Development of Vision and Voice Assistive Technology for Visually Impaired Using Convolutional Neural Network (CNN)

Ralph Aaron M. Panahon[1], Jayem M. Navarro[1], Frenly S. Castro[1], Bryan Joseph C. Feliciano[1], Ronel Q. David[1]

[1]*Computer Engineering Department, St Nicolas College of Business and Technology, Pampanga, Philippines*

Corresponding Author: arkyudee.snc@gmail.com

***Abstract*: Vision impairment affects approximately 2.2 billion people globally, presenting significant challenges to autonomous movement and spatial awareness. This study developed a wearable vision and voice assistive device designed to improve the mobility, safety, and independence of visually impaired individuals. The system integrates a Raspberry Pi 5 central processing unit with a Pi Camera Module 3 for real-time visual capture and a TF-Luna LiDAR sensor for precise distance measurement and obstacle avoidance. The research followed an Evolutionary Prototyping Model, incorporating descriptive investigation and laboratory testing. The software architecture utilizes a YOLOv11 Convolutional Neural Network (CNN) model capable of detecting over 80 object classes, the Vosk speech engine for offline voice commands, and Pyttsx3 for real-time text-to-speech feedback.**

***Keywords*: Assistive Technology, Convolutional Neural Network (CNN), LiDAR Sensor, Object Detection, Voice Feedback, Wearable Device, Real-Time, Processing.**

1. Introduction

The global prevalence of vision impairment affects approximately 2.2 billion people, with nearly half of these cases remaining untreated or preventable (World Health Organization, 2023). Common conditions such as cataracts, glaucoma, and refractive error, age-related macular degeneration and diabetic retinopathy continue to hinder the ability of many individuals to navigate daily environments safely. While advancements in deep learning, particularly Convolutional Neural Networks (CNNs), have significantly improved image processing and object classification, existing assistive technologies often face limitations regarding portability, real-time integration, and practical utility in unpredictable real-world settings (Bai et al., 2019; Yu et al., 2019). Previous research has explored wearable sensors and lightweight CNN models to assist with specific tasks like obstacle classification and pedestrian traffic light recognition (Hsieh et al., 2021). However, a significant gap remains: few systems offer a cohesive, high-performance solution that combines real-time object recognition with natural, voice-activated feedback in a portable device. Many current prototypes are confined to controlled environments or require excessive processing power, limiting their use for daily independent navigation. To address these challenges, this study presents a comprehensive vision and voice assistive technology designed for real-time environmental awareness. The research focuses on implementing a voice-activated object detection system and a distance-integrated collision avoidance tool. By combining CNN-based computer vision with LiDAR technology and a hands-free audio interface, the proposed system aims to provide a practical and efficient solution for increasing the independence and safety of visually impaired individuals.

2. Methodology

A. Approach Used in Implementation and Process Overview

The research methodology for the "Vision and Voice Assistive Technology" project utilizes a combination of descriptive investigation and laboratory testing. To ensure a flexible and iterative development process, the study adopts the Evolutionary Prototyping Model, which involves a continuous cycle of requirement gathering, analysis, design, and testing. This approach allows the researchers to refine the prototype based on ongoing evaluations, ensuring the final device effectively meets the mobility needs of visually impaired users.

1) *Development Phases*

The implementation follows a systematic cycle that moves from conceptualization to a refined prototype. Requirement Gathering the researchers conducted interviews with 20 visually impaired respondents to understand navigation habits and obstacle recognition needs. These insights highlighted the necessity for rapid object recognition and clear audio guidance. Analysis the technical requirements were established based on user needs, focusing on functional features like image capture, CNN-based object detection, voice command processing, and audio output. Non-functional requirements emphasized low latency, portability, and dependability. Design and programming the system integrates hardware—including a Raspberry Pi 5, Pi Camera Module 3, and TF-Luna LiDAR sensor with software powered by Python. Prototype Development A wearable prototype was constructed using the central processing unit and sensors, designed to launch automatically upon startup via a system service.

2) Technical Implementation

The system's intelligence relies on the integration of computer vision and sensor technology Object Detection phase the device utilizes a YOLOv11 Convolutional Neural Network (CNN) model, which can detect over 80 object classes without retraining. Distance Measurement phase the LiDAR sensor provides continuous distance readings to alert users when objects are within 150 centimeters. Voice Interface the Vosk speech engine is used for offline voice commands, while Pyttsx3 handles text-to-speech output to provide audio instructions.

3) Testing and Evaluation

Integration and testing the hardware and software were combined into a functional prototype and subjected to functional tests (for detection accuracy), performance tests (for processing speed), and usability tests (for clear voice feedback). Client Evaluation the user testing verified the system's accuracy and ease of use, helping identify areas for improvement in audio guidance and detection responsiveness. Maintenance the final phase involves continuous optimization and debugging to address issues such as false detections or hardware stability.

4) Statistical Treatment

To ensure the study's feasibility and accuracy, data is processed using Slovin's Formula to determine an appropriate sample size. A Weighted Mean is also calculated to evaluate system performance metrics, including object detection accuracy and navigation guidance reliability.

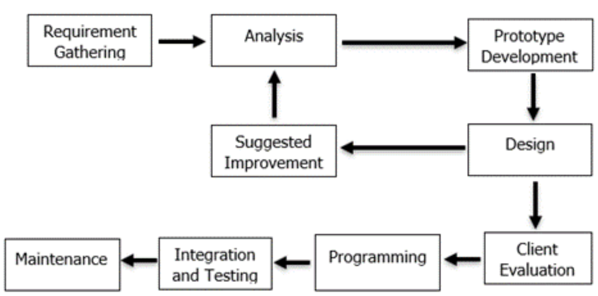

Fig.1. Show the Evolutionary Prototyping Model of Vision and Voice Assistive Technology for Visually Impaired Using Convolutional Neural Network (CNN)

Table.1. The overall evaluation results from the 20 expert respondents

ISO/IEC 2510 Category	Weighted Mean	Descriptive Rating
System Usability and Voice Interaction	3.95	Agree
Object Detection Accuracy	3.75	Agree
Distance Measurement & Proximity Awareness	4.00	Agree
Navigation Guidance	3.80	Agree
Overall System Performance & Reliability	3.75	Agree

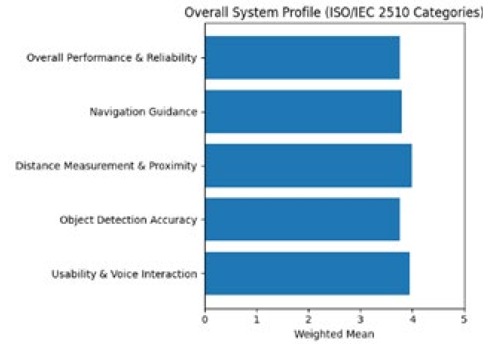

Fig.2. The system's performance across five ISO/IEC 2510 categories using weighted mean scores

All results fall under the "Agree" rating, indicating positive user evaluation. Distance measurement and proximity awareness scored the highest (4.00), showing strong reliability in detecting obstacles. Usability and voice interaction (3.95) and navigation guidance (3.80) also performed well, suggesting the system is easy to use and helpful in guiding users. Object detection accuracy and overall performance both scored 3.75, indicating good but slightly lower performance compared to other features. Overall, the system is reliable, user-friendly, and effective, with its strongest capability in environmental awareness.

3. Results and Discussion

The system evaluation was done with the involvement of all 30 expert respondents. Before the actual conduct of the system evaluation, all the respondents were made aware of the nature and purpose of the study and voluntarily agreed to be part of it.

Laboratory and usability testing demonstrated that the system provides rapid object recognition and clear auditory guidance. The integrated approach successfully translates complex environmental data into actionable audio instructions, significantly reducing a user's dependency on others. The study concludes that combining lightweight CNN models with LiDAR technology offers a portable, affordable, and effective solution for navigating everyday environments, bridging the gap between theoretical deep learning and practical assistive tools.

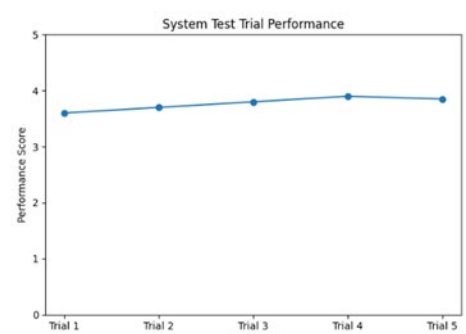

Fig.3. A gradual improvement in system performance from Trial 1 to Trial 4, followed by a slight stabilization in Trial 5.

This indicates that the system became more consistent and reliable after repeated testing.

4. Related Literature

The reviewed literature highlights the development and application of intelligent assistive technologies designed to improve the mobility, communication, and independence of visually impaired individuals. Local studies emphasize the integration of affordable embedded systems such as Raspberry Pi, ultrasonic sensors, Optical Character Recognition (OCR), and text-to-speech (TTS) technologies. These systems, including smart text readers, audio-based smart canes, and assistive smart sunglasses, demonstrate how voice feedback, object detection, and communication features (e.g., GPS and emergency messaging) can support daily navigation and learning. Overall, local literature focuses on accessibility, affordability, and practical assistive solutions that enhance user independence and learning experiences. Foreign literature further strengthens these findings by emphasizing the use of advanced deep learning techniques, particularly Convolutional Neural Networks (CNNs), combined with depth sensing and real-time computer vision systems. Studies show that CNN-based models such as YOLO and Fast-SCNN are effective for obstacle detection, semantic segmentation, and real-time path guidance. These systems convert visual environmental data into audio instructions, improving situational awareness and safety for users. However, challenges such as real-time processing limitations, device portability, and cognitive load from frequent audio feedback are also identified. Across both local and foreign studies, there is a consistent trend toward combining computer vision, AI-based object detection, and audio feedback systems to enhance assistive technology. These findings strongly support the development of vision and voice assistive systems using CNN, as they demonstrate the effectiveness of integrating intelligent image processing with real-time auditory guidance to promote independence, safety, and accessibility for visually impaired individuals.

5. Summary

The research focuses on the development of a wearable "Vision and Voice Assistive Technology" designed to enhance the independence of visually impaired individuals. The project identifies the significant mobility challenges faced by the 2.2 billion people globally with vision impairments and proposes a solution using a Raspberry Pi 5 integrated with a YOLOv11 Convolutional Neural Network (CNN) for real-time object detection and a TF-Luna LiDAR sensor for precise distance measurement. Through an evolutionary prototyping methodology, the researchers gathered requirements from 20 visually impaired respondents, leading to a design that provides natural audio feedback and responds to offline voice commands via the Vosk engine. The study concludes that this hands-free, portable system effectively assists users in navigating their environments and avoiding obstacles within a 150-centimeter range.

6. Conclusion

The prototype performed well and met its objectives, showing that combining CNN-based image processing, voice feedback, and distance sensing provides a more complete assistive solution than systems with limited functions. The developed hardware-based assistive device successfully improved accessibility for visually impaired users. All components were properly integrated, and the system now runs smoothly without major bugs. The Raspberry Pi successfully processes and sends sensor data, enabling real-time distance measurement, object detection, and audio/text-to-speech feedback.

A. Recommendation

The following suggestions for future researchers can be added to enhance and improve the study:

- To provide a better alternative power supply
- To enhanced the Convolutional Neural Networks' (CNN) model efficiency to speed up processing and boost real-time performance on portable hardware.
- To enhance real-world testing, thoroughly test in a variety of settings (such as streets, shopping centers, and low-light areas) to guarantee accuracy and dependability.
- Integration of facial recognition
- GPS navigation
- Performance optimization for extreme environmental conditions such as total darkness or heavy rain.

References

[1] Weisstein, E. W. (n.d.). *Slovin's formula*. Wolfram MathWorld.
[2] Adam, Y. M., et al. (2022). A deep learning-based smart assistive framework for visually impaired people. NSF Public Access Repository.
[3] Amorado, J., Calimlim, M., Iscala, J., Lim, A., & Misola, R. (2022). Smart assistive sunglasses for the visually impaired.
[4] Bai, Y., Liu, Y., Lin, J., et al. (2019). A wearable device combining CNN with depth sensing for obstacle detection in visually impaired users.
[5] Baldonado, J., & Faelangca, M. (2024). Enhanced audio-based smart cane. Ramon Magsaysay Memorial Colleges.
[6] Baldovino, R., Roxas, J., Bugtai, M., et al. (2024). A visual aid system using image processing and deep learning with audio-haptic feedback for the blind and visually impaired.
[7] Bermudo, J., et al. (2020). Assistive device for the visually impaired using ultrasonic sensor.
[8] Caballero, J., Catli, M., & Babierra, A. (2020). Object recognition and hearing assistive technology mobile application using convolutional neural network.
[9] Dahiya, R., & Singh, M. (2020). Wearable electronics: Sensors and systems.
[10] De Souza, R., Francisco, M., Tavares, J., & Barbosa, J. (2024). Intelligent environments and assistive technologies for visually impaired people: A review.
[11] Gillis, A. S., et al. (2024). Convolutional neural network (CNN). [12] TechTarget.
[12] Hsieh, Y., Cheng, C., Ke, R., Chen, Y., & Wang, C. (2021). A CNN-based wearable assistive system for visually impaired people walking outdoors.
[13] Krishna, G. V. N., et al. (2025). Object and text detection and recognition for blind people.

[14] Laguna State Polytechnic University. (2023). Smart text reader for blind/visually impaired students.
[15] Lin, Y., Wang, K., Yi, W., & Lian, S. (2019). Deep learning-based wearable assistive system for visually impaired people.
[16] NOAA Office for Coastal Management. (2024). What is LiDAR.
[17] Raspberry Pi Foundation. (2022). Pi camera module documentation.
[18] Raspberry Pi Foundation. (2023). Raspberry Pi documentation.
[19] Raspberry Pi Ltd. (2025). Raspberry Pi hardware documentation.
[20] Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.
[21] SD Association. (n.d.). SD memory card standards.
[22] ScienceDirect. (n.d.). Audio devices and wearable communication systems.
[23] Shan, J., & Toth, C. (2018). Topographic laser ranging and scanning: Principles and processing. CRC Press.
[24] Tanenbaum, A. S., & Bos, H. (2015). Modern operating systems (4th ed.). Pearson.
[25] World Health Organization. (2023). World report on vision.
[26] Yu, C., et al. (2019). Lightweight CNN model for pedestrian traffic light detection using MobileNetV3.
[27] Texas Instruments. (n.d.). *Buck converter basics*. Retrieved March 27, 2026.