

Towards Human-Centric and Resilient Hospital Ecosystems: An AI-Driven Facemask Compliance Detection Framework for Healthcare Workers and Patients

Anup .

Sharda University

Manik Batra

manikbatra8@gmail.com

Sharda University

Parul Saxena

Sharda University

Himanshi Puri

Sharda University

Systematic Review

Keywords: Human-centricity, resilience, digitalization, facemask detection, deep learning, hospital ecosystems, healthcare-associated infections, YOLOv8, MobileNetV2, PPE compliance, Industry 4.0-aligned healthcare systems, SDG 3

Posted Date: August 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10255921/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Towards Human-Centric and Resilient Hospital Ecosystems: An AI-Driven Facemask Compliance Detection Framework for Healthcare Workers and Patients

Anup¹, Manik Batra^{1*}, Parul Saxena^{1*}, Himanshi Puri^{1*}

1: Sharda School of Business Studies, Sharda University, Greater Noida, India

Email Id (Corresponding author): manikbatra8@gmail.com

Abstract:

The hospitals around the globe are undergoing a triple transformation driven by digitalization, resilience, and human-centricity. Hospitals are critical service-delivery sectors, where digital technologies need to be embedded in clinical operations to protect patients and healthcare workers (HCWs). This study reviews artificial intelligence (AI) and deep learning (DL)-based facemask detection techniques, examining how they can be embedded into hospital infrastructure to serve as personal protective equipment (PPE) compliance for both patients and HCWs. A systematic review was conducted of deep learning architectures for facemask detection including ResNet-50, MobileNetV2, CNN, the YOLO family (v1–v5), VGG-19, and Inception-V3 and evaluated against the transformation dimensions: (i) *digitalization*, assessed through real-time inference capability and integration with existing hospital CCTV; (ii) *resilience*, assessed through robustness to clinical-environment variability (occlusion, lighting, patient positioning); and (iii) *human-centricity*, assessed through the system's sensitivity to the differentiated needs of HCWs, patients, and visitors across hospital zones. The review shows that no single architecture optimally satisfies all three transformation dimensions simultaneously. YOLOv5-based models deliver the strongest digitalization and resilience performance, achieving precision and recall above 97% under occlusion and variable lighting, making them suitable for high-acuity zones such as ICUs and operating theatres. Embedding AI-based facemask detection into hospital infrastructure operationalizes the triple transformation at the level of frontline care delivery: it digitalizes a previously manual compliance process, strengthens institutional resilience against healthcare-associated infections (HAIs) and focuses on the wellbeing of HCWs and patients. This study extends the integration of triple transformation to hospital service ecosystems, an environment where the same transformation pressures apply but have received comparatively little systematic attention from an AI-engineering perspective. It provides one of the first frameworks to evaluate facemask detection architectures explicitly against these three transformation dimensions.

Keywords: Human-centricity, resilience, digitalization, facemask detection, deep learning, hospital ecosystems, healthcare-associated infections, YOLOv8, MobileNetV2, PPE compliance, Industry 4.0-aligned healthcare systems, SDG 3

1. Introduction

Hospitals have been recognized as the most infection-prone settings, where millions of people are affected by Healthcare-associated infections (HAIs) spread by patients or hospital staff during healthcare services. According to the World Health Organization (WHO), out of 100 patients in acute care in hospitals, 7 in high-income countries and 15 in low and middle-income countries suffer from at least one HAI during their hospital stay, where one in every 100 patients dies from their HAI [57].

According to CDC guidelines (2024) [9], facemasks are mandated as a personal protection matter and control in clinical settings, to control the spread of pathogens responsible for HAIs. Despite this compliance, consistently enforcing mask-wearing across all hospital zones has been a significant operational challenge. The clinical and non-clinical hospital staff, including doctors, para-medical staff, and biomedical technicians, should wear appropriate masks on their faces while they are present in clinical areas, yet the practice of facemask adoption is not being carried out seriously, and the compliance rates are far below the recommended thresholds, especially in non-clinical areas [42]. Inpatients, outpatients, and visitors present more compliance challenges, as they don't receive the same professional training and don't consider the risks associated with improper mask use (Figure 1). Hospital departments such as reception, emergency departments, and hospital corridors have high outpatient-to-staff ratios, and infection control is severely compromised manual monitoring of mask compliance is practically infeasible.

The insistent deficit in facemask compliance across both clinical and non-clinical hospital environments is not just an operative lapse but a direct obstacle to the realisation of Sustainable Development Goal 3 (SDG 3), which aims to ensure healthy lives and promote well-being for all at all ages [55]. Within this principal goal, the control of communicable and healthcare-associated infections occupies a central position, and the correct, consistent use of personal protective equipment represents one of its most immediate frontline instruments [55],[58].

The term “facemask detection” refers to the automated identification of whether someone is wearing a facemask or ignoring it. A typical face mask detection method involves training a model on face datasets. The face is then recognized and the final image frames are classified using the mask detection model. Deep learning models have been generally categorized into six phases. (1) The image dataset is passed to the system (dataset input), (2) Image pre-processing stage where the input is pre-processed so that images are arranged in a list, and then that list is converted to an array, (3) The data augmentation phase where images are rotated and flipped into the vertical or horizontal format, (4) Feature extraction stage for feature extraction of dataset images, (5) Training and Testing of images are done using the CNN model [7] and (6) inference using face detector model to identify whether people have worn the facemask or not.

The Single Shot Detector (SSD) framework is commonly used for object detection. This facemask detector can be used in various public settings, including hospitals to keep an eye on the public and monitoring compliance with mask-wearing regulations. DL based facemask detector using MobileNetV2 is used to for image processing as it offers computationally fast and accurate detection, making the detector suitable for a range of camera environments [18]. A multi-stage DL (CNN) framework, integrated with pre-installed security cameras, can discriminate between masked and

unmasked individuals. This multi-stage system can help in maintaining the health of the individuals and promotes facemask use [5].

The integration of AI with computer vision for facemask detection offers a transformative solution to this challenge in hospitals. Automated systems can be developed by integrating with the existing CCTV infrastructure in hospitals to monitor facemasks compliance in several hospital zones simultaneously and detect violations in real time. This technique falls under the broader category of object detection, specifically for human faces (with or without a mask) [40]. Deep Learning models which are developed by getting trained on large datasets of masked and unmasked faces can be used in flagging violators in real time who are without masks or not wearing them properly in hospital premises [45].

This study conducts a systematic literature review of different AI-based detection methods in varied environment. It takes into account the exhaustive analysis of facemask recognition approaches and examine the accuracy and efficiency of these methods for clinical deployment. It examines majority of the deep learning architectures and their application in facemask detection like ResNet-50, Convolutional Neural Networks (CNN), MobileNetV2, YOLO, VGG-19, and Inception-V3 along with their hybrid implementation. Finally the study discusses available open research challenges, and future directions for hospital level AI mask surveillance systems.

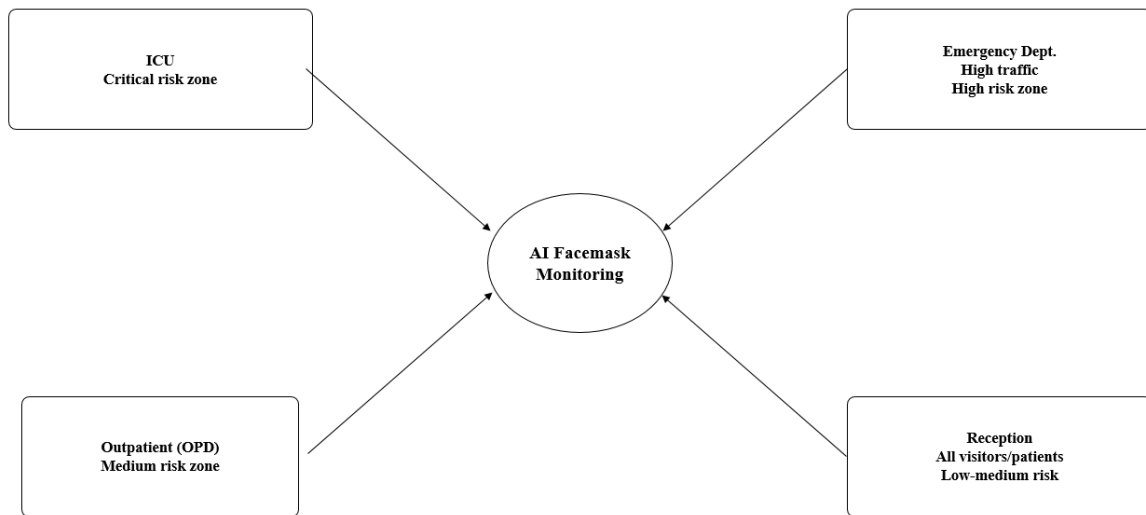


Figure 1: AI-based facemask monitoring across hospital zones (ICU, ED, OPD, and Reception)

The study has been undertaken with the following objectives:

- To examine the literature on facemask detection, which uses a variety of techniques like including ResNet-50, CNN, YOLO, VGG-19, and Inception-V3 and determined and compared their accuracy, efficiency, and scalability
- To discuss the advantages and limitations of each facemask detection techniques
- To propose a hospital zone-specific method for mask detection deployment, differentiating requirements by risk level and user group (Healthcare staff, patients, visitors).

The remaining sections are discussed in below. The research methodology for this work is covered in Section 2. The current state of the art in facemask detection research is covered in Section 3, along with recent techniques. The research challenges and issues from earlier studies are presented in Section 4. Finally, summarize this study and review the article's future directions.

2. Research methodology

This study conducts a systematic literature review by examining the research papers to extract the most appropriate techniques used in facemask detection. The proposed research methodology is divided into three subsections described below.

2.1 Identification of RQ

The following research questions (RQs) are developed to guide the review:

- **RQ1:** Which facemask detection method demonstrate the highest accuracy and real-time performance in hospital or healthcare-adjacent environments?
- **RQ2:** What are the principal technical challenges of deploying facemask detection systems for patients and HCWs in hospital settings?
- **RQ3:** What future research directions are needed to advance AI mask detection systems to hospital deployment readiness?

The exclusion/inclusion criteria for studies are mentioned in Table 1.

Exclusion Criteria	• Studies that don't have experimental results. • When the dataset is not clear. • Discuss topics beyond facemask detection images. • Studies not related to images.
--------------------	--

Inclusion Criteria	• Facemask detection studies include datasets, problems, topics, and techniques used. • This research consisted of reputed journals/ conferences. • Discuss topic-related facemask detection images. • The selection of research articles was studied from 2011 to 2025.
--------------------	---

Table 1: Exclusion and Inclusion Criteria for selection of research articles related to facemask detection

2.2 Search strategy for research articles

The data sources that will be analyzed for facemask detection techniques include research articles from Elsevier, IEEE Xplore, Scopus, ScienceDirect, Springer, Google Scholar, and ACM. The search strings for research articles were based on keywords extracted from the research topic: (facemask recognition OR facemask detection OR 2D facemask recognition OR 3D facemask recognition OR linear facemask detection OR non-linear facemask detection) AND (methods OR techniques OR approach).

2.3 Selection of research articles

Once the initial search for the research articles is completed, the articles that meet the inclusion criteria were selected. This was done according to the exclusion and inclusion criteria listed in Table 1. In addition, we applied a PRISMA style mechanism (Figure 2) to collect limited research paper articles that meet the research topic for further review.

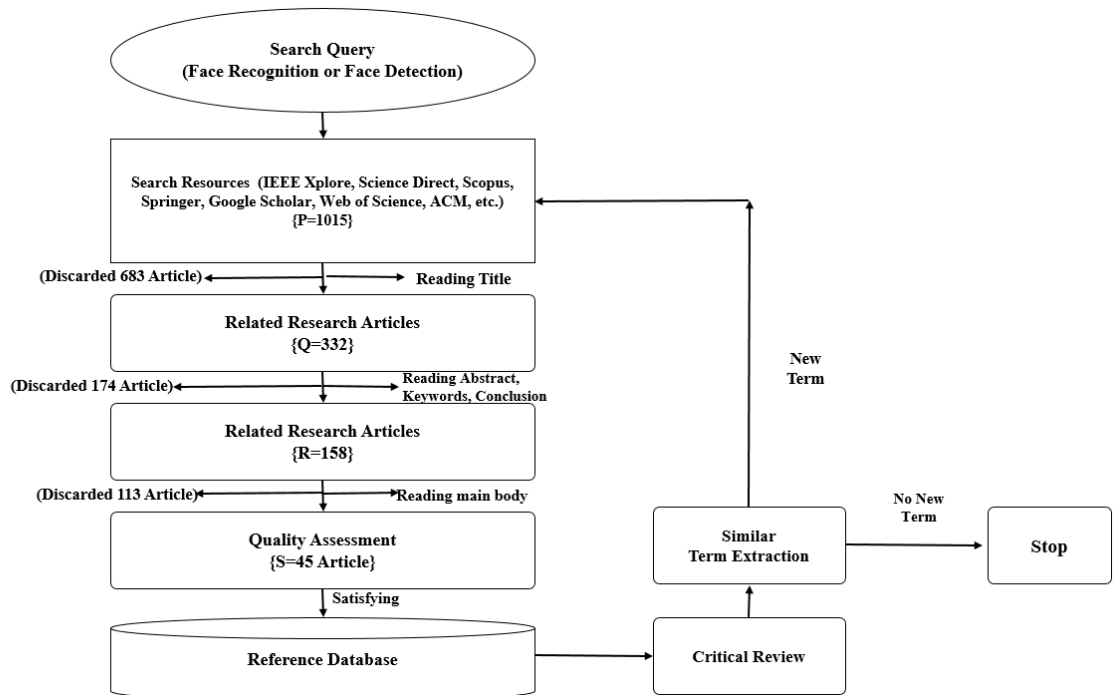


Fig. 2: Search and selection of research article process

The initial research article search was done on database like IEEE Xplore, ScienceDirect, Scopus, Google Scholar, Web of Science, and ACM. These resources yielded 1015 research articles. After reading and reviewing the existing research papers, we discarded 683 research articles out of 1015 articles based on title, 174 research articles out of 332 based on abstract and keywords and 113 research articles based on the main body of the research article. At last, we were left with 45 research articles suitable for final review.

3. Taxonomy of facemask detection techniques

The taxonomy of facemask detection techniques discussed after conducting review of literature is classified in several ways as mentioned in Figure 3. These classifications are systematically discussed for their advantages and disadvantages in the following subsections:

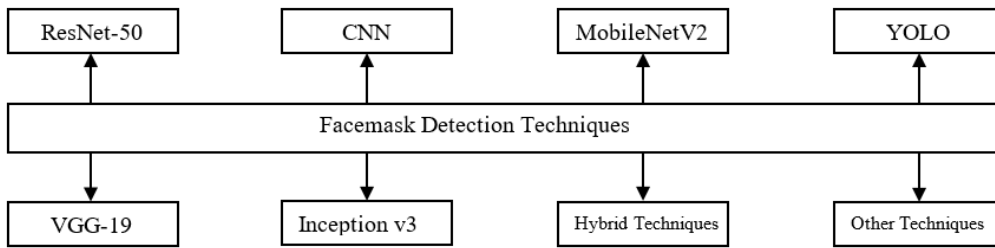


Figure 3: Taxonomy of facemask detection techniques

3.1 ResNet-50

ResNet-50 is a convolutional neural network (CNN) and is widely used to extract features during object identification and image categorisation. ResNet [51] introduced skip connections which are used to address the vanishing gradient issues in photos. ResNet enables deep training of DNNs with more than 150 layers. Prior to ResNet, the issue of disappearing components made it difficult to train very few DNNs. This is feasible because the connections that ResNet skips prevent the weights from dropping to a very low value.

Loey et al. [32] discussed a hybrid design that combined deep and traditional machine learning for facemask detection. There were two parts to the hybrid model. The first component performs extracting features using ResNet 50 while the second component identifies facemasks using SVM ensemble methods and decision trees. The authors used three datasets: The LFW, the SMFD, and the RMFD and got 100% testing accuracy in LFW, 99.47% in SMFD, and 99.63% in RMFD. Snyder and Husari [49] further implemented an automated detection system to identify unmasked individuals in public settings and then provided them with a surgical mask. Three major components make up

the proposed strategy. The first component employs a deep neural network model that combines Feature Pyramid Network (FPN) with ResNet-50 to recognize people in video feeds. In the second component, human faces are extracted from these videos using Multi-Task CNN (MT-CNN), which can identify and recognize faces. The third component creates and trains a CNN classifier to discriminate between human beings wearing masks and those not. The proposed model was tested using a collection of videos gathered in public locations. Mandal et al. [34] introduced a domain specific facemask identification system based on ResNet-50 across source and targeted domain (unmasked and masked faces). The study was conducted in two components: In the first, the model was trained and evaluated solely on the source domain. In the second case, the model was evaluated on the target domain and a part of the given dataset. Thareja et al. [52] discussed facemask detection using ResNet-50 and developed a transfer learning model. These techniques detect whether people are wearing facemasks. Deng et al. [13] proposed the MFCosface MFV algorithm based on high-margin cosine loss and employed an attention-aware technique to distinguish important facial characteristics. The author created a module to direct the model's attention to regions not covered by the mask (also known as the convolution block attention and the Inception-Resnet module). Any unoccupied area has its weight in the feature map enhanced, which increases its contribution to the detection stage. Baalaji et al. [56] discussed the ResNet-50 and single-shot multi-box detector for facemask detection. The model can identity face retrieval based on matching the face through the deep Siamese neural network. Naseri et al. [39] introduced an optimized face detector-based intelligent facemask detection model in IoT using DL techniques. These techniques identify accurate facemask detection of whether people wear face masks or not. Table 2 briefly discusses the existing work related to ResNet techniques.

Sno	Authors & References	Techniques	Datasets	Performance
1	Loey et al. [32]	SVM, DT, EM	RMFD, SMFD, LFW	• For the RMFD dataset: The accuracy of the SVM classifier achieved 99.64%. • For the SMFD dataset: The accuracy of the SVM classifier achieved 99.49%. • For LFW dataset: Accuracy of SVM classifier achieved 100%.
2	Snyder and Husari [49]	CNN	COCO, WIDER FACE, CelebA, CMCD	• F₁-score=87.7% • Recall=99.2%.
3	Mandal et al. [34]	ResNet-50	RMFRD	• The model has achieved the accuracy of mask and unmask, 44.73% and 88.70%, respectively.
4	Thareja et al. [52]	ResNet-50 and transfer	Kaggle dataset	• The proposed model achieved 99.70% accuracy.

		learning techniques		
5	Deng et al. [13]	Learning large margin cosine loss	MFR2, VGGFace2_m, RMFD	• The model achieved 98.54% detection accuracy.
6	Baalaji et al. [56]	ResNet-50 using single-shot multi-box detector	Collected from RMFD dataset and Kaggle dataset	• Trained classifier accuracy=99.74%. • Identity retrieval model accuracy=98.24%.
7	Naseri et al. [39]	ResNet-50, ASMFO-HResMobileNet	Three Datasets (Live create dataset ¹ , Dataset ² , and Dataset ³)	• ASMFO-HResMobileNet accuracy exceeded 17.56%, 18.57%, 19.20% than VGG16-LSTM, SVM, ResNet50-MobileNetv2

Table 2: ResNet techniques discussed in existing work

Applicability in Hospital Environment: ResNet-50's high accuracy levels make it well-suited for hospital areas like ICUs where false negatives (missed mask violations) can lead to significant clinical risk [39].

3.2 CNN

The CNN-based facemask identification system uses a dataset comprising both masked and unmasked face images of distinct individuals. Rajasekhar et al. [43] discussed facemask detection using active learning via CNN techniques. Kumar et al. [27] proposed a hybrid DL, ML, and image processing approach to spot facemasks in a live video broadcast. The model was tested using open and brand-new (fresh) datasets. The outcomes showed that the CNN classifier performs better than the DNN classifier; it has an error rate of only 1.12% and almost perfect face-identification abilities in facemask-wearers. Sambhare et al. [46] developed a binary face classifier unaffected by the faces' alignment in an image. The authors describe how to create realistic face segmentation masks from any sized image. This method extracts features using the VGG-16 architecture with predefined training weights, starting from an RGB image of any size. Fully CNN (FCNN) is utilized in the training procedure to differentiate the faces in this image semantically applying Gradient Descent as the training function and Binomial Cross-Entropy (BCE) is the loss function. Kaur et al. [24] discussed a MobileNet-based model for facemask image classification. Additionally, it can recognize movement of a human face wearing a mask in a video. This approach has excellent accuracy of 86%.

Golwalkar and Mehendale [16] used the FaceMaskNet-21 network to recognize masked faces fast, which was trained with quadruplets and deep metric learning (DML) [16]. Every person's facemask is stored in one dataset, i.e., the live video stream and the input image, and it is encoded with 128 bits. FaceMaskNet-21 translates photos from the dataset (during training) and live video or input

images using 128-bit encodings (during testing). The encoding that results (from the test image) is then contrasted with every other encoding in the dataset. Using the dataset's minimum error decoding, the name of the identified person is selected. Chavda et al. [10] presented a multi-stage CNN Architecture for facemask detection. The authors demonstrated a CNN-based framework for detecting incorrect facemask use cases. This system can work with CCTV cameras and can recognize both masked and unmasked faces using CNN architecture. This has aided in maintaining a safety-focused workplace, encouraging the usage of facemasks, and tracking safety violations. Lin et al. [31] introduced a real-time technique that automatically combines human posture recognition with a CNN to distinguish facemask wearers. Images clipped to the ostensibly facemask-covered region are used to train the CNN. This technology was tested in various scenarios using photographs shot in public settings at multiple times of day and from various angles. This system has a daytime recognition accuracy of 95.8% and a nighttime recognition accuracy of 94.6%. Al Amin et al. [3] discussed facemask detection using CNN techniques. CNN identified the people wear facemasks or not, where the author used a public dataset, i.e., 1300 images (without facemask 650 images and with facemask 650 images). Gupta et al. [19] suggested facemask detection based on machine learning techniques using public surveillance cameras dataset to detect whether peoples wear facemasks or not. Zhang et al. [61] gave a "context-attention R-CNN" and architecture for detecting facemasks. This approach enhances intraclass distance while reducing the interclass distance by recognizing distinctive characteristics. Gathani and Shah [15] examined occluded facemask detection or facial expression. The author suggested a CNN-based model that detects masked and unmasked faces using region proposals. There weren't many datasets available for this project and thus a secondary goal was to compile a high-resolution "non-facemask and facemask" dataset for DL architecture. With a detection accuracy of 98.62% for non-facemask detection, the results are significantly superior to the 68.73% of mAP for facemask detection. Umer et al. [54] proposed facemask detection based on CNN and multi-stage image processing. Inamdar and Mehendale [21] discussed a real-time facemask identification method [21]. Rao et al. [44] described a facemask identification system that can be used to identify someone wearing a facemask by separating them from others who weren't. Facemasks were classified using a cascade classifier based on around 1500 pictures, with and without the facemasks. It was fairly precise, with a 91.25% level of accuracy. Meenpal et al. [36] provided instructions for building a binary face classifier that can identify every face in the frame despite alignments. The procedure uses the Predefined Training Weights of the VGG-16 Framework to extract features from an RGB image of any size. Fully CNN is trained to differentiate the faces visible in the photograph semantically. BCE is the loss function, and gradient descent is the training function. The segmented facemasks on the MPH dataset have a mean pixel-level accuracy of 93.88 %. . Table 3 briefly discusses the existing work related to CNN techniques

Sno	Authors & References	Techniques	Datasets	Performance
1	Rajasekhar et al. [43]	CNN	Kaggle dataset	The model achieved accuracy at the range of 72% to 79%.
2	Kumar et al. [27]	CNN	Kaggle and Witkowski's Medical Mask datasets, Real-time dataset	The model has an accuracy rate of 99% and a 1% error rate.
3	Sambhare et al. [46]	VGG-16, CNN	Custom dataset	Facemask accuracy is 98.7%.
4	Kaur et al. [24]	CNN-based approach	Sample dataset (with and without facemasks from Kaggle)	More effective outcomes may be obtained with a lightweight DWS-based CNN and gain accuracy of 86%.
5	Golwalkar and Mehendale [16]	Deep metric learning	Collected dataset	The model has achieved 91.67% accuracy.
6	Chavda et al. [10]	CNN	Custom dataset	The NASNetMobile model performs slightly better than the other models.
7	Lin et al. [31]	CNN	Custom dataset	The system achieves 95.8% and 94.6% accuracy in the daytime and nighttime recognition.
8	Al Amin et al. [3]	CNN	Public dataset	The proposed approached achieved an accuracy of 97.5%.
10	Gupta et al. [19]	CNN	Public surveillance cameras	The proposed system detects live facemasks using surveillance cameras.
11	Zhang et al. [61]	R-CNN	Basic dataset	The model improves the accuracy from 83.7% to 84.1%.
12	Gathani and Shah [15]	Region-based CNN	Custom dataset	- Total mAP detection accuracy is 85.84%. - Non-facemask with a detection accuracy (mAP) is 98.62%, - Facemask with accurate detection accuracy (mAP) is 68.712%.
13	Umer et al. [54]	CNN	RILFD dataset	Accuracy (Facemask)=97.5%.
14	Inamdar and Mehendale [21]	Facemasknet model	RWFCD, MFDD, SMFRD	The model achieved 98.6% accuracy.
15	Rao et al. [44]	DL model	Sample dataset	The DL achieved 91.21% accuracy.
16	Meenpal et al. [36]	VGG-16, Fully CNN	Multi Human Parsing Dataset	The accuracy of segmented facemasks is 93.884 %.

Insert Table 3: CNN techniques discussed in existing work

Applicability in Hospital Environment: The ease of integration of CNN architectures with CCTV make it perfectly suited for hospital environment, as they can be integrated with minimal hardware investment. Thus the key advantage of CNN architecture is its adaptability.

3.3 MobileNetV2

MobileNetV2 enables faster and precise detection of facemask among a bunch of people. As input, images of people with and without facemasks are provided. As output, the same image is segmented into multiple images. Hussain et al. [20] proposed a multi sensor Smart Screening and Disinfection Walkthrough Gate (SSDWG) for public safety based on the Internet of Things (IoT). The SSDWG is designed for rapid screening by measuring temperature with a contact-free sensor and storing the suspected individual record for subsequent management and monitoring. Facemask classification and detection are performed using real-time deep learning models [48] in the proposed IoT-based surveillance system. Shamrat et al. [47] investigated the detection performance of three DNN approaches for facemask recognition: Average pooling, Max pooling, and MobileNetV2 framework. A dataset of 1845-images from various sources as well as 120-images of the co-authors taken with a web camera and a mobile camera was used to train a DNN framework. The accuracy rate was 98.690%, while the accuracy score was 96.490% using max pooling. Additionally, the average pooling accuracy for training and validation was 95.290% and 96.240%, respectively. The accuracy of training and testing using the MobileNetV2 framework is 99.74% and 99.840%, respectively. Chen et al. [11] described a lightweight Face Mask Screening Device (FSSD) using a mobile phone. The authors presented a cell phone-based detection method. The facemask micro photos were examined using the Grey Level Co-occurrence Matrixes (GLCMs) to examine the spatial orientation of the pixels in the image using texture analysis. Nagrath et al. [37] developed SSDMNV2, a method for detecting facemasks using the Single Shot Multibox Detector as the face detector and the MobileNetV2 framework as the classifier's framework in the SSDMNV2 (Figure 4). This works for both images and real-time datasets. It is reasonably portable and even allows for real-time facemask detection in embedded systems (such as the Raspberry Pi and NVIDIA Jetson Nano). This model generated good accuracy score and f_1 -score i.e., 0.9264 and 0.93, respectively.

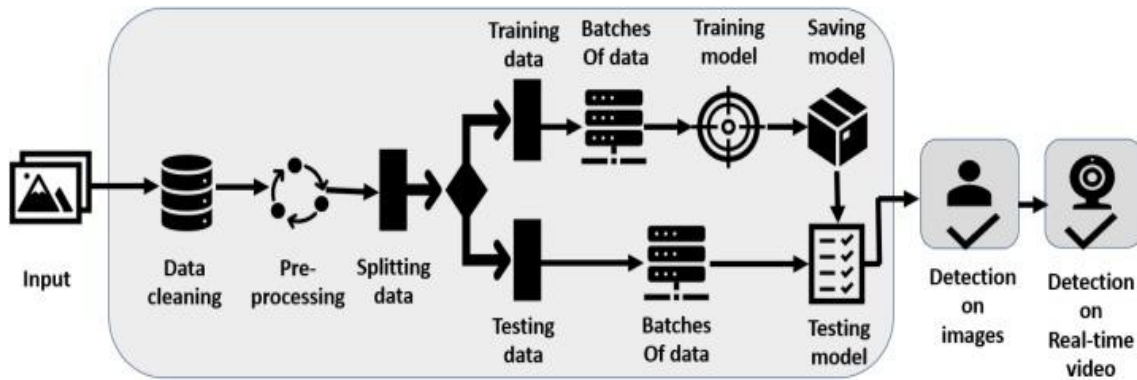


Figure 4: SSDMNv2 model for facemask detection

Addagarla et al. [1] provided several ways for multi-scale facemask categorization and real-time identification for the medical business. They created two separate detection models using two different facemask datasets of 1400 and 680 photos. The mean average precision (mAP) was 34%, and the recall rate was 91.7% on the FMV3 model, while the accuracy and recall rates were 98% and 99%, respectively. Chen et al. [12] created a technique for determining the service stage of a facemask using a mobile microscope. The proposed methodology extracts four attributes from micro-photos of facemasks using grey level co-occurrence matrices: correlation, contrast, energy, and homogeneity (GLCMs). The mask's service stage is then determined using the k-NN approach. According to the empirical results, the proposed model detects facemask with 98.6% accuracy. Table 4 briefly discusses the existing work related to MobileNet2 techniques

Sno	Authors & References	Techniques	Datasets	Performance
1	Hussain et al. [20]	SSDWG	Masked Face-Net, MAFA	• Accuracy (VGG-16) = 99.81% • Accuracy (MobileNetV2) = 99.6%
2	Shamrat et al. [47]	CNN	RMFD, SMFD, Custom dataset	• Accuracy for MobileNetV2 architecture: • 99.72% for training • 99.82% for validation
3	Chen et al. [11]	k-NN	Real-Time Image	• The system (testing dataset) has an accuracy of 82.87%. • Precision (Type-I) = 92.01% that is not recommended. • Recall (Type-III) = 92.58% that is not recommended.
4	Nagrath et al. [37]	MobileNet-V2	Kaggle's Medical Mask	• The accuracy of deploy technique is 0.9264%. • F1- score is 0.93%.

6	Addagarla et al. [1]	Real-Time DLT model	The models were trained using 1400 and 680 images	• The DTL model achieved 98% accuracy.
7	Chen et al. [12]	Noisy SoftMax	Facemask micro-photos	• The model accuracy is 82.87%

Table 4: MobileNetV2 techniques for facemask detection, i.e., used in existing work

Applicability in Hospital Environment: MobileNetV2 is arguably the most hospital-practical architecture because of its low computational power and its integration with IoT temperature sensors [20].

3.4 You Only Look Once (YOLO)

One of the quickest object detection techniques, “You only look once” (YOLO) gained widespread adoption due to its effectiveness and performance. The algorithms used to this point identify items in the image by selecting specific regions and then classifying the objects within them [6], [28]. This has been an extra overhead because each region selection must be done separately for each object in the image. The YOLO approach eliminates this stage by analyzing the entire image once. For each object in the image, the bounding boxes and class probabilities are predicted in a single forward pass. Thus in comparison to previous object detection algorithms, the YOLO model is significantly faster and more effective. Table 5 briefly discusses the existing work related to YOLO techniques.

Yolo- V_1 : This is the first generation of YOLO, commonly referred to as YOLO-v1. YOLO-v1 splits the image into an $S \times S$ grid for object detection. However, due to the large localization error in this method and a relatively low recall rate, it is challenging to locate close objects [28], [29]. The 24 convolutional layers of YOLO-v1 are modified for detection by adding 4 CNN layers and two fully connected layers. A fully connected layer is used by YOLOv1 to regress bounding boxes, which imposes significant spatial constraints on their predictions and makes it difficult to combine small items.

Yolo- V_2 : YOLOv2 was developed to address the flaws of the previous version, specifically to reduce localization error and increase recall by leveraging high-resolution classifiers, batch normalization, and convolutional layers with anchor boxes [28], [29]. Wu et al. [59] proposed the FMD-Yolo model to detect whether people are wearing facemasks. Im-Res2Net-101, a deep residual network, and deep residual network Res2Net module hybrid is used by the feature extractor.

Yolo- V_3 : The network predicts four coordinate values for each bounding box. During the training phase, predictions are made using binary cross-entropy loss. Bhuiyan et al. [29] measured performance with a strong GPU, but the “YOLO-v3 darknet” command is suitable for real-time

computing with minimum memory. The author discovered that pictures of people with and without facemasks are utilized for training people. After processing real-time video data, the experimental setup performed detection, localization, and identification. In the following 4000 epochs of training, the overall average is 0.0730. The mAP score after 4000 epochs of training is 0.96. Vinh and Anh [57] discussed a real-time facemask detector. It is used the YOLO-v3 algorithm and the HAAR cascade classifier to distinguish between the face and the mask. Jiang et al. [22] suggested using the YOLO-v3-slim method, an enhanced version of the original YOLO-v3 technique, for facemask recognition. Al-Tameemi et al. [4] introduced the Bi-path Combination You Only Look Once (BC-YOLO) for FSOD. The author categorized into transfer learning and training scheme. Kristo et al. [26] investigated the issue of human detection in IR thermal images using the model of YOLOv3, Faster R-CNN, and many more. Here, author conclude that IR thermal images are more effective than VIS images for detecting objects at night in inclement (bad) weather.

Yolo-V₄: This model's accuracy has increased than the previous versions but the prediction time has remained the same [62]. In comparison to the third generation YOLO model, the average precision is 10% higher [29]. Yu and Zhang [61] used YOLO-v4 for facemask identification demonstrating its effectiveness in the fields of computer vision and image processing (IP). This is extremely light and might be used in embedded systems to provide real-time facemask detection.

Yolo-V₅: YOLO-V₅ brought about the addition of mosaic data and auto learning bounding box anchors for significant enhancement of speed and accuracy. Zhou et al. [63] demonstrated the strong performance of YOLOv5 over previous versions on facemask detection datasets and found it to be more accurate with an increase in mean average precision (mAP) of 14.97% and 4.75%, respectively, as compared to YOLOv3 and YOLOv4.

Sno	Authors & References	Techniques	Datasets	Performance
1	Bhuiyan et al. [6]	YoLov3	Custom dataset	• After training 4000 epochs: • The mAP score is 0.96. • The average loss is 0.0730.
2	Wu et al. [48]	FMD Yolo	Public dataset	• AP50 at IoU=0.50, Precision is 88.4% and 92%, respectively. • AP75 at IoU=0.75, Improved the performance by 3.9% and 5.5%, respectively.
3	Yu and Zhang [61]	YOLO-v4 and CSPDarkNet53	Masked Face-Net and RMFD	• According to the comparisons, facemask recognition has a mAP of 98.3% and a frame rate of 54.57 fps.
4	Al-Tameemi et al. [4]	YOLO	PASCAL VOC 2007, MS COCO 2014 datasets	• The model accuracy's 82%.

Table 5: YOLO techniques discussed in existing work

Applicability in Hospital Environment: YOLOv5 is the recommended detection model for high-priority hospital zones (ICU, operating theatres, and emergency departments) where detection accuracy is critical. Its architecture provide better handling of partial occlusion of masks, for example, a mask partially covered by surgical loops, a common scenario in clinical environments.

3.5 VGG-19

The Visual Geometry Group (VGG) is made up of 2D Convolution and Max Pooling layers. Kim et al. [25] introduced the edge map and IPN (EIPNet) for face super-resolution networks. It reduced the distortion by utilizing lightweight identity information and edge block. Xiao et al [60] developed the upgraded CNN VGG-19 method to identify workers without facemasks in workplaces. They collected over 3000-photos for model training and testing. Using the VGG-19 network model, three FC layers are replaced with a single plain layer and two FC layers with fewer features. The SoftMax classification sheet of the original model has been replaced by a 2-label SoftMax classifier. The experimental data show that the model has a recall of 96.31% and an accuracy of 97.62% of identifying employees without masks. Carvalho et al. [8] studied the inconsistencies in the eye region where graphic design simulation still has to be enhanced using highly realistic computer-generated images. These differences are identified by investigating the expressive power of features obtained from the VGG-19 DNN model and a transfer learning (TL) strategy. Two distinct datasets with highly realistic photos were used in the experiments, and the results showed accuracy of 0.80 and AUC of 0.88, respectively. Diviya et al. [14] discussed the VGG-19 and YOLO. It is helpful for to detect people wear facemask or not. Marjan et al. [35] described the facemask detection techniques through extended VGG-19 that performed better than the CNN and VGG-19 variants. Table 6 briefly discuss existing research papers related VGG-19.

Sno	Authors & References	Techniques	Datasets	Performance
1	Kim et al. [25]	EIPNet	VGGFace2 and elebA	• It outperforms preliminary approaches in terms of FPS while yielding promising quantitative findings. • It improves the perceived quality of SR images.
2	Xiao et al. [60]	VGG-19	3000 images	• The model achieved accuracy is 96.31%.
3	Carvalho et al. [8]	VGG-19 and CNN	DSCG1 and DSCG2	• Accuracy=80% • AUC=88%
4	Diviya et al. [14]	YOLO and VGG-19	Image dataset	• Accuracy (VGG-19) = 99.80%

5	Marjan et al. [35]	Extended VGG-19	LFW and CASIA	• Accuracy=96%
---	--------------------	-----------------	---------------	----------------

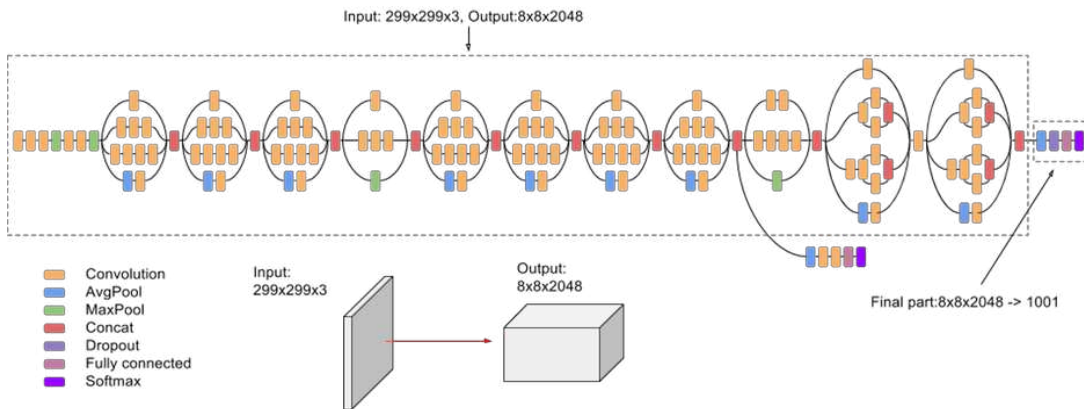
Table 6: VGG-19 techniques discussed in existing work

Applicability in Hospital Environment: VGG-19 is developed for highly secured areas where hospital deployments with centralised computing is available, for example, server-based monitoring of operating theatre entrances or secured entrances of pharmaceutical areas where computational resources are available and vigilance of false negative rates must be determined.

3.6 Inception-v3

Inception-V3 model is a popular image recognition model which is architected using various inception blocks containing diverse layer combinations: convolution, concatenation, dropout, AvgPool, MaxPool, fully connected, and SoftMax output.(Figure 5)

Figure 5: The architecture of Inception-V3 Model



Li [30] talked about facemask detection through Inception-V3. These methods detect three types of classes without facemasks, with facemasks, and masks worn incorrectly. Nainan et al. [38] explained the real time facemask detection based on two techniques like Inception-v3 and MobileNetV2. Agarwal and Susan [2] described the Emotion Recognition from facemasks using Inception-v3 techniques. Nyarko et al. [41] discussed the hybrid techniques analysis of facemask recognition. These techniques identified many types of masks, such as N95, fabric, and surgical masks. Chowdary et al. [23] talked about Google-Net. It is an alternative CNN model built on the Inception framework. It uses “Inception modules” which allow the network to analyze visuals at different scales within each frame. These modules are stacked by Google-Net using a feed-forward framework with sporadic max-pooling layers to minimize the dimensionality of the data. Tab17e briefly discusses the existing work related to InceptionV3 techniques.

Sno	Authors & References	Techniques	Datasets	Performance
-----	----------------------	------------	----------	-------------

1	Li [30]	Inception-v3	Public area dataset	• Accuracy (Without mask) = 97.11% • Accuracy (With mask) = 94.52% • Accuracy (Wear mask incorrect) = 94.26%
2	Nainan et al. [38]	CNN model	Real-time dataset	• Accuracy (With mask) = 97% • Accuracy (Without mask) = 98%
3	Agarwal and Susan [2]	Inception-v3	FER-2013 dataset	• Accuracy (With mask) = 54.82% • Accuracy (without mask) = 68.80%
4	Nyarko et al. [41]	Inception-V3, Resnet-50, and AlexNet	Statistical dataset	• Accuracy (AlexNet)= 95.7%, • Accuracy (Resnet-50) = 97.5% • Accuracy (Inception-V3) = 95.5%
5	Chowdary et al. [23]	CNN Model	Kaggle Dataset	• The model achieved 96% detection accuracy.

Table 7: InceptionV3 techniques discussed in existing work

Applicability in Hospital Environment: Inception-V3's capability of determining multi-type mask and its ability to distinguish between different types of masks makes it uniquely valuable for hospital areas requiring compliance of specific mask types (e.g., ensuring N95 masks are used in ICUs instead of surgical masks).

3.7 Hybrid Techniques

Hybrid techniques incorporates two more techniques (including MobileNet-V2, VGG-19, DenseNet-121, Inception-V3, and numerous others) to recognize people who are wearing a facemask. Table 8 briefly discuss the Hybrid techniques for facemask detection. Goyal et al. [17] introduced a facemask identification algorithm for static images and real-time videos, categorizing the images as “without a mask” or “with a mask”. The acquired data set consists of around 4,000 images and has a 98% performance accuracy rate. When compared to MobileNet-V2, VGG-19, DenseNet-121, and Inception-V3, the newly proposed model outperforms them in terms of effectiveness and accuracy. Tamang et al. [50] described YOLOv5 and DL techniques for facemask detection that may be used to test hybrid dataset (without facemask, improper/partial mask, and with facemask). In the hybrid dataset used, there are 1000 images of each class which means a total of 3000 images. Loey et al. [33] developed a special paradigm for medical facemask identification that focuses on the medical-mask. To get incredibly good image detection results, the author used the ResNet-50 model, which is based on YOLO v2, i.e., represented in Figure 6. The developed model improves detection performance by using mean IoU to calculate the ideal number of anchor boxes. The author combines two publicly accessible facemask datasets to generate a brand-new dataset used to train and test

detectors in a supervised setting. The author demonstrated the developed framework strategy of YOLOv2 with ResNet-50 to be a successful model for detecting a medically masked face.

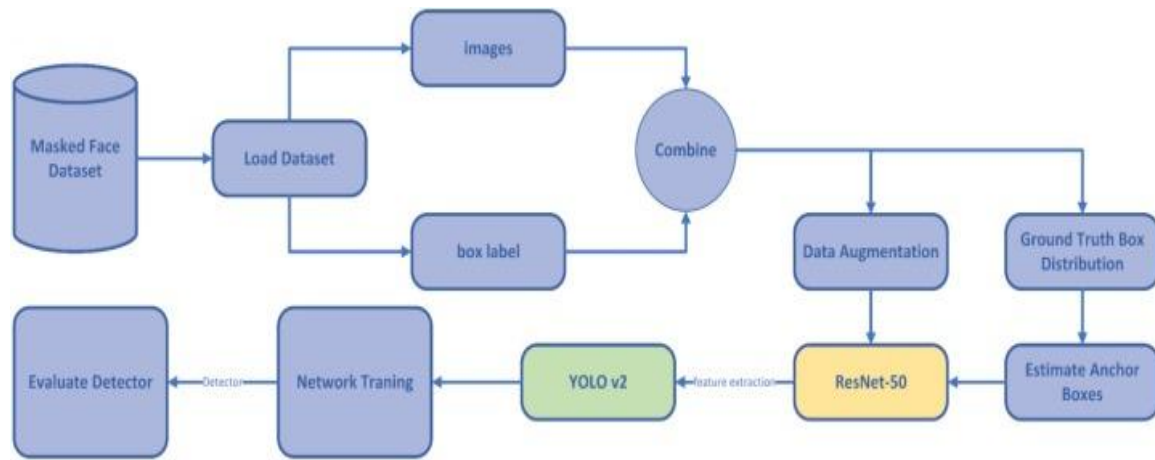


Figure 6: Facemask detector model used to YOLOv2 with ResNet-50

Sno	Authors & References	Techniques	Datasets	Performance
1	Goyal et al. [17]	CNN	Kaggle dataset	• Around 98.2% of the time, the trained model is accurate.
2	Tamang et al. [50]	YOLOv5 and DL	hybrid datasets (3000-images)	• Accuracy = 90%.
4	Loey et al. [72]	YOLOv2 with ResNet-50	FMD, MMD dataset	• Accuracy (precision) = 81%.

Table 8: Hybrid techniques discussed in existing work

Table 9 depicts the hospital suitability assessment of each reviewed facemask detection technique, providing recommendations of their deployment in hospital zone.

Architecture	Best Accuracy	Inference Speed	Multi-Mask Type	Recommended Hospital Zone
MobileNetV2	99.63 %	Real-time	Partial	Ward entrances, OPD,

				Reception
ResNet-50 (Hybrid)	100% (LFW)	Near real-time	Partial	ICU, Isolation Wards
VGG-16/19	99.85 %	Moderate	Partial	OT, High-security zones
CNN (Multi-stage)	98.62 %	Real-time	Partial	Wards, Corridors, ED
Inception-V3	~96–98%	Moderate	Yes (N95/Surgical)	ICU, Isolation, OPD
YOLOv5	~97–98%	Fast	Yes (3-class)	ED, OPD, General Wards

Table 9: Facemask detection technique for hospital deployment

4. Open issues and research challenges of facemask detection

The deployment of facemask detection in different hospital departments introduces a set of challenges that are not otherwise encountered in general public places. These challenges are listed below:

4.1 Technical Challenges

- **Medical Equipment Occlusion:** Hospital subjects are often required to wear additional medical equipment like oxygen masks, surgical loupes, and head covering layers that can affect mask detection. The facemask detection models are still not trained on these occlusion patterns and thus these models require fine-tuning on hospital-specific datasets.
- **Variable Clinical Lighting:** Hospital subjects are sometimes exposed to bright operating theatres with surgical lights and sometimes they have to stay in dim ICU rooms. Thus the facemask detection models in hospitals should maintain sufficient accuracy and performance across this luminance range.

- **Correct vs. Incorrect Mask Wear:** It is very important in hospital environment that the facemask detection technique mask worn correctly (covering nose and mouth) and incorrectly (below nose, at chin, on forehead) is clinically critical — particularly for HCWs in high-risk zones. This requires fine-grained detection models that assess mask positioning relative to facial landmarks, a more demanding task than binary presence/absence detection.

4.2 Operational and Ethical Challenges

- **Patient Privacy:** Hospital-based detection architectures need to comply with healthcare data protection agencies that impose strict regulations on analysing the biometric data collected from patients and staff. Thus facemask detection systems that perform only facial recognition (detection-only, without individual identity linkage) can be conducted with existing regulatory frameworks, but detection systems that timestamp and zone-identify individuals require stringent privacy impact assessment.
- **Clinical Dataset Scarcity:** There has been no publicly available dataset that has been collected in a hospital environment with patients. Thus there is a critical need of developing a robust training datasets in hospital environment.

6. Conclusion and Future Directions

This study presents a systematic review of AI-based facemask detection techniques evaluated specifically in the context of hospital environments, addressing the needs of both healthcare workers (HCWs) and patients. Healthcare-associated infections remain a profound global burden, and facemask compliance is one of the most evidence-based interventions for their reduction.

The review demonstrates that significant progress has been made in AI facemask detection, with the latest architectures particularly YOLOv5 [58a], [9] and hybrid MobileNetV2-based systems [20], [37] achieving precision and recall rates exceeding 97% in real-time configurations compatible with embedded clinical deployment. MobileNetV2-based systems, by contrast, offer the most favourable accuracy-to-deployment-cost ratio for distributed hospital-wide monitoring across ward entrances, reception areas, and outpatient facilities.

Critical gaps identified include the absence of hospital-specific annotated training datasets, and the challenge of detecting correct versus incorrect mask usage under the complex occlusion conditions of clinical environments. Future research should prioritise the development of ethically cleared, institution-specific hospital datasets; the integration of mask detection with complementary clinical sensor data (temperature, personnel ID); and the evaluation of AI mask detection systems in live clinical environments through randomised controlled trials measuring HAI incidence as a primary outcome.

As hospitals invest in AI-driven clinical intelligence platforms, facemask detection represents an immediately deployable, evidence-based application of computer vision technology with a measurable impact on patient safety. The integration of AI mask monitoring into hospital infection control protocols offers the prospect of continuous, objective, and scalable enforcement of PPE compliance — a capability that could meaningfully reduce the global burden of healthcare-associated infections.

Declaration

Conflict of Interest

All authors declare that they have no conflict of interest.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sector.

Data Availability

All relevant data are included in the paper. Other data will be made available on request.

Ethical approval

Not applicable

Consent to Participate

Not applicable

Consent to Publish

Not applicable

References

- [1] Addagarla, S. K., Chakravarthi, G. K., & Anitha, P. (2020). Real time multi-scale facial mask detection and classification using deep transfer learning techniques. *International Journal*, 9(4), 4402-4408.
- [2] Agarwal, A., & Susan, S. (2023, March). Emotion Recognition from Masked Faces using Inception-v3. In *2023 5th International Conference on Recent Advances in Information Technology* (pp. 1-6). IEEE.
- [3] Al Amin, I. H., Marinda, D. E., Winarno, E., UN, D. H., & Lusiana, V. (2023). Real-Time Detection of Face Mask Using Convolutional Neural Network. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 7(3), 697-704.

- [4] Al-Tameemi, M. I., Hasan, A. A., & Oleiwi, B. K. (2023). Design and implementation monitoring robotic system based on you only look once model using deep learning technique. *IAES International Journal of Artificial Intelligence*, 12(1), 106.
- [5] Bhambani, K., Jain, T., & Sultanpure, K. A. (2020, October). Real-time face mask and social distancing violation detection system using yolo. In *2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC)* (pp. 1-6). IEEE.
- [6] Bhuiyan, M. R., Khushbu, S. A., & Islam, M. S. (2020, July). A deep learning based assistive system to classify COVID-19 face mask for human safety with YOLOv3. In *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-5). IEEE.
- [7] Cao, Z., Shao, M., Xu, L., Mu, S., & Qu, H. (2020). MaskHunter: real-time object detection of face masks during the COVID-19 pandemic. *IET Image Processing*, 14(16), 4359-4367.
- [8] Carvalho, T., De Rezende, E. R., Alves, M. T., Balieiro, F. K., & Sovat, R. B. (2017, December). Exposing computer generated images by eye's region classification via transfer learning of VGG19 CNN. In *2017 16th IEEE international conference on machine learning and applications (ICMLA)* (pp. 866-870). IEEE.
- [9] CDC report, 2025. <https://www.cdc.gov/infection-control/hcp/viral-respiratory-prevention/index.html>
- [10] Chavda, A., Dsouza, J., Badgujar, S., & Damani, A. (2021, April). Multi-stage cnn architecture for face mask detection. In *2021 6th International Conference for Convergence in Technology (I2CT)* (pp. 1-8). IEEE.
- [11] Chen, Y., Hu, M., Hua, C., Zhai, G., Zhang, J., Li, Q., & Yang, S. X. (2021). Face mask assistant: Detection of face mask service stage based on mobile phone. *IEEE Sensors Journal*, 21(9), 11084-11093.
- [12] Chen, B., Deng, W., & Du, J. (2017). Noisy softmax: Improving the generalization ability of dcnn via postponing the early softmax saturation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5372-5381).
- [13] Deng, H., Feng, Z., Qian, G., Lv, X., Li, H., & Li, G. (2021). MFCosface: a masked-face recognition algorithm based on large margin cosine loss. *Applied sciences*, 11(16), 7310.

- [14] Diviya, M., Hemapriya, N., Charulatha, B. S., & Subramanian, M. (2023). An intuitive and structured detection system for facial mask using YOLO and VGG-19 to limit COVID-19. In *Recent Advances in Material, Manufacturing, and Machine Learning* (pp. 485-492). CRC Press.
- [15] Gathani, J., & Shah, K. (2020, November). Detecting masked faces using region-based convolutional neural network. In *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)* (pp. 156-161). IEEE.
- [16] Golwalkar, R., & Mehendale, N. (2022). Masked-face recognition using deep metric learning and FaceMaskNet-21. *Applied Intelligence*, 1-12.
- [17] Goyal, H., Sidana, K., Singh, C., Jain, A., & Jindal, S. (2022). A real time face mask detection system using convolutional neural network. *Multimedia Tools and Applications*, 1-17.
- [18] Gupta, C., & Gill, N. S. (2020). Coronamask: a face mask detector for real-time data. *Int. J. Adv. Trends Comput. Sci. Eng.*
- [19] Gupta, S., Sreenivasu, S. V. N., Chouhan, K., Shrivastava, A., Sahu, B., & Potdar, R. M. (2023). Novel face mask detection technique using machine learning to control COVID'19 pandemic. *Materials Today: Proceedings*, 80, 3714-3718.
- [20] Hussain, S., Yu, Y., Ayoub, M., Khan, A., Rehman, R., Wahid, J. A., & Hou, W. (2021). IoT and deep learning-based approach for rapid screening and face mask detection for infection spread control of COVID-19. *Applied Sciences*, 11(8), 3495.
- [21] Inamdar, M., & Mehendale, N. (2020). Real-time face mask identification using facemasknet deep learning network. Available at SSRN 3663305.
- [22] Jiang, X., Xiang, F., Lv, M., Wang, W., Zhang, Z., & Yu, Y. (2021, February). YOLOv3_slim for face mask recognition. In *Journal of Physics: Conference Series* (Vol. 1771, No. 1, p. 012002). IOP Publishing.
- [23] Jignesh Chowdary, G., Pun, N. S., Sonbhadra, S. K., & Agarwal, S. (2020, December). Face mask detection using transfer learning of inceptionv3. In *International Conference on Big Data Analytics* (pp. 81-90). Springer, Cham.
- [24] Kaur, G., Sinha, R., Tiwari, P. K., Yadav, S. K., Pandey, P., Raj, R. & Rakhra, M. (2021). Face mask recognition system using CNN model. *Neuroscience Informatics*, 100035.
- [25] Kim, J., Li, G., Yun, I., Jung, C., & Kim, J. (2021). Edge and identity preserving network for face super-resolution. *Neurocomputing*, 446, 11-22.

- [26] Krišto, M., Ivasic-Kos, M., & Pobar, M. (2020). Thermal object detection in difficult weather conditions using YOLO. *IEEE access*, 8, 125459-125476.
- [27] Kumar, T. A., Rajmohan, R., Pavithra, M., Ajagbe, S. A., Hodhod, R., & Gaber, T. (2022). Automatic face mask detection system in public transportation in smart cities using IoT and deep learning. *Electronics*, 11(6), 904.
- [28] Kumar, A. (2022). A cascaded deep-learning-based model for face mask detection. *Data Technologies and Applications*, (ahead-of-print), 1-24.
- [29] Kumar, A., Kalia, A., Verma, K., Sharma, A., & Kaushal, M. (2021). Scaling up face masks detection with YOLO on a novel dataset. *Optik*, 239, 166744.
- [30] Li, Y. (2021, September). Facemask detection using inception V3 model and effect on accuracy of data preprocessing methods. In *Journal of Physics: Conference Series* (Vol. 2010, No. 1, p. 012052). IOP Publishing.
- [31] Lin, H., Tse, R., Tang, S. K., Chen, Y., Ke, W., & Pau, G. (2021, January). Near-realtime face mask wearing recognition based on deep learning. In *2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)* (pp. 1-7). IEEE.
- [32] Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2021). A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. *Measurement*, 167, 108288.
- [33] Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2021). Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. *Sustainable cities and society*, 65, 102600.
- [34] Mandal, B., Okeukwu, A., & Theis, Y. (2021). Masked face recognition using resnet-50. *arXiv preprint arXiv:2104.08997*.
- [35] Marjan, M. A., Hasan, M., Islam, M. Z., Uddin, M. P., & Afjal, M. I. (2022, December). Masked Face Recognition System using Extended VGG-19. In *2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)* (pp. 1-4). IEEE.
- [36] Meenpal, T., Balakrishnan, A., & Verma, A. (2019, October). Facial mask detection using semantic segmentation. In *2019 4th International Conference on Computing, Communications and Security (ICCCS)* (pp. 1-5). IEEE.

- [37] Nagrath, P., Jain, R., Madan, A., Arora, R., Kataria, P., & Hemanth, J. (2021). SSDMNv2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. *Sustainable cities and society*, 66, 102692.
- [38] Nainan, N. A., Jeevika, H., Jalan, R., Karthik, V. C., & Shankar, S. P. (2021, October). Real Time Face Mask Detection Using MobileNetV2 and InceptionV3 Models. In *2021 IEEE Mysore Sub Section International Conference (MysuruCon)* (pp. 341-345). IEEE.
- [39] Naseri, R. A. S., Kurnaz, A., & Farhan, H. M. (2023). Optimized face detector-based intelligent face mask detection model in IoT using deep learning approach. *Applied Soft Computing*, 134, 109933.
- [40] Nowrin, A., Afroz, S., Rahman, M. S., Mahmud, I., & Cho, Y. Z. (2021). Comprehensive review on facemask detection techniques in the context of covid-19. *IEEE access*.
- [41] Nyarko, B. N. E., Bin, W., Zhou, J., Agordzo, G. K., Odoom, J., & Koukoyi, E. (2022, June). Comparative analysis of Alexnet, resnet-50, and inception-V3 models on masked face recognition. In *2022 IEEE World AI IoT Congress (AIIoT)* (pp. 337-343). IEEE.
- [42] Oumina, A., El Makhfi, N., & Hamdi, M. (2020, December). Control the covid-19 pandemic: Face mask detection using transfer learning. In *2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS)* (pp. 1-5). IEEE.
- [43] Rajasekhar, A. H. N., Kavya, G., Sriya, K., & Belwal, M. (2023, June). Face Mask Detection Using CNN via Active Learning. In *2023 3rd International Conference on Intelligent Technologies (CONIT)* (pp. 1-7). IEEE.
- [44] Rao, T. S., Devi, S. A., Dileep, P., & Ram, M. S. (2020). A Novel Approach to Detect Face Mask to Control Covid Using Deep Learning. *European Journal of Molecular & Clinical Medicine*, 7(6), 658-668.
- [45] Sachdeva, R. (2020). Face Mask Detection System. Available at SSRN 3755508.
- [46] Sambhare, P. B., Raut, H. W., & Joshi, V. N (2022). Live Face Mask Detection using CNN. *International Research Journal of Modernization in Engineering Technology and Science*, 1852-1857.
- [47] Shamrat, F. J. M., Chakraborty, S., Billah, M. M., Al Jubair, M., Islam, M. S., & Ranjan, R. (2021, June). Face Mask Detection using Convolutional Neural Network (CNN) to reduce the

- spread of COVID-19. In 2021 5th International Conference on Trends in Electronics and Informatics (pp. 1231-1237). IEEE.
- [48] Shoeibi, A., Khodatars, M., Jafari, M., Ghassemi, N., Sadeghi, D., Moridian, P. & Gorriz, J. M. (2024). Automated detection and forecasting of covid-19 using deep learning techniques: A review. *Neurocomputing*, 127317.
- [49] Snyder, S. E., & Husari, G. (2021, March). Thor: A deep learning approach for face mask detection to prevent the covid-19 pandemic. In *SoutheastCon 2021* (pp. 1-8). IEEE.
- [50] Tamang, S., Sen, B., & Pradhan, A. (2023, March). Face Mask Wearing Condition Classification using YOLOv5 trained with Hybrid Dataset. In *2023 4th International Conference on Computing and Communication Systems (I3CS)* (pp. 1-7). IEEE.
- [51] Thakur, N., Bhattacharjee, E., Jain, R., Acharya, B., & Hu, Y. C. (2024). Deep learning-based parking occupancy detection framework using ResNet and VGG-16. *Multimedia Tools and Applications*, 83(1), 1941-1964.
- [52] Thareja, R., Thareja, R., & Singh, S. (2022). Face mask detection system using resnet-50. *International Journal of Applied Engineering Research*, 17(6), 588-592.
- [53] Truong, T., Lalseta, D., Ittyipe, R., & Yanushkevich, S. (2020, December). Detecting proper mask usage with soft attention. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1745-1750). IEEE.
- [54] Umer, M., Sadiq, S., Alhebshi, R. M., Alsubai, S., Al Hejaili, A., Nappi, M., & Ashraf, I. (2023). Face mask detection using deep convolutional neural network and multi-stage image processing. *Image and Vision Computing*, 133, 104657.
- [55] United Nations, "Goal 3: Ensure healthy lives and promote well-being for all at all ages," Sustainable Development Goals Knowledge Platform, United Nations, 2025. [Online]. Available: <https://www.un.org/sustainabledevelopment/health>
- [56] Vignesh Baalaji, S., Sandhya, S., Sajidha, S. A., Nisha, V. M., Vimalapriya, M. D., & Tyagi, A. K. (2023). Autonomous face mask detection using single shot multibox detector, and ResNet-50 with identity retrieval through face matching using deep siamese neural network. *Journal of Ambient Intelligence and Humanized Computing*, 1-11.

- [57] Vinh, T. Q., & Anh, N. T. N. (2020, November). Real-time face mask detector using YOLOv3 algorithm and Haar cascade classifier. In 2020 International Conference on Advanced Computing and Applications (ACOMP) (pp. 146-149). IEEE.
- [58] World Health Organization, "SDG Target 3.3: End the epidemics of AIDS, tuberculosis, malaria and neglected tropical diseases and combat hepatitis, water-borne diseases and other communicable diseases," Global Health Observatory, WHO, 2024. [Online]. Available: https://www.who.int/data/gho/data/themes/topics/sdg-target-3_3-communicable-diseases
- [58a] WHO report, 2022. <https://www.who.int/news/item/06-05-2022-who-launches-first-ever-global-report-on-infection-prevention-and-control>
- [59] Wu, P., Li, H., Zeng, N., & Li, F. (2022). FMD-Yolo: An efficient face mask detection method for COVID-19 prevention and control in public. *Image and vision computing*, 117, 104341.
- [60] Xiao, J., Wang, J., Cao, S., & Li, B. (2020, April). Application of a novel and improved VGG-19 network in the detection of workers wearing masks. In *Journal of Physics: Conference Series* (Vol. 1518, No. 1, p. 012041). IOP Publishing.
- [61] Yu, J., & Zhang, W. (2021). Face mask wearing detection algorithm based on improved YOLO-v4. *Sensors*, 21(9), 3263.
- [62] Zhang, J., Han, F., Chun, Y., & Chen, W. (2021). A novel detection framework about conditions of wearing face mask for helping control the spread of covid-19. *Ieee Access*, 9, 42975-42984.
- [63] Zhou, F., Zhao, H., & Nie, Z. (2021, January). Safety helmet detection based on YOLOv5. In 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA) (pp. 6-11). IEEE.