Govt. Model Engineering College, Kochi
Department of Computer Engineering
B.Tech Computer Science & Engineering
CSD415 Project Phase 1

S.A.G.E - Personalized AI-Powered Smartglass

LITERATURE SURVEY

Group 05

September 2025

Group Members:

- Ananya P (22CSB08 MDL22CS036)
- M Gayathri (22CSB36 MDL22CS126)
- Navneet Ranjish Pillai (22CSB41 MDL22CS138)
- Nikhil M (22CSB44 MDL22CS144)

1 Introduction

The AI-powered Smart Glass is an innovative wearable designed to enhance real-world interaction through advanced vision and voice-based features. Built on the lightweight yet powerful Raspberry Pi Zero 2 W, the device emphasizes software-driven functionality with minimal hardware complexity, ensuring portability and efficiency.

A custom Flutter mobile application is integrated to manage AI tasks, provide user controls, and enable real-time communication, creating a seamless bridge between the user and the smartglass. Key features include a voice-triggered assistant, HUD-based feedback, facial recognition, object detection, and real-time translation, all aimed at augmenting user awareness and interaction with their surroundings.

At its core lies S.A.G.E (Situational Awareness and Guidance Engine), a context-aware companion that empowers users with intelligent insights, guidance, and support, wherever they go. This makes the AI-powered Smart Glass not just a wearable, but a situational companion for smarter living.

2 Articles

2.1 "Visual Information Translator Using Smart Glasses for Blind" (2023), by T.P. Rani, T. Vignesh, S. Susila Sakthy, M. Priyadharshan, and P. Kalaichelvi[1]

This paper presents the development of a smart glasses system designed to assist visually impaired individuals by translating visual information into audio output. By integrating Internet of Things (IoT) and Machine Learning (ML) methodologies, the system aims to improve the daily lives of people with vision-related issues, fostering greater independence, safety, and self-reliance. The primary goal is to provide real-time auditory feedback about the user's surroundings, enabling them to read text, recognize faces, and navigate their environment more effectively.

Methodology: The system's methodology is based on a wearable, hands-free device that captures visual data and processes it through a connected mobile application. A camera mounted on the smart glasses captures real-time images or video of the user's environment. This visual data is transmitted wirelessly via Bluetooth or Wi-Fi to a dedicated mobile app. The app then utilizes machine learning models to process the data for tasks such as text extraction, object detection, and facial recognition. Finally, the processed information is converted into speech using a text-to-speech engine and relayed to the user as audio through a headset or speaker.

Implementation Details: The hardware components include a microcontroller (such as a Raspberry Pi), a Pi camera, and a Bluetooth/Wi-Fi module mounted on the glasses. The software consists of a mobile application developed using the **Flutter** framework. This app serves as the primary interface and processing hub, employing on-device ML frameworks like **TensorFlow Lite** and **Google ML Kit** for its core functions. The system is built around four main modules: **Text-to-Speech** for reading text from images, **Object/Obstacle Detection** for identifying objects in the user's path, **Facial Recognition** for identifying known individuals, and an **SOS Message** feature that sends an emergency alert with the user's location to predefined contacts. The app also includes features for user personalization and a health report database.

Advantages:

- **Enhanced Independence:** Allows users to navigate their environment, read text, and recognize objects autonomously, reducing reliance on others.
- **Improved Safety:** The object detection module helps users avoid obstacles, while the SOS button provides a quick way to call for help in emergencies, enhancing the user's sense of security.
- **Multi-functional Solution:** Integrates several assistive technologies (text reader, obstacle detector, facial recognizer) into a single wearable device.
- **Cost-Effective Design:** Utilizes readily available components and open-source software to create an affordable solution.
- **User-Friendly Interaction:** A dedicated mobile application provides a simple interface for controlling the smart glasses and managing features.
- **Better Social Engagement:** The facial recognition feature can help users identify friends and family, making social interactions easier.

Disadvantages:

- **Dependency on a Smartphone:** The system is not standalone; it requires a constant wireless connection to a smartphone where all the ML processing occurs, which could be a point of failure.

- **Battery Life Concerns:** The continuous operation of a camera, microcontroller, and wireless module on the glasses, coupled with intensive processing on the paired phone, may lead to significant battery drain for both devices.
- **Reliance on ML Accuracy:** The system's effectiveness is entirely dependent on the accuracy of the ML models. Errors in object, text, or face recognition could provide misleading or incorrect information to the user.
- **Potential for Latency:** The wireless data transmission between the glasses and the phone could introduce latency, potentially delaying the audio feedback for the user in critical real-time navigation scenarios.

2.2 "Smart Glasses Embedded with Facial Recognition Technique" (2023), by M. Kumar, B. Bharti, and U. Chauhan[2]

This paper presents the design of smart glasses for visually impaired individuals, integrating facial recognition and voice assistance to enhance their daily lives. The system's primary goal is to help users recognize people around them, detect obstacles, and measure distances, providing all information as spoken audio output. Built on a Raspberry Pi, the project aims to be an affordable, lightweight, and user-friendly digital assistant that also includes features like web browsing and music playback.

Methodology: The system's operation begins when a camera mounted on the glasses captures an image of a person or object in the user's vicinity. This visual input is processed by a **Raspberry Pi**, which runs the facial recognition algorithm. The system compares the captured face against a pre-loaded database to find a match. Simultaneously, an ultrasonic sensor detects obstacles and calculates their distance from the user. The identified name (or "unknown" if no match is found) and the distance information are then converted into speech. This audio is relayed to the user through a connected speaker or headphone, providing real-time feedback about their environment.

Implementation Details: The core of the prototype is a **Raspberry Pi 4**, which serves as the central processing unit. The hardware setup includes a camera for visual input, an ultrasonic sensor for obstacle and distance detection, and a speaker or headphone for audio output. Internet connectivity is managed via a lane (Ethernet) cable or a Wi-Fi hotspot. The software is developed in the **Python** programming language and utilizes the **Google Text-to-Speech** service for audio conversion. A tool called **VNC Viewer** is used to monitor the system's performance and to add new faces to the database. Notably, new faces and objects must be manually added to the system by uploading an image and updating the code with the corresponding name.

Advantages:

- **Affordable:** The design is described as "extremely affordable," centered around the low-cost Raspberry Pi microprocessor.
- **Multifunctional:** It combines facial recognition for social interaction with ultrasonic obstacle detection for safe navigation.
- **User-Friendly and Lightweight:** The paper emphasizes that the device is designed to be simple to use and physically light.
- **Digital Assistant Features:** Beyond its primary functions, the system includes extra capabilities like web browsing and playing music, making it a more versatile tool for the user.

Disadvantages:

- **Static Database:** The system can only recognize people or objects that have been manually pre-loaded into its database and code; it cannot identify unfamiliar objects or people on its own.
- **Network Dependency:** Future work aims to make the project run without Wi-Fi, which implies the current version requires an internet connection to function, likely for the text-to-speech service.
- **Manual Updates:** Adding a new person to the system is not a simple process, as it requires the user to upload an image and manually edit the code to include the person's name.
- **Hardware Limitations:** The device's capabilities are constrained by the storage capacity of the Raspberry Pi's SD card.

2.3 "Facial Recognition Smart Glasses for Visually Impaired People" (2023), by A. Ghodake, H. Sale, A. Kamble, K. Mankari, O. Lad, and V. Mishra[3]

This paper details the development of smart glasses designed to improve the quality of life for visually impaired individuals by enhancing their independence and safety. The core objective is to create a discreet, wearable tool that assists with navigation and recognition tasks. The system is engineered to identify people from a known database, detect and label multiple objects in real-time, read text using Optical Character Recognition (OCR), and measure the distance to nearby obstacles, providing all feedback to the user through audio.

Methodology: The system's methodology is built around a dual-sensor approach controlled by a **Raspberry Pi**. A webcam continuously captures the user's surroundings. If an object is detected within the camera's frame, the system employs the **YOLO** algorithm for real-time object identification and labeling. It also performs facial recognition by comparing faces against a pre-existing database. For text, it uses the **Tesseract OCR** engine to extract written information. If the camera frame is clear of objects, an ultrasonic sensor takes over to measure the distance to the nearest obstacle. If the distance is below a safe threshold, an audio warning is triggered. All outputs—whether object names, recognized person's name, read text, or obstacle warnings—are converted to audio via a Text-To-Speech (TTS) module and relayed to the user through earphones. The device also includes a specialized mode for currency detection.

Implementation Details: The hardware is centered on a **Raspberry Pi**, which processes all data. The input devices consist of a webcam and an ultrasonic sensor, which has a sensing range of 2 to 400 cm. Audio output is delivered through earphones. For software, the project leverages several powerful algorithms: the **YOLO** (You Only Look Once) algorithm for its speed and accuracy in real-time object detection, and the **Tesseract** engine (implemented via the **Pytesseract** Python module) for OCR. A separate trained **Keras** model is used for currency denomination. The facial recognition database is created by storing named image files (e.g., "Person Name.jpg") in a designated folder.

Advantages:

- **Comprehensive Functionality:** Integrates multiple advanced features, including real-time object detection, facial recognition, text-to-speech, and currency recognition.

- **Multiple Detections:** The system is capable of identifying and labeling multiple distinct objects and people within a single camera frame.
- **Hybrid Sensing System:** It combines computer vision for identification with an ultrasonic sensor for reliable distance measurement, creating a robust assistive tool.
- **Real-time Performance:** Utilizes the fast and efficient YOLO algorithm to provide immediate feedback, which is crucial for navigation.

Disadvantages:

- **Fragility:** The paper explicitly states that the device can be easily broken or damaged, as the glass frames may not be strong enough to support the attached components.
- **Power Dependency:** The reliance on battery power is a significant concern. A dead battery while the user is in an unfamiliar environment could create a difficult situation.
- **Potential for High Cost:** Depending on the processor and other components used, the final device could be expensive.
- **Risk of Malfunction:** Any failure in the processor or sensors could lead to incorrect feedback or a complete lack of warnings, posing a safety risk to the user.

2.4 "Single-Shot Image Recognition Using Siamese Neural Networks" (2023), by A. Malhotra[4]

This paper addresses the challenge of one-shot image recognition, a scenario where machine learning models must accurately classify new categories after seeing only a single example of each. Traditional deep learning models often fail in such data-scarce situations. This work proposes a method using deep convolutional Siamese Neural Networks, which are trained to learn a similarity function between images, enabling them to generalize to entirely new classes without retraining.

Methodology: The core of the methodology is to train a Siamese Neural Network on a verification task rather than a direct classification task. A Siamese network consists of two identical "twin" subnetworks that share the same weights and architecture. The network is presented with a pair of images and processes each through one of the twins to extract high-level feature vectors. The network's goal is to learn a metric that determines whether the two images belong to the same class or different classes. By training on a large dataset of "same" and "different" pairs, the network becomes an excellent feature extractor. Once this training is complete, the learned feature representation can be used for one-shot classification tasks on new, previously unseen categories by comparing a test image to the single provided example of each new class.

Implementation Details: The architecture is a deep convolutional Siamese network featuring multiple convolutional layers with **ReLU** activation functions and max-pooling, followed by fully-connected layers. The model was trained using standard backpropagation, and its hyperparameters, such as learning rate and momentum, were optimized using a Bayesian optimization tool called **Whetlab**. To improve the model's robustness, the training process included data augmentation through random **affine distortions** like rotation, scaling, and shear. The primary dataset used for experiments was the **Omniglot** dataset, which contains a large variety of handwritten characters from 50 different alphabets.

Advantages:

- **High Accuracy on One-Shot Tasks:** The proposed convolutional Siamese network achieved 92 percent accuracy on the Omniglot one-shot classification task, outperforming other deep learning models.
- **No Retraining Needed for New Classes:** The most significant benefit is that once the model is trained on the verification task, it can generalize to new classes without any additional training or fine-tuning.
- **Relies Only on Pixel Data:** Unlike some competing models like HBPL, this approach does not require any additional metadata, such as stroke order or symbol structure, making it more broadly applicable.
- **Robust Feature Learning:** Training with affine distortions helps the network learn features that are invariant to common transformations, making it more robust in real-world scenarios.

Disadvantages:

- **Complex Training Process:** The model is not trained for simple classification but on a verification task, which requires creating image pairs and using sophisticated Bayesian methods to optimize a large set of hyperparameters.
- **Indirect Classification:** The network learns a similarity function, not direct class labels. For one-shot tasks, a test image must be compared against every available new class example, which can be less efficient than direct classification if the number of new classes is large.
- **Slightly Sub-Par Performance:** While highly effective, its 92 percent accuracy was slightly lower than the state-of-the-art Hierarchical Bayesian Program Learning (HBPL) model on the same dataset.

2.5 "Smart Glasses for Visually Impaired People using Machine Learning" (2023), by S. Siva Priyanka, A. Sai Kumar, M. V. Nagabhushanam, D. Vennela, and P. Divya Tulasi[5]

This paper proposes a smart glasses prototype designed to assist visually impaired individuals by providing real-time information about their surroundings. The system aims to enhance user safety and independence by integrating object recognition and text-to-speech technologies. By identifying objects and conveying their names and positions through audio, the glasses are intended to be a user-friendly tool that helps users navigate their environment with greater confidence.

Methodology: The system's workflow begins with a camera module capturing live images from the user's perspective. These images are processed on a **Raspberry Pi** using the **OpenCV** computer vision library. The core of the object detection is the **YOLOv3** (You Only Look Once) algorithm, a high-accuracy, real-time Convolutional Neural Network (CNN) pre-trained on the COCO dataset. YOLOv3 analyzes the image frames, identifies objects, and places bounding boxes around them with corresponding class labels. The names and positions of the detected objects are then converted into a natural-sounding audio description using the **Google Cloud Text-to-Speech API**. This spoken output is played back to the user through a speaker, providing them with timely warnings and environmental awareness.

Implementation Details: The hardware for this system consists of a **Raspberry Pi 3** as the main processor, a compatible camera module for capturing images, and a speaker for audio output. The entire software stack runs on the Raspberry Pi. Key software components include the **OpenCV** library for image processing tasks and the **YOLOv3** algorithm for object detection. The pre-trained YOLOv3 model enables the recognition of approximately 80 different object classes. For the speech synthesis component, the system relies on an external **Google API**.

Advantages:

- **High Precision:** The system demonstrates high class precision in object detection, with reported results of 99% for objects like keyboards and an overall precision of 0.99 percent.
- **Real-Time Processing:** The use of the efficient YOLOv3 algorithm allows the device to detect objects relatively quickly, providing immediate feedback to the user.
- **User-Focused Design:** The technology is specifically designed to be user-friendly, aiming to increase the confidence, ease of movement, and independence of visually impaired individuals.
- **Multi-Object Detection:** The implementation is capable of detecting and identifying multiple different objects within a single frame and providing a comprehensive audio description.

Disadvantages:

- **Internet Dependency:** The system's text-to-speech functionality is dependent on the Google API, which requires a constant internet connection to work.
- **Limited Object Recognition:** The model is limited to the approximately 80 object classes pre-defined in the COCO dataset, and cannot identify any objects outside of this scope.
- **Lack of Distance Measurement:** The design does not include hardware like an ultrasonic sensor, so it can identify an object and its position in the frame but cannot provide crucial information about its distance from the user. - **Older Hardware:** The use of a Raspberry Pi 3, an older model, may impose limitations on processing speed and the ability to run more advanced or larger machine learning models in the future.

2.6 "Indoor Navigation Glasses for the Visually Impaired" (2024), by P. J. Loresco, R. J. C. Cruz, K. Z. Ramones, J. A. D. Zafra, and K. R. G. Ramirez[6]

This paper introduces a novel indoor navigation system for visually impaired individuals, utilizing smart glasses equipped with deep learning models and an audio-based guidance system. The research addresses the limitations of existing assistive technologies, which often lack comprehensive navigation and obstacle avoidance capabilities. The developed prototype aims to provide a complete, infrastructure-independent solution for localization, navigation, and obstacle management to promote greater independence and spatial awareness for users in unfamiliar indoor environments.

Methodology: The system's methodology is centered around four key deep learning-driven functions: localization, navigation, obstacle detection, and obstacle avoidance. Users interact with the device through an **Audio User Interface (AUI)**, using voice commands like "location" to identify their current room or "go to [destination]" to initiate turn-by-turn navigation. For localization and navigation, a **MobileNetV2** model classifies the room based on the live feed from a camera. It then provides simple audio cues like "slowly forward" or "slowly left" to guide the user. For obstacle detection, a **MobileNet-SSD** model, specifically trained to identify chairs, is employed. When an obstacle is detected, the avoidance algorithm analyzes its position within the camera's view and instructs the user to move left or right accordingly to safely bypass it.

Implementation Details: The hardware consists of 3D-printed glasses made from lightweight **PETG** material, mounting a **Raspberry Pi v2 camera**. The core processing is handled by a separate housing unit containing a **Raspberry Pi 4B** and a **Google Coral Edge TPU**, which accelerates the inference speed of the deep learning models for real-time performance. The software stack uses a **MobileNetV2** model for room classification and a **MobileNet-SSD** model for obstacle detection. The voice-controlled AUI is powered by the **Vosk** speech recognition toolkit. The entire system is built using Python with key libraries such as **OpenCV**, **Pyaudio**, and **PyCoral**.

Advantages:

- **Excellent Performance Metrics:** The system demonstrated outstanding results in testing, with accuracy scores of approximately 99% for navigation and 99.6% for obstacle detection.
- **Real-Time Processing:** The integration of the **Google Coral Edge TPU** ensures extremely fast, low-latency processing, with an average user interface response time of only 0.65 milliseconds.
- **High User-Rated Usability:** Based on the standardized System Usability Scale (SUS), the device received an average score of 85, placing it in the "Best Imaginable" category for usability and intuitiveness.
- **Infrastructure-Free Navigation:** As a purely vision-based system, it does not depend on pre-installed hardware like RFID tags or beacons, allowing it to function in any indoor space.

Disadvantages:

- **Limited Obstacle Detection Scope:** The obstacle detection model was trained to recognize only chairs and would fail to identify other common indoor obstacles.
- **Constrained Testing Environment:** The system's validation was conducted in a single, controlled condominium unit, so its effectiveness in larger, more complex, or dynamic public spaces remains unproven.
- **Narrow Field of View:** The camera's limited viewing angle means the system cannot detect hazards outside the user's direct line of sight, posing a risk from peripheral obstacles.
- **Bulky Two-Part Design:** While the glasses are lightweight, the main processing components are located in a separate housing unit that must be carried by the user, which could be inconvenient.

2.7 "Google Pi Using Raspberry Pi for Visually Impaired" (2024), by G. V. Katkar, K. Abirami, A. M. Posonia, B. Ankayarkanni, and D. Ushanandini[7]

This paper introduces "Google Pi," a voice assistant system specifically designed to enhance the independence and security of visually impaired individuals. Built upon the low-cost and versatile **Raspberry Pi**, the system integrates the familiar **Google Assistant API** with custom deep learning models for real-time object and currency recognition. The project aims to address the shortcomings of traditional voice assistants by providing essential, vision-oriented functionalities that promote greater financial autonomy and environmental awareness for its users.

Methodology: The system operates through a voice-driven interface. A user issues a command, which is captured by a microphone and processed by the **Google Assistant API**. Based on the user's request, the system activates one of its core modules. For general queries, it functions as a standard voice assistant. For visual tasks, it triggers one of two deep learning models. For currency recognition, a camera captures an image of a banknote, which is then analyzed by a custom-trained **Convolutional Neural Network (CNN)** to identify its denomination. For object recognition, the system uses a pre-trained **YOLOv3** model to detect and identify objects in the camera's live feed. In all cases, the result is converted into a synthesized audio response and delivered to the user.

Implementation Details: The hardware foundation of the Google Pi is a **Raspberry Pi** single-board computer, connected to a microphone for input and a speaker or headphones for audio output. A Pi Camera or a standard webcam is used to capture visual data. On the software side, the voice interaction is managed through the **Google Assistant API**. The currency recognition module employs a custom **CNN** model trained on a dataset of 1,786 Indian currency images, achieving a test accuracy of approximately 92.5%. The object recognition module utilizes a pre-trained **YOLOv3** model implemented with the **OpenCV** library. The system is also designed to be user-friendly, allowing for personalized custom voice commands.

Advantages:

- **Integrated Core Functions:** The device successfully combines a standard, powerful voice assistant with essential, specialized tools for currency and object recognition on a single platform.
- **Cost-Effective and Accessible:** By using the affordable Raspberry Pi, the project makes advanced assistive technology financially accessible to a wider audience.
- **High Accuracy for Currency Recognition:** The custom-trained CNN model provides reliable and accurate identification of Indian currency notes, with a test accuracy of 92.47%.
- **Familiar and Customizable Interface:** The use of the Google Assistant API offers an intuitive user experience, which is further enhanced by the ability to create custom voice commands.

Disadvantages:

- **Requires Significant Computing Power:** The paper notes that a drawback of the system is its need for substantial computing power, which could be a challenge for the resource-constrained Raspberry Pi.
- **Internet Connectivity is Essential:** The core voice assistant functionality is entirely dependent on the Google API, meaning the device cannot function without a stable internet connection.
- **Lacks Navigational Assistance:** The system focuses on identification tasks and does not include any features for navigation or obstacle avoidance, which are critical for user mobility.
- **No Miniaturization or Wearable Form Factor:** The project is presented as a proof-of-concept system, and the paper suggests future work could explore miniaturization and integration into a more discrete, wearable device.

2.8 “Smart Glass for Visually Impaired Person” (2025), by R. Sabitha, S. Senthil Pandi, J. Rathi Devi, and G. Jegan[8]

This paper presents a ”Smart Glass” system designed to enhance the safety, mobility, and emergency response capabilities for visually impaired individuals. The system integrates IoT components into a wearable device that communicates with a dedicated mobile application. Its core innovation lies in its context-aware emergency alert system, which differentiates between familiar and unfamiliar environments to reduce unnecessary notifications while ensuring a robust safety net for the user.

Methodology: The system is built around an **Arduino** microcontroller that orchestrates data from various sensors. An **ultrasonic sensor** continuously scans for obstacles in the user’s path, providing real-time audio feedback through an earphone for navigation. Simultaneously, a **GPS module** tracks the user’s location. The system’s unique logic checks if the current location is a known, ”familiar” area. If the user enters an ”unfamiliar” location, the device automatically activates an ”Emergency Mode.” In this mode, the **ESP32-CAM** camera begins streaming live video, which, along with the user’s real-time GPS coordinates, is sent via a **Bluetooth module** to a paired smartphone. The smartphone application then instantly forwards this critical information to pre-configured emergency contacts, enabling prompt and informed assistance.

Implementation Details: The hardware implementation uses an **Arduino** as the central controller. Key components include an **ultrasonic sensor** for obstacle detection, a **GPS module** for location tracking, an **ESP32-CAM** for emergency video recording, and a **Bluetooth module** for wireless communication. The software consists of code written in **C/C++** for the Arduino and a dedicated **mobile application** that serves as the user interface and the hub for the emergency notification system. While the primary function is obstacle detection and location-based alerts, the system also incorporates the **YOLO** (You Only Look Once) algorithm for object recognition, which is processed on the connected mobile app rather than the Arduino itself.

Advantages:

- **Context-Aware Emergency System:** The ability to distinguish between familiar and unfamiliar locations is a key advantage, preventing ”notification fatigue” by only triggering alerts in new or potentially unsafe environments.
- **Rich Emergency Data:** In an emergency, the system sends both live video and precise GPS coordinates, giving responders valuable context that is superior to a simple location ping.
- **Integrated Navigation and Safety:** The device combines practical navigation assistance (obstacle detection) with a powerful emergency response system in a single wearable solution.
- **Seamless Connectivity:** The use of Bluetooth and a dedicated mobile app creates a reliable communication channel between the user, the device, and their emergency contacts.

Disadvantages:

- **Dependency on Smartphone:** The system offloads computationally intensive tasks, such as YOLO object recognition, to a paired smartphone, making the device’s full functionality dependent on the phone’s presence and performance.
- **Limited On-board Processing:** The use of an Arduino microcontroller limits the device’s standalone processing power, preventing on-board execution of advanced AI tasks.

- **Potential for False Alarms:** The emergency mode's trigger is based on entering any "unfamiliar" location, which could lead to unnecessary alerts when the user visits new but safe places.
- **No Reading or Face Recognition:** The system is focused on navigation and safety alerts and does not include common assistive features like text reading (OCR) or facial recognition.

Table 1: Comparison of Literature on Smart Glasses for the Visually Impaired

Year	**Paper**	**Advantages**	**Disadvantages**
2023	*Visual Information Translator Using Smart Glasses for Blind*	- Comprehensive features: text-to-speech, object detection, facial recognition, and an SOS button. - Cost-effective design using open-source software and a user-friendly Flutter app.	- Heavily dependent on a paired smartphone for all processing. - Potential issues with battery life and network latency between the glasses and the phone.
2023	*Smart Glasses Embedded with Facial Recognition Technique*	- Multifunctional design combining facial recognition with obstacle and distance detection. - Includes digital assistant features like web browsing and music playback.	- Relies on a static database; new faces and objects must be manually added to the code. - Requires an internet connection to function.
2023	*Facial Recognition Smart Glasses for Visually Impaired People*	- Integrates a wide array of features including object detection (YOLO), OCR, and currency recognition. - Hybrid sensing using both a camera and an ultrasonic sensor for robust detection.	- The paper states the device may be fragile and easily damaged. - High power dependency and risk of malfunction are noted as key concerns.
2023	*Single-Shot Image Recognition Using Siamese Neural Networks*	- Achieves high accuracy (92%) on one-shot learning tasks. - A key benefit is that it does not require retraining to recognize new classes, relying only on pixel data.	- The training process is complex, as it learns a similarity metric rather than direct classification. - Performance is slightly lower than the state-of-the-art HBPL model.
2023	*Smart Glasses for Visually Impaired People using Machine Learning*	- High precision in object detection (up to 99%) using the YOLOv3 algorithm. - Capable of detecting multiple objects within a single frame in real-time.	- Lacks distance measurement, as it does not include an ultrasonic sensor. - Dependent on Google's API for text-to-speech, requiring internet access.
2024	*Indoor Navigation Glasses for the Visually Impaired*	- Excellent real-time performance and accuracy due to the use of a Google Coral Edge TPU. - Received a high user usability score (SUS score of 85) and is infrastructure-independent.	- Obstacle detection is limited in scope, as it was only trained to recognize chairs. - The design is a bulky two-part system with a separate processing unit.
2024	*Google Pi Using Raspberry Pi for Visually Impaired*	- Integrates the familiar Google Assistant with custom models for currency and object recognition. - Cost-effective and accessible due to its Raspberry Pi foundation.	- Lacks any navigational or obstacle avoidance features for mobility. - Requires a constant internet connection for the Google Assistant API to work.
2025	*Smart Glass for Visually Impaired Person*	- Features a context-aware alert system that distinguishes between familiar and unfamiliar locations to reduce false	- Relies on an Arduino, which has limited on-board processing power, requiring a smartphone for AI tasks.

3 Problem Statement

Most personal finance management (PFM) tools focus solely on expense tracking, lacking comprehensive budgeting, future expense planning, and AI-driven insights. They fail to support long-term financial goals, recurring expenses, and proactive financial behavior due to the absence of wishlists, savings goals, and timely notifications. **Wyzo** bridges these gaps by offering a holistic solution with advanced dashboards, in-depth budget planning, goal setting, future expense management, and AI-powered reports. It not only helps users manage their finances end-to-end but also keeps them engaged and motivated for long-term financial efficiency.

4 Proposed Solution

Wyzo is a comprehensive personal finance management (PFM) solution designed to bridge the gaps in existing tools by offering advanced financial planning, expense tracking, and insightful analytics.

- Users can create an account or sign in to access their personalized financial dashboard.
- Users can organize and monitor expenses across customizable categories for better financial control.
- Users can set monthly spending limits and savings goals to promote disciplined financial management.
- Users can manage a wishlist and savings goals to stay engaged and actively work toward financial aspirations.
- The application analyzes income, expenses, and savings trends, offering AI-generated reports & insights for better financial decisions.
- Users receive timely reminders to stay on track with budgets, prevent overspending, and keep them motivated.

5 Conclusion

In conclusion, the comparative analysis of existing literature on personal finance management reveals notable advancements, yet uncovers several gaps in current tools. While many applications effectively handle expense tracking and promote financial literacy, they often lack comprehensive features such as robust budget planning, personalized goal-setting, and insightful financial reports. Our proposed web application **Wyzo** addresses these limitations by offering a super insightful dashboard, advanced budgeting tools, and AI-generated reports that provide actionable financial insights. Additionally, features like financial wishlists and motivational notifications will help users stay on track and achieve their financial goals. This holistic solution is designed to enhance financial management, encourage proactive behavior, and support long-term financial stability and well-being.

References

[1] T.P. Rani, V.T. Vignesh, S.Susila, and P.Kalaichelvi, in *"Visual Information Translator Using Smart Glasses for Blind"*, IEEE, 2023, doi:10.1109/ICCEBS58601.2023.10449230 Available: `https://ieeexplore.ieee.org/document/10449230`

[2] M.Kumar, B.Bharti, and U.Chauhan, in *"Smart Glasses Embedded with Facial Recognition Technique"*, IEEE, 2023, doi:10.1109/AISC56616.2023.10085337 Available: `https://ieeexplore.ieee.org/document/10085337`

[3] A.Ghodake, H.Salé, A.Kamble, K.Mankari, O.Lad, and V.Mishra, in *"Facial Recognition Smart Glasses for Visually Impaired People"*, doi: 10.1109/ICCUBEA58933.2023.10391977 Available: `https://ieeexplore.ieee.org/abstract/document/10391977`

[4] A.Malhotra, in *"Single-Shot Image Recognition Using Siamese Neural Networks"*, IEEE, 2023, doi:10.1109/ICACITE57410.2023.10182466 Available: `https://ieeexplore.ieee.org/document/10182466`

[5] S.PriyankaS., A.S. Kumar, M.V. Nagabhushanam, D.Vennela, and P.D. Tulasi, in *"Smart Glasses for Visually Impaired People using Machine Learning"*, IEEE, 2023, pp.1–6. doi:10.1109/ICCCNT56998.2023.10307374. Available: `https://ieeexplore.ieee.org/document/10307374`

[6] P.J. Loresco, R.J.C. Cruz, K.Z. Ramones, J.A.D. Zafra, and K.R.G. Ramirez, in *"Indoor Navigation Glasses for the Visually Impaired with Deep Learning and Audio Guidance Using Google Coral Edge TPU"*, IEEE, 2024, doi:10.1109/TENCON61640.2024.10902929 Available: `https://doi.org/10.1109/TENCON61640.2024.10902929`

[7] G.Vinod Katkar, K.Abirami, A.Mary Posonia, B.Ankayarkanni, and D.Ushanandini, in *"Google Pi using Raspberry Pi for Visually Impaired"*, IEEE, 2024, doi: 10.1109/AIIoT58432.2024.10574549 Available: `https://ieeexplore.ieee.org/document/10574549`

[8] R.Sabitha, S.PandiS., J.Rathi Devi, and G.Jegan, in *"Smart Glass for Visually Impaired Person"* , IEEE, 2023, doi:10.1109/ICCCT63501.2025.11019089 Available: `https://ieeexplore.ieee.org/document/11019089`