The Critical Review of Social Sciences Studies

Online ISSN: 3006-7170
Print ISSN: 3006-7162
Volume 3, Number 3, 2025, Pages 2532 – 2545
Journal Home Page
https://thecrsss.com/index.php/Journal/about

Real-World Mapping Model for Visually Impaired People

Arjumand Rehan[1], Sapna Fazal[2], Taiba Naseem[3] & Aneesa Ahmad Zeb[4]

[1]*Govt. Frontier College For Women, Peshawar, Email: Arjumand.rehan@gmail.com*
[2]*Govt. Frontier College For Women, Peshawar, Email: sapnaafazal@gmail.com*
[3]*Govt. Frontier College For Women, Peshawar, Email: taibanaseem743@gmail.com*
[4]*Govt. Frontier College For Women, Peshawar, Email: Moon19042002@gmail.com*

ARTICLE INFO

Article History:

Received:	*July*	*11, 2025*
Revised:	*August*	*06, 2025*
Accepted:	*August*	*25, 2025*
Available Online:	*September*	*10, 2025*

Keywords:

A-EYE model, visually impaired

Corresponding Author:
Arjumand Rehan
Email:
Arjumand.rehan@gmail.com

ABSTRACT

The A-EYE model is an advanced assistive technology designed to empower visually impaired users by enabling them to perform daily tasks independently, reducing their reliance on sighted individuals. This AI-powered system provides real-time voice guidance, ensuring seamless interaction with its features. The object detection feature uses YOLOv4 on the COCO dataset to accurately identify everyday objects, enhancing users' spatial awareness. The face recognition feature utilizes Google ML Kit for detection and MobileFaceNet for recognition, allowing users to identify familiar faces effortlessly. For financial independence, the system incorporates currency recognition, employing a VGG19 model trained on a Pakistani currency dataset to distinguish different denominations accurately. Additionally, text recognition is powered by Firebase ML Kit, enabling users to read printed material with ease. The emergency shake feature ensures safety by automatically sharing the user's location with selected emergency contacts when the phone is shaken. The model's accuracy will be enhanced by refining training datasets and implementing advanced AI techniques, while the currency recognition model will be expanded to support a global currency dataset for greater versatility and usability worldwide.

Introduction

Visually impaired people face a lot of difficulties in their daily lives, especially when navigating unfamiliar surroundings. They often rely on sighted individuals for assistance, which can be limiting. The main idea of this research is to help these visually impaired people perform their daily activities without depending on someone else, developing a realtime-based vision system using object detection, face recognition, currency recognition, and text recognition for visually

impaired people's assistance. A deep learning model is trained with multi-class indoor and outdoor objects highly relevant to visually impaired people. To bring more robustness to the training images, training images are augmented and manually labeled. To ensure accuracy, the training images have been enhanced and manually labeled.

The main purpose is to develop a low-cost model that can detect and recognize surrounding obstacles and try to estimate the risk to the user, to facilitate visually impaired people in a challenging environment.

Background

Recent studies have made significant progress in the field of object detection, especially in applications related to assistive technology and real-time surveillance. A notable contribution is the paper titled "Object Recognition and Currency Detection for Visually Impaired People" (May 2024), which leverages the YOLO model to develop a real-time system that aids visually impaired individuals in recognizing objects and currencies; despite its efficiency, environmental factors and damage to currency notes can affect recognition accuracy .Similarly, the study "Currency Recognition For The Visually Impaired People" (2022) focuses on a mobile application that utilizes machine learning and image processing techniques to identify Indian banknotes, providing valuable assistance but facing difficulties in accurately recognizing damaged or faded notes, especially under varying lighting conditions

In another significant work, "VISION TXT: Visual Aid using Computer Vision" (December 2024), employs high-resolution cameras combined with YOLOv11n for real-time object detection, achieving an error rate of just 8%, although its performance diminishes in poor lighting or adverse weather conditions . The research "Realtime Object Detection Using OpenCV" (April 2024) explores the integration of deep learning algorithms with OpenCV for assistive applications, offering promising real-time detection and voice-based guidance systems; however, such approaches demand substantial data volumes and computational power . Additionally, "Object Detection and Classification using YOLOv3" (February 2021) emphasizes YOLOv3's capability for rapid and accurate detection but notes limitations in handling small objects and complex scene scenarios. Furthermore, a review by Ahmad & Arnd (2020) highlights that temporal context significantly improves object detection in videos and discusses the benefits of combining deep learning techniques with recurrent neural networks like LSTMs, which effectively capture spatial and temporal dependencies, although they introduce challenges related to model complexity and interpretability . These studies collectively illustrate the ongoing efforts to enhance object detection systems across various domains, emphasizing both the advancements and the persistent challenges faced in achieving robust and efficient

Proposed System

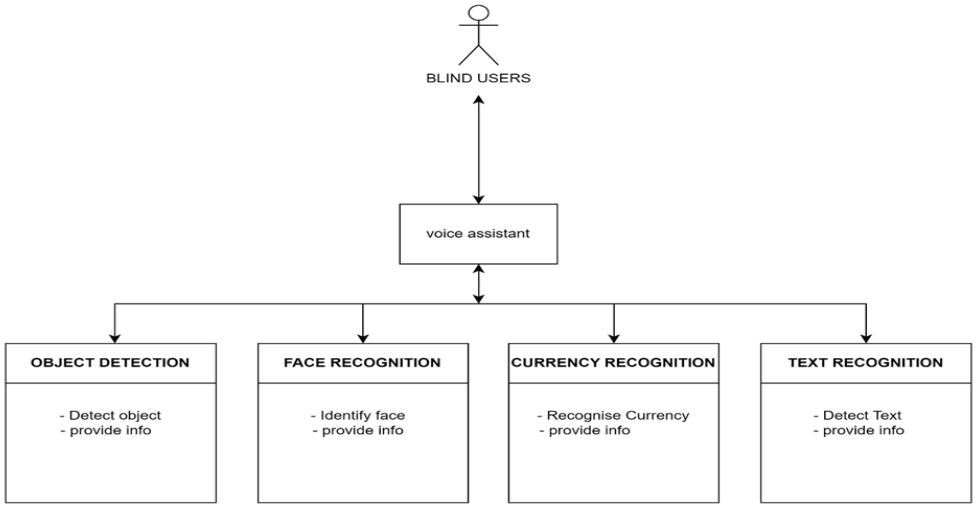
The proposed system A-EYE is designed to develop a whole solution for a visually impaired person through a multi-modal recognition system.

It consists of four core components: Face recognition, Currency recognition, Text recognition, and Object recognition. The module Face Recognition identifies people's faces by comparing them to a database of known Images, the Currency Recognition module recognizes money values through image processing techniques, and the Object Recognition module applies deep learning in the recognition of objects. Each of these modules is trained on an image dataset, thus enhancing recognition. It provides audio responses or vibrations so users are able to perceive and interact with their surroundings, countering those limitations of the current technologies.

Objectives of the Proposed system

1. To acquire knowledge and understanding of various ML algorithms, which would be applied to the design & deployment of "Real World Mapping System for Visually Impaired People".
2. Design an optimized, cost-effective, scalable, highly available, and secure model Using the best-fitting ML algorithm for the task.
3. The model should be used for objects, face, text as well as currency note recognition.
4. This system should bring ease and comfort into the lives of visually impaired people.

Methodology



Use Case Diagram

This Use Case diagram helps understand the system's behavior and requirements from a user's perspective.

Figure Use Case Diagram

Workflow

This System has the following six phases of development as shown in Figure 4.1.

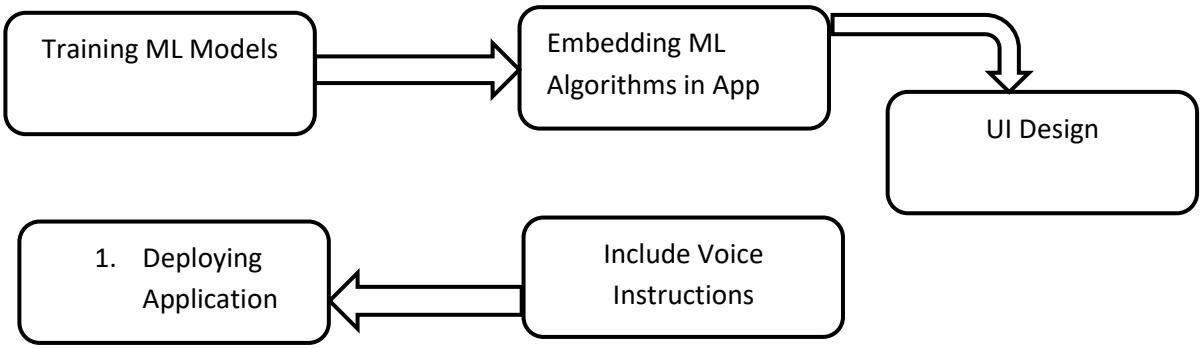


Figure: Workflow

Machine Learning (ML) Models

This research employs AI and machine learning models to enhance the quality of life for individuals with visual impairments. These models facilitate the identification of common objects like laptops, bottles, TVs, beds, and tables, enable facial recognition of family members and friends, and aid in recognizing different currencies.

Object Detection

Overview

This YOLO algorithm is trained on the COCO Dataset. COCO is a large-scale object detection, segmentation, and captioning dataset. It contains around 330,000 images out of which 200,000 are labeled for 80 different object categories.

Working Principle

The working principle of YOLOv4 involves these steps:

- **Input Image Acquisition:** The model Captures real-time input from a camera.
- **Image Division:** The input image is divided into a grid (e.g., a 3x3 grid).
- **Classification and Localization:** Image classification and localization are performed for each grid cell. The model predicts bounding boxes and their corresponding class probabilities for objects found within the grid cell.

This process allows YOLOv4 to simultaneously predict multiple bounding boxes and class probabilities in real-time.

Face Recognition

Overview

This project utilizes the MobileFaceNet model, specifically designed for real-time face recognition tasks on mobile and embedded devices. Before recognizing a face, it must first be detected. For this purpose, Google ML Kit's Face Detector. The detected face is then processed by the MobileFaceNet model to generate face embeddings.

Working Principle

The operational principle of MobileFaceNet includes the following steps:

- **Image Acquisition:** Initially, the system captures real-time input from a camera and utilizes a face detector to locate faces within it.
- **Face Embedding Generation:** MobileFaceNet processes the detected face to generate a vector of 128 values, which encapsulates the critical features of the face. This vector is referred to as an embedding.
- **Face Recognition:** For identifying a person, the embedding of an unknown face is computed and compared with previously stored embeddings in the local database. If the computed embedding is sufficiently like one of the stored embeddings, the system recognizes the face as that of a known individual.

Currency Recognition

Overview

Identifying currency involves the task of differentiating between various types of banknotes. This project uses a dataset that includes images of Pakistani currency notes, sourced from Kaggle. For currency recognition, the VGG19 model is used, a deep convolutional neural network (CNN) well-known for its effective image classification capabilities. VGG19 is structured with a total of 19 layers, with 16 layers designated for extracting features and 3 layers for classification purposes.

Working Principle

The working principle of VGG19 for currency recognition involves:

- **Input Image Acquisition:** The model takes an input image Capture real-time input from a camera of size 224x224 pixels.
- **Feature Extraction:** The first 16 layers of VGG19 extract features from the image using multiple 3x3 filters in each convolutional layer.
- **Classification:** The final 3 layers classify the extracted features to identify the currency note.

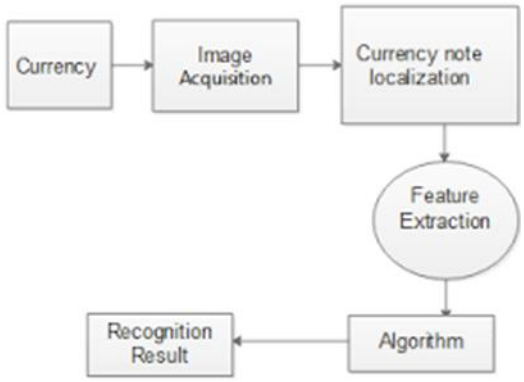


Figure: Currency Recognition Process

Text Recognition

Overview

This project utilizes Firebase ML Kit for on-device, real-time text recognition. Machine Learning SDK is optimized for offline usage and fast processing. That is why it is so good at recognizing and extracting text from every type of document and surface with its enhanced user interaction

Working Principle

With the Firebase ML Kit, the process of text recognition is as follows:

- **Image Capture:** The system captures a live image from the camera.
- **Text Detection:** Here, Firebase ML Kit scans the image and identifies the parts that contain text, thus breaking down the images into elements to identify areas with written text.
- **Text Recognition:** Once these text regions are identified the system converts those text into a digitized version. For this digitalized text, further processes may be carried out, such as voice out or showing in the screen.

- **Output:** The digitized text will either be saved to a device or used directly to provide audio with real time feedback, thus making the activity accessible for the user.

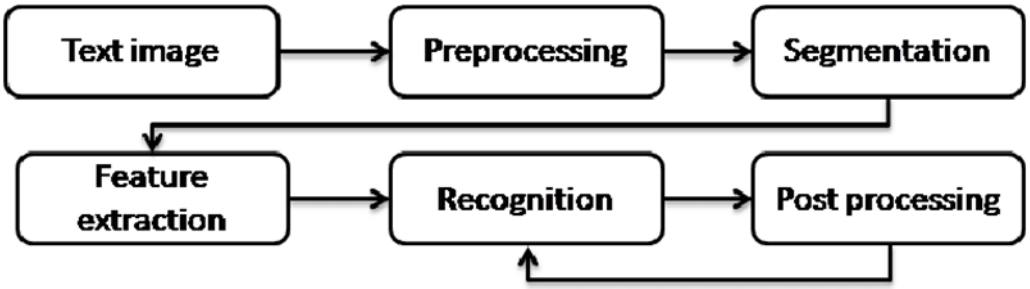
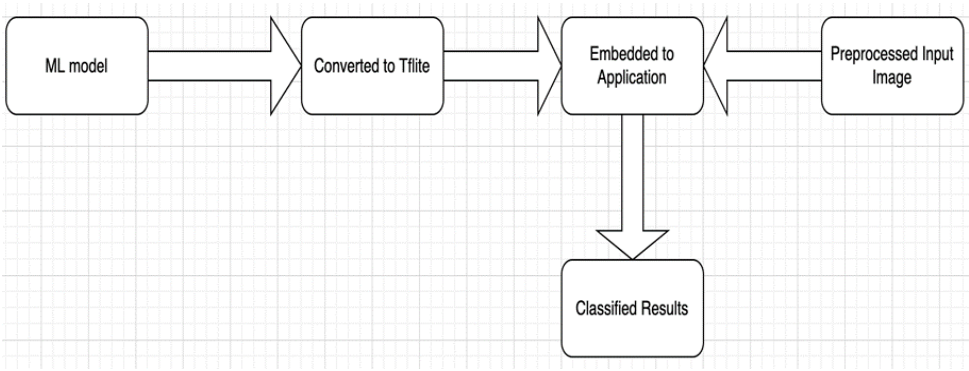


Figure: Text Recognition Process

Embedding ML Models in Application

The trained ML models are embedded in the Application as shown in Figure 5.3. In this way, the Application will be able to recognize Objects, Faces, Pakistani Currency Notes, and Text for visually impaired people, helping them become more independent of their sighted friends and family members. The Object Detection model can classify 80 common objects from the COCO dataset. For the face recognition model to work, it first need to save a few faces in the phone's local storage. The model compares the embeddings of an unseen image with the embeddings stored in the local storage. If any embeddings match those in the local database, the model will recognize that face,



User Interface Development

Before developing the front end, the basic screens of the user interface for A-EYE were designed using Adobe XD. Figure 5.1 shows the basic prototype of the system.

Voice Instructions

Since the system's target audience is visually impaired, a medium is needed to easily convey the results recognized by the ML models and assist users in navigating different parts of the application. For this purpose, the flutter_tts package from Flutter is utilized. Flutter_tts is a Flutter plugin for text-to-speech conversion, supported on iOS, Android, Web, and macOS. This plugin helps users navigate through the app and provides information about detected objects.

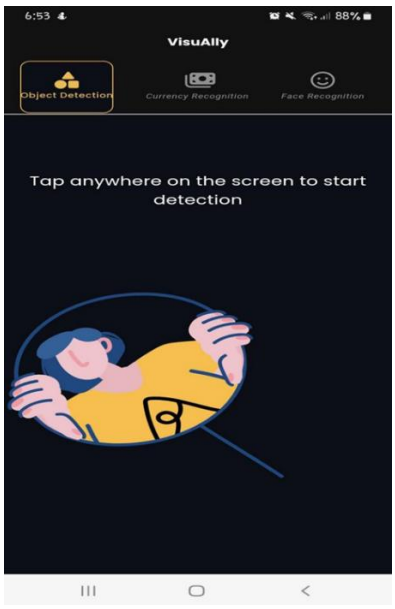
and the App will announce the name of the person. Similarly, the currency recognition model is trained to identify all seven Pakistani currency notes—10, 20, 50, 100, 500, 1000, and 5000 PKR. The ML model embedded in the mobile application can easily classify each of these notes. Furthermore, the Application uses Firebase ML Kit for one of text recognition feature whereby people can read printed and hand-written text from sources. This capability enhances functionality in the Application, since the visually impaired can read signs, labels, etc.

Results

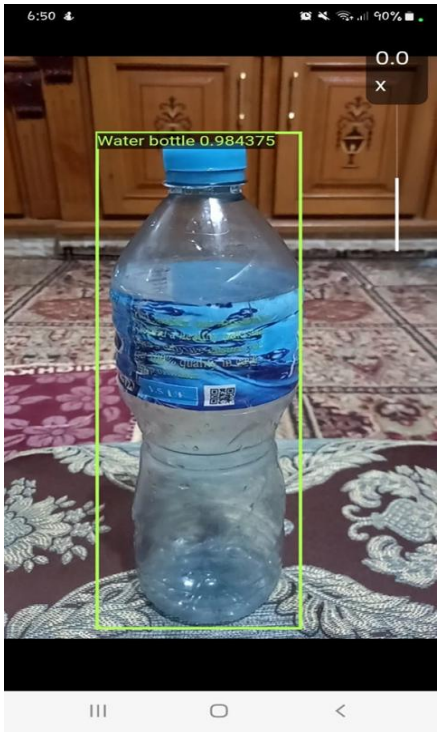
Testing Results

The section that opens when the application is launched is the Object Detection Screen. The app informs the visually impaired user about the current screen and how to navigate to another screen using Voice instructions. It also tells the user how to activate the Object Detection Model.

View



Camera View

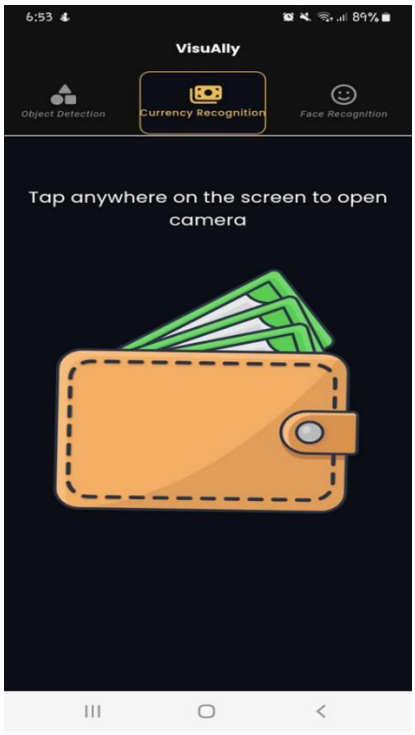


Voice Instructions

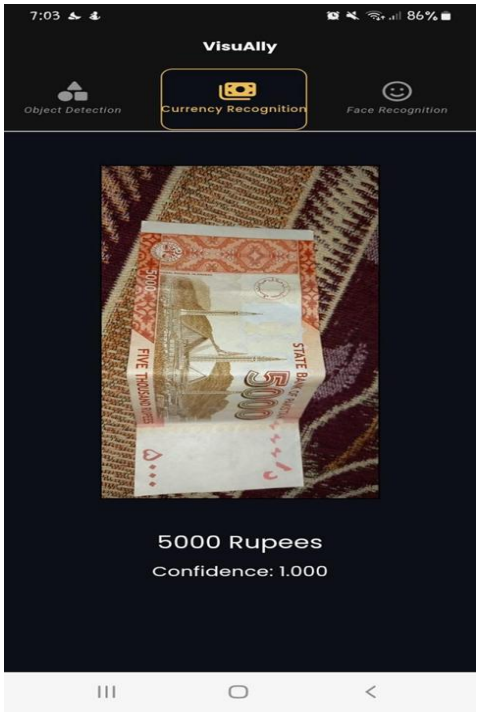
Object Detection Screen, tap anywhere to start detection or swipe left for Currency

Currency Recognition

View



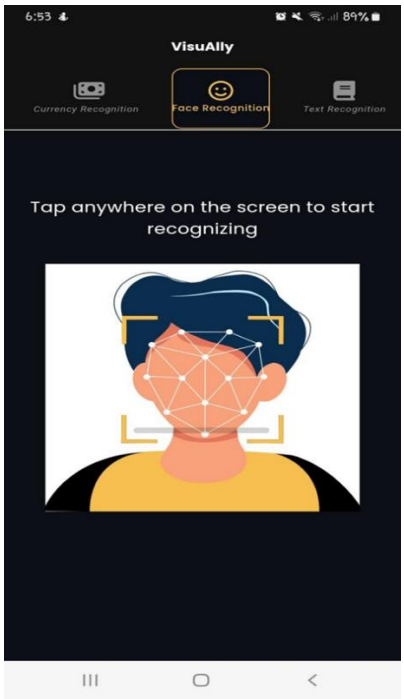
Camera View



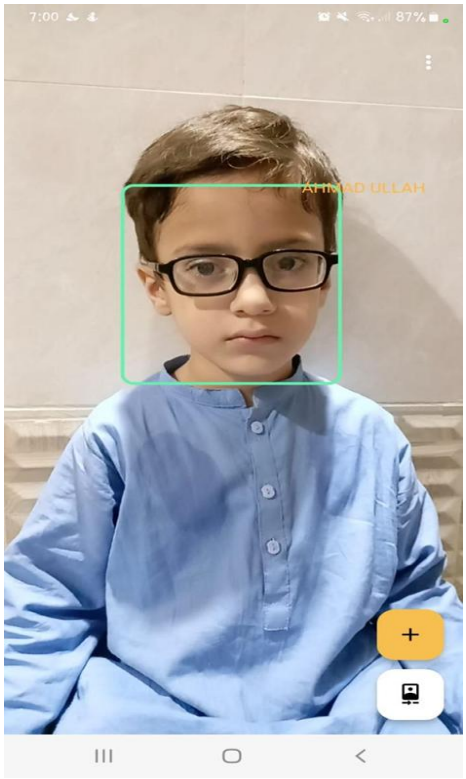
Voice Instructions: Face Recognition Screen, tap anywhere to start recognizing or swipe right for Currency Recognition.

Face Recognition

View



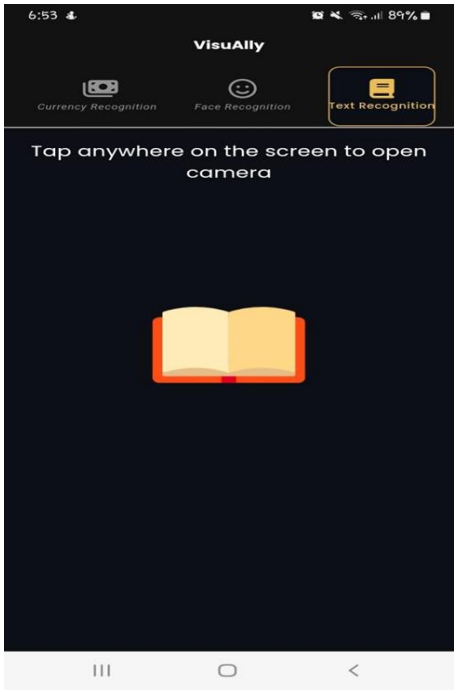
Camera View



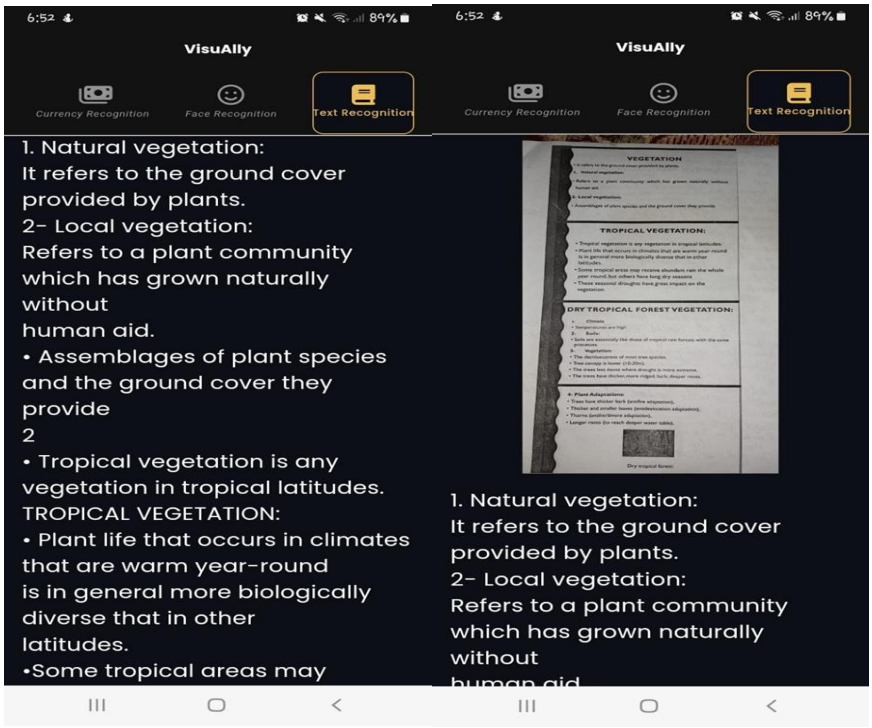
Voice Instructions: Face Recognition started, press, and hold anywhere to exit Face Recognition.

Text Recognition

View



Camera View

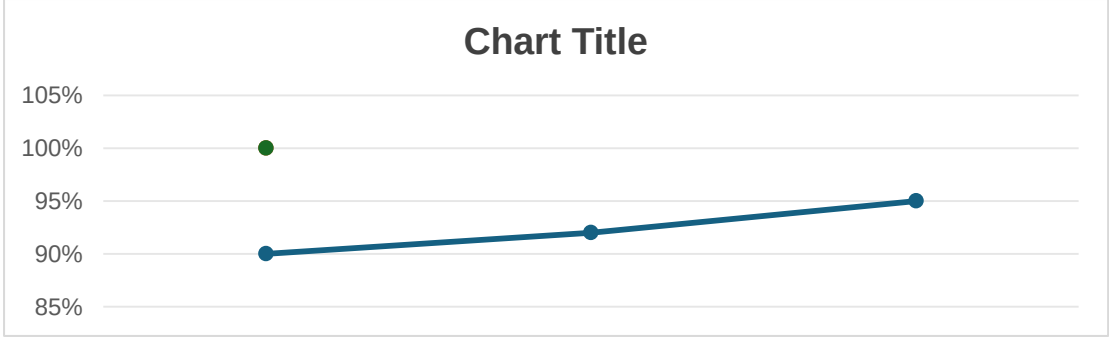


Voice Instructions: Text Recognition started, press and hold anywhere to exit Text Recognition.

Time Complexity and Efficiency

A-EYE employs sophisticated machine learning and deep learning techniques to aid users who are blind or visually impaired in real-time. Object detection is done simultaneously and accurately, enabled using YOLOv4. MobileFaceNet allows face recognition specifically geared towards mobile and embedded systems. VGG-19 provides high precision and accuracy in recognizing foreign currency. All these components work together in helping visually impaired users gain independence and quality of living by providing all round and effective help.

Component	Purpose	Time Complexity	Efficiency	Accuracy
YOLOv4	Object Detection	≈0.2s per image	High speed and accuracy	90%
MobileFaceNet	Face Recognition	≈0.5s per face	Lightweight and efficient	92%
VGG-19	Currency Recognition	≈1.0s per note	High accuracy in classification	95%
Firebase ML Kit	Text Recognition	≈0.7s per text block	Fast and reliable	90%



Discussion

The paper presents a comprehensive approach to developing a real-world mapping system tailored for visually impaired individuals, leveraging advanced machine learning (ML) and deep learning techniques. The system integrates several modules—including object detection, face recognition, currency recognition, and text recognition—to assist users in navigating and interacting with their environment more effectively. The use of models such as YOLOv4 for object detection, MobileFaceNet for facial recognition, VGG19 for currency identification, and Firebase ML Kit for text recognition ensures a multi-modal recognition approach that addresses a wide range of real-world scenarios faced by visually impaired users. The methodology emphasizes real-time processing, affordability, and user-friendliness, aiming to replace or supplement current assistive tools like TapTapSee, Seeing AI, and others with higher accuracy and responsiveness.The inclusion of voice commands and audio feedback enhances accessibility, enabling users to receive immediate information about their surroundings through intuitive interactions. The development process incorporates training ML models on augmented datasets to improve robustness and recognition accuracy. The system's modular design facilitates ease of updates and potential integration of additional functionalities in future iterations.However, some challenges persist, such as dependency on camera quality, computational resource limitations, and battery consumption, which can influence system reliability in real-world environments. Despite these limitations, the proposed system shows significant promise in improving autonomy and quality of life for visually impaired users, emphasizing scalability and cost-effectiveness.

This research successfully demonstrates the feasibility of a multi-modal, machine learning-based real-world mapping system designed to assist visually impaired individuals. By integrating object detection, face recognition, currency identification, and text recognition modules, the system provides comprehensive assistance that enhances environmental awareness and independence. The use of lightweight and scalable models ensures that the system can operate efficiently in real-time on mobile devices, making it accessible and practical for everyday use. The focus on audio feedback and voice instruction further aligns with user needs for hands-free operation.

The system addresses critical gaps in existing assistive technologies, particularly in accuracy, speed, and multi-modal recognition capabilities. While there are challenges related to hardware limitations and energy consumption, these do not overshadow the system's potential to significantly improve navigational safety and social interaction for visually impaired users.

Future Work

The future work of this research involves enhancing the system's capabilities through the integration of additional sensors like ultrasonic or LiDAR for improved obstacle detection, optimizing machine learning models for better efficiency and lower power consumption, and incorporating gesture and advanced voice commands to make user interaction more intuitive. Further, developing context-aware features tailored to different environments, conducting user trials for feedback-driven improvements, and exploring cloud-edge hybrid processing will help boost performance and responsiveness. Additionally, efforts to reduce hardware costs and promote accessible, open-source solutions aim to make the system more affordable and widely available, ultimately advancing the independence and safety of visually impaired users in diverse real-world scenario.

References

1. Ashraf, M. M., Hasan, N., Lewis, L., Hasan, M. R., & Ray, P. (2016). A systematic literature review of the application of information communication technology for visually impaired people. International Journal of Disability Management, 11, e6.
2. TapTapSee. (2024). TapTapSee: Helping the blind and visually impaired. Retrieved from https://taptapseeapp.com/
3. Kavya, S., & Shetty, M. (2019). Assistance system for visually impaired using AI. International Journal of Engineering Research & Technology (IJERT). Retrieved from https://www.ijert.org/assistance-system-for-visually-impaired-using-ai
4. aczmirek, L., & Wolff, K. G. (n.d.). Survey design for visually impaired and blind people.
5. Seeing AI - Talking camera for the blind. (2024, May 6). Retrieved from https://www.seeingai.com/
6. Noteify: Indian currency recognition app. (n.d.). Retrieved from https://github.com/chandran-jr/Noteify
7. Samrgandi, N. (2021). User interface design & evaluation of mobile applications. Retrieved from https://www.researchgate.net/publication/349087972_User_Interface_Design_Evaluation_of_Mobile_Applications
8. huyang. (2022, January 4). How machine learning can help visually impaired people: An outdoor obstacle identification project proposal for blind people with DeepLabv3+. Towards Data Science. https://towardsdatascience.com/how-machine-learning-can-help-visually-impaired-people-an-outdoor-obstacle-identification-project-proposal-for-blind-people-with-deeplabv3-61baf5d907ce
9. Real-time object detection with speech recognition using TensorFlow Lite. (2024, May 6). Retrieved from https://www.researchgate.net/publication/359393141_REAL_TIME_OBJECT_DETECTION_WITH_SPEECH_RECOGNITION_USING_TENSORFLOW_LITE
10. Ahmad, Z. M., Balogun, A. O., Muazu, A. A., & Mamman, H. (2024). MobileFaceNet-based facial recognition system for contactless access control. PLATFORM: A Journal of Science & Technology, 7(1). https://doi.org/10.61762/pjstvol7iss1art27053
11. Bardhi, M. (2021, January 2). Image detection using the VGG-19 convolutional neural network. Medium. https://medium.com/@melisa.bardhi/image-detection-using-the-vgg-19-convolutional-neural-network-3e4d973b7bc1
12. GeeksforGeeks. (n.d.). Text detector in Android using Firebase ML Kit. GeeksforGeeks. Retrieved February 18, 2025, from https://www.geeksforgeeks.org/text-detector-in-android-using-firebase-ml-kit/
13. Neoh, D. K. S., Lock, E. C. M., Lew, K. H., Tang, H. Y., Tong, D. L., & Tan, T. T. (2021). PicToText: Text recognition using Firebase Machine Learning Kit. In Proceedings of the 2021 2nd International Conference on Artificial Intelligence and Data Sciences (AiDAS) (pp. 1-6). IEEE. https://doi.org/10.1109/AiDAS53897.2021.9574187
14. Codecademy. (2025). Catalog: Flutter articles and resources. Codecademy. https://www.codecademy.com/resources/flutter
15. Coursera Staff. (2024, September 10). What is Kaggle and what is it used for? Coursera. https://www.coursera.org/articles/what-is-kaggle-and-what-is-it-used-for
16. Google. (n.d.). ML Kit: Machine learning for mobile developers. Google Developers. https://developers.google.com/ml-kit
17. Purdila, A. (2024, January 2). What is Adobe XD? Envato Tuts+. https://tutsplus.com/tutorials/what-is-adobe-xd

18. Ganesan, Veeramani. (2022). Machine Learning in Mobile Applications. International Journal of Computer Science and Mobile Computing. 11. 110-118. 10.47760/ijcsmc.2022.v11i02.013. https://www.researchgate.net/publication/358908273_Machine_Learning_in_Mobile_Applications
19. PowerMapper. (2022). Apps for people with visual impairment. Retrieved from https://www.powermapper.com/resources/guides/apps-vision-impairment/
20. Ye, H., Oh, U., & Findlater, L. (2014). Current and future mobile and wearable device use by people with visual impairments. Retrieved from https://dl.acm.org/doi/abs/10.1145/2556288.2557085