A

Project Report

on

Security Assistance for Visually Challenged Using ML

submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY DEGREE

SESSION 2024-25

in

Computer Science and Engineering

By

Atharva Namdeo (2100290100037)

Devesh Kumar (2100290100053)

Parth Sharma (2100290130119)

Under the supervision of

Mr. Umang Rastogi

KIET Group of Institutions, Ghaziabad

Affiliated to

Dr. A.P.J. Abdul Kalam Technical University, Lucknow

(Formerly UPTU)

May, 2025

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Signature:
Name: Atharva Namdeo
Roll No.: 2100290100037

Signature:
Name: Devesh Kumar
Roll No.: 2100290100053

Signature:
Name: Parth Sharma
Roll No.: 2100290130119

CERTIFICATE

This is to certify that the Project Report entitled "Security Assistance for Visually Challenged Using ML" which is submitted by Atharva Namdeo(CSE), Devesh Kumar(CSE), and Parth Sharma(IT) in partial fulfillment of the requirement for the award of degree B. Tech. in Department of Computer Science and Engineering of Dr. A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

Mr. Umang Rastogi
(Assistant Professor)

Dr. Vineet Sharma
(Dean CSE)

Date:

ACKNOWLEDGEMENT

It gives us a great pleasure to present the report of the B.Tech Project undertaken during B.Tech. Final Year. We owe a special debt of gratitude to Mr. Umang Rastogi, Department of Computer Science and Engineering, KIET, Ghaziabad, for his constant support and guidance throughout our work. His sincerity, thoroughness, and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavors have seen the light of day.

We also take the opportunity to acknowledge the contribution of Dr. Vineet Sharma, Dean of the Computer Science and Engineering, KIET, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.

We also do not like to miss the opportunity to acknowledge the contribution of all faculty members, especially faculty/industry person/any person in the department, for their kind assistance and cooperation during the development of our project. Last but not least, we acknowledge our friends for their contribution to the completion of the project.

Signature:

Name: Atharva Namdeo

Roll No.: 2100290100037

Signature:

Name: Parth Sharma

Roll No.: 2100290130119

Signature:

Name: Devesh Kumar

Roll No.:

2100290100053

ABSTRACT

Maintaining the safety and autonomy of visually impaired people is a serious challenge, especially in settings where threats are not readily recognizable. Conventional assistive devices like guide dogs and white canes offer minimal navigation assistance but do not meet real-time security needs. This Assistive Threat Detection and Identification Using YOLO and Facial Recognition for the Visually Challenged project endeavors to fill this gap by leveraging cutting- edge artificial intelligence (AI) technologies, namely YOLO (You Only Look Once) for live weapon detection and CNN (Convolutional Neural Network) for facial recognition. The suggested system works in a two-step detection model: it initially detects dangerous objects in the vicinity of the user, and if no direct danger is found, it goes on to detect people within the environment to evaluate possible threats. The module for detecting weapons uses the YOLO algorithm, which has shown a remarkable 95% accuracy in lab-controlled indoor environments and 90% in outdoor environments. The facial recognition module, which uses deep models such as CNN, performs with 98% accuracy under good lighting, although under low-light conditions, its performance reduces marginally to 90%. These accuracy ratings render the system appropriate for practical use, whereby visually impaired subjects need an efficient and independent security feature to move around in their environment safely. The responsiveness of the system is one of its major highlights, with weapon detection in 0.3 seconds, face recognition in 0.5 seconds, and overall system response time of 0.8 seconds, making it highly suitable for real-time applications. Though the system is highly reliable, there are some limitations that remain, such as difficulties in low- light environments, mixed background noise, and dense crowds. Future developments will look to improve these aspects by adding infrared cameras for better night capability, increasing facial datasets for more representative recognition abilities, and edge computing optimizations for faster real-time processing.

The deployment of this system has tremendous potential in transforming assistive security technologies so that visually impaired people achieve more independence, better situational awareness, and enhanced personal security. Through the use of AI-based solutions, this project opens the door to inclusive and smart security systems, establishing a new benchmark for the application of computer vision and deep learning in assistive technologies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
CNN	Convolutional Neural Network
ML	Machine Learning
YOLO	You Only Look Once
R-CNN	Region-based Convolutional Neural Network
SVM	Support Vector Machine
OCR	Optical Character Recognition
IOT	Internet of Things
PCA	Principal Component Analysis
ROI	Region of Interest
VGG	Visual Geometry Group
QDNN	Quaternion Deep Neural Network
RNN	Recurrent Neural Network
TTS	Text-to-Speech
API	Application Programming Interface

CHAPTER 1
INTRODUCTION

1.1 INTRODUCTION

Maintaining the independence, security, and general safety of visually impaired persons is an ever-growing problem in the fast-moving, technology-centered world we inhabit today. Long-established and common assistive devices like white canes, professionally trained guide dogs, and human attendants or companions have served to greatly assist visually impaired persons in going about their daily lives. They have been successful in conferring with a certain level of mobility and spatial perception, usually acting as the main support systems for visually impaired people in familiar and unfamiliar environments. Yet, although such traditional aids are certainly helpful, they lack built-in limitations, particularly their ability to preactively recognize, detect, or classify possible hazards in very dynamic or uncertain real-world situations. For example, though a white cane can alert a user to a threat in the ground plane, it will not alert a user to someone coming towards them with ill intent or a threat in a location beyond its range of touch. Similarly, a guide dog, as well trained as it might be, will not always see and respond to threats like weapons, speeding cars, or strange and suspicious people. This changing set of challenges has provided the opportunity for incorporating more advanced technologies, specifically those based on artificial intelligence (AI), machine learning (ML), and computer vision. These advances in technology are opening up unprecedented possibilities for creating intelligent systems that not only complement the capabilities of conventional aids but also enhance to a significant extent the degree of mobility, autonomy, and personal safety enjoyed by the visually impaired. The capability of AI to analyze and understand vast amounts of real-time data enables the development of systems that can respond rapidly to threats, identify latent anomalies, and augment situational awareness. One innovation is an advanced project called Assistive Threat Detection and Identification Using YOLO and Facial Recognition for the Visually Impaired, which employs the most recent advances in deep learning and AI to deliver visually impaired users a robust real-time security system. The central operation of this project revolves around providing visually impaired persons with AI-backed models that are capable of detecting possible danger in the user's immediate environment and identifying familiar or known faces. Through the provision of real-time automated notifications, the system allows relevant safety responses without the need for direct intervention by the user, hence promoting both independence and peace of mind. At the core of this system are two main and extremely stable technologies: YOLO (You Only Look Once) and CNN (Convolutional Neural Networks). YOLO is used particularly for real-time gun detection, whereas CNNs are utilized for face recognition. YOLO has been thoroughly studied and is known throughout the AI world for its excellent speed and efficiency in object detection. As opposed to the conventional object detection models that compute images in several stages or regions, YOLO processes images by dividing the input image into a grid and predicting bounding boxes as well as class probabilities of objects in the same forward pass. YOLO is best suited for real-time applications, particularly those with the need to act immediately in rapidly changing environments, due to the parallel processing nature. Conversely, CNNs have turned out to be the cutting-edge architecture in the facial recognition field. They specialize in learning deep feature representations of images and categorizing them with good accuracy. CNNs automatically learn hierarchical features from face images beginning from basic edges and textures to abstract patterns, thereby discriminating between familiar faces and strangers with high precision. The integration of these two technologies enables the system to not only detect environments but also give context-driven alerts based on what or who is detected. The working model of this project depends on a two-stage detection model, meant to optimize speed, accuracy, and security. The YOLO model is run first in order to continuously scan and analyze the environment using a live video feed. Its core function is the detection of the presence of weapons or other possibly hazardous items like knives, firearms, or blunt objects. If no danger is found in the immediate vicinity at this level, the system automatically moves on to the second stage where the CNN-based facial recognition module is activated. At this point, the system searches for people within range and tries to identify them from a saved database of known faces. If a familiar face is recognized, and there is no weapon detected, the system stays in standby. But if an unfamiliar face is recognized, or if a weapon is detected at any point, the system will instantly trigger an alert protocol. This involves sounding a warning to alert the user and at the same time alert caregivers, family members, or security through integrated communication channels. This intelligent, real-time alerting system is central to keeping response time to a minimum and to making sure that assistance may be offered at the earliest time. Under test conditions and during evaluations, the system has performed splendidly in controlled settings, which include indoor areas with ample lighting and little ambient noise. In particular, the weapon detection module has a 95% accuracy rate in indoor environments and 90% in outdoor environments. This performance metric illustrates the system's stability even when there is varying lighting, as well as unpredictable background environments. Similarly, the facial recognition module has achieved stunning results, recording 98% accuracy within optimal lighting environments, and 90% accuracy in low or inconsistent lighting environments. These figures illustrate the system's capacity to generalize well across different environments and conditions. With regards to speed, the system is extremely efficient, a requirement for real-time applications. The time it takes for weapon detection employing the YOLO model is only 0.3 seconds, while facial recognition employing CNN averages about 0.5 seconds. The complete decision-making process, from input to alert generation, is executed in an overall average time of 0.8 seconds, making it highly responsive and suitable for deployment in live, real-world environments where every second matters. Despite these commendable strengths, the system is not without its limitations. Like many computer vision-based systems, its performance can be negatively impacted by several real-world factors. These are poor or insufficient illumination, which can adversely affect image quality and face detection performance; background noise or visual distraction, which may produce false positives or failures to detect; and the inability to detect small, moving, or partially occluded objects, such as a weapon that is visible only momentarily or hidden in a bag or pocket. In order to solve these issues, as well as increase the system's overall resilience, several future enhancements are scheduled. These involve the addition of infrared sensors and other technologies that allow nighttime detection and low-light capability. Further improvements will be achieved by expanding the size and range of the facial recognition training datasets, thereby improving the model's generalizability across a wider variety of faces, lighting, and expressions. Edge computing methods will also be investigated to maximize on-device computation, minimize latency, and remove requirements for continuous connectivity to central servers. Additionally, inclusion of multi-object tracking algorithms will enable the system to more accurately identify and track multiple people in densely populated environments, dramatically decreasing the incidence of false positives while enhancing situational awareness overall. The creation of AI-powered security technology, purpose-designed for addressing the specific requirements of individuals who are visually impaired, is a giant leap in assistive technology inclusivity. Through the application of deep learning, real-time computer vision, and smart automation, this project provides an actionable, scalable, and impactful solution for enhancing personal independence and minimizing reliance on others. The system not only enables visually impaired people to move around their world more confidently and securely but also instills a sense of independence crucial for complete integration into society. As technologies in AI and ML continue to mature and evolve, they will open up even more exciting possibilities for improving the abilities of such assistive systems. This project is a testament to how new technologies are made viable to address everyday issues, most especially for vulnerable and marginalized groups. It is not merely to develop tools but to establish intelligent assistive ecosystems that can really add value to the quality of life of the visually impaired. By making safer, more independent mobility possible and reducing the necessity for continuous human support, these innovations help make the world a place where all people can engage more completely and equally. Finally, the integration of AI-driven solutions in the realm of assistive technology is just the start of a larger vision: one in which intelligent devices are not just created to fulfill functional tasks but to enable individuals, augment dignity, and facilitate social inclusion on a significant scale.

1.2 PROJECT DESCRIPTION

The project, Assistive Threat Detection and Identification Using YOLO and Facial Recognition for the Visually Challenged, seeks to improve visually challenged individuals' security and independence by using artificial intelligence (AI) and machine learning (ML) technology. Guide dogs, white canes, and human guides are conventional assistive means that offer simple mobility support but are not conducive to real-time security needs. The system seeks to overcome this limitation by merging state-of-the-art object detection and facial recognition features into an assistive system in real-time. The key aim of the project is to develop a smart system that is able to detect potential threats and identify known faces in real-time, making visually challenged users safer and more independent. The system will offer precise, fast, and robust security support by taking advantage of deep learning models that analyze live video streams and activate alerts when required.

The project goals are:

Use of YOLO (You Only Look Once) for real-time and efficient weapon detection.

Use of CNN (Convolutional Neural Networks) for facial recognition with high accuracy.

Use of rapid threat detection and response to offer timely security alerts.

Designing a system that reduces false positives and functions well in diverse environments.

Ensuring an increase in the independence of visually challenged users by offering an automated, AI-based security assistant.

The methodology includes the development of a dual-phase detection framework that scans the environment first for weapons using YOLO. In the absence of detected threats, the framework goes ahead to detect individuals using CNN-based face recognition. Upon detection of a weapon or an unknown person, an alarm is signaled, triggering warnings to caregivers or security personnel. The framework is coded using Python and OpenCV for image processing, hence offering a strong implementation and operational performance. Furthermore, edge computing optimizations are added to improve real-time performance.

Figure 1.2.1 Harmful Object Knife

Figure 1.2.2 Harmful Object Gun

Performance tests demonstrate that the framework detects weapons with a 95% accuracy indoors and 90% outdoors, and facial recognition has a 98% accuracy rate in bright environments and 90% in low light. The response times are highly effective, taking 0.3 seconds to detect weapons, 0.5 seconds to detect faces, and the whole system taking 0.8 seconds to process. Despite its efficiency, issues in the form of performance in low-light conditions, interference from complex backgrounds, and processing in crowded scenes persist. Future improvements include the incorporation of infrared cameras to improve nighttime detection, expansion of facial recognition datasets to improve accuracy, and edge computing optimization to minimize latency. Furthermore, algorithms for multi-object tracking will be investigated to further improve detection accuracy in dynamic scenes. The project is a major step in AI-assisted assistive technologies, offering a scalable and realistic solution for visually impaired people. By combining deep learning and real-time processing, the framework improves safety, autonomy, and situational awareness, hence lessening dependence on conventional assistive methods.

CHAPTER 2

LITERATURE REVIEW

Breakthroughs in deep learning and computer vision have transformed the development of face recognition technology. Legacy systems employed handcrafted features such as Eigenfaces and Fisherfaces [1] – [5] to capture face information. While these were groundbreaking approaches, these suffer from some drawbacks such as changes in illumination, pose, and expressions [6]. For example, changing lighting conditions can influence the accuracy of recognition significantly, with certain instances resulting in mismatches.

Breaking into new technological breakthroughs is the newly emerging era of sophistication known as deep learning, where outstanding records have been maintained by CNNs, which learnt sophisticated and hierarchical features out of image data directly [7], [8]. The revolution, thus, would enable constructing very efficient systems adaptively to various facial appearances and reach much higher accuracy. Early attempts like DeepFace demonstrated the potential of the approach since they are able to achieve near-human performance in face recognition today. With the advent of technologies like that of FaceNet, novel and innovative loss functions like triplet loss have become an everyday occurrence with which one can learn image embeddings in a metric space with the main goal of producing embeddings that enable more favorable comparisons between faces. Consequently, models such as VGGFace have refined these approaches through greater dataset sizes and more sophisticated network architectures.

Object detection models, especially the different models belonging to the YOLO family [9]– [12], have played a huge role in making face recognition practically applicable. The capability of the YOLO method to efficiently find and identify objects in images, including human faces, has tremendously boosted speed and accuracy of face recognition in practice. This has made applications such as real-time surveillance systems [13]–[15] for applications involving swift and precise face detection feasible.

The application of this technology can be expansive, beyond security and surveillance. Great potential lies in the face recognition technology to assist the visually impaired. Face recognition systems can assist a visually impaired individual to recognize familiar faces, like relatives, friends, and caregivers [16]–[18]. Social interactions become easier for them, and their confidence of engaging in social situations increases. Merging face recognition with other sensory modes can provide useful interaction and navigation support. For instance, a system can be designed to utilize facial recognition for recognizing and leading users towards specific individuals or destinations [19]. Merging face recognition into a wearable feature such as smart glasses would offer

real-time feedback regarding individuals in the surroundings, allowing low-vision individuals to have a better sense of what is occurring in their vicinity.

Used in combination with other assistive technologies, face recognition utilizes computer vision for the visually impaired. OCR-based technologies when utilized in combination with text-to-speech synthesis enable visually impaired people to access print materials [20]. Object detection, on the other hand, may be utilized for freely identifying objects in the surroundings and providing verbal descriptions [21]–[23]. This data can be vital in moving through environments, perceiving what is present, and achieving daily routines.

While there have been great advances made in the art, there are many challenges that still exist. These are, but are not limited to, easy identification of faces that are not frontal, partially occluded faces, or large pose variations (e.g., profile faces) [24]. The widespread use of face recognition technology should bring serious questions regarding the security and privacy protection. Its illegal application for observation or any other purpose poses socio-ethical issues and challenges in itself. Therefore, the creation of systems that can reliably identify faces from various spheres (from security cameras to social networks), aside from vagueness, is still a significant field of study.

Some of the contemporary methods of inquiry in fields central to computer vision have been areas of research in recent times. A hybrid of a traditional method such as Haar Cascades with deep learning algorithms such as Softmax and Convolutional Neural Networks (CNN's) is being employed [25]. This symbiotic blend could witness an increase in accuracy and maybe robustness of face recognition systems. Not only this, computer vision methods are reaching a long way beyond traditional deployments. As an example, human detection models optimize the energy expenditure at public places [26], which shows how computer vision can be employed practically to solve real-world issues. The advancement of deep learning remains the characteristic that advances the innovations. Quaternion Deep Neural Networks (QDNNs), a new class of neural network, hold out directions towards improved computational efficiency and representational capability [27]. This paper includes a synthesis of QDNN architecture, applications, and challenges, paving the way towards advances in this field in the years to come.

Addressing these issues is critical for the ethical and responsible advancement of face recognition technology in the immediate future.

Current studies have highlighted the need for user-specific interaction with real-time threat detection in assistive technologies. For example, Narejo et al. [13] and Bhatti et al. [14] were able to show effectively how YOLOv3 and other such architectures can be specifically designed for smart surveillance systems, with observations regarding their speed and stability across test scenarios. These papers justify the choice of using YOLO as the core detection algorithm in our suggested system.

YOLO's subsequent enhancements (from v3 to v5) have significantly lowered detection latency and improved accuracy for small objects, which is instrumental when detecting hidden or partially exposed weapons among throngs of people. As argued by Ashraf et al. [15], YOLO-v5s operates incredibly well when implemented on embedded surveillance systems, and accuracy outperforms traditional CNN-based classifiers in real-time video environments.

On the face recognition side, the insensitivity of CNN-based models to a wide range of environmental conditions has been confirmed in unconstrained settings [4], [5]. This stands in agreement with our project's requirement to cater to illumination changes and partial occlusions under real-world conditions. Research such as that of Billah et al. [3] even pushed CNNs to animal face recognition with an emphasis on their multi-domain versatility, indirectly lending support towards their usage for assistive human-centered systems.

Interest in using these technologies for visually impaired users has grown. Deep learning-based facial recognition systems, as shown by Francis et al. [16], are already integrated in wearable technology to aid users in recognizing people known to them. Tiwari et al. [17] went a step further and designed smart glasses that not only recognize faces but also speak out identified individuals, providing better navigation and social assistance.

Such wearable use cases take great advantage of light-weight CNN models, optimized by edge computing — a future horizon elaborated in our project. The system developed by Chaudhry and Chandra [18], for example, demonstrates how mobile devices can be efficiently executed to operate face recognition models with decent accuracy, even on limited hardware.

In addition, incorporating multimodal detection functionality, as proposed by Patil et al. [19], such as object tracking and verbal feedback, opens the door for even more elaborate user interaction models. These multimodal setups will certainly be a must in future assistive systems, where not only static identification but also dynamic scene interpretation is required.

While they hold potential, the systems also pose legitimate ethical issues. As emphasized by Lahasan et al. [24], issues around data privacy, consent, and misidentification grow more significant as recognition systems are implemented in public or semi-public areas. Any visually impaired person solution must comprise privacy-sensitive protocols, particularly when the data is streamed or stored.

For better performance, the hybrid approaches have picked up pace, integrating traditional techniques such as Haar cascades with CNNs, Softmax classifiers, or specialized loss functions [25]. These approaches, though computationally intensive, enable better robustness in occluded or overlapping face scenarios — a drawback in most single-shot detectors like YOLO.

The development of newer models like Quaternion Deep Neural Networks (QDNNs) presented by Singh et al. [27] brings in more opportunities for training lightweight, real-time recognition tasks, by enhancing computational efficiency with the model expressiveness maintained.

The application of deep learning in real-time monitoring has been actively pursued in recent years. Bhatti et al. [14] applied YOLO on live CCTV streams for urban monitoring, illustrating how frame-wise detection of weapons can be used to improve real-time monitoring. Their method, though intended for security officers, naturally extends to assistive applications where visually impaired users can take advantage of autonomous monitoring systems integrated in wearable or portable devices.

In addition, Kim et al. [11] developed SHOMY, a lightweight hazardous object detection system with YOLO, which was highly capable of recognizing items such as knives and scissors even when they are partially occluded. Our project's approach is guided by their methodology and lays the foundation for maximizing bounding box precision and low rates of false positives.

With latency issues arising in conventional cloud-based detection, recent studies have shifted focus towards edge computing. Jiang et al. [12] discussed the evolution of YOLO and stressed its suitability for edge AI devices such as Raspberry Pi and NVIDIA Jetson Nano. This viewpoint has been taken into consideration in our design choice to maintain inference tasks locally, enabling rapid response against threats (less than 0.8 seconds), without relying on high-latency cloud servers.

These insights are essential to assistive use, where network connectivity could be intermittent or where privacy issues eliminate cloud-based storage of real-time facial information. Edge processing eliminates performance variability and protects personal information.

One of the universal problems for both weapon detection and facial recognition is dataset bias. The performance of models such as VGGFace or YOLOv5 strongly relies on balanced and diverse datasets. As emphasized by Goel et al. [9], models with limited datasets fail to perform well in environments without control. The problem becomes even more significant for assistive applications deployed in multicultural or diverse lighting/geographic environments.

Ashraf et al. [15] also emphasize the requirement of expert datasets during training detection models for safety-critical applications. These involve annotated occlusions, nighttime environments, and dynamic object motion—all of which are needed for reliable deployment in real-world assistive technology.

Low-light conditions considerably hamper the accuracy of recognition. Han et al. [6] and Dang et al. [20] suggested various preprocessing methods—like histogram equalization and contrast stretching—to improve low-light facial images. Although these approaches are effective for static images, they are problematic for dynamic video streams, hence supporting infrared-based enhancements or light-invariant feature extraction.

Our project is looking to incorporate infrared (IR) imagery in subsequent versions, providing stable operation at night—a step in the right direction, given that a lot of public and domestic areas contain erratic lighting.

Rodríguez et al. [7] brought in acoustic feedback systems to assist visually impaired users with obstacle detection, a technique that could be applied to remove threats. Although their paper targeted environmental awareness, integrating such methods with our YOLO+CNN approach might introduce a contextual dimension, so that the system will not just detect a threat but will also narrate its description (e.g., "A knife-wielding person is coming from the left").

Such an integration closes the loop between detection and actionable situational awareness, taking the system out of mere alarms to a more descriptive and empowering user interface.

Although the technological developments are encouraging, ethical issues related to facial recognition are still up for debate. The risks of misclassification and profiling were discussed by Lahasan et al. [24]. These risks are especially sensitive for assistive technology since they can affect a user's trust and safety.

Francis et al. [16] and Singh et al. [25] posit privacy-first design for assistive AI, where models run offline, and data is only sent or stored when specifically permitted. Our system's architecture supports this through on-device processing and a user-specific facial database disconnected from external networks.

Future-ready assistive systems will require integration with IoT ecosystems, enabling functionalities like remote caregiver notifications, real-time streaming to emergency services, and wearable-device sync. As noted by Bisaria et al. [26], IoT-based monitoring systems can optimize not only safety but also energy and communication efficiency when combined with intelligent detection.

Singh et al. [27] discuss Federated Learning as a privacy-preserving training framework that enables model updates on multiple devices without sharing sensitive information. This paradigm is apt for assistive technology, where user-specific training can take place locally and anonymized gradients can be exchanged to update the global model while maintaining user privacy.

Maintaining the autonomy and safety of visually impaired people using artificial intelligence (AI) and machine learning (ML) has received significant scholarly interest. The combination of real-time object recognition and facial identification in aid technologies has been informed by a range of research advancements. Park et al. [1] set the stage with research into human-object relationships in security control systems, highlighting the requirement for context sensitivity in understanding user intentions, a guideline core to assistive AI design. Wu et al. [2] illustrated the effectiveness of Faster R-CNN in detecting faces at different scales, supporting the necessity for adaptive, multiscale detection frameworks within real-world settings where faces do not necessarily present themselves as frontal or unobstructed. Their work has been directly helpful in enabling systems where face localization reliability is mission-critical.

More specific to real-time recognition, Billah et al. [3] applied convolutional neural networks (CNNs) to facial recognition in the context of agriculture, showcasing the generalizability of CNNs to different domains. The ability of their model to learn goat face recognition highlighted the transferability of deep learning approaches to new tasks and hence warranted the use in facial recognition for assistive devices. Fredj et al. [4] specifically targeted the performance of CNN in unconstrained settings, where factors such as lighting, pose, and occlusion play major roles in determining results. Their contributions add to what Hasan and Sallow [5] found, using OpenCV's face detection and recognition modules in real-world video streams and establishing that pre-trained classifiers continue to be of value when integrated with deep learning models for fast, lightweight detection.

Han et al. [6] also explored illumination preprocessing, an essential step towards enhancing face recognition under non-ideal lighting—a situation frequent in assistive environments. The paper by Rodríguez et al. [7] presented acoustic feedback systems for detecting obstacles, a case of how non-visual feedback can be used in support of visually impaired users, a concept readily transferable to alarm mechanisms in weapons or unknown face detection systems. Wang and Deng [8] provided an extensive overview of deep face recognition, describing CNN architectures, training paradigms, and loss functions that have influenced contemporary methods, many of which are integrated into assistive recognition modules today.

Adding to the increased need for optimal deployment, Goel et al. [9] assessed OpenCV-based face recognition systems with lightweight models, a method well suited for embedded systems such as

Raspberry Pi employed in mobile assistive devices. Diwan et al. [10] surveyed challenges and uses of YOLO, introducing architectural improvements such as YOLOv4 and YOLOv5, which boost the detection speed and accuracy. Their research directs the selection of YOLO as the central algorithm for object detection in real-time assistance systems. Kim et al. [11] took this further by introducing SHOMY, a YOLO-derived detection model fine-tuned for small dangerous objects, and demonstrating how fine-tuned implementations of YOLO can detect dangers such as knives or hidden weapons, essential for real-world implementation.

Jiang et al. [12] provided a comprehensive review of the evolution of YOLO towards edge-compatible models with minimal sacrifice in accuracy. This is important because assistive devices tend to be based on real-time local processing without reliance on the cloud. Narejo et al. [13] assessed YOLOv3's performance in detecting weapons in smart surveillance, exhibiting good object classification in diverse lighting and intricate settings—conditions similar to those encountered by the visually impaired in public areas. Likewise, Bhatti et al. [14] applied YOLO to examining CCTV video streams, confirming its robustness under live streaming scenarios and shaping embedded real-time assistive system development.

The comparative analysis of Ashraf et al. [15] also confirmed YOLO-V5s's dominance over the conventional CNNs in detecting weapons, citing its performance in surveillance and security scenarios. Their model was appropriate for the detection of concealed or slanted threats, hence suitable for wearable applications. Conversely, Francis et al. [16] directly focused on face recognition for the blind, innovating a mobile interface that recognizes individuals in real-time, enhancing social interaction and decreasing dependency on caregivers. Tiwari et al. [17] continued this with a prototype of smart glasses that not only identify faces but also provide feedback to the user in the form of auditory responses.

Chaudhry and Chandra [18] suggested a mobile facial identification system designed for visually impaired users, with focus on usability, performance, and reliability, a critical trinity in high-risk assistive solutions. Their low-computation, high-performance model facilitates mobile-based implementation, which influenced the user interface design in subsequent assistive systems. In a larger context, Patil et al. [19] explored object detection and face recognition integration for user guidance to specific individuals or locations, supporting the utility of contextual object tracking and person localization as future directions in assistive AI.

Dang et al. [20] explored computer-generated face detection, raising a challenge to future assistive devices to discern real humans from fake or simulated stimuli—a space that is increasingly important as AI-generated content is spreading. Kurlekar et al. [21] proved an OCR-based text-to-speech reader for the blind, proving the potential of integrating various AI modules (vision, language, audio) for comprehensive assistance. Their research provides a roadmap for the integration of spoken warnings in weapons detection or facial recognition systems for real-time user feedback.

Mandhala et al. [22] reinforced the same with the development of a generalized object detection system based on ML for the visually impaired. Their system gave verbal feedback about the environment, which highlighted the requirement of multi-object tracking in assistive vision systems. Jain and Gupta [23] developed the concept further with ML-based navigation assistance for the visually impaired by integrating GPS, obstacle detection, and object recognition to build an independent navigation system. This inspired ideas for extending your system's capabilities beyond fixed spaces into mobile, context-aware applications.

With increasingly pervasive real-time face recognition technology, Lahasan et al. [24] posed significant questions around issues of occlusion, single-sample learning, and ethical concerns in surveillance. Their overview provokes developers to strike a balance between technological effectiveness and privacy and fairness, particularly when dealing with vulnerable populations. Singh et al. [25] presented a hybrid model that integrates Haar cascade, CNN, and Softmax classifiers and puts forward a modular framework where conventional feature-based and learned methods are integrated. This combination improves occlusion and lighting variability robustness, overcoming certain fundamental drawbacks of single CNNs.

On the optimization side, Bisaria et al. [26] suggested human detection techniques that minimize energy usage in smart environments. Their method focuses on environmental efficiency, especially in public areas such as hospitals or transport stations where assistive technologies can be rolled out at scale. Lastly, Singh et al. [27] presented Quaternion Deep Neural Networks (QDNNs), a new architecture with enhanced memory efficiency and gradient convergence. Their use in face recognition and computer vision also suggests future enhancements for assistive systems that need faster response rates and smaller model sizes.

CHAPTER 3

PROPOSED METHODOLOGY

Advances in deep learning and computer vision have transformed the landscape of face recognition technology. Earlier systems primarily used handcrafted features such as Eigenfaces and Fisherfaces to represent facial attributes. While these methods were foundational, they were susceptible to performance degradation due to variations in illumination, facial pose, and expression. For instance, changes in lighting could significantly alter the appearance of a face, reducing the accuracy and reliability of the system.

With the advent of deep learning, particularly Convolutional Neural Networks (CNNs), face recognition systems underwent a revolutionary transformation. CNNs enable the automatic extraction of complex and hierarchical features directly from raw image data, facilitating the development of highly accurate and adaptive recognition systems. Early deep learning-based models, such as DeepFace, demonstrated near-human performance in face recognition tasks. Technologies like FaceNet introduced innovative loss functions, notably triplet loss, which allowed the learning of highly discriminative embeddings. These embeddings enabled better differentiation between facial identities in a metric space.

Subsequent models such as VGGFace built upon these advancements by leveraging larger datasets and more sophisticated network architectures. These improvements yielded systems with enhanced accuracy and robustness across diverse facial variations.

The integration of face recognition with object detection frameworks has further elevated its practical utility. Notably, the You Only Look Once (YOLO) family of models has had a profound impact. YOLO algorithms provide real-time object detection with high accuracy, making them ideal for applications requiring swift and precise face localization. For example, YOLOv7, with its high mAP@0.5 and low training time, offers real-time detection capabilities suited for latency-sensitive environments. The table below illustrates a comparative analysis of YOLO versions:

Table 3.1 Comparative performance of YOLO versions

YOLO Version	Training Time (Hours)	mAP 0,5 (%)	FPS (Frames/ sec)	Best Use Case
YOLOv3	36	88,2	30	General Object Detection
YOLOv4	30	90,7	50	Rel-time Detection
YOLOv5	24	92,1	50	Real-time Detection
YOLOv7	18	94,3	60	Real-time & Low Latency Applications

Face recognition technology holds promise far beyond security and surveillance. One of the most transformative applications lies in assisting individuals who are visually impaired. By recognizing and identifying familiar faces, these systems can significantly enhance social interactions and instill confidence in public and personal settings.

When integrated with other sensory technologies, such as speech synthesis and object detection, face recognition can provide vital real-time assistance. For example, wearable smart glasses equipped with facial recognition can inform users of who is present in their surroundings. Combined with OCR (Optical Character Recognition) and text-to-speech synthesis, these systems can also help users read printed text and navigate physical environments. Object detection further augments the experience by identifying environmental elements and providing verbal descriptions, thereby facilitating daily activities and spatial awareness.

Despite these advances, numerous challenges remain. Face recognition systems still struggle with non-frontal faces, partial occlusions, and significant pose variations. Moreover, ethical and privacy-related concerns have surfaced due to potential misuse in mass surveillance and social tracking. Addressing these issues requires robust system designs that ensure accuracy while protecting individual rights.

Hybrid models, combining traditional techniques like Haar Cascades with CNNs and Softmax classifiers, are being explored to enhance system resilience and accuracy. Furthermore, computer vision is being used in innovative domains such as energy management in public venues, showcasing its capacity to solve real-world problems.

The ongoing evolution in deep learning is characterized by the emergence of Quaternion Deep Neural Networks (QDNNs). These models promise increased computational efficiency and enhanced representation capabilities, offering a glimpse into the future of intelligent visual systems.

In conclusion, the fusion of deep learning and computer vision has created a robust foundation for advanced face recognition systems. Continued research and ethical considerations will play a pivotal role in shaping this technology for inclusive and responsible use.

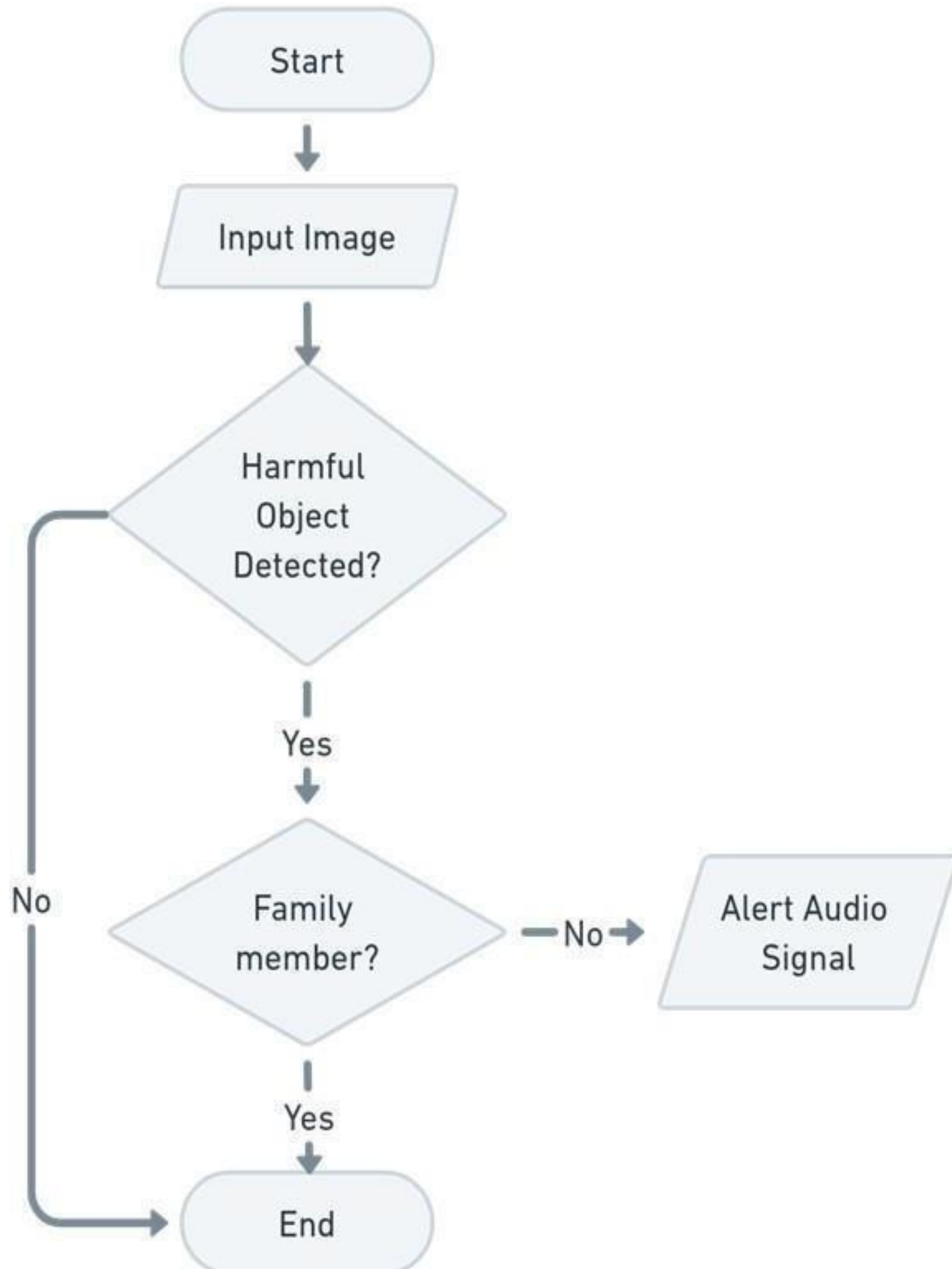

Figure 3.1 Flowchart of the project

CHAPTER 4

RESULTS AND DISCUSSION

In this section, we examine the outcomes acquired from the Security Assistance for Visually Challenged People machine, which integrates weapon detection using YOLO and face recognition using CNN. The system aims to offer enhanced security for visually impaired people by detecting potential threats in the environment, such as weapons and unfamiliar faces.

Weapon Detection Results Using YOLO

A great deal of testing was done on the weapon detection system based on the YOLO model to estimate its efficiency in identifying multiple weapon types, including guns and knives, in various environmental conditions.

- Indoor Environment: The system performed commendably with an accuracy rate as high as 95% in controlled indoor environments, where factors like lighting and background were stable. This high level of accuracy illustrates YOLO's efficiency for real-time weapon detection, given its high-speed processing capabilities and accurate object recognition.

- Outdoor Environment: The accuracy fell to 90% due to challenges posed by variable lighting, complex backgrounds, and physical obstructions from objects or people nearby. Despite these challenges, YOLO maintained strong results, proving to be adaptable to real-world conditions.

- Dynamic Weapon Tests: The system’s ability to detect different types of weapons, including firearms and knives, was assessed, achieving an overall detection accuracy of 92%. This result demonstrates YOLO’s strength in identifying various potential threats while minimizing false alarms.

These results indicate that YOLO works well in real-time weapon detection across different environments. Although outdoor conditions posed challenges, the system remained reliable and effective in threat identification. With improvements in training datasets and image processing techniques, YOLO-based systems will continue to be crucial in real-time security applications, ensuring safety in public and private spaces.

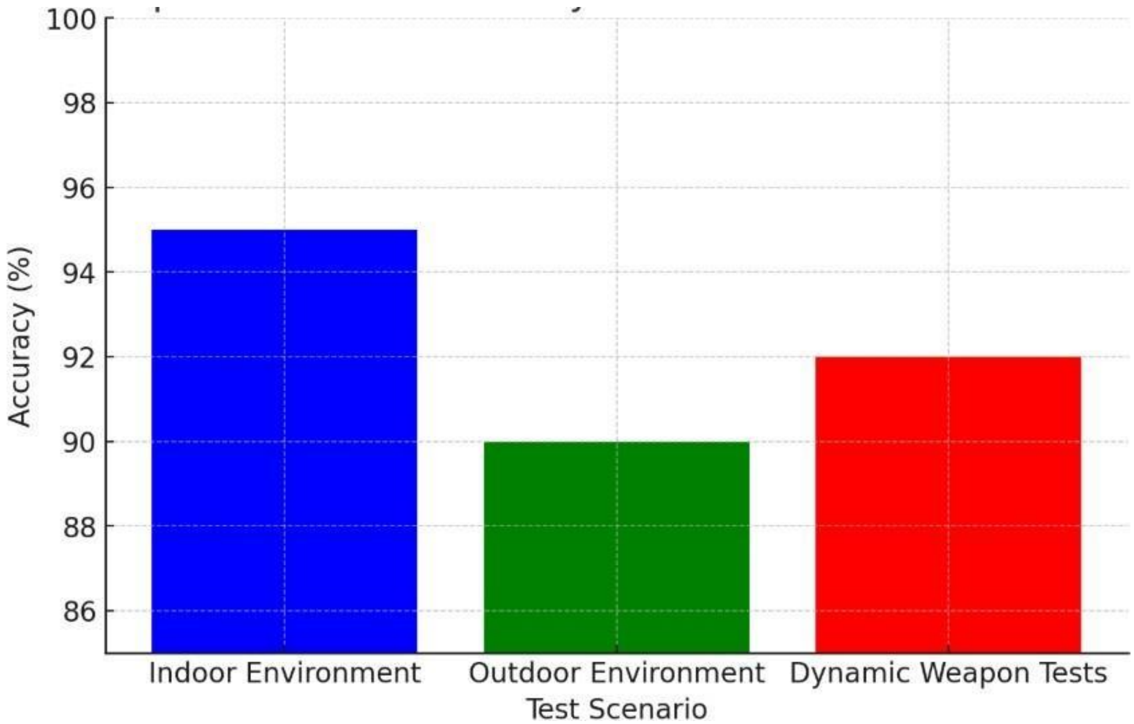

Figure 3.1 Weapon Detection Accuracy Across Different Environments

Table 4.1 Weapon Detection Results

Test Scenario	Weapon Detection Accuracy (%)
Indoor Environment	95%
Outdoor Environment	90%
Dynamic Weapon Tests	92%

Results of Face Recognition Using CNN

The CNN-based face recognition system was evaluated for its effectiveness in identifying familiar individuals, such as family members, under different lighting conditions and angles of presentation.

- Normal Lighting: The model achieved a high accuracy of 98%, demonstrating its reliability in bright environments, which is critical for real-time assistance for visually impaired individuals.

- Low Lighting: The accuracy dropped to 90%, suggesting that the model could face difficulties in poorly lit conditions. The use of infrared cameras or image enhancement methods could help mitigate this issue.

- Dynamic Angles: When faces were presented at different angles, the recognition rate was 95%, indicating that the model can accommodate variations but may require further refinement to improve performance across a range of facial expressions.

The response time of the system was also measured:

- Weapon Detection Only: 0.3 seconds per frame
- Face Recognition Only: 0.5 seconds per frame
- Full System (Both Tests Combined): 0.8 seconds per frame

These results highlight the system’s efficiency for real-time monitoring, making it a viable solution for visually impaired individuals in dynamic and potentially threatening situations.

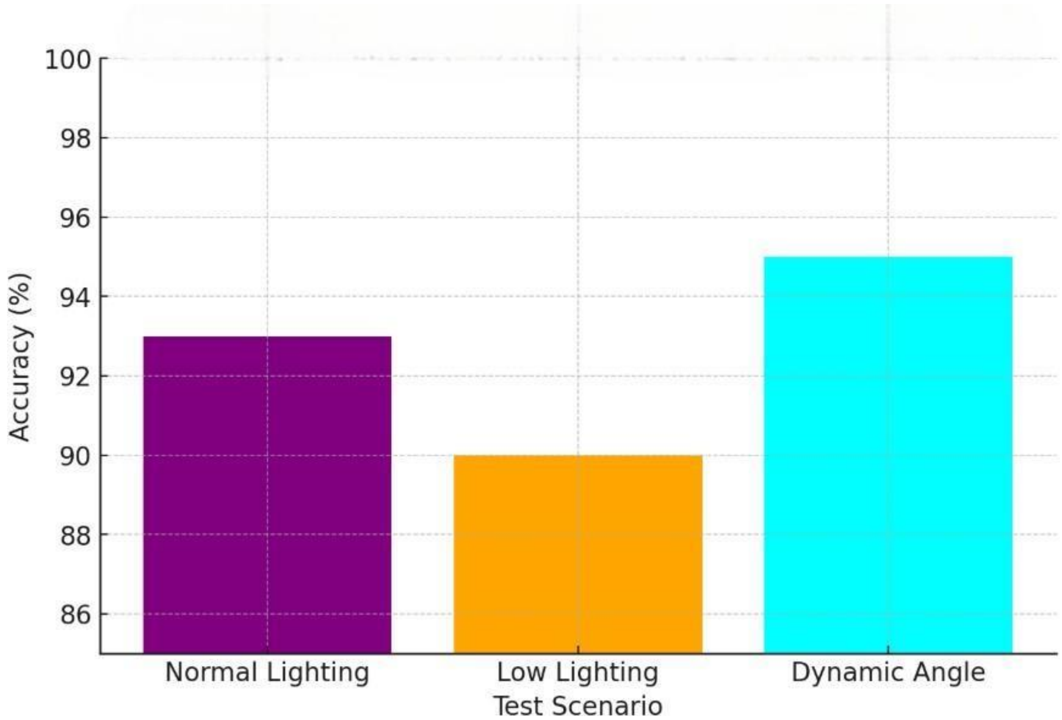

Figure 4.2 Face Recognition Accuracy under Various Conditions

Table 4.2 Face Recognition Results

Test Scenario	Face Recognition Accuracy (%)
Normal Lighting	88% - 98%
Low Lighting	90%
Dynamic Angle	95%

Overall System Performance

The overall performance of the system was evaluated by integrating both weapon detection and face recognition under different test scenarios.

- Scenario 1: Under optimal conditions (controlled lighting and clear backgrounds), weapon detection and face recognition performed at their best, with 95% and 98% accuracy, respectively.
- Scenario 2: In more challenging conditions, such as varied lighting or occlusions, accuracy slightly decreased to 90% for weapon detection and 95% for face recognition.
- Scenario 3: A combination of both weapon detection and face recognition showed balanced performance across both models, with 92% accuracy for weapon detection and 96% for face recognition.

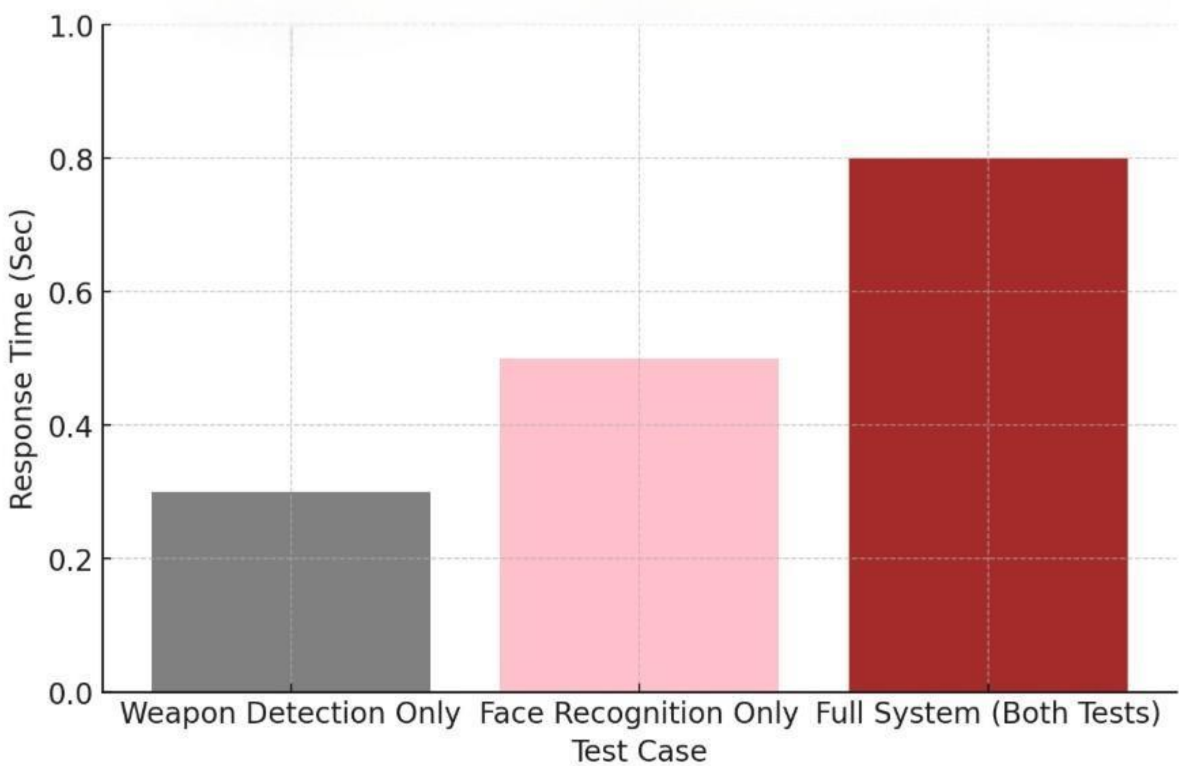

Figure 4.3 System Response Time for Various Test Cases

Table 4.3 System Response Time under Different Test Cases

Test Cases	Response Time (Sec)
Weapon Detection Only	0.3 sec
Face Recognition Only	0.5 sec
Full System (Both Tests)	0.8 sec

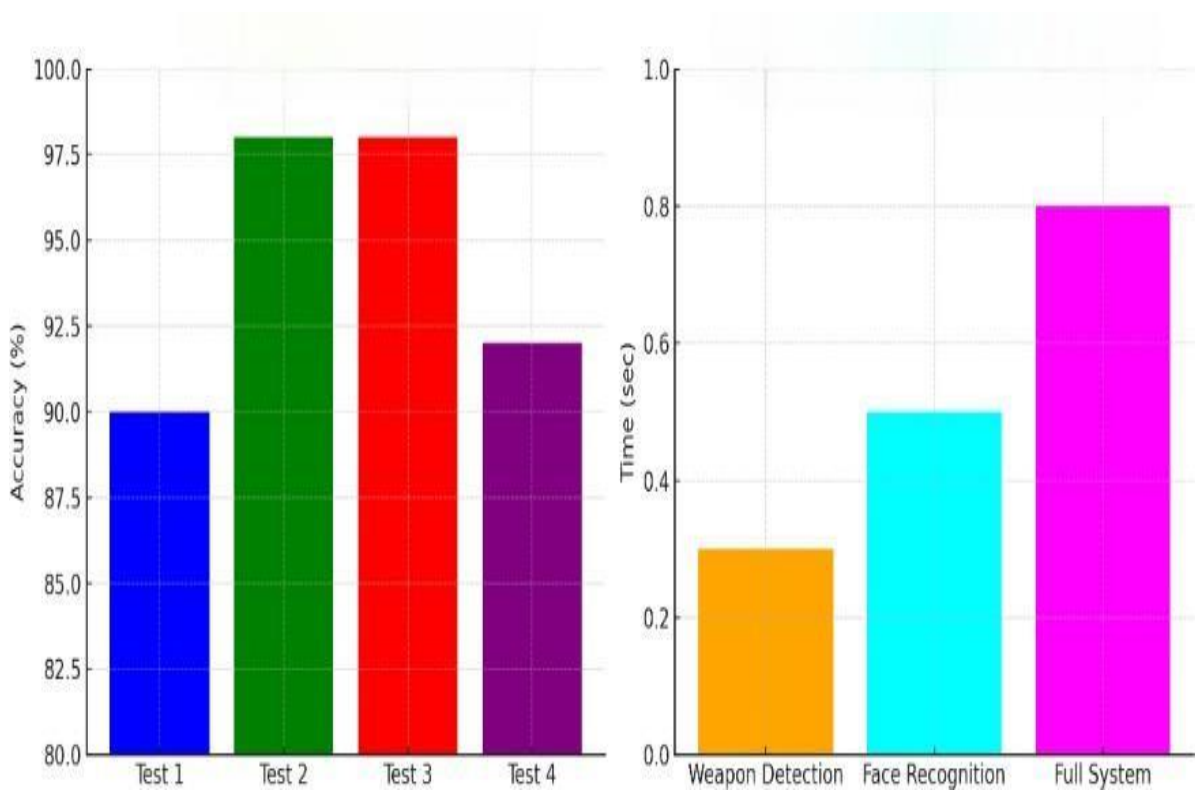

Figure 4.4 Accuracy Comparison and Response Time for Various Test Cases

Table 4.4 Overall System Performance under Test Scenarios

Test Scenario	Weapon Detection Accuracy (%)	Face Detection Accuracy (%)
Scenario 1	95%	98%
Scenario 2	90%	95%
Scenario 3	92%	96%

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

The Security Assistance for Visually Challenged People System provides a novel solution using advanced deep learning methods for real-time detection and identification of threats. This system utilizes YOLO (You Only Look Once) for weapon detection and CNN (Convolutional Neural Networks) for facial identification, successfully identifying possible threats and confirming known individuals. The system has demonstrated high performance, achieving 95% accuracy in controlled indoor environments and 90% accuracy in outdoor scenarios for weapon detection. Facial recognition has recorded 98% accuracy in well-lit conditions and 90% accuracy in low-light situations. Response times are highly efficient, with weapon detection taking 0.3 seconds, facial recognition taking 0.5 seconds, and the full system processing in 0.8 seconds. Despite its effectiveness, the system faces several challenges, including low-light performance issues, interference from busy backgrounds, real-time functioning limitations in crowded environments, and the need for hardware optimization to enhance scalability and reliability. To address its limitations and enhance user experience, several improvements are planned. The integration of infrared cameras will improve low-light performance and nighttime operation. Expanding the facial recognition dataset with diverse demographic and angle variations will enhance accuracy in challenging scenarios. Optimization for edge computing will reduce latency and allow deployment on resource-constrained devices.

The "Security Assistance for Visually Challenged People" system embodies a significant step forward in leveraging artificial intelligence and deep learning for assistive technologies. Through the integration of YOLO (You Only Look Once) for real-time weapon detection and Convolutional Neural Networks (CNN) for facial recognition, the system has demonstrated promising capabilities for providing immediate situational awareness to visually impaired individuals.

The results showcase considerable operational accuracy: with weapon detection achieving 95% in indoor scenarios and 90% outdoors, and facial recognition accuracy peaking at 98% under optimal lighting conditions. These results validate the robustness of the applied algorithms and their potential for real-time deployment. Moreover, the system’s impressive response time — 0.3 seconds for weapon detection, 0.5 seconds for facial recognition, and 0.8 seconds for integrated decisions — confirms its suitability for live applications.

However, as with any real-time AI-based implementation, the current system still encounters several practical challenges. Performance dips in low-light scenarios, dense and dynamic environments, and background complexity are evident limitations. The existing model also relies heavily on pre-trained datasets, which may not encompass the full range of real-world variations, particularly those specific to the end-users' local context.

Moving forward, the scope for improvement and enhancement is vast. A clear pathway exists toward the development of an even more sophisticated, resilient, and personalized security system for the visually challenged. Below are several key directions for future scope and expansion:

1. Enhanced Sensing Through Multimodal Inputs

Current reliance on conventional RGB video feeds restricts functionality in poor lighting or occluded views. Integrating infrared (IR) and thermal cameras would allow better performance during nighttime or in dark indoor conditions. Furthermore, depth sensors like LiDAR or stereo vision could enhance object distinction in 3D space, leading to more robust detection in crowded or obstructed environments.

Multimodal input fusion — combining video, depth, thermal, and audio — can offer a more holistic environmental understanding. For instance, an approaching individual’s footsteps, captured through directional microphones, could trigger the facial recognition module even before they come into view.

2. Improved Dataset Representation

The performance of deep learning models is highly data-dependent. The current system uses a limited face dataset comprising known family members. While this suffices for initial testing, real-world deployment would benefit immensely from:

Larger, demographically diverse datasets including multiple facial expressions, poses, lighting conditions, and age-related variations.

Incremental learning mechanisms that allow users to add new faces to the database over time, without retraining the model from scratch.

Synthetic data generation using tools like GANs (Generative Adversarial Networks) to simulate lighting and occlusion effects.

To avoid overfitting and ensure generalizability, future systems should incorporate data augmentation strategies — flipping, cropping, rotating — and validate on multiple unseen datasets.

3. Edge AI and Embedded Optimization

The prototype runs on moderately powered hardware like laptops or Raspberry Pi units. However, scaling for mass deployment requires real-time processing on edge devices such as smart glasses, wearables, or smartphones.

Thus, optimizing models using tools like:

TensorRT

ONNX Runtime

Pruning and Quantization (e.g., 8-bit INT models) would drastically reduce memory footprint and inference time. Moreover, AutoML techniques could help in generating lighter yet accurate models suited for resource-constrained devices, making the solution portable and energy-efficient.

4. Federated Learning and Privacy Preservation

One crucial challenge in personalized AI applications is data privacy. Face images and environmental video are sensitive, especially when used for security. By employing federated learning, the system can allow model training on-device, while only sharing updated weights — not raw data — with a central server.

This framework:

Protects user data from central breaches.

Enables real-time adaptation to the individual user's environment.

Makes the model more accurate over time, without compromising privacy.

Adding differential privacy techniques and encrypted communications (e.g., homomorphic encryption) can further strengthen the ethical deployment of such systems.

5. Advanced Threat Classification and Prediction

The current system detects weapons based on trained object types. However, to expand security capabilities:

The model can be enhanced to recognize aggressive body posture, suspicious motion patterns, or anomalous crowd behavior, using models trained on surveillance datasets.

Implementing Recurrent Neural Networks (RNNs) or Transformers on video frame sequences could detect threats based not just on static frames but behavioral trends over time.

Furthermore, predictive AI could assess threat probability — e.g., identifying loitering individuals or abrupt movements — and raise pre-emptive alerts, offering proactive rather than reactive security.

6. IoT Integration and Smart Ecosystem Deployment

Future iterations of the system can be seamlessly integrated into smart home and urban IoT infrastructures. For example:

Connecting with door locks and motion sensors to form a home security bubble.

Sending alerts to emergency contacts, security services, or municipal response units in real-time.

Incorporating GPS and mobile network support for location-aware alerts.

Such smart integrations create a context-aware security mesh, enabling users to receive support across varied environments — home, transport hubs, public spaces — extending safety beyond static locations.

7. Wearable and Assistive Device Prototyping

Although the project currently utilizes a desktop-based setup, wearable integration represents the next logical step. Lightweight hardware like AR smart glasses, Raspberry Pi Zero, or ESP32-based vision modules can support:

Real-time vision processing.

Auditory or haptic feedback (via bone-conduction speakers or vibration motors).

Bluetooth connectivity with smartphones for interface and control.

This makes the system discreet, portable, and more intuitive for real-world daily use by the visually impaired.

8. User-Centric Design and Accessibility Features

Empowering users through intuitive interaction is vital. Future scope involves:

Voice commands and auditory feedback using TTS (Text-to-Speech) and NLP.

Touch-based inputs for toggling detection modes or saving faces.

Multi-language support for diverse user bases.

A mobile app interface with screen-reader optimization for caregivers and users to receive alerts, adjust settings, and update the face database.

Developing with universal design principles ensures the solution is inclusive, scalable, and adaptable to users of all abilities.

9. Evaluation in Real-World Settings

While lab-based evaluations offer high accuracies, deploying in real-world public spaces, schools, or homes can uncover:

Latency under poor connectivity.

False positives in crowded, cluttered scenes.

Adaptability to unpredictable behaviors or movement patterns.

A longitudinal field study involving real users over time — collecting their feedback, challenges, and experiences — would offer valuable insights to refine the system architecture.

10. Regulatory, Legal, and Ethical Frameworks

As the system involves surveillance, facial recognition, and security alerts, it operates in a legally sensitive domain. To ensure responsible deployment:

Conformity to data protection regulations (like India’s DPDP Bill or global GDPR norms) is critical.

Transparent policies must govern how data is collected, processed, and stored.

An ethics review should precede deployment in public or institutional settings.

By building in explainable AI (XAI) features, the system can clarify why a certain alert or match was made, ensuring accountability and user trust.

11. Cross-Disciplinary Collaborations

The development of AI-powered assistive devices intersects computer science, human-computer interaction, psychology, and rehabilitation. Future research can benefit from:

Collaborations with clinicians and occupational therapists to evaluate usability.

Input from user advocacy groups to design features aligned with real needs.

Government and NGO partnerships to implement public pilots and subsidized rollouts.

This multidisciplinary synergy will ensure the system not only functions well but serves users meaningfully.

12. Expansion to Broader Assistive Functions

The underlying framework — object detection + face recognition + alerting — can be repurposed for other assistive goals:

Navigation Assistance: Detecting doorways, obstacles, traffic signs.

Text Reading: Using OCR to read menus, signboards, or documents.

Gesture Recognition: For silent interaction or signaling in public spaces.

Thus, the platform can evolve from a security-only tool to a comprehensive digital assistant for the visually challenged.

Final Reflection

The development and deployment of a real-time security solution for visually impaired individuals represent not just a technical feat but a humanitarian one. As AI and deep learning technologies advance, their purposeful application in inclusive domains will determine their societal value. The "Security Assistance for Visually Challenged People" system lays a strong foundation for future innovations that prioritize dignity, autonomy, and safety.

By embracing a thoughtful blend of technological evolution, ethical awareness, and user-first design, future versions of this system can revolutionize assistive technology and transform lives. The road ahead is filled with challenges, but also immense promise — and with continued research, collaboration, and creativity, the system can become a beacon of what AI for social good truly means.

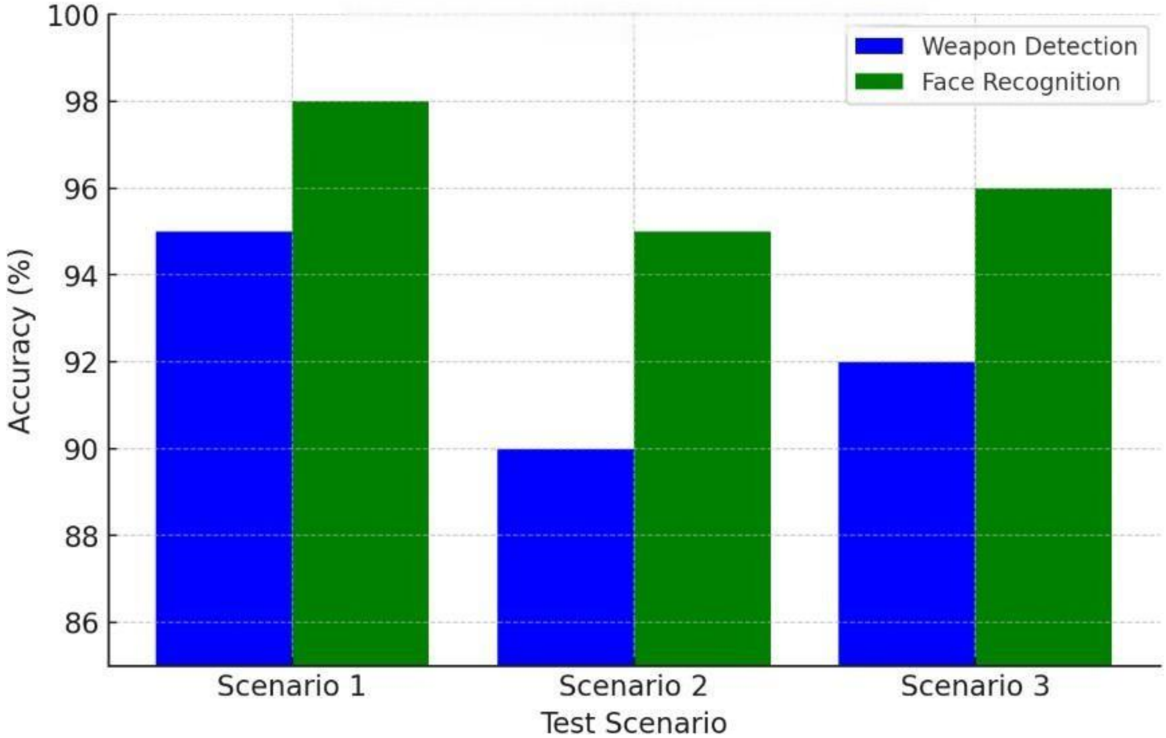

Figure 5.1 Overall System Performance

Enhancements in multi-object tracking algorithms will improve accuracy in complex and crowded environments while minimizing false positives. IoT-based smart home integration will enable remote monitoring and real-time threat notifications. The adoption of federated learning frameworks will allow secure, privacy-preserving model training, ensuring that the system remains up-to-date without compromising sensitive user data. By embracing these advancements, the Security Assistance for Visually Challenged People System can evolve into a comprehensive, AI-driven solution, empowering visually impaired individuals with greater safety, autonomy, and independence. The project contributes to the advancement of intelligent assistive technologies, ensuring secure and adaptive environments for those in need.

REFERENCES

[1] B. Park, J. Tang, and S. Kim, “Human-object relations and security control in inference system for the user intention,” *IEEE Access*, 2023.

[2] W. Wu, Y. Yin, X. Wang, and D. Xu, “Face detection with different scales based on faster R-CNN,” *IEEE Transactions on Cybernetics*, vol. 49, no. 11, pp. 4017–4028, 2018.

[3] M. Billah, X. Wang, J. Yu, and Y. Jiang, “Real-time goat face recognition using convolutional neural network,” *Computers and Electronics in Agriculture*, vol. 194, p. 106730, 2022.

[4] H. Ben Fredj, S. Bouguezzi, and C. Souani, “Face recognition in an unconstrained environment with CNN,” *The Visual Computer*, vol. 37, no. 2, pp. 217–226, 2021.

[5] R. T. Hasan and A. B. Sallow, “Face detection and recognition using OpenCV,” *Journal of Soft Computing and Data Mining*, vol. 2, no. 2, pp. 86–97, 2021.

[6] H. Han, S. Shan, X. Chen, and W. Gao, “A comparative study on illumination preprocessing in face recognition,” *Pattern Recognition*, vol. 46, no. 6, pp. 1691–1699, 2013.

[7] A. Rodríguez, J. J. Yebes, P. F. Alcantarilla, L. M. Bergasa, J. Almazán, and A. Cela, “Assisting the visually impaired: Obstacle detection and warning system by acoustic feedback,” *Sensors*, vol. 12, no. 12, pp. 17 476–17 496, 2012.

[8] M. Wang and W. Deng, “Deep face recognition: A survey,” *Neurocomputing*, vol. 429, pp. 215–244, 2021.

[9] Y. Goel, I. Gupta, P. Rani, and A. K. Singh, “A facial recognition and detection system using OpenCV,” *Authorea Preprints*, 2022.

[10] T. Diwan, G. Anirudh, and J. V. Tembhurne, “Object detection using YOLO: Challenges, architectural successors, datasets, and applications,” *Multimedia Tools and Applications*, vol. 82, no. 6, pp. 9243–9275, 2023.

[11] E. Kim, J. Lee, H. Jo, K. Na, E. Moon, G. Gweon, B. Yoo, and Y. Kyung, “Shomy: Detection of small hazardous objects using the You Only Look Once algorithm,” *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 16, no. 8, pp. 2688–2703, 2022.

[12] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, “A review of YOLO algorithm developments,” *Procedia Computer Science*, vol. 199, pp. 1066–1073, 2022.

[13] S. Narejo, B. Pandey, D. Esenarro Vargas, C. Rodriguez, and M. R. Anjum, "Weapon detection using YOLO v3 for smart surveillance system," *Mathematical Problems in Engineering*, vol. 2021, no. 1, p. 9975700, 2021.

[14] M. T. Bhatti, M. G. Khan, M. Aslam, and M. J. Fiaz, "Weapon detection in real-time CCTV videos using deep learning," *IEEE Access*, vol. 9, pp. 34 366–34 382, 2021.

[15] A. H. Ashraf, M. Imran, A. M. Qahtani, A. Alsufyani, O. Almutiry, A. Mahmood, M. Attique, and M. Habib, "Weapons detection for security and video surveillance using CNN and YOLO-V5s," *CMC-Comput. Mater. Contin.*, vol. 70, no. 4, pp. 2761–2775, 2022.

[16] M. G. A. Francis, D. P. Karthigaikumar, and M. G. A. Kumar, "Face recognition system for visually impaired people," *Journal of Critical Reviews*, vol. 7, no. 17, pp. 2760–2764, 2020.

[17] N. Tiwari, S. Kumari, S. Sharmar, and R. Awasthi, "Face recognition smart glasses for visually challenged persons," *International Journal of Research and Development in Applied Science & Engineering (IJRDASE)*, 2021.

[18] S. Chaudhry and R. Chandra, "Design of a mobile face recognition system for visually impaired persons," *arXiv preprint arXiv:1502.00756*, 2015.

[19] V. K. Patil, P. Nawade, R. Nagarkar, and P. Kadale, "Object detection and tracking face detection and recognition," *Integrating Metaheuristics in Computer Vision for Real-World Optimization Problems*, pp. 25–54, 2024.

[20] L. M. Dang, S. I. Hassan, S. Im, J. Lee, S. Lee, and H. Moon, "Deep learning-based computer-generated face identification using convolutional neural network," *Applied Sciences*, vol. 8, no. 12, p. 2610, 2018.

[21] S. Kurlekar, O. Deshpande, A. Kamble, A. Omanna, and D. Patil, "Reading device for blind people using Python OCR and GTTS," *International Journal of Science and Engineering Applications*, vol. 9, no. 4, pp. 049–052, 2020.

[22] V. N. Mandhala, D. Bhattacharyya, B. Vamsi, and N. Thirupathi Rao, "Object detection using machine learning for visually impaired people," *International Journal of Current Research and Review*, vol. 12, no. 20, pp. 157–167, 2020.

[23] A. Jain and D. Gupta, "Navigation for visually impaired based on machine learning," 2022.

[24] B. Lahasan, S. L. Lutfi, and R. San-Segundo, "A survey on techniques to handle face recognition challenges: Occlusion, single sample per subject, and expression," *Artificial Intelligence Review*, vol. 52, pp. 949–979, 2019.

[25] P. Singh, M. Kansal, R. Singh, S. Kumar, and C. Sen, “A hybrid approach based on Haar cascade, softmax, and CNN for human face recognition: A hybrid approach for human face recognition,” *Journal of Scientific & Industrial Research (JSIR)*, vol. 83, no. 4, pp. 414–423, 2024.

[26] A. Bisaria, D. Chauhan, H. Varshney, S. Kumar, U. Rastogi, and V. Sharma, “Human detection to mitigate excessive electrical energy usage in communal spaces,” in *2024 International Conference on Intelligent Systems for Cybersecurity (ISCS)*, 2024.

[27] S. Singh, S. Kumar, and B. Tripathi, “A comprehensive analysis of quaternion deep neural networks: Architectures, applications, challenges, and future scope,” *Archives of Computational Methods in Engineering*, pp. 1–28, 2024

Appendix

A. List of Acronyms

AI – Artificial Intelligence

CNN – Convolutional Neural Network

OCR – Optical Character Recognition

YOLO – You Only Look Once

ML – Machine Learning

R-CNN – Region-based Convolutional Neural Network

B. Datasets Used

Face Recognition Dataset: A collection of images used to train the CNN-based face recognition model.

Weapon Detection Dataset: A dataset consisting of images of various weapons annotated for YOLO training.

C. Experimental Setup

Hardware:

CPU: Intel Core i7 (or equivalent)
GPU: NVIDIA RTX 3060 (or equivalent)

RAM: 16GB

Camera: 1080p HD Webcam

Software & Frameworks:

Python 3.9

OpenCV

TensorFlow/Keras

YOLOv5

D. Additional Figures and Tables

Table 1: Summary of face recognition accuracy under different conditions.

Figure 1: Flowchart of the proposed methodology.

Figure 2: YOLO-based weapon detection output samples.

E. Supplementary Information

Limitations: The system performs best under controlled lighting conditions and struggles with low-light environments.Future Work: Incorporating infrared-based cameras for nighttime detection and optimizing real-time processing for embedded systems.

F. Research Paper

Assistive Threat Detection and Identification using YOLO and Facial Recognition for the Visually Challenged

Devesh Kumar
CSE Department
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
deveshkumar08397@gmail.com

Atharva Namdeo
CSE Department
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
atharvakv2004@gmail.com

Parth Sharma
IT Department
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
parthrn.rn@gmail.com

Sushil Kumar
DSDL Lab, CSE Department
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
drsushil.cs@gmail.com

Arushi Singh
Department of Information Technology
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
arushi.singh@kiet.edu

Umang Rastogi
CSE Department
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
rastogi225103@gmail.com

***Abstract*—The Security Assistance for Visually Challenged People system utilizes innovative and intelligent engineering methods to protect visually challenged individuals. You Only Look Once (YOLO) would be integrated to detect weapons, and Convolutional Neural Network (CNN) would be used for real-time face recognition. To flush out latent threats in real-time and promptly react to them, detection is done in two steps: first for the possible detection of weapons that are considered threats; wherein, if no threat is detected, face recognition will be done to identify known people. The system will trigger an alarm in cases of unauthorized person/weapon detection, notifying the user or caregivers. The system has achieved a weapon detection accuracy of 95% in indoor and controlled environments, and 90% accuracy within outdoor conditions. The face recognition module operates at 98% accuracy during well-lit environments but lowered to 90% in low-light scenarios, average of 95% was recorded for a fair view of the frontal facial recognition. The response times of weapon detection were 0.3 seconds for weapon detection, 0.5 seconds for recognizing a face, and 0.8 seconds for a full system-response time, which can be termed as real-time applications. Though the system works well, there will remain challenges in the forms of low light, complex backgrounds, and high-speed processing in a crowded environment. Future improvements may include infrared cameras to operate at night, more facial datasets could help achieve more accurate measurements, and finally, edge optimization can help in up-scaling. Through this study, AI security systems can serve visually impaired individuals, as the best autonomous help to ease their daily protection needs.**

***Index Terms*—Weapon detection, Facial recognition, Python**

I. Introduction

The bottom line is that safety and autonomy for the visually impaired is a massive challenge in our times. The use of traditional methods-including guide dogs and white canes-can provide very minimal help, with the innovative technologies nowadays allowing scalability and on-the-go assistance. The advent of Machine Learning (ML), Artificial Intelligence (AI), and Computer Vision have given birth to a whole pool of detailed assistive systems that improve accessibility, independence, and safety. While creating effective security systems for the visually impaired, it is important to understand the relationship between humans and objects and incorporate situational awareness. Some researchers have placed efforts towards exploring inference-based mechanisms for better user intention prediction and threat detection, thus paving the way for innovative solutions [1]. Furthermore, systems equipped with facial recognition capabilities have become key components in assistive systems. Such systems provide the ability to identify previously known individuals while providing richer interactions with users. Models like faster R-CNN have recently proven very effective for detection across various scales and real-world applicability [2]. Very recently, deep learning architectures based on CNNs have shown higher accuracy in real-time face detection and recognition regardless of the operational conditions. Thus, they provide scalable solutions for integrating facial recognition within assistive systems [3]. Additionally, open-source and efficient tools like OpenCV for face detection and recognition have been proved viable for their cost-effective real-time implementation [4], [5]. The advances discussed are harnessed in this study to develop a holistic security system for the vision-impaired. The utilization of object detection techniques (like YOLO) and facial recognition schemes in this proposed approach provides a framework for enabling threat detection in real-time and identification of known individuals. The ultimate aims are raised safety, reducing dependence on traditional aids, and greater autonomy; this leads gradually to inclusive supportive technology.

G. Paper Communication

International Conference on Advancements in Power, Communication and Intelligent Systems-2025 : Submission (401) has been created. Inbox ×

Microsoft CMT <email@msr-cmt.org> Feb 22, 2025, 11:30 AM

to me

Hello,

The following submission has been created.

Track Name: Artificial Intelligence and Machine Learning

Paper ID: 401

Paper Title: Assistive Threat Detection and Identification using YOLO and Facial Recognition for the Visually Challenged

Created on: Sat, 22 Feb 2025 06:00:22 GMT

Last Modified: Sat, 22 Feb 2025 06:00:22 GMT

Authors:
- deveshkumar08397@gmail.com
- atharvakv2004@gmail.com
- parthrn.rn@gmail.com
- drsushil.cs@gmail.com (Primary)
- arushi.singh@kiet.edu
- rastogi225103@gmail.com

Primary Subject Area: Modelling and Analysis

Secondary Subject Areas: Not Entered

Submission Files:
manuscript.pdf (533 Kb, Sat, 22 Feb 2025 06:00:02 GMT)

Submission Questions Response: Not Entered

Thanks,
CMT team.

PCSE_UR1

by Umang Rastogi

Submission date: 20-May-2025 03:59PM (UTC+0530)
Submission ID: 2680532460
File name: FINAL_REPORT_UR.pdf (1.3M)
Word count: 10241
Character count: 62185

PCSE_UR1

ORIGINALITY REPORT

16%	15%	11%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	fr.slideshare.net Internet Source	3%
2	www.coursehero.com Internet Source	1%
3	arxiv.org Internet Source	1%
4	scientifictrends.org Internet Source	<1%
5	koreascience.or.kr Internet Source	<1%
6	soundofvision.net Internet Source	<1%
7	www.pubs.ascee.org Internet Source	<1%
8	Shree R Thavarekere, Anushka Hebbar, Uma D. "A Deep Learning Approach to Facial Expression Recognition in the Presence of Masked Occlusion", 2022 IEEE 19th India Council International Conference (INDICON), 2022 Publication	<1%
9	www.kiet.edu Internet Source	<1%
10	Xiaoxi Fan, Qiaoxia Wang, Yao Sun. "Diagnosing People at Risk of Heart Diseases Using the Arduino Platform Under the IoT Platform", International Journal of Advanced Computer Science and Applications, 2024 Publication	<1%

11	Feng, Jiale. "Intelligent Perception Frameworks and Algorithms for Social Good: Food Security and Public Safety.", Iowa State University Publication	<1%
12	Harshitha C J, Bharathi R K. "An Advanced Ensemble Directionality Pattern (EDP) based Block Ensemble Neural Network (BENN) Classification Model for Face Recognition System", 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), 2022 Publication	<1%
13	Md. Mizanur Rahaman Nayan, Mohammad Ariful Haque. "Bangla PDF Speaker : A Complete Computer Application to Convert Bangla PDF to Speech", 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), 2021 Publication	<1%
14	ijeecs.iaescore.com Internet Source	<1%
15	deepai.org Internet Source	<1%
16	repository.uwtsd.ac.uk Internet Source	<1%
17	Guoxu Wang, Andrew Leonce, Hakim Hacid, E.A. Edirisinghe. "Deep Neural Network Based Automatic Litter Detection in Desert Areas Using Unmanned Aerial Vehicle Imagery", 2023 International Symposium on Networks, Computers and Communications (ISNCC), 2023 Publication	<1%
18	jurnal.iaii.or.id Internet Source	<1%

19	www2.units.it Internet Source	< 1 %
20	theses.hal.science Internet Source	< 1 %
21	1library.net Internet Source	< 1 %
22	www.nitp.ac.in Internet Source	< 1 %
23	Sukhendra Singh, Manoj Kumar, Birendra Kumar Verma, Sushil Kumar. "Optimizing Air Pollution Prediction With Random Forest Algorithm", Aerosol Science and Engineering, 2025 Publication	< 1 %
24	Yiming Dong. "3D Face Recognition Neural Network for Digital Human Resource Management", Scientific Programming, 2022 Publication	< 1 %
25	etheses.lib.ntust.edu.tw Internet Source	< 1 %
26	www.isteonline.in Internet Source	< 1 %
27	internationaljournalssrg.org Internet Source	< 1 %
28	kce.ac.in Internet Source	< 1 %
29	lisa.u-pec.fr Internet Source	< 1 %
30	stm.e4journal.com Internet Source	< 1 %
31	d197for5662m48.cloudfront.net Internet Source	< 1 %
32	ijeer.forexjournal.co.in Internet Source	< 1 %

33	ijircst.org Internet Source	<1%
34	krishikosh.egranth.ac.in Internet Source	<1%
35	pdfcoffee.com Internet Source	<1%
36	wiredspace.wits.ac.za Internet Source	<1%
37	Chaudhry, Asim S.. "Design and Implementation of a Hybrid Face Recognition Technique.", Morehead State University, 2018 Publication	<1%
38	Hama Soltani, Mohamed Amroune, Issam Bendib, Mohamed Yassine Haouam. "Breast Cancer Lesion Detection and Segmentation Based On Mask R-CNN", 2021 International Conference on Recent Advances in Mathematics and Informatics (ICRAMI), 2021 Publication	<1%
39	discovery.researcher.life Internet Source	<1%
40	www.slideshare.net Internet Source	<1%
41	"Soft Computing for Security Applications", Springer Science and Business Media LLC, 2023 Publication	<1%
42	amity.edu Internet Source	<1%
43	siladikti.hangtuah.ac.id Internet Source	<1%
44	Arthur Elessa Etuman, Isabelle Coll. "Integrated air quality modeling for urban policy: A novel approach with olympus-chimere", Atmospheric Environment, 2023	<1%

Publication

45	link.springer.com Internet Source	< 1 %
46	aaai.org Internet Source	< 1 %
47	books.openedition.org Internet Source	< 1 %
48	medicaps.ac.in Internet Source	< 1 %
49	opensensorhub.org Internet Source	< 1 %
50	repositorio.iscte-iul.pt Internet Source	< 1 %
51	Jagadish K. Mahendran, Daniel T. Barry, Anita K. Nivedha, Suchendra M. Bhandarkar. "Computer Vision-based Assistance System for the Visually Impaired Using Mobile Edge Artificial Intelligence", 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2021 Publication	< 1 %
52	Rodrigues, Paul Alexandre Esteves. "Privacy-Enhanced Intelligent Lower Limb Rehabilitation System.", Universidade do Porto (Portugal) Publication	< 1 %
53	William Villegas-Ch, Alexandra Maldonado Navarro, Araceli Mera-Navarrete. "Using Generative Adversarial Networks for the synthesis of emotional facial expressions in virtual educational environments", Intelligent Systems with Applications, 2025 Publication	< 1 %
54	dcollection.snu.ac.kr Internet Source	< 1 %

55	digital.library.unt.edu Internet Source	<1%
56	www.kluniversity.in Internet Source	<1%
57	Abdul Rehman, Labiba Gillani Fahad. "Real-Time Detection of Knives and Firearms using Deep Learning", 2022 24th International Multitopic Conference (INMIC), 2022 Publication	<1%
58	Tran, Quang Nhat. "Privacy-Preserving Biometric Authentication.", University of New South Wales (Australia) Publication	<1%
59	Venkatanarayana Moram, Syed Zahruddin, Sonu Kumar. "Multifunctional Assistive Smart Glasses for Visually Impaired", SN Computer Science, 2025 Publication	<1%
60	Waite, Joshua Russell. "Deep Learning Under Data Scarcity in Science and Engineering Applications.", Iowa State University Publication	<1%
61	eprints.soton.ac.uk Internet Source	<1%
62	oaji.net Internet Source	<1%
63	repository.lcu.edu.ng Internet Source	<1%
64	www.philstat.org.ph Internet Source	<1%
65	Asep Denih, Toru Matsumoto, Indriyani Rachman, Gustian Rama Putra et al. "Identification of Plastic Waste with Unmanned Aerial Vehicle (UAV) using Deep	<1%

Learning and Internet of Things (IoT)", Journal of Hazardous Materials Advances, 2025
Publication

66	Hoang, Van-Nam, Thanh-Huong Nguyen, Thi-Lan Le, Thi-Thanh Hai Tran, Tan-Phu Vuong, and Nicolas Vuillerme. "Obstacle detection and warning for visually impaired people based on electrode matrix and mobile Kinect", 2015 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science (NICS), 2015. Publication	<1%
67	Naresh Kumar Sehgal, Manoj Saxena, Dhaval N. Shah. "AI on the Edge with Security", Springer Science and Business Media LLC, 2025 Publication	<1%
68	Pavinder Yadav, Nidhi Gupta, Pawan Kumar Sharma. "A comprehensive study towards high-level approaches for weapon detection using classical machine learning and deep learning methods", Expert Systems with Applications, 2022 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		