2021, pp. 1-5, doi: 10.1109/ICSCAN53069.2021.9526494.

[12] 12. S. M. Felix, S. Kumar and A. Veeramuthu, "A Smart Personal AI Assistant for Visually Impaired People," 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2018, pp. 1245-1250, doi: 10.1109/ICOEI.2018.8553750.

[13] 13. S. M. R. Bagwan and L. J. Sankpal, "VisualPal: A mobile app for object recognition for the visually impaired," 2015 International Conference on Computer, Communication and Control (IC4), Indore, India, 2015, pp. 1-6, doi: 10.1109/IC4.2015.7375665.

[14] 14. Y. Li, M. Li, S. Li and Y. Li, "Improved YOLOv5 for Remote Sensing Rotating Object Detection," 2021 6th International Conference on Communication, Image and Signal Processing (CCISP), Chengdu, China, 2021, pp. 64-68, doi: 10.1109/CCISP52774.2021.9639292.

[15] 15. H. Chen, S. Tan, Z. Xie and Z. Liu, "A new method based on YOLOv5 for remote sensing object detection," 2022 China Automation Congress (CAC), Xiamen, China, 2022, pp. 605-610, doi: 10.1109/CAC57257.2022.10055729.