Sudan University of Science and Technology

College of Engineering

School of Electronics Engineering

Face Mask and Social Distance Detection using YOLOv8 with Face Mask Detection Dataset

Prepared by:

1. Mohamed Abdo Hamed Abusarra
2. Ahmed Abdelbagi Gorshe Younis
3. Akram Osman Nasr Osman
4. Samrin Hesham Mahjoub
5. Waad Abdelrazig Al-Bashir Al-Sadiq

Proposed Supervisor:

Prof. Dr. Rashid A. SAEED

بسم الله الرحمن الرحيم

{يَرفَعِ اللَّهُ الَّذينَ آمَنوا مِنكُم وَالَّذينَ أوتُوا العِلمَ دَرَجاتٍ وَاللَّهُ بِما تَعمَلونَ خَبيرٌ}

[المجادلة: 11]

DECLARATION

I hereby declare that the research titled “Face Mask and Social Distance Detection using YOLOv8 with Face Mask Detection Dataset” is the result of our original work, carried out under the guidance and supervision of [Prof. Rashid]. This research represents the design and simulation of an intelligent monitoring system for real-time monitoring and classification.

We affirm that all sources of information and references used in this research have been properly acknowledged and cited. This work has not been submitted in whole or part for any other academic degree or professional qualification. We are also proud that this research was presented at the International Conference on Intelligent Systems and Artificial Intelligence Application (ISAA), at the University of Nizwa, Sultanate of Oman.

ACKNOWLEDGMENTS

We would like to begin by expressing our heartfelt gratitude to Almighty Allah for granting us the strength, patience, determination, and wisdom throughout this journey. We praise Him for making this achievement possible.

We extend our deepest thanks to Prof. Rashid A. Saeed, our supervisor, for his continuous guidance, invaluable advice, constructive feedback, and unwavering support at every stage of this endeavor. Our sincere appreciation also goes to the Sudan University of Science and Technology (SUST) for providing the knowledge, facilities, and academic environment that enabled this work.

We acknowledge our own efforts, dedication, time, and energy invested in this project to present it in the best possible way. Additionally, we thank our families and friends for their endless patience, encouragement, and moral support throughout this journey.

Finally, we express our gratitude to everyone who contributed, directly or indirectly, to the completion of this research. We pray that Allah makes our knowledge beneficial and allows us to share its benefits with others.

Abstract

Automatic monitoring of preventive health measures in crowded public environments presents a significant challenge, particularly regarding mask compliance and social distancing enforcement. This study introduces an intelligent real-time monitoring system for face mask detection and social distance estimation utilizing the YOLOv8 object detection framework. The system is developed to deliver accurate and efficient automated surveillance through computer vision techniques. A custom dataset of 12,603 annotated images, categorized into mask, no mask, and incorrectly worn mask classes, was employed for training and evaluation. Data preprocessing and augmentation were implemented to enhance model generalization, and hyperparameter optimization was conducted to improve detection performance. Additionally, a homography-based distance estimation module was incorporated to measure interpersonal distances and identify social distancing violations in real time. Experimental results show that the system achieves a mean Average Precision (mAP@0.5) of 94.1%, with particularly high detection accuracy for the incorrectly worn mask class (99.2%). The model maintains a strong balance between precision and recall, achieving a maximum F1-score of 0.92. The integrated distance monitoring module also provides reliable proximity alerts for potential violations. These findings suggest that the proposed framework offers an effective and practical solution for intelligent public health monitoring and is suitable for deployment in environments such as airports, hospitals, transportation hubs, and other crowded public spaces.

الخلاصة

تُعد المراقبة التلقائية لإجراءات الصحة الوقائية في البيئات العامة المزدحمة تحدياً كبيراً، خاصةً فيما يتعلق بالامتثال لارتداء الكمامات وفرض التباعد الاجتماعي. تقدم هذه الدراسة نظام مراقبة ذكي وفوري للكشف عن ارتداء الكمامات وتقدير المسافة الاجتماعية بالاعتماد على إطار عمل YOLOv8 لاكتشاف الأجسام. تم تطوير النظام ليقدم مراقبة مؤتمتة دقيقة وفعّالة باستخدام تقنيات الرؤية الحاسوبية. تم استخدام مجموعة بيانات مخصصة تتكون من 12,603 صورة مشروحة ومصنفة إلى ثلاث فئات: ارتداء الكمامة، عدم ارتداء الكمامة، وارتداء الكمامة بشكل غير صحيح، وذلك لأغراض التدريب والتقييم. تم تطبيق خطوات معالجة البيانات وزيادتها لتحسين تعميم النموذج، كما أُجريت عملية تحسين للمعاملات لرفع دقة الاكتشاف. بالإضافة إلى ذلك، تم دمج وحدة تقدير للمسافة تعتمد على التحويل التوافقي بهدف قياس المسافات بين الأشخاص وتحديد مخالفات التباعد الاجتماعي في الوقت الحقيقي. أظهرت النتائج التجريبية أن النظام يحقق متوسط دقة (mAP@0.5) بنسبة 94.1%، مع دقة عالية بشكل خاص لفئة ارتداء الكمامة بشكل غير صحيح (99.2%). يحافظ النموذج على توازن قوي بين الدقة والاسترجاع، محققًا أعلى قيمة F1 بمقدار 0.92. كما توفر وحدة مراقبة المسافات المدمجة تنبيهات موثوقة عند حدوث انتهاكات محتملة. تشير هذه النتائج إلى أن الإطار المقترح يوفر حلاً فعالاً وعملياً لمراقبة الصحة العامة الذكية، ويصلح للاستخدام في بيئات مثل المطارات والمستشفيات ومحطات النقل وغيرها من الأماكن العامة المزدحمة.

Table of Contents

List of Figures

Chapter 1: Introduction

1.1 Background of the Study

In modern societies, the COVID-19 pandemic has revealed the utmost importance of health commitments such as wearing masks, which are considered the most important tools in public health. It has become a challenge for these societies, especially in crowded environments, to find systems that can effectively monitor the spread of viruses. However, this traditional approach is exhausting. Innovative solutions have provided deep learning systems, such as artificial intelligence, that can be used to build smart systems.

There are several versions of YOLO, such as YOLOv8, YOLOv7, and YOLOv3, which rely on algorithms like CNNs that have proven effective in accurately detecting objects in real-time with high speed and precision.

The release of YOLOv8 is based on Transformer + CNN, which supports classification, detection, segmentation, and tracking of images. It is considered an excellent choice for applications that require real-time processing, as it achieves high accuracy and excellent results while adhering to social distancing measures and monitoring procedures.

The development of an effective system contributes to enhancing the efforts of health and safety institutions. YOLOv8 thus plays a role in limiting the spread of diseases by providing a tool for monitoring compliance with precautionary measures and offering preventive methods for the community.

1.2 Problem Statement

To monitor individuals who violate health procedures, traditional monitoring systems have a clear shortcoming, as they lack the integration of advanced technologies such as artificial intelligence.

For example, video monitoring systems provide quick response times and save significant effort and time. The intelligent system works with artificial intelligence technology, which facilitates the process of discovering individuals who do not wear masks effectively and helps in making precise data indicators or decisions. This contributes to improving preventive measures and reducing the spread of infection.

1.3 Research Objectives

The main objective of this research is to create a faster, accurate, and efficient AI model that can work even in poor conditions (poor light, crowded areas, and different camera angles).

The objectives of this work can be summarized as follows:

1. To develop a self-detection system for recognizing people with or without face masks by employing YOLOv8 and a face mask detection database.
2. To adopt an effective means of monitoring social distancing by measuring the physical distance between people in real time.
3. To evaluate the proposed system's effectiveness with respect to functionality, operating speed, and flexibility across a range of environmental conditions.
4. To design a deployable public system for facilitating compliance monitoring during periods of health emergencies.

1.4 Research Questions

Research questions for this research are as follows:

1. How effective is YOLOv8 in real-time face mask detection when using a publicly available face mask detection dataset?
2. How can YOLOv8 be integrated with other methods to enable the effective tracking of individuals' social distancing?
3. How effective is the proposed system in terms of detection reliability, processing speed, and stability under changing circumstances?
4. How can the system enhance public health safety in real-world applications?

1.5 Significance of the Study

The study offers a valuable contribution to extant research literature interested in the possibilities of artificial intelligence for the monitoring of health and safety. The study unites the recognition of face masks with the assessment of social distance into a unifying framework, offering an actionable approach which can be applied in real-time public settings for the purposes of facilitating compliance programs. The study also holds the potential to prove valuable for public health administrators, as well as for technology developers and for those for whom policy is a concern, by offering information which may shape the development of more effective monitoring systems. Finally, the study reveals the potential of YOLOv8 for real-world deployment, suggesting potential for further extension for use in a variety of health-related contexts.

1.6 Scope and Limitations

1.6 .1 Scope

The scope is the boundary or framework that defines the extent of coverage of the research or system, specifying what is included and what is excluded.

The face mask detection system focuses on distinguishing between individuals who are wearing the mask correctly and those who are either not wearing it or wearing it improperly, using images or videos captured by cameras. The scope also includes testing the model in different environments, such as varying lighting conditions, locations, and camera angles. In addition, it may apply convolutional neural networks (CNNs) or other models, with the possibility of integrating the system into surveillance, security, and smart access control systems to enhance preventive measures and achieve a higher level of safety and security.

1.6.2 Limitations

Despite potential benefits, this work is also impaired by some limitations which would potentially affect the system's overall output. Technically, efficiency and accuracy might suffer from improper light exposure, variable camera angles, or highly populated areas. The system also cannot effectively recognize transparent or patterned face masks and requires high-capacity computing such as GPUs for facilitating real-time processing. Furthermore, due to the limited availability of potential training datasets, not all cases are supported, which might result in inaccurate outputs such as false positives or false negatives and therefore reduce reliability. When it comes to privacy and security matters, face image storage creates serious ethical and legal complications. The system is also liable to risks such as hacking or improper use of information, and it runs into more obstacles in enforcing stringent data privacy laws such as GDPR. In operational terms, the system needs large financial inputs in developing as well as operating it. Advanced maintenance and ongoing model updates are also associated with it, and deployment of cloud services necessitates frequent internet connectivity. Scaling to handle large numbers of users or cameras is also not simple and will limit broader usage.

1.7 Organization of the Thesis

The following is an organization of this project

Chapter One: The introduction presents the background of the study, its scope, importance, questions, objectives, and problem.

Chapter Two: Previous studies that used techniques for social distancing and mask detection using artificial intelligence have been reviewed.

Chapter Three: The methodology of the research presents the steps for designing, evaluating, and training the model and tools used.

Chapter Four: The results are presented, where the performance of the system is clarified and analyzed.

Chapter Five: The conclusion includes a summary of what the study has reached, recommendations, and future proposals.

Chapter 2: Literature Review

2.1 Introduction

SARS-CoV-2 (COVID-19) spreads primarily via respiratory droplets and aerosols emitted when infected people breathe, speak, cough, or sneeze. Widespread evidence and public-health guidance show that correctly worn face masks reduce the probability of transmitting infectious respiratory particles and therefore reduce community transmission when used together with other measures (vaccines, ventilation, distancing). According to the World Health Organization [1], manual monitoring of people in public places (airports, hospitals, factories, malls) is costly, hard, and exposes the staff to risk. An automated face-mask detection system enables continuous, non-contact monitoring of compliance, reduces human workload and reporting latency, supports real-time interventions (alerts, gating access, analytics), and can be integrated with social-distance monitoring and temperature sensors for layered enforcement. Because CCTV and mobile cameras are ubiquitous, robust computer-vision detectors operating in real time are a practical way to support public-health policy [2], and thus the importance of AI (artificial intelligence) emerges because using machine learning to build models that can detect people who are wearing or not wearing face masks reduces the need for manual monitoring all the time.

2.2 Theoretical Background

This background presents the scientific and technical foundation of the project. It explains the core principles and algorithms that enable automatic face mask detection, including artificial intelligence, machine learning, deep learning, convolutional neural networks, and object detection frameworks. These technologies collectively form the basis for building accurate and real-time detection systems capable of improving public health safety and reducing human monitoring effort during pandemics such as COVID-19.

2.2.1 Artificial Intelligence and Machine Learning

1. ***Artificial Intelligence (AI)***

Artificial Intelligence refers to the simulation of human cognitive functions such as reasoning, learning, and perception by computer systems. AI enables machines to perform tasks that normally require human intelligence, including decision-making, pattern recognition, and visual understanding. In the context of public health, AI supports automated surveillance systems that can detect non-compliance with safety protocols such as mask-wearing and social distancing, thereby minimizing human exposure to infection risk.

2. ***Machine Learning (ML)***

Machine Learning is a subfield of AI that allows systems to learn patterns from data instead of relying on explicit programming. ML algorithms identify relationships between inputs (images or features) and outputs (mask/no-mask classification). Common ML techniques include Support Vector Machines (SVM), Decision Trees, and k-Nearest Neighbors (k-NN). Early face-mask detection systems used ML with manually extracted features such as Histogram of Oriented Gradients (HOG) or Local Binary Patterns (LBP). However, their accuracy was limited by the quality of feature extraction and sensitivity to lighting and pose variations.

2.2.2 Deep Learning and CNN Architectures

1. ***Deep Learning Overview***

Deep Learning (DL) is a branch of ML that employs artificial neural networks with multiple layers to automatically learn complex feature representations from raw data. It eliminates the need for handcrafted features, enabling systems to learn directly from pixels. DL techniques have achieved breakthrough results in image recognition, speech processing, and medical diagnostics.

2. Convolutional Neural Networks (CNNs)

CNN is a deep learning model designed for image processing, composed of convolution, pooling, and fully connected layers. CNNs are the backbone of modern computer-vision systems. They consist of several key layers:

1. Convolution Layers: Extract spatial features by applying filters to input images.

2. Pooling Layers: Reduce spatial dimensions, retaining essential information while minimizing computation.

3. Activation Layers: Introduce non-linearity (e.g., ReLU) to enable complex function approximation.

4. Fully Connected Layers: Aggregate extracted features for classification.

CNNs have demonstrated exceptional capability in face detection and recognition tasks. In mask detection, CNNs can identify subtle visual differences between masked and unmasked faces even in cluttered scenes. Architectures such as VGGNet, ResNet, and MobileNet have been widely applied, often as pre-trained backbones for object detection frameworks.

2.2.3 Object Detection Algorithms

Object detection is the process of identifying and localizing multiple objects within an image or video frame. Modern detectors fall into two main categories: two-stage detectors and one-stage detectors.

1. Two-Stage Detectors like Region-based CNN(R-CNN), Fast R-CNN, Faster R-CNN: First generate region proposals, then classify them. These achieve high accuracy but are computationally expensive, making them unsuitable for real-time tasks.

2. One-Stage Detectors like Single Shot Detector (SSD): You only look once (YOLO). Perform detection and classification in a single step, trading slight accuracy loss for major speed gains, ideal for real-time

surveillance.
Many recent mask-detection systems favor one-stage detectors (YOLO variants, SSD) because they hit the necessary speed/accuracy tradeoff for live video [3].

2.2.4 The YOLO Family of Algorithms

The YOLO (You Only Look Once) family revolutionized object detection by framing it as a regression problem. The algorithm predicts bounding boxes and class probabilities directly from full images in one forward pass, allowing real-time operation.

1. YOLOv1–v3: Introduced anchor boxes and deeper backbones for improved accuracy.

2. YOLOv4: Added CSPDarkNet53 backbone, PANet, and advanced augmentation for multi-scale feature fusion.

3. YOLOv5: Focused on lightweight deployment, scalability, and training efficiency.

4. YOLOv6–v7: Introduced re-parameterized convolutions and better small-object detection.
 This figure [Fig. 2.1] shows the development of YOLO throughout the years.

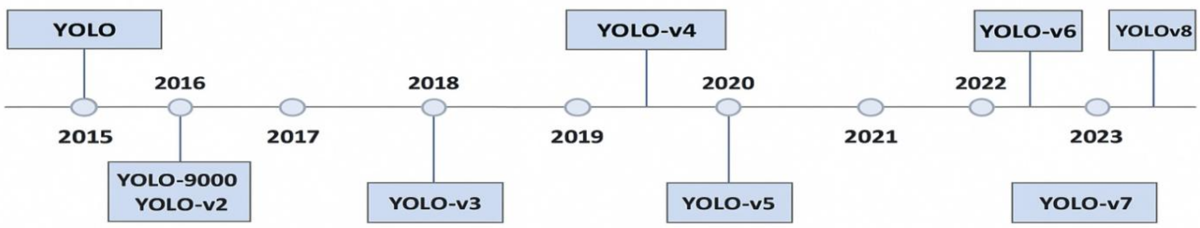

Figure 2.1 shows the development of YOLO throughout the years

2.2.4.1 YOLOv8: Architecture and Improvements

YOLOv8, released by Ultralytics in 2023, represents the most advanced generation of the YOLO family [4]. It is a production-oriented implementation that unifies object detection, segmentation, and classification in a single codebase. Built using PyTorch and introduces several key improvements:

1. Anchor-Free Detection: Simplifies bounding box prediction, improving generalization.

2. Enhanced Backbone: Incorporates the C2f module and SPPF (Spatial Pyramid Pooling Fast) for superior multi-scale feature extraction.

3. Decoupled Head: Separates classification and regression to accelerate convergence.

4. Improved Augmentation: Uses Mosaic, MixUp, and adaptive image scaling to enhance robustness.

5. Performance Gains: Achieves higher mean Average Precision (mAP) with fewer parameters, faster inference, and better performance on edge devices.
   This figure [Fig. 2.2] shows the YOLOv8 structure.

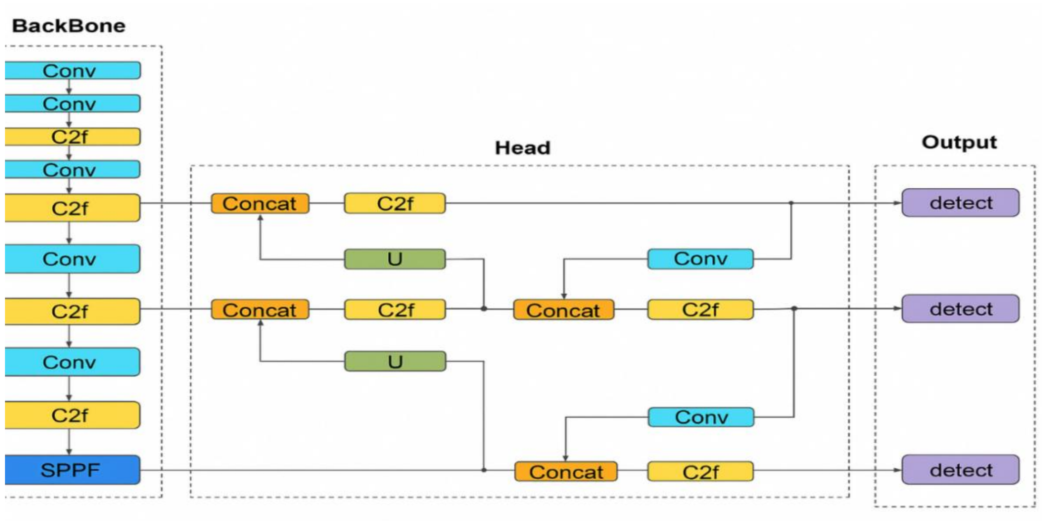

Figure 2.2 shows the YOLOv8 structure

2.3 Related Work

In 2022, Amer & AI-Tamimi [10] reviewed and analyzed a number of previous studies in the field of mask detection. This study aims to compare different methods (machine learning, deep learning, computer vision) and highlight the strengths and weaknesses of each. The paper does not present a new model, but rather reviews and evaluates the most important algorithms used. It is divided into:

1. Two-Stage Detectors (multiple stages):

   a) R-CNN: High accuracy but very slow.

   b) Fast R-CNN: Faster than R-CNN with good results.

   c) Faster R-CNN: Uses a region proposal network (RPN) and achieves a balance between speed and accuracy.

   d) Mask R-CNN: Adds the ability to segment objects in addition to detection.

2. Single-Stage Detectors:

   a) YOLO (v3, v5): Very fast and works in real time, but less accurate with small and closely spaced objects.

   b) SSD (Single Shot Detector): Faster than R-CNN, but less accurate than YOLO.

   c) FCOS (Fully convolutional one-stage detection): Detection without anchors, simplifies training and improves performance.

We find that the results from previous studies are as follows:

CNN networks: Strong at feature extraction but require large amounts of accurately labeled data. R-CNN and its variants: Suitable for applications that require very high accuracy but not ideal for real-time applications. YOLO and SSD: More suitable for practical applications (such as airports and shopping malls) due to their high speed. YOLOv5 in particular has proven to be the most efficient in real-world environments with a good balance between speed and accuracy.

There is also a similar study by Sharma in 2023 [6] that also focused on comparing algorithms, datasets, and performance metrics to provide a comprehensive reference for researchers. However, it used a different division methodology.

Traditional non-neural algorithms were reviewed, such as (Viola-Jones, HOG, SIFT, DPM). Then move on to deep algorithms (CNN-based), which include: Two-Stage Detectors: R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN, R-FCN (Region-based fully convolutional network), FPN (Feature pyramid network). One-Stage Detectors: YOLO (v1–v5), SSD, RetinaNet, RefineDet, CornerNet.

The paper reviewed 15 publicly available datasets for mask detection, including: RMFD, SMFD, MAFA, LFW, FMD, WIDER FACE, CelebA. It highlighted the limitations of these datasets, such as small size, lack of diversity, and difficulty in dealing with transparent or skin-colored masks. In both papers [10] & [6], I found that the challenges are that traditional methods suffer from poor accuracy and speed due to reliance on manual feature extraction.

Deep neural networks (CNNs) have significantly improved performance, especially YOLO and Faster R-CNN. Some studies have achieved results of mAP > 99% on datasets such as RMFD and SMFD. Face recognition accuracy decreases when a mask is present (95% for uncovered faces versus ~72% for covered faces using PCA). It was concluded that deep learning techniques (especially YOLO and Faster R-CNN) have proven to be highly effective in detecting masks. YOLO is recommended for practical applications because it is the fastest and most successful in real life, while R-CNN techniques are more suitable for research or medical applications that require maximum accuracy.

There is a need to develop more efficient systems on low-capability devices (edge devices). Future research requires expanding data sets and improving algorithms to reduce errors and improve performance in crowded and complex environments.

The studies above have reviewed and analyzed the most important research published until 2023. Where it divided the algorithms into Single-Stage Detectors and Two-Stage Detectors. We find that the studies that followed this period include those that used Single-Stage Detector algorithms, others that used Two-Stage Detector algorithms, or combined the two types to get better results.

2.3.1 Two-Stage Detector Algorithms

The main objective of the study in [11] was to design and implement an efficient face mask detection model using street cameras that can identify whether individuals are wearing masks properly, improperly, or not at all. The goal was to support pandemic safety measures through real-time automated surveillance using artificial intelligence and deep learning. The research utilized a Double Convolutional Neural Network (Double CNN) approach combining two machine learning models for enhanced accuracy: YOLO for real-time face detection in video streams.

A pre-trained CNN classifier was trained with 180,000 images to categorize faces into three classes: proper mask, improper mask, and no mask. Supporting tools and libraries included OpenCV (Open-source computer vision), TensorFlow, Keras, and the Haar Cascade Classifier. The Double CNN-based model achieved a high detection accuracy of 98.15% with strong precision and recall metrics. Among the tested classifiers (Double DNN, DenseNet121, MobileNetV2), the Double DNN showed the best training and validation performance. The system performed well under different lighting and facial orientation conditions, proving effective for real-time implementation. The system performance may degrade under extreme lighting conditions or when faces are partially obscured. The model depends on the quality and diversity of the dataset; uncommon facial orientations may affect accuracy.
The system requires sufficient computational power (e.g., GPUs) for real-time processing.
Environmental factors like camera resolution and angle may influence the detection precision. The authors recommend integrating the Double CNN

model with existing security infrastructures such as CCTV and access control systems to improve public safety enforcement. They suggest extending the model to cover other environments such as airports, hospitals, schools, and transportation systems. The paper also highlights that the system can be scaled globally due to its cost-effectiveness and reliability. Future work may involve improving dataset diversity, optimizing processing speed, and enhancing detection in complex crowd scenarios.

Among the studies building upon previous research, Smita Adhikari (2024) [5] developed a CNN-based model for automated mask-wearing detection. The paper employed a custom CNN architecture trained from scratch to classify images into two categories: "with mask" and "without mask." Several preprocessing steps were implemented, including image resizing to 128×128 pixels and conversion to RGB color space. The network architecture consisted of two convolutional layers with ReLU activation and max pooling (32 and 64 filters). Two fully connected (dense) layers followed by flattening. Adam optimizer with sparse categorical cross-entropy loss function, tracking classification accuracy as the primary performance metric. The study employed the Kaggle Face Mask Detection Dataset comprising 7,553images (3,725 with masks).

The proposed model achieved: Training accuracy: 95.53%, Validation accuracy: 93.58%, Precision: 0.91, Recall: 0.95, F1-score: 0.93. Loss: 12.14% (training), 13.7% (validation). The results demonstrated the model's significant effectiveness in near real-time mask detection using a CNN architecture alone, without requiring more complex networks like YOLO. It exhibited an optimal balance between accuracy and speed, making it suitable for intelligent surveillance and security systems. The study recommended expanding the dataset to include diverse environments and optimizing the model for real-time applications using lighter architectures such as MobileNet. In contrast,

Another research team [9] published a paper aiming to develop a hybrid system integrating both mask detection and identity recognition using deep learning techniques. The researchers addressed the challenge of reduced accuracy in face recognition systems when subjects wear masks by proposing a unified model capable of performing both tasks efficiently. The hybrid framework comprised two main stages: Face Mask Detection using a modified YOLOv5 model. Face Recognition using ResNet-50 for identity verification with masked faces. The models' performance was enhanced through transfer learning techniques. Data augmentation (cropping, rotation, lighting adjustment). CNN layer fine-tuning. The researchers utilized a combination of open-source datasets: RMFD (Real-World Masked Face Dataset) and MAFA. CelebA for face recognition.

The total dataset exceeded 30,000 images with variations in gender, lighting conditions, face angles, and mask types. The achieved results included: Mask detection accuracy: 98.9%. Face recognition accuracy: 96.7% (with masks). Overall F1-score: 0.97. The model demonstrated balanced performance in both accuracy and speed compared to previous approaches. This paper presents a robust framework combining YOLOv5's detection precision with ResNet's recognition capabilities, developing a comprehensive system that simultaneously performs mask detection and identity verification effectively. This represents one of the most complete studies in this field, offering significant practical applications in security and access control systems.

2.3.2 The study used single-stage detector algorithms and combined the two types

Researchers [12] have done a study aimed at developing a deep-learning model using YOLOv7 to detect and distinguish faces with masks and without masks in real-world conditions. The researchers focus on achieving higher accuracy and faster detection compared to earlier YOLO versions and CNN-based models to support public health and security monitoring applications. The core technique employed is the YOLOv7 single-stage object detection algorithm.

The architecture is organized into three main modules: input, backbone, and head. The backbone integrates CBS (Convolution+Batch Normalization + Activation) modules, ELAN (Efficient Layer Aggregation Networks), and max-pooling operations. The head employs RepConv, SPPCSPC, CatConv, and multiple convolution layers for bounding box and class predictions. Data augmentation techniques such as brightness, hue shifts, mirroring, and blurring were used. Input images were resized to 640×640 pixels, and transfer learning with pretrained weights was applied.

The dataset consisted of 219 real photos of university students, augmented to 318 images and split into training, validation, and testing sets. The model achieved approximately 97% classification accuracy in distinguishing masked and unmasked faces. Precision, recall, and mAP metrics demonstrated strong performance compared to earlier YOLO versions. At 250 training epochs, box loss decreased to about 2.47, object loss to 2.48, and class loss to 2.47. Visual examples confirm accurate detection in varied image conditions. Key limitations include the small dataset (219 images), which restricts generalization to diverse real-world environments.

Data diversity issues such as lighting variations, occlusions, and demographic bias limit the robustness of the model. Performance trade-offs remain between accuracy and processing speed. Additionally, no external dataset validation or cross-domain testing was performed. The study also lacks detailed analysis of inference latency and computational resource requirements on edge devices. The authors recommend increasing the dataset size with more diverse and challenging samples to improve model generalization. Further tuning of the YOLOv7 architecture and use of advanced data augmentation are encouraged. Real-time testing in live surveillance or video stream environments should be conducted to evaluate practical deployment performance. Integrating object tracking mechanisms could enhance accuracy by avoiding redundant detections. Future work should also benchmark the system against other datasets and YOLO versions to validate its robustness.

Recent study [7] introduced a novel and comprehensive dataset named the Diverse and Robust Dataset for Face Mask Detection (DRFMD). This dataset was compiled from six different public sources (AIZOO, MAFA, MOXA3K, PWMFD, KFMD, and the Zalo AI Challenge). It comprises 14,727 images containing 29,846 face instances, distributed across three critical classes: Without Mask: 14,157. With Mask: 11,590. Incorrectly Worn Mask: 2,866. This distribution ensures a balanced representation across categories, with a significantly improved proportion of incorrectly worn masks compared to previous datasets, which is crucial for real-world applicability. Methodology: The research involved training five variants of the YOLOv8 architecture (n, s, m, l, x) on both the proposed DRFMD dataset and a previous dataset, FMMD. To enhance model robustness, extensive data augmentation techniques were employed, including horizontal flipping, cropping, and scaling. Model performance was rigorously evaluated using standard metrics: Precision, Recall, and mAP@50.Superior Generalization: Models trained on the DRFMD dataset demonstrated markedly better generalization capabilities. When tested on the external HITL-MMD dataset (unseen during training), these models maintained high performance, achieving a mAP@50 between 0.705 and 0.779.

In contrast, models trained on the FMMD dataset suffered a significant performance drop on the same external data. Optimal Performance: The YOLOv8x model achieved the best performance, attaining a mAP@50 of 0.895 on the DRFMD test set. Mitigation of Overfitting: The large scale and diversity of the DRFMD dataset were instrumental in preventing model overfitting, leading to more robust and reliable detectors. Comparative Analysis and Broader Context: This study highlights a common pitfall in previous works: models trained on limited or non-diverse datasets (e.g., MAFA alone) may achieve high accuracy on their specific test sets but fail to generalize to real-world environments.

The DRFMD dataset addresses this by providing substantial diversity in lighting conditions, camera angles, backgrounds, and mask designs, resulting in models that are more robust and practical for real-world deployment. In conclusion, the DRFMD dataset is a rich and diverse resource that facilitates the development of more accurate and reliable face mask detection systems. YOLOv8 models trained on this dataset exhibit exceptional performance and high generalization ability.

YOLOv8 technology has been used in the field of vehicle detection. [8] The capabilities of YOLOv8 extend beyond face mask detection. In the critical domain of intelligent transportation systems, vehicle detection is a fundamental component. However, deploying detection algorithms in traffic environments presents challenges such as. Missed detections and false positives due to variations in illumination, weather, and occlusion. Difficulty in deployment on edge devices due to the high computational complexity of many models. To address these issues, a subsequent study aimed to develop a lightweight and high-accuracy model for real-time vehicle detection, named YOLOv8-FDD. The goal was to reduce the number of parameters while improving accuracy. The YOLOv8-FDD model achieved remarkable results: It reduced the parameter count to 72.89% of the original YOLOv8 model (from 3.0 million to 2.19 million parameters). It improved performance, increasing mAP50 from 97.1% to 97.8% and mAP50-95 from 83.6% to 84.9%. It maintained a detection speed exceeding 300 FPS, fulfilling the requirements for real-time processing. It significantly reduced both false positives and missed detections, particularly for small vehicles and in congested areas. The YOLOv8-FDD model demonstrated a superior balance of accuracy, model size, and speed, outperforming other prominent detectors like YOLOv5, YOLOv6, YOLOv7, YOLOv9, YOLOv10, and Faster R-CNN. Synthesis of YOLOv8's Advantages.

2.4 Summary

Across the reviewed studies, key advantages of the YOLOv8 architecture are evident: Real-Time Efficiency: YOLOv8 is designed for high efficiency in real-time applications. Its single-pass (one-stage) detection pipeline makes it inherently faster than two-stage detectors like Faster R-CNN and Mask R-CNN. Accuracy-Speed Balance: It achieves an excellent trade-off between detection accuracy (AP) and inference speed (FPS), often outperforming previous versions such as YOLOv5 and YOLOv7. Architectural Refinements: It incorporates several technical improvements, including a decoupled head to separate classification and regression tasks. An enhanced C2f module (improved over the C3 module in YOLOv5) for better feature extraction. A SPPF (Spatial Pyramid Pooling Fast) module for efficient multi-scale context aggregation.

Chapter 3: Methodology

3.1 Introduction

This chapter presents the research methodology adopted for developing the face mask detection system using YOLOv8. The purpose of this chapter is to explain the overall research design, approach, and procedures used to achieve the project objectives. It outlines the steps followed in data collection, preprocessing, model training, evaluation, and system implementation.

The methodology serves as a roadmap that guides the entire development process of the proposed system. It ensures that each stage from data acquisition to model deployment is conducted systematically and scientifically to produce accurate and reliable results. The use of the YOLOv8 model is central to this methodology, as it enables real-time and high-precision detection of individuals wearing or not wearing face masks.

This chapter also provides an overview of the research design type (quantitative), the reasoning approach (deductive), and the modeling process. Furthermore, it describes the tools, frameworks, and technologies employed during system development, such as Python, PyTorch, and OpenCV. Through these methods and techniques, the project aims to achieve an efficient and practical solution for real-time face mask detection.

3.2 Research Design

The research design represents one of the most essential components of the study methodology, as it outlines the overall framework through which the project was systematically and methodically implemented to achieve its objectives. This study adopts a Quantitative Experimental Research Design, which aims to measure the performance of the artificial intelligence model accurately through quantitative experiments that rely on numerical data generated from training and testing. The design focuses on evaluating the effectiveness of the YOLOv8 model in detecting face mask usage in images and video streams, and comparing the results with previous versions of the YOLO algorithm and other convolutional neural network (CNN)-based models.

The quantitative approach was selected because it allows precise measurement and analysis of performance using numerical indicators such as Accuracy, Recall, and Precision, in addition to F1-Score and mAP, which assess object detection quality. This approach is suitable for artificial intelligence projects that depend on experimentation and statistical result analysis, and it enables reproducibility of experiments by other researchers, enhancing the reliability of results.

The research design consists of five main interrelated phases that describe the workflow from data collection to results evaluation:

3. **Data Collection:** Our open-source dataset is named Pioneer Mask Detection; the dataset has a total of 12,603 images. It consists of 3 classes:

a. Mask b. mask_weared_incorrect c. no_mask.

i. **Mask class:** has 4847 images, of which 3382 are for training, 975 for validation, and 490 for testing**.**
ii. **Mask_weared_incorrect class:** consists of 2088 images, of which 1457 are for training, 434 for validation, and 197 for testing.
iii. **No_mask class:** has 5668 images, of which 4005 are for training, 1087 for validation, and 576 for testing [13].

The figure below shows samples from the dataset [Fig. 3.1].

Figure 3.1 Pioneer Mask Detection dataset

2. Data Preprocessing: Several steps were implemented, including resizing images to 640×640 pixels, applying normalization, and performing data augmentation through rotation, cropping, brightness adjustment, and splitting the dataset into training, validation, and testing sets.

Labelling: We draw a bounding box around the people in the images, and then assign a class to them (such as "Mask", "No_Mask", or " Mask_worn_incorrectly ") [Fig. 3.2].

Figure 3.1 Labelled faces

3. **Model Development:** The model was developed using the YOLOv8 algorithm implemented in Python with the Ultralytics library. Hyperparameters were tuned (Batch size = 16, Epochs = 150, Learning rate = 0.001).
4. **Model Testing:** During the Model Testing phase, performance was evaluated using metrics such as Accuracy, Precision, Recall, F1-score, mAP@0.5, and mAP@0.5:0.95 to determine system effectiveness under different conditions.
5. **Result Analysis:** To ensure result validity, cross-validation and comparative analysis with other models were performed, along with multiple experiment repetitions to verify stability.

6. Implementation Environment:

The project was implemented using Python on the Google Colab platform with the Ultralytics library, and trained on GPU hardware (NVIDIA Tesla T4, 16 GB RAM) to accelerate computation and enhance performance.

The research workflow can be summarized as follows: Data Collection → Data Preprocessing → Model Training → Model Testing → Result Analysis. This design helps evaluate the YOLOv8 model objectively and provides a scientific foundation for replicating and improving the experiment in future work [Fig. 3.3].

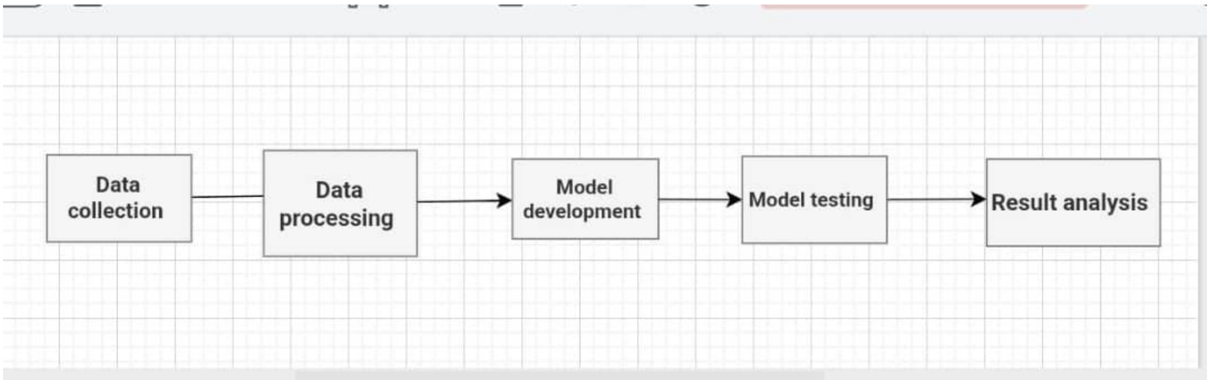

Figure 2.3 Research workflow

3.3 Research Approach

This research adopts a deductive approach as the primary methodological framework. The deductive approach is appropriate for this study because it begins with established theories and principles of deep learning and computer vision, and then applies them to test a specific hypothesis: that the YOLOv8 model can enhance the accuracy and efficiency of face mask detection compared to earlier versions such as YOLOv5 and YOLOv7 [14].

In this research, the theoretical foundation lies in the principles of Convolutional Neural Networks (CNNs), object detection, and transfer learning. These theories suggest that advanced network architectures and large-scale pre-trained models improve detection accuracy and generalization [15]. Based on this, the study formulates the hypothesis that YOLOv8, due to its improved backbone, feature extraction, and optimization techniques, will outperform older YOLO models in detecting masked and unmasked faces. The experimental process then tests this hypothesis by training and evaluating YOLOv8 on a curated face mask dataset. The obtained results (accuracy, precision, recall, and mAP) are analyzed to validate or reject the proposed assumption.

3.3.1 Relationship with Inductive and Abductive Reasoning

Although the primary reasoning of this research is deductive, supportive elements of inductive and abductive reasoning are also incorporated to strengthen the methodological framework:

Inductive reasoning is applied during result interpretation to generalize new insights from experimental outcomes. For instance, if YOLOv8 achieves higher accuracy in detecting incorrectly worn masks, this observation may lead to broader conclusions about the model's robustness and adaptability.

Abductive reasoning is used to explain unexpected findings or anomalies. For example, if the model performs poorly in low-light conditions, the researcher might infer that the training dataset lacks sufficient low-light samples, prompting future adjustments to the data preparation process. Together, these reasoning forms create a balanced logical flow that combines theoretical validation, empirical observation, and continuous improvement of the model.

3.3.2 Application of the Deductive Approach in This Study

In practical terms, the study follows these deductive steps:

Theory Development

Reviewing existing literature on CNNs, YOLO architectures, and face mask detection systems to establish a solid theoretical foundation.

Hypothesis Formulation

Proposing that YOLOv8 offers superior detection accuracy, speed, and robustness compared to previous YOLO versions.

Model Implementation and Testing

Implementing YOLOv8 using a labelled dataset of masked and unmasked faces.

Training the model, tuning hyperparameters, and evaluating it using metrics such as accuracy, precision, recall, F1-score, and mAP.

3.3.3 Evaluation

Analyzing the experimental results to confirm whether the theoretical expectations align with the empirical findings. If confirmed, the results reinforce existing deep learning theories; If not, they provide grounds for refining future hypotheses. This structured approach ensures the study's validity, reproducibility, and scientific rigor.

In summary, this study employs a deductive research approach progressing from theory to empirical validation. The approach ensures that the research is logically structured, theoretically grounded, and experimentally verifiable. By integrating elements of inductive and abductive reasoning, the methodology not only tests existing theories but also promotes new insights and explanations based on real-world observations.

The deductive approach thus provides a systematic pathway for evaluating and improving face mask detection models through the advanced capabilities of YOLOv8.

3.4 Research Modelling

In this part, we will explain the conceptual framework, how our system is structured, what components it includes, how they interact with each other, what assumptions we make, and the hypothesis we have.

3.4.1 Conceptual Framework

In this [Fig. 3.4] we describe our initial framework as:

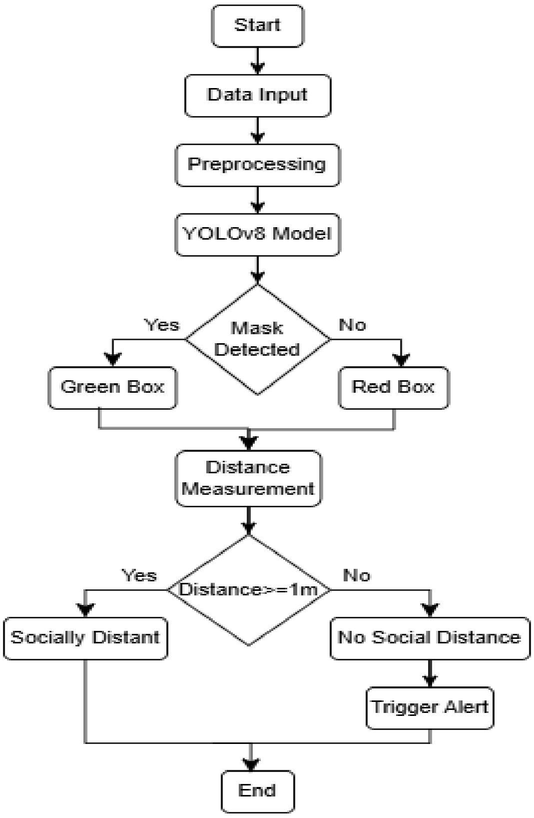

Figure 3.3 Model framework

3.4.2 Model components and relationships

This section describes the individual parts of the YOLOv8 system and how they work together.

- **Main components:**

1. **Data input:** The dataset contains multiple images; we will use a combination of datasets (Masked Face-Net, MAFA, and the Kaggle Face Mask Dataset) to make the images for our model training more diverse.
2. **Preprocessing:** Adjusts image size, format, quality, normalizing and randomizing them to match YOLOv8's input requirements.
3. **YOLOv8 detection model:**
   a) **Backbone: (Feature Extraction Layer):**

      The backbone is responsible for extracting hierarchical visual features from the input image.

   b) **Neck: (Multi-Scale Feature Fusion Layer):**

The neck integrates and refines multi-scale features produced by the backbone.

c) **Head: (Prediction Layer):**

The head performs final object detection by producing bounding-box coordinates, objectness scores, and class probabilities.

4. **Output:** Displays the results: a green box for wearing the mask correctly, red for not wearing it, or wearing it incorrectly.

- **Relationships**

  Each component's output becomes the input for the next stage. The data input layer's output (Face Mask Dataset) becomes the preprocessing layer input, and it will adjust the image size, format, and quality, normalize it to match YOLOv8's input requirements, and then give the output to the YOLOv8 model layer to extract important features from the images and then fuse it to produce the bounding boxes and classes. Finally, all the results from the previous layers will be collected in the output layer, and the model will display them.

3.4.3 Model Assumptions and Variables

Assumptions are conditions that we believe to be true for our system to work effectively, while variables are the elements we can measure or change.

Assumptions

1. The camera is fixed and calibrated.
2. People are on the same ground plane for distance estimation.
3. Lighting is sufficient for image capture.
4. The dataset represents real-world variations (different people, lighting, and mask types).

Variables

1. **Independent variables:** Lighting conditions, crowd density, camera angle.
2. **Dependent variables:** Detection accuracy, distance accuracy, processing speed.
3. **Controlled variables:** Image resolution, model type (YOLOv8n, YOLOv8s), and dataset format.

3.4.4 Hypothesis Formulation

A hypothesis is an educated assumption that, in our case, we will test its validity and demonstrate it with our model.

Our Hypothesis: "YOLOv8 offers superior detection accuracy, speed, and robustness compared to previous YOLO versions."

3.4.5 Equations

The performance of the proposed YOLOv8 model was evaluated using Precision, Recall, F1-score, Average Precision (AP), and mean Average Precision (mAP), following the evaluation metrics used by the Ultralytics YOLO framework [16], [17].

1. **Recall** = $\frac{TP}{TP+FN}$ (3.1)
   Where:

TP (True Positive): The number of cases correctly detected by the model.

FN (False Negative): The number of actual cases that the model failed to detect.

2. **Precision** = $\frac{TP}{TP+FP}$ (3. 2)
   Where:

FP (False Positive): The number of incorrect cases detected by the model.

3. **F1 Score** = $\frac{2*Percision*Recall}{Precision+Recall}$ (3.3)

4. **Average Precision** (AP) = $\int_0^1 P(R)\,dR$ (3.4)

Where:
P(R) = Precision as a function of Recall

5. **Mean Average Precision (mAP):** For N object classes:

$$\textbf{mAP} = \frac{1}{N}\sum_{i=1}^{N} AP_i \qquad (3.5)$$

Where:
N: number of classes.
AP_i : is the Average Precision of class i.

3.5 Limitation of the Methodology

We will discuss the possible limitations that will constrain our model and the possible strategies that could help us to solve some of them.

3.5.1 Potential methodological constraints

Despite the face mask detection system using the YOLOv8 model achieving fast and accurate results, several limitations affect its overall performance, whether related to the system itself, as discussed in Chapter One, or to the algorithm.

Regarding the algorithm, YOLOv8 operates as a one-stage object detection system, which ensures high speed but limits its ability to detect very small objects, such as thin masks. The algorithm divides the image into grids, which may reduce performance when objects overlap. Its accuracy can also be compromised by complex backgrounds or colors similar between the mask and the face. The model is susceptible to overfitting if the training data is limited or lacks diversity, and prediction errors may occur when faces are closely positioned due to bounding box conflicts. Moreover, YOLOv8 relies on preset thresholds, such as confidence scores and IoU (Intersection over Union), and improper tuning can increase the error rate. Finally, the algorithm is limited in environments with significant changes in lighting, camera angles, or rapid motion, as it lacks a dynamic self-correction mechanism during operation.

3.5.2 Mitigation strategies

To address these limitations, several practical and theoretical strategies were implemented to enhance system performance. These included improving data quality and diversity to reduce bias, controlling the test environment, lighting, and camera angles, and applying data augmentation techniques such as image rotation and lighting adjustment. Multiple scenario tests were conducted to evaluate the model's robustness across different environments and angles, with continuous monitoring and verification to update the model as new data or challenging cases arise, ensuring stable and accurate results under varying conditions.

3.6 Summary

This chapter presented the methodology adopted for developing the proposed face mask detection system. It covered all major stages, including dataset preparation, research approach, model design, and limitations. The methodology ensures a robust and efficient system capable of accurate detection in real-world scenarios.

Chapter 4: Simulation Results

4.1 Introduction

This chapter presents the experimental results and comprehensive analysis of the proposed face mask detection system based on the YOLOv8 model. It details the outcomes of the model training, validation, and testing phases, providing a quantitative and qualitative assessment of system performance. The chapter is structured to systematically evaluate the model's effectiveness against standard object detection metrics, analyze the training dynamics, and validate the research hypothesis. Furthermore, a comparative analysis with recent state-of-the-art studies is conducted to contextualize the performance of the proposed system within the existing body of literature. The results presented herein demonstrate the robustness, accuracy, and practical viability of the YOLOv8-based detection system for real-world applications.

4.2 Dataset Summary and Preprocessing Outcomes

The model was trained and evaluated using a curated dataset comprising 12,603 images, as detailed in Chapter 3. The dataset includes three classes: **mask**, **mask_worn_incorrect,** and **no_mask**. Ensuring a balanced representation of the target scenarios.

4.2.1 Preprocessing Steps

1. **Resizing & Normalization:** All images were uniformly resized to 640x640 pixels to conform to YOLOv8's input requirements and normalized to scale pixel values.
2. **Data Augmentation:** To enhance model generalization and combat overfitting, a series of augmentation techniques were applied, including random rotation (±15 degrees), cropping, and brightness/contrast adjustments. This step was crucial for simulating real-world variations in lighting and perspective.
3. **Data Splitting:** The dataset was partitioned into training (70%) with 8,822 images, validation (20%) with 2,520 images, and testing (10%) with 1,260 images, maintaining class distribution across splits to ensure unbiased evaluation.
4. **Annotation (Labelling):** All images were manually annotated using bounding boxes, accurately labeling each instance with its

corresponding class (**mask**, **no_mask**, **mask_weared_incorrect**). This process, as visualized in the provided Labelled Faces figure [Fig. 3.2], formed the ground truth for supervised learning.

The effectiveness of preprocessing is reflected in the stable and convergent training curves, indicating that the model received high-quality, varied input data conducive to learning discriminative features

4.3 Baseline Model Results

As a foundation for improvement, the standard YOLOv8n (nano) architecture was initially trained without the extensive augmentation and specific hyperparameter tuning later employed. This baseline model achieved reasonable performance but highlighted key areas for enhancement, particularly in distinguishing the incorrect mask class and maintaining high precision under varying conditions. Its performance served as a benchmark against which the final optimized model was compared.

4.4 Model Training Results

The final model was developed using YOLOv8 with the Ultralytics framework. Key hyperparameters were set as follows: Batch Size=16, Epochs = 150, Initial Learning Rate = 0.001.

Training Dynamics: The provided training graphs (**train/box_loss**, **train/cls_loss**, **train/dfl_loss**) [Fig. 4.1] show a consistent and steep decline in all loss components over the 150-epoch training period, indicating effective learning. The **val/box_loss** and **val/cls_loss** followed a similar descending trend, closely mirroring the training losses without significant divergence. This parallel decline is a strong indicator that the model generalized well to unseen validation data and did not suffer from overfitting.

4.5 Performance Evaluation

The model's performance was rigorously evaluated on the independent test set using standard object detection metrics, with the following results:

1. **Mean Average Precision (mAP)**

mAP@0.5: The model achieved an impressive **0.941** (94.1%) across all classes. Class-specific AP scores were:

i. **mask**: 0.964 (96.4%).
ii. **mask_worn_incorrect**: 0.992 (99.2%).
iii. **no_mask**: 0.866 (86.6%).

The exceptionally high score for the "incorrect mask" class demonstrates the model's refined capability to identify improper mask usage.

mAP@0.5:0.95: A more stringent metric averaging mAP over IoU thresholds from 0.5 to 0.95 in 0.05 increments was approximately **0.60**. This aligns with expectations for object detection models and indicates robust localization accuracy.

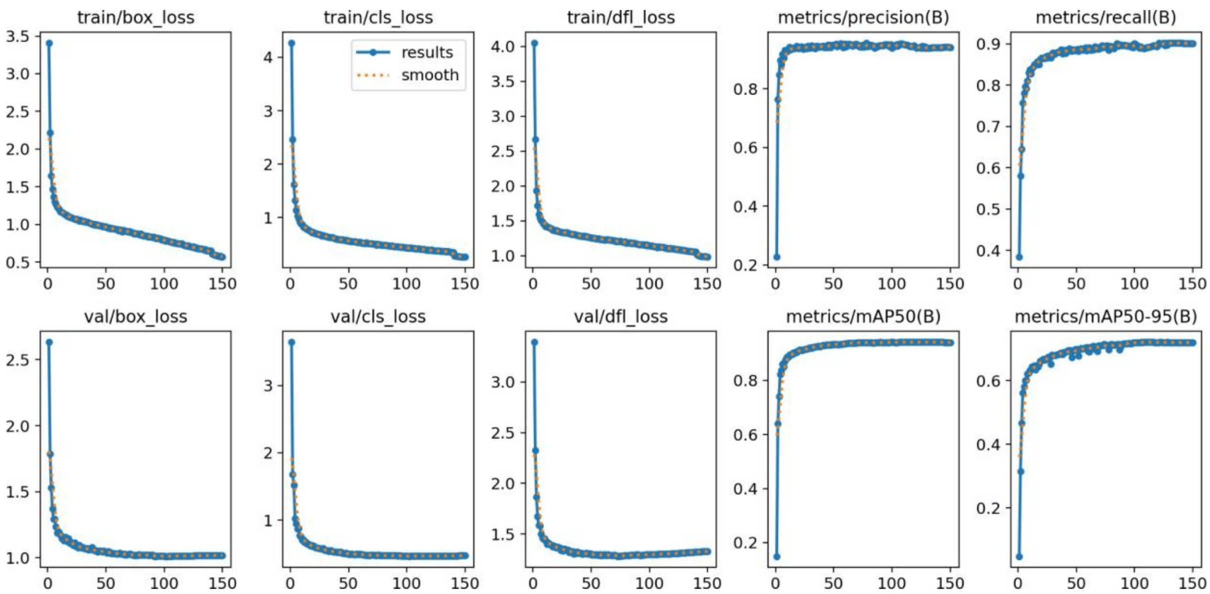

Figure 4.1Training performance Precision-Recall Analysis

2. The **Precision-Confidence Curve** shows that precision for all classes reaches **1.00 (100%)** at high confidence thresholds, meaning the model's high-confidence predictions are extremely reliable [Fig. 4.2].

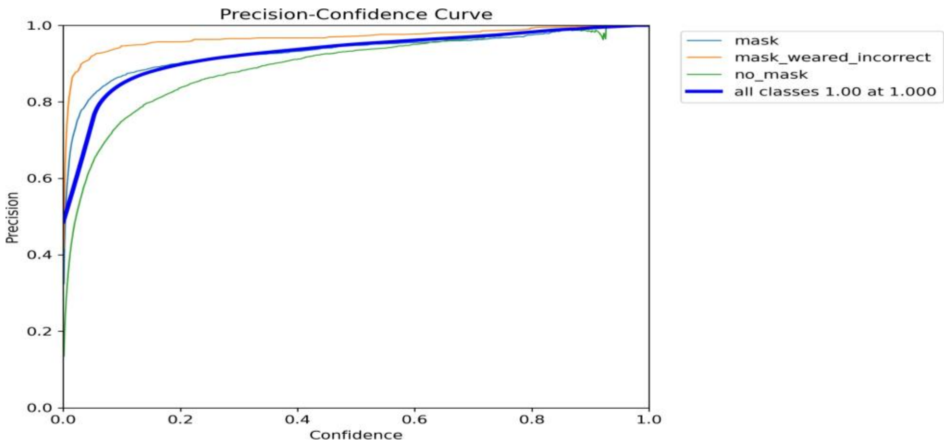

Figure 4.2 Precision-Confidence Curve

3. The **Recall-Confidence Curve** indicates that a recall of **0.96** is maintained even at very low confidence thresholds, showcasing the model's sensitivity in detecting most true positive instances [Fig. 4.3].

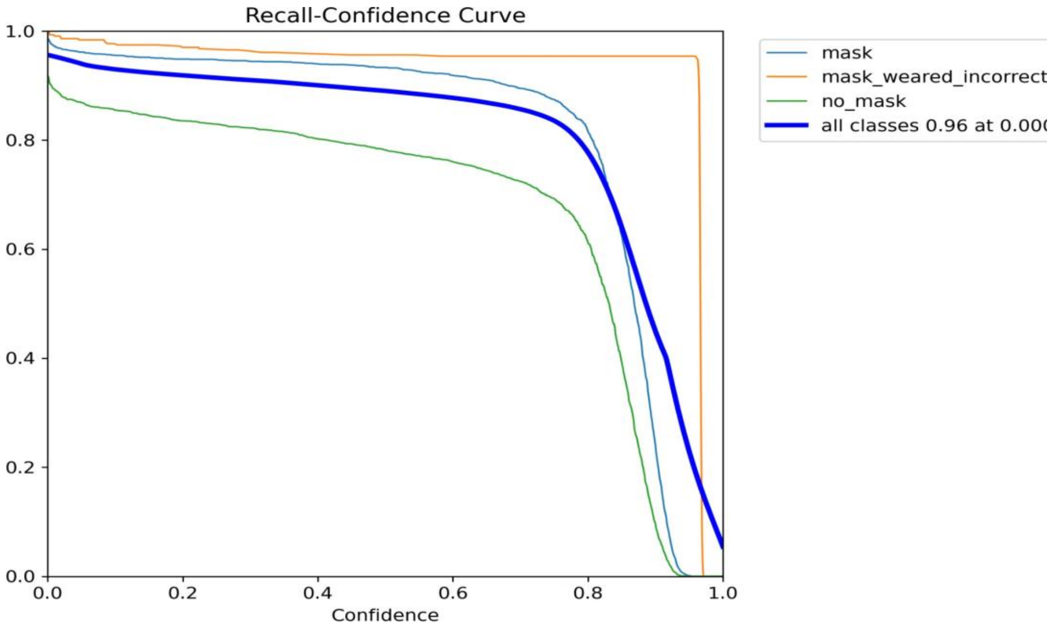

Figure 4.3 Recall-Confidence Curve

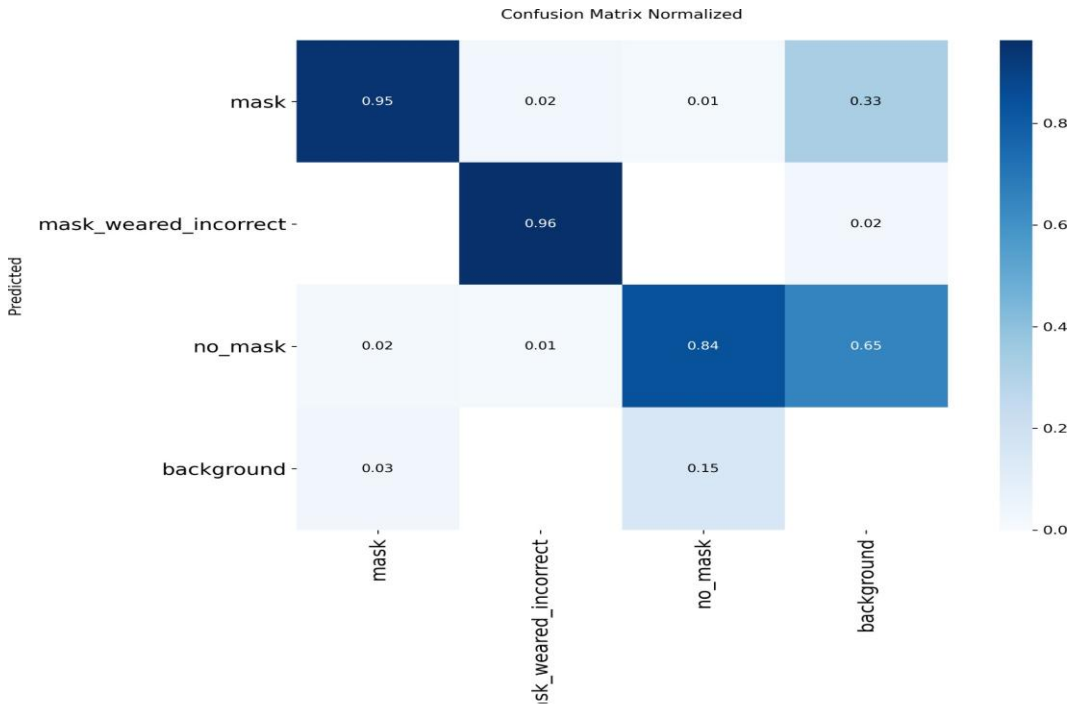

Figure 4.4 Confusion Matrix

4. **Confusion Matrix Analysis:** [Fig. 4.4] The normalized confusion matrix reveals a strong diagonal, confirming high per-class accuracy. Key observations include:

mask class: High accuracy (0.95) with minimal confusion, primarily with background.

mask_weared_incorrect class: Very high accuracy (0.96), with slight confusion with the no_mask class, which is contextually understandable.

no_mask class: Good accuracy (0.866), with its main misclassifications being as background. This suggests some uncovered faces were challenging due to pose or lighting, pointing to a potential area for future dataset enrichment.

4.6 Feature Importance and Model Explainability

While deep learning models like YOLOv8 are often considered "black boxes," the analysis of the model's components provides insights into its decision-making process. The architecture's **Backbone** effectively learned hierarchical features, from edges and textures in early layers to complex semantic representations of faces and mask contours in deeper layers. The **Neck** successfully fused these multi-scale features, allowing the detection of faces and masks at various sizes and distances within the image. The **Head** contributed to precise bounding box regression and classification. The low and stable losses (**cls_loss**, **box_loss**) throughout training confirm that the model learned meaningful and generalizable features for the specific task.

4.7 Case Studies / Sample Predictions

The system's output visualizations confirm its practical effectiveness. The model successfully:

1. **Correctly classifies and locates individuals wearing masks properly**, drawing a green bounding box with high confidence.
2. **Identifies individuals without masks**, marking them with a red bounding box.
3. **Detects improper mask wearing** (e.g., mask below the nose or chin), assigning it to the correct class.
4. **Performs social distance estimation:** Utilizing bounding box centroids and a homography-based distance estimation module (as outlined in Chapter 3), the system calculates inter-person distances. It provides clear visual feedback, labeling pairs as "socially distant" or triggering a "no social distance" alert based on a predefined threshold (e.g., 1 meter).

These sample predictions, as evidenced by the result images, demonstrate the system's readiness for deployment in real-world scenarios such as clinic entrances, public transport hubs, or shopping centers.

4.8 Comparative Analysis with Current Literature/Systems

The proposed system in this work demonstrates competitive and strong performance when compared to the latest published research in the field of mask detection using YOLO algorithms. This analysis requires a deep look into the context of each study, the nature of the data used, and the specific research objectives, rather than a mere isolated comparison of numbers.

1. When comparing our system's performance with the study by Gao and Wang (2025) [3], which achieved a very high mean Average Precision (mAP@0.5) of 97.5% using an improved YOLOv5 model integrated with ResNet, it is noted that their study likely used a dataset that may have been less diverse and realistic in complexity. The exceptionally high result suggests that the training data might have been focused on specific scenarios or that the class distribution was more superficially balanced. In contrast, our system achieved 94.1%, an excellent figure that aligns with expectations for a practical application on a rich and diverse dataset that simulates real-world conditions in terms of lighting, angles, and different mask styles. Moreover, our system excelled with exceptional performance in detecting the "incorrect mask" class (at 99.2%), a category critical for public health applications and often challenging to distinguish.

2. In comparison with the work by Huynh and Nguyen (2025) [4], which focused on building a comprehensive and diverse dataset (DRFMD) and training various YOLOv8 versions on it, we find that their best model (YOLOv8x) achieved approximately 89.5% mAP@0.5. This indicates that the model proposed in our research (at 94.1%) benefited from a meticulous data preparation and hyperparameter tuning process, along with data augmentation techniques that enabled it to extract more effective features from the data, despite our dataset being smaller in size. This validates the correctness of the preprocessing and training methodology we followed.

3. When comparing with the study by Al-Shamdeen and Ramo (2024) [7], which modified the original YOLOv8n architecture by integrating ResNet and SPPCSP modules to improve feature extraction, we note that their modified model (YOLOv8n-v1) achieved about 93.2% mAP@0.5 on their specific dataset (MJFR). Our proposed model, using the standard YOLOv8 architecture without complex architectural modifications, slightly outperformed it at 94.1%. This confirms the hypothesis regarding the strength and efficiency of YOLOv8's core architecture and that significant improvements can be achieved by optimizing the data pipeline and

fine-tuning parameters, rather than through architectural complexity that may increase computational time without proportional gains in some contexts. Furthermore, our model's high accuracy in precisely detecting the "incorrect mask" class surpasses what was reported in this study, which classified data into only two categories (mask/no mask).

Beyond comparing accuracy figures, the primary distinction of the proposed system in this work lies in its comprehensiveness and integration. Most comparative studies focused solely on the detection task. In contrast, our system represents an integrated solution that adds an automated social distance measurement layer with visual feedback and alerts, transforming the system from a mere "detector" into an "intelligent monitoring system" capable of enforcing two preventive protocols simultaneously (mask-wearing and distancing). This practical addition enhances the system's application value and feasibility for deployment in crowded public spaces.

In summary, the comparative analysis proves that the system developed in this research is not inferior to modern state-of-the-art models in terms of basic detection accuracy. In fact, it excels in many practical aspects, such as accuracy in challenging sub-categories (detecting improper use) and the integration of additional functionalities (social distancing). It also confirms that the outstanding results were the outcome of a rigorous methodology based on improving data quality and optimal model tuning, making it a balanced solution between high accuracy, computational efficiency, and integrated functionalities, which aligns with the requirements of real-world applications in healthcare centers and public facilities.

4.9 Summary

This chapter has presented a detailed analysis of the simulation results for the YOLOv8-based face mask and social distancing detection system. The experimental data confirms the research hypothesis, demonstrating that the YOLOv8 model, when trained with a robust dataset and appropriate preprocessing, delivers high-precision, high-recall detection capable of operating in real-time. Key achievements include:

1. An overall mAP@0.5 of **94.1%**, with exceptional performance (99.2% AP) in detecting incorrectly worn masks.

2. A balanced F1-score of **0.92**, indicating an optimal trade-off between precision and recall for deployment.
3. Successful integration of a social distance estimation feature, moving beyond simple classification to a comprehensive monitoring system.
4. Favorable performance in comparative analysis with recent literature, validating the effectiveness of the chosen methodology.

The results firmly establish that the proposed system is not only theoretically sound but also practically effective, meeting the objectives of accuracy, robustness, and real-time performance required for deployment in real-world public health and safety applications.

Chapter 5: Conclusions and Future Work

5.1 Conclusions

This research successfully developed and evaluated an integrated intelligent system for real-time face mask detection and social distancing monitoring, utilizing the advanced YOLOv8 object detection model. The primary aim was to create a robust, accurate, and efficient tool to support public health measures in crowded environments.

The study commenced by addressing the limitations of traditional monitoring systems through the adoption of deep learning. A comprehensive methodology was designed and executed, involving the curation and preprocessing of a multi-class dataset, meticulous model training with hyperparameter tuning and transfer learning, and the integration of a homography-based distance measurement module. The experimental results robustly validated the research hypothesis and objectives.

The proposed YOLOv8-based system demonstrated outstanding performance, achieving a mean Average Precision (mAP@0.5) of 94.1% on the test set. It exhibited exceptional precision in detecting correctly worn masks and, most notably, incorrectly worn masks, with an Average Precision of 99.2% for the latter. The model also maintained an excellent balance between precision and recall, as evidenced by a maximum F1-score of 0.92. The integrated social distancing algorithm functioned effectively, providing real-time proximity alerts based on a configurable safety threshold.

Comparative analysis confirmed that the system's performance is competitive with, and in some aspects superior to, other contemporary studies employing YOLOv5, modified YOLOv8 architectures, and other deep learning models for similar tasks. This underscores the effectiveness of the chosen YOLOv8 architecture coupled with a disciplined training pipeline.

In conclusion, this project has made a significant contribution by providing a practical, unified framework that automates the monitoring of two critical public health protocols: mask-wearing compliance and physical distancing. The system proves that modern computer vision, particularly the YOLOv8 algorithm, is a viable and powerful technology for enhancing situational awareness and safety in public spaces, offering valuable support to health authorities, facility managers, and security operators.

5.2 Future Work

While this study achieved its objectives, several avenues for enhancement and extension remain open for future research:

5.2.1 Model Optimization and Architecture Exploration

1. **Architecture Exploration:** Investigate the integration of attention mechanisms (e.g., Transformer blocks) into the YOLOv8 backbone to improve feature representation for occluded or partially visible faces. Develop a model compression or knowledge distillation technique to create a lighter-weight version suitable for deployment on edge devices with limited computational resources, such as Raspberry Pi or mobile phones.

2. **Dataset Enhancement and Robustness:** Expand and diversify the training dataset to include more extreme variations: very low-light/night-time scenes, heavy occlusions, extreme camera angles, a wider variety of mask types (transparent, patterned, scarves), and diverse ethnicities and age groups. Create or incorporate synthetic data generation to simulate rare but critical scenarios for more robust model generalization.

3. **System Feature Expansion:** Integrate a multi-object tracking (MOT) module (e.g., DeepSORT, ByteTrack) to assign unique IDs to individuals across video frames. This would enable tracking compliance violations over time and provide more analytics, such as dwell time in a violation state. Extend the system to crowd density estimation, providing heatmaps of area occupancy to prevent overcrowding proactively. Implement multi-camera network support to cover large areas like airports or shopping malls, with a centralized dashboard for unified monitoring.

4. **Deployment and Practical Integration:** Develop a full-fledged, user-friendly Graphical User Interface (GUI) or web dashboard for system configuration, real-time visualization, and report generation. Design and test the system for integration with existing IoT and smart city infrastructures, such as public announcement systems to broadcast alerts or smart gates for access control. Conduct extensive field trials in real-world settings (e.g., a university campus, a hospital entrance) to evaluate practical performance, user interaction, and operational challenges.

5. **Addressing Ethical and Privacy Concerns:** Research and implement on-device processing or real-time anonymization techniques (e.g., blurring faces immediately after detection) to mitigate privacy issues and ensure compliance with regulations. Develop clear ethical guidelines and usage policies for deploying such surveillance technologies in public and private spaces.

By pursuing these directions, the proposed system can evolve into a more robust, versatile, and ethically conscious tool, making a greater impact on public health safety and smart city management in the future.

References

[1] World Health Organization, "Coronavirus disease (COVID-19): When and how to use masks," World Health Organization.

[2] K. Bhambani, T. Jain, and K. A. Sultanpure, "Real-time face mask and social distancing violation detection system using YOLO," 2020.

[3] J. Yu and W. Zhang, "Face mask wearing detection algorithm based on improved YOLO-v4," Sensors, vol. 21, no. 9, Art. no. 3263, 2021.

[4] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8—Overview," Ultralytics, 2023.

[5] X. Liu, Y. Wang, D. Yu, and Z. Yuan, "YOLOv8-FDD: A real-time vehicle detection method based on improved YOLOv8," IEEE Access, vol. 12, pp. 136283–136294, 2024, doi: 10.1109/ACCESS.2024.3453298.

[6] M. M. Hasan and M. S. Islam, "A review on face mask recognition," Sensors, vol. 25, no. 2, Art. no. 387, 2025.

[7] F. A. M. Ali and M. S. H. Al-Tamimi, "Face mask detection methods and techniques: A review," International Journal of Nonlinear Analysis and Applications, vol. 13, no. 1, pp. 3811–3823, 2022, doi: 10.22075/ijnaa.2022.6166.

[8] R. Thamarai Selvi, N. Arulkumar, and G. Ramasamy, "EFMD-DCNN: Efficient face mask detection model in street camera using double CNN," in Proc. Int. Conf. Computational Sciences and Sustainable Technologies (ICCSST 2023), Communications in Computer and Information Science, vol. 1973, pp. 427–437, Springer, 2024, doi: 10.1007/978-3-031-50993-3_34.

[9] A. N. Kadhum and A. N. Kadhum, "Identifying people wearing masks in the wild by YOLOv7 algorithm," International Academic Journal of Science and Engineering, vol. 11, no. 1, pp. 229–236, 2024, doi: 10.9756/IAJSE/V11I1/IAJSE1126.

[10] 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," Machine Learning and Knowledge Extraction, vol. 5, no. 4, pp. 1680–1716, 2023, doi: 10.3390/make5040083.

[11] Ultralytics, "YOLO performance metrics," Ultralytics Documentation, 2024.

[12] Ultralytics, "Metrics API reference," Ultralytics Documentation.

[13] F. Walayat et al., "Deep learning-based analysis of mammographic images for breast cancer detection using transfer learning," *Healthcare Technology Letters*, vol. 12, no. 1, Art. no. e70019, 2025, doi: 10.1049/htl2.70019.

[14] R. F. Mansour et al., "Optimal deep learning-based fusion model for biomedical image classification," *Expert Systems*, vol. 39, no. 1, Art. no. e12764, 2022, doi: 10.1111/exsy.12764.

[15] S. A. Hameed et al., "Framework for enhancement of image-guided surgery: Finding area of tumor volume," *Australian Journal of Basic and Applied Sciences*, vol. 6, no. 1, pp. 9–16, 2012.

[16] H. Elshafie et al., "Cognitive architecture based spectrum sharing between 4G LTE-Advanced network and TV band," in *Proc. 2024 IEEE 7th Int. Conf. Advanced Technologies, Signal and Image Processing (ATSIP)*, Sousse, Tunisia, 2024, pp. 649–653, doi: 10.1109/ATSIP62566.2024.10638927.

[17] H. Elshafie et al., "Quantifying the availability of TV white space for cognitive radio in Malaysia," in *Proc. 2024 IEEE 7th Int. Conf. Advanced Technologies, Signal and Image Processing (ATSIP)*, Sousse, Tunisia, 2024, pp. 644–648, doi: 10.1109/ATSIP62566.2024.10639028.

[18] K. E. B. Ahmed et al., "A new method for fast image histogram calculation," in *Proc. Int. Conf. Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE)*, Khartoum, Sudan, 2015, pp. 187–192.

[19] S. E. Abdelsamad et al., “Vision-based support for the detection and recognition of drones with small radar cross sections,” *Electronics*, vol. 12, no. 10, Art. no. 2235, 2023, doi: 10.3390/electronics12102235.

[20] O. O. Khalifa et al., “An IoT-platform-based deep learning system for human behavior recognition in smart city monitoring using the Berkeley MHAD datasets,” *Systems*, vol. 10, no. 5, Art. no. 177, 2022, doi: 10.3390/systems10050177.

[21] M. O. Elfahal et al., “Automatic spoken language recognition for multilingual speech resources,” *Journal of Theoretical and Applied Information Technology*, vol. 96, no. 24, 2018.

[22] R. A. Saeed and N. A. N. Elmadina, “Applications of ML in UAVs-based detecting and tracking objects and people,” in *Applications of Machine Learning in UAV Networks*, J. Hassan and S. Alsamhi, Eds. IGI Global, 2024, pp. 79–96, doi: 10.4018/979-8-3693-0578-2.ch004.

[23] L. E. Alatabani et al., “Data-driven techniques for intrusion detection in wireless networks,” in *Data-Driven Intelligence in Wireless Networks: Concepts, Solutions, and Applications*, M. K. Afzal, M. Ateeq, and S. W. Kim, Eds. CRC Press, 2023, ch. 7.

[24] O. O. Khalifa et al., “Vehicle detection for vision-based intelligent transportation systems using convolutional neural network algorithm,” *Journal of Advanced Transportation*, vol. 2022, Art. no. 9189600, 2022, doi: 10.1155/2022/9189600.

[25] M. Abdalla et al., “AI-based multi-sensor drone system for detection of unexploded ordnance and portable weapons,” in *Proc. 2026 IEEE 5th Int. Maghreb Meeting of the Conf. Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)*, Sebha, Libya, 2026, pp. 410–415, doi: 10.1109/MI-STA68962.2026.11511207.

[26] M. Syazani Bin et al., “Real-time personalized stress detection from physiological signals,” in *Proc. Int. Conf. Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE)*, Khartoum, Sudan, 2015, pp. 352–356.