Enhancing Disaster Response Efforts with YOLOv8-based Human Detection in Mobile Robotics

M.B. Badawi[1,2]
[1]*Mechanical Engineering Department, Faculty of Engineering, Alexandria University, El-Chatby, Alexandria 21544, Egypt.;*

[2]*Department of Artificial Intelligence, College of Information Technology, Misr University for Science and Technology (MUST), 6th of October City 12566, Egypt.*

Abd-EL-Rahman Ahmed[2]
Department of Artificial Intelligence, College of Information Technology, Misr University for Science and Technology (MUST), 6th of October City 12566, Egypt

Khaled Fadi[2]
Department of Artificial Intelligence, College of Information Technology, Misr University for Science and Technology (MUST), 6th of October City 12566, Egypt

Rania Elgohary[3]
Department of Information Systems, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt
Rania.elgohary@cis.asu.edu.eg

***Abstract*— In the aftermath of natural disasters, swiftly detecting individuals trapped beneath debris is crucial for successful rescue operations. This paper presents a Mobile Controlled Robot with advanced human detection capabilities designed to expedite search and rescue missions, emphasizing the importance of rapid response to save lives. Utilizing a YOLOv8 model with 90% accuracy, the robot analyzes real-time images captured by a webcam to detect human forms and movements, triggering a buzzer alert to notify rescue teams upon identifying potential victims. The robot's remote operation via a mobile interface enhances flexibility and adaptability in complex terrains, allowing rescue personnel to control it from a safe distance. Rigorous testing has demonstrated the system's efficacy and reliability in accurately locating trapped individuals, offering a promising solution to improve the efficiency and effectiveness of disaster response operations.**

***Keywords*— Robotics; YOLO; Deep learning; Fire rescue**

I. INTRODUCTION

The use of technology in firefighting has evolved significantly over the past century, driven by the need to improve the safety and effectiveness of firefighting efforts. Early firefighting relied heavily on manual methods, with firefighters using basic tools such as buckets, axes, and ladders. The introduction of steam-powered fire engines in the 19th century marked a significant technological advancement, allowing for more efficient water delivery and greater reach.[1]

In the 20th century, the development of motorized fire engines and the widespread use of two-way radios transformed firefighting. Motorized engines enhanced mobility and response times, while radios improved communication and coordination among firefighting teams. In the latter half of the century, innovations such as thermal imaging cameras enabled firefighters to see through smoke, locate victims, and identify hotspots, greatly enhancing their ability to conduct search and rescue operations safely and efficiently.

In recent years, integrating advanced technologies such as drones, GPS, and real-time data analytics has further revolutionized firefighting. Drones equipped with thermal cameras and sensors can provide aerial views of fire scenes, offering critical information that can guide firefighting strategies. GPS and data analytics help in mapping out fire-prone areas, predicting fire behavior, and optimizing resource allocation[2].

1.1. History of Deep Learning

Deep learning, a subset of machine learning, has its roots in early work on artificial neural networks in the mid-20th century. Warren McCulloch and Walter Pitts introduced the concept of a neural network, inspired by the human brain, in 1943. However, it wasn't until the 1980s and 1990s that significant progress was made, thanks to the development of algorithms such as backpropagation, which allowed for the training of multi-layer neural networks.

The term "deep learning" gained prominence in the 2000s with the advent of more powerful computational resources and the availability of large datasets. Pioneering work by researchers such as Geoffrey Hinton, Yann LeCun, and Yoshua Bengio laid the foundation for deep learning as we know it today. Hinton's work on deep belief networks and LeCun's development of convolutional neural networks (CNNs) were particularly influential.

Deep learning achieved significant breakthroughs in various fields, including image and speech recognition, natural language processing, and game playing. For instance, in 2012, a CNN known as AlexNet, developed by Hinton's team, won the ImageNet Large Scale Visual Recognition Challenge, demonstrating the superior performance of deep learning models over traditional machine learning techniques. This success spurred widespread interest and investment in deep learning research and applications.[3]

1.2. Integration of Deep Learning with Robotics

The integration of deep learning with robotics represents a major advancement in the field of artificial intelligence and automation. Early robotic systems relied on pre-programmed instructions and simple sensor inputs, limiting their ability to adapt to dynamic and complex environments. The advent of deep learning has enabled robots to perceive, learn, and make decisions based on vast amounts of sensory data, mimicking human-like intelligence[4].

One of the first significant applications of deep learning in robotics was in computer vision, where CNNs allowed robots to recognize and interpret visual information with high accuracy. This capability is crucial for tasks such as object detection, navigation, and manipulation. For example, robots equipped with deep learning-based vision systems can identify and grasp objects with precision, navigate through cluttered environments, and interact with humans in more natural ways.[5]

Beyond computer vision, deep learning has been applied to various aspects of robotics, including natural language processing, reinforcement learning, and sensor fusion. Reinforcement learning, in particular, has enabled robots to learn complex behaviors through trial and error, improving their performance over time[6]. Notable achievements include Google's DeepMind training a robot to play video games at a superhuman level and OpenAI's robotic hand learning to solve a Rubik's Cube.[7]

1.3. Importance of Our Robotic System

The integration of deep learning with robotics holds immense potential for enhancing disaster response and rescue operations. Natural disasters such as earthquakes, tsunamis, and hurricanes often result in large-scale devastation, with many individuals trapped beneath debris and rubble. Traditional search and rescue methods, while effective, can be slow and labor-intensive, often putting rescuers at significant risk.

Our Mobile Controlled Robot represents a significant leap forward in the application of deep learning and robotics to disaster response. By combining advanced human detection capabilities with autonomous navigation and remote operation, our system aims to expedite search and rescue missions, thereby increasing the chances of saving lives. The use of a YOLOv8 model with 90% accuracy in detecting human forms ensures high reliability in identifying trapped individuals, while the real-time image analysis and buzzer alert system facilitate prompt response by rescue teams.

The remote operation feature, accessible via a user-friendly mobile interface, allows rescue personnel to control the robot from a safe distance, reducing the risk to human life and enhancing operational flexibility. This capability is particularly crucial in unstable and hazardous environments where direct human intervention may be dangerous.

1.4. Aim of the Project

The primary aim of our project is to develop a robust and reliable robotic system that can assist in search and rescue operations following natural disasters. By leveraging deep learning for accurate human detection and integrating it with advanced robotics and remote control capabilities, we seek to improve the speed and effectiveness of rescue missions. Our goals include:

1- Enhancing Detection Accuracy: Utilizing state-of-the-art deep learning models to achieve high accuracy in identifying human presence amidst debris.

2- Improving Safety: Minimizing the risk to rescue personnel by enabling remote operation and autonomous navigation of the robot in hazardous environments.

3- Increasing Efficiency: Streamlining search and rescue operations through real-time data analysis and prompt alert systems, thereby reducing response times.

4- Flexibility and Adaptability: Designing a versatile robotic system that can adapt to various disaster scenarios and evolving conditions, ensuring optimal performance in diverse environments.

Through this project, we aim to contribute to the broader field of disaster response technology, providing a valuable tool that can significantly enhance the capabilities of rescue teams and ultimately save more lives in the critical aftermath of natural calamities.

II. RELATED WORKS

This study [8] explores the significance of incorporating hand gesture recognition in natural human-robot interaction (HRI) to overcome communication barriers and enhance robotics development. It examines the process of hand gesture recognition using both monocular cameras and RGB-D cameras, encompassing data acquisition, hand gesture detection, segmentation, feature extraction, and gesture classification. The paper provides a comprehensive analysis of algorithms for hand gesture recognition in HRI and discusses experimental evaluations. Additionally, it addresses the need for advancements to improve current hand gesture recognition systems, aiming for more effective and efficient human-robot interaction[9]. Overall, the study highlights the potential of hand gestures as natural, intuitive communication methods in robotics and emphasizes the importance of further research in this area for advancing HRI capabilities.

This study [10] highlights the critical advancements in face recognition technology driven by deep learning models, enabled by the availability of large and complex training datasets. However, existing datasets sourced from news sites or social media platforms are limited in their applicability to advanced security, forensics, and military domains due to resolution, range, and viewpoint constraints. To address this gap, the study introduces the creation of the first and second subsets of a large multi-modal biometric dataset tailored for research and development (R&D) of biometric recognition technologies in challenging conditions. The dataset comprises over 350,000 still images and 1,300 hours of video footage from approximately 1,000 subjects, collected using various cameras including Nikon DSLR, commercial surveillance cameras, and specialized long-range R&D cameras mounted on UAV platforms. The primary objective is to facilitate the development of algorithms capable of accurately recognizing individuals at ranges of up to 1,000 meters and from elevated viewpoints, supporting critical applications in security, forensics, and military domains[11].

This study [12] provides an in-depth analysis of the evolution of YOLO (You Only Look Once) as a central real-time object detection system widely used in robotics, driverless cars, and video monitoring applications. The analysis spans from the original YOLO to subsequent iterations such as YOLOv8, YOLO-NAS, and YOLO with transformers. The study delves into the innovations and contributions introduced in each iteration, covering changes in network architecture, training techniques, and post-processing methods. Standard metrics for evaluation are described, and the study discusses the evolution of YOLO in terms of performance and efficiency. Finally, the study offers insights into the future of real-time object detection systems, highlighting potential research directions for further advancements in the field.

This study [13] provides a comprehensive overview of object detection in the realm of computer vision, highlighting its significance and various applications across fields such as security, military, transportation, and medical sciences. Deep Convolutional Neural Networks (DCNNs) are recognized for their remarkable performance in object detection tasks, alongside other applications like video processing, image segmentation, and speech recognition. The review covers different aspects of object detection, including frameworks, backbone convolutional neural networks, common datasets, and evaluation metrics. It acknowledges the evolution of deep learning algorithms and their significant impact on improving object detection model performance, while also recognizing the continued relevance of conventional methods. The study identifies future research challenges in designing deep neural networks for object detection and concludes with a comparison of object detection model performance on standard datasets like PASCAL VOC and MS COCO, drawing insightful conclusions from the analysis[14].

This study [15] proposes a method to measure the size of a predefined region in video footage and count the number of people within that area in real-time, aiming to ensure compliance with capacity rules in indoor spaces, particularly relevant during the COVID-19 pandemic. The method involves predetermining the borders of the region, identifying and counting people within it, and estimating the maximum capacity based on the size of the area. The You Only Look Once (YOLO) object detection model is utilized for this purpose, with pre-trained weights from the Microsoft COCO dataset used to identify and label individuals. The study evaluates the performance of different YOLO models, analyzing metrics such as mean average precision (mAP), frames per second (fps), and accuracy rate for person detection within the specified region. Results indicate that the YOLO v3 model achieves the highest accuracy rate and mAP scores, while the YOLO v5s model achieves the highest fps rate among non-Tiny models.

This [16] study addresses the critical task of fallen person detection (FPD) for ensuring individual safety, recognizing the limitations of existing deep-learning models such as poor feature extraction, inadequate utilization of contextual information, and high computational requirements. To overcome these challenges, the study proposes a novel lightweight detection model called Global and Local You-Only-Look-Once Lite (GL-YOLO-Lite), which integrates global and local contextual information using transformer and attention modules within the YOLOv5 framework. The model features a stem module to replace the original focus module, rep modules with re-parameterization technology, and a lightweight detection head to reduce redundant channels. Additionally, the study introduces a large-scale, well-formatted FPD dataset (FPDD). Experimental evaluation on both FPDD and Pascal VOC datasets demonstrates that GL-YOLO-Lite outperforms state-of-the-art models, achieving significant improvements in mean average precision (mAP) ranging from 2.4 to 18.9 on FPDD and 1.8 to 23.3 on Pascal VOC. Furthermore, GL-YOLO-Lite maintains real-time processing speeds of 56.82 frames per second (FPS) on a Titan Xp and 16.45 FPS on a HiSilicon Kirin 980, validating its effectiveness in real-world scenarios.

III. Dataset

The dataset used for training, validation, and testing of the Mobile Controlled Robot’s human detection capabilities consists of annotated images from various locations, each labeled with human presence. It is divided into three subsets: a train set comprising 4407 images (80%), a validation set with 1071 images (20%), and a minimal test set of 5 images. All images are resized to 640x640 pixels for uniformity. Data augmentation techniques are applied to enhance dataset diversity, including 90° rotations (clockwise and counter-clockwise), exposure adjustments between -20% and +20%, and adding noise to up to 8% of the pixels. These augmentations ensure the model can generalize well to various disaster scenarios and perform reliably under different conditions.[17]

IV. Methodology

we employ YOLOv8 as the core object detection algorithm for human detection in the Mobile Controlled Robot. The YOLOv8 model is trained on the annotated dataset of images containing humans in diverse environments. During training, the model learns to predict bounding boxes and class probabilities for human instances within the input images. We utilize transfer learning by initializing the YOLOv8 model with pre-trained weights on large-scale datasets such as COCO or ImageNet, allowing the model to leverage knowledge learned from generic object detection tasks. The training process involves optimizing the model's parameters using backpropagation and stochastic gradient descent (SGD) algorithms to minimize the detection loss function, which combines localization loss and classification loss. Once trained, the YOLOv8 model is evaluated on a separate validation dataset to assess its performance using evaluation metrics such as precision, recall, F1 score, IoU, and mAP. Finally, the trained YOLOv8 model is deployed onto the Mobile Controlled Robot, where it performs real-time human detection using input from onboard cameras. The detected humans are then communicated to rescue teams via auditory alerts, facilitating prompt response and rescue operations in disaster scenarios.

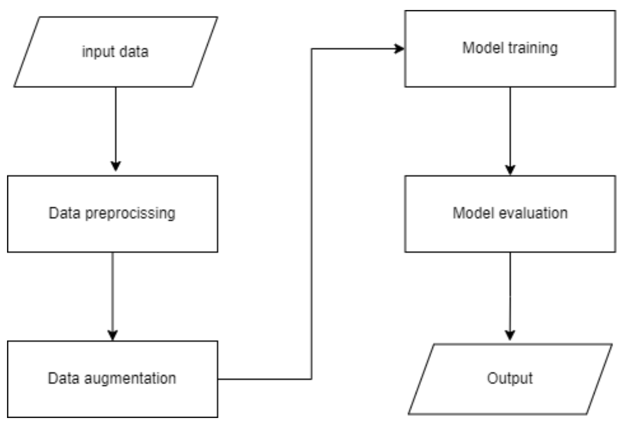

Fig.1. Our Proposed Methodology.

4.1. YOLO Algorithm

The YOLO (You Only Look Once) object detection algorithm represents a breakthrough in computer vision, offering real-time detection of objects in images or videos with impressive accuracy. Unlike traditional object detection methods that require multiple passes over an image, YOLO frames object detection as a single regression problem, directly predicting bounding boxes and class probabilities from a single neural network. YOLO divides the input image into a grid and assigns each grid cell responsibility for predicting bounding boxes and class probabilities within its spatial region. This grid-based approach enables YOLO to achieve remarkable speed while maintaining high accuracy, making it suitable for real-time applications. However, earlier versions of YOLO suffered from limitations in detecting small objects and precise localization due to their coarse grid granularity and reliance on a single-scale feature map.

4.2. YOLOv8

YOLOv8, an evolution of the YOLO algorithm, addresses these limitations by introducing several architectural improvements and optimizations. YOLOv8 utilizes a larger backbone network, typically based on variants of Darknet or ResNet, to extract more detailed features from the input image. This allows YOLOv8 to better capture fine-grained information necessary for accurate object detection, especially for small objects or objects with intricate features. Additionally, YOLOv8 employs feature pyramid networks (FPN) to generate multi-scale feature maps, enabling the model to detect objects of various sizes more effectively. Furthermore, YOLOv8 incorporates advanced training techniques such as focal loss and data augmentation to enhance model robustness and generalization capabilities. These enhancements result in significant improvements in detection accuracy and localization precision compared to earlier versions of YOLO.

4.3. Evaluation Metrics

To assess the performance of the YOLOv8-based human detection model, several evaluation metrics are employed:

4.3.1. Precision: Precision measures the proportion of true positive detections among all positive detections. It is calculated as the ratio of true positives (TP) to the sum of true positives and false positives (FP).

$$Precision = \frac{TP}{TP + FP}$$

Where: (1)

- TP: True Positives
- FP: False Positives

4.3.2. Recall: Recall, also known as sensitivity, measures the proportion of true positive detections among all actual positives in the dataset. It is calculated as the ratio of true positives (TP) to the sum of true positives and false negatives (FN).

$$Recall = \frac{TP}{TP + FN}$$

where: (2)

FN is False Negatives: Ground truth bounding boxes that have an IoU less than a certain threshold with all predicted bounding boxes.

4.3.3. Mean Average Precision (mAP): mAP is the average of the precision-recall curves across all classes. It quantifies the overall performance of the object detection model across different classes and detection thresholds.

$$mAP = \frac{1}{C}\sum_{i=1}^{C} AP_i$$

where: (3)

- C is the number of classes.
- AP_i is the Average Precision for class i.

V. RESULTS

The YOLOv8-based human detection model achieved promising results across multiple evaluation metrics. The mean Average Precision (mAP) score, indicative of the model's overall performance, was 95.3%. Furthermore, the model demonstrated high Precision of 92.6% and Recall of 86.7%, indicating its ability to accurately detect humans while minimizing false positives.

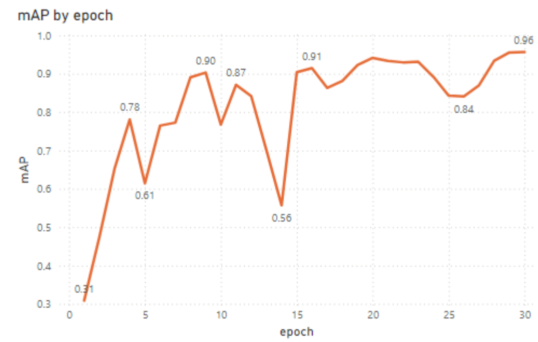

Fig.2. mAP results over epochs.

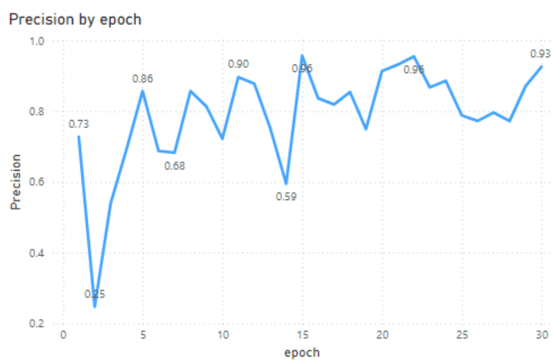

Fig.3. Precision results over epochs.

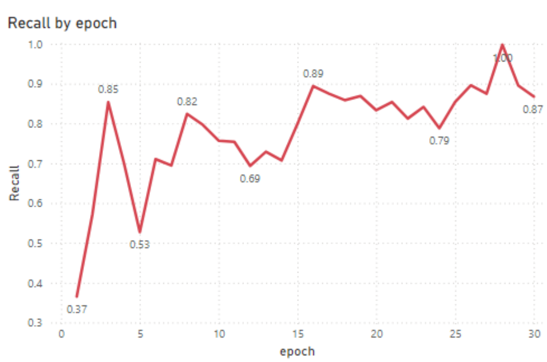

Fig.4. Recall results over epochs.

Visual inspection of the model's inference on the test set further affirmed its effectiveness. Sample images from the test set showed the YOLOv8 model accurately detecting humans in various environmental conditions, including cluttered scenes and occluded individuals. The model exhibited robustness in detecting humans of different scales and orientations, highlighting its capacity to generalize well to diverse real-world scenarios.

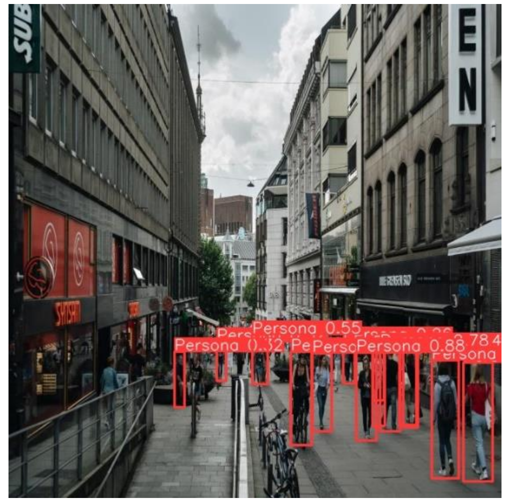

Fig.5. Inference results on the test set.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to all those who have contributed to the successful completion of this project. We extend our heartfelt thanks to our supervisor, mentors, colleagues, and collaborators for their guidance, support, and valuable insights throughout the research process. Additionally, we acknowledge the funding agencies and institutions that provided resources and support for this endeavor. Lastly, we express our appreciation to our families and friends for their understanding and encouragement. Their unwavering support has been a source of motivation and inspiration.

VI. CONCLUSION

The development and evaluation of the YOLOv8-based human detection model represents a significant step forward in leveraging advanced technology to enhance search and rescue operations in disaster scenarios. The high accuracy and robust performance demonstrated by the model underscores its potential to facilitate prompt and effective response efforts, ultimately saving more lives in times of crisis. While further refinements and optimizations may be needed to address challenges in real-world deployments, the promising results obtained thus far highlight the transformative impact of integrating deep learning with robotics for humanitarian purposes. Moving forward, continued research and innovation in this field holds the promise of further improving the efficiency and effectiveness of disaster response efforts, ultimately contributing to a safer and more resilient society.

REFERENCES

[1] C. H. Kang and S. Y. Kim, "Real-time object detection and segmentation technology: an analysis of the YOLO algorithm," *JMST Advances*, vol. 5, no. 2, pp. 69–76, 2023.

[2] A. Alqerem, H. Attar, W. Alomoush, and M. Deif, "The Ability of Ultra Wideband to Differentiate Between Hematoma and Tumor Occur in The Brain," in *2022 International Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI)*, IEEE, 2022, pp. 1–7.

[3] M. Krichen, "Convolutional neural networks: A survey," *Computers*, vol. 12, no. 8, p. 151, 2023.

[4] M. A. Deif, M. A. A. Eldosoky, A. M. El-Garhy, H. W. Gomma, and A. S. El-Azab, "Parasympathetic Nervous Signal Damping Using the Adaptive Neuro-Fuzzy Inference System Method to Control Overactive Bladder," *J Clin Eng*, vol. 40, no. 4, pp. 197–201, 2015.

[5] L.-D. Quach, K. N. Quoc, A. N. Quynh, and H. T. Ngoc, "Evaluating the effectiveness of YOLO models in different sized object detection and feature-based classification of small objects," *Journal of Advances in Information Technology*, vol. 14, no. 5, pp. 907–917, 2023.

[6] F. R. Ahmed, S. A. Alsenany, S. M. F. Abdelaliem, and M. A. Deif, "Development of a hybrid LSTM with chimp optimization algorithm for the pressure ventilator prediction," *Sci Rep*, vol. 13, no. 1, p. 20927, 2023.

[7] M. Rizk and I. Bayad, "Human Detection in Thermal Images Using YOLOv8 for Search and Rescue Missions," in *2023 Seventh International Conference on Advances in Biomedical Engineering (ICABME)*, IEEE, 2023, pp. 210–215.

[8] J. Qi, L. Ma, Z. Cui, and Y. Yu, "Computer vision-based hand gesture recognition for human-robot interaction: a review," *Complex & Intelligent Systems*, vol. 10, no. 1, pp. 1581–1606, 2024.

[9] H. Attar et al., "Modeling and computational fluid dynamics simulation of blood flow behavior based on MRI and CT for Atherosclerosis in Carotid Artery," Multimed Tools Appl, vol. 83, no. 19, pp. 56369–56390, 2024.

[10] D. Cornett *et al.*, "Expanding accurate person recognition to new altitudes and ranges: The briar dataset," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 593–602.

[11] M. A. Deif, W. Alomoush, H. Atta, O. A. Khashan, A. Solyman, and R. ELGohary, "PID Controller Tuning Using Multi-Objective Ant Colony Optimization for Blood Glucose Level of a Diabetic Patient," in *2024 2nd International Conference on Cyber Resilience (ICCR)*, IEEE, 2024, pp. 1–10.

[12] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas," *Mach Learn Knowl Extr*, vol. 5, no. 4, pp. 1680–1716, 2023.

[13] R. Kaur and S. Singh, "A comprehensive review of object detection with deep learning," *Digit Signal Process*, vol. 132, p. 103812, 2023.

[14] A. Alrosan, W. Alomoush, M. Youssef, A. W. Nile, M. A. Deif, and R. E. L. Gohary, "Hyper-Parameters Tuning Using Meta-Heuristic Algorithms for Nurses Stress Detection," in *2024 2nd International Conference on Cyber Resilience (ICCR)*, IEEE, 2024, pp. 1–5.

[15] M. Ş. Gündüz and G. Işık, "A new YOLO-based method for real-time crowd detection from video and performance analysis of YOLO models," *J Real Time Image Process*, vol. 20, no. 1, p. 5, 2023.

[16] Y. Dai and W. Liu, "GL-YOLO-Lite: a novel lightweight fallen person detection model," *Entropy*, vol. 25, no. 4, p. 587, 2023.

[17] "Khaled Fadi, 'Human Detection Dataset.' [Online]. Available: https://universe.roboflow.com/titulacin/person-detection-9a6mk".