

# Face Mask Detection in Smart Cities Using Deep and Transfer Learning: Lessons Learned From the COVID-19 Pandemic

Yassine Himeur<sup>1,2,\*</sup>, Somaya Al-Maadeed<sup>1</sup>, Iraklis Varlamis<sup>3</sup>, Noor Al-Maadeed<sup>1</sup>, Khalid Abualsaud<sup>1</sup> and Amr Mohamed<sup>1</sup>

<sup>1</sup> Department of Computer Science and Engineering, Qatar University, Doha P.O. Box 2713, Qatar

<sup>2</sup> College of Engineering and IT, University of Dubai, Dubai 4343, United Arab Emirates

<sup>3</sup> Department of Informatics and Telematics, Harokopio University of Athens, Omirou 9, Tavros, 17778 Athens, Greece

\* Correspondence: yhimeur@ud.ac.ae

**Abstract:** After different consecutive waves, the pandemic phase of Coronavirus disease 2019 does not look to be ending soon for most countries across the world. To slow the spread of the COVID-19 virus, several measures have been adopted since the start of the outbreak, including wearing face masks and maintaining social distancing. Ensuring safety in public areas of smart cities requires modern technologies, such as deep learning and deep transfer learning, and computer vision for automatic face mask detection and accurate control of whether people wear masks correctly. This paper reviews the progress in face mask detection research, emphasizing deep learning and deep transfer learning techniques. Existing face mask detection datasets are first described and discussed before presenting recent advances to all the related processing stages using a well-defined taxonomy, the nature of object detectors and Convolutional Neural Network architectures employed and their complexity, and the different deep learning techniques that have been applied so far. Moving on, benchmarking results are summarized, and discussions regarding the limitations of datasets and methodologies are provided. Last but not least, future research directions are discussed in detail.



**Citation:** Himeur, Y.; Al-Maadeed, S.; Varlamis, I.; Al-Maadeed, N.; Abualsaud, K.; Mohamed, A. Face Mask Detection in Smart Cities Using Deep and Transfer Learning: Lessons Learned From the COVID-19 Pandemic. *Systems* **2023**, *11*, 107. <https://doi.org/10.3390/systems11020107>

Academic Editor: Fernando De la Prieta Pintado

Received: 30 December 2022

Revised: 27 January 2023

Accepted: 9 February 2023

Published: 17 February 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** face mask detection; deep learning; deep transfer learning; deep domain adaptation; YOLO; MobileNet

## 1. Introduction

### 1.1. Preliminary

In early 2020, the world health organization (WHO) considered Coronavirus disease (COVID-19) as a spreading epidemic [1]. It has seriously threatened the life of individuals worldwide since its outbreak. Moving on, different variants and mutations of COVID-19 have worsened the situation nowadays [2]. To that end, the focus of the global scientific community has been on proposing adequate mechanisms to handle this calamity, and a significant research effort has been made. This includes developing different kinds of vaccines, introducing timely treatment approaches for severely infected patients, etc. Many studies, e.g., [3–5] show that COVID-19 spreads by droplet and aerosol transmission, mainly occurring during social interactions with infected persons. Thus, pushing individuals to wear protective face masks properly has been established as a pressing and scientific solution to slow down the spread of COVID-19. Typically, face masks help filter and block virus particles in the air [6]. This approach has been recommended by WHO and adopted by most governments across the globe. However, ensuring an efficient adoption of the masks wearing strategy and compliance of people with the requested norms requires developing automatic face mask detection (FMD) techniques. The latter helps check whether individuals are wearing masks or not and whether they are wearing them correctly or not in public areas. In doing so, intelligent control and prevention of the COVID-19 epidemic

are facilitated. This comes in addition to insisting people respect social distancing norms and take routine temperature screenings [3].

FMD and masked face recognition (MFR) has become a key challenge due to the growing numbers of people wearing masks to prevent the infection of the COVID-19 virus [7,8]. However, the lack of efficient technological solutions to detect face masks and non-adherence to COVID-19 policy measures leave the door open for new infection waves [9]. Additionally, the detection of face masks with the use of image processing techniques is challenging due to (i) the diversity and forms of masks, (ii) the camera pixel variation, (iii) the variation of obstruction degrees (e.g., illumination, rotation, resolution, angle of view, etc.), and (iv) the computational complexity of AI-based video analytics. In addition to the above, maintaining people's privacy during the mask detection task is another critical issue [10]. To that end, various studies have been proposed to develop face detection systems using AI-based video analytics, ensuring that only the proper wearing of masks is detected and individual privacy is not exposed [11]. The success of artificial intelligence (AI) and machine learning (ML) models in object detection and face recognition makes this technology suitable for developing face mask detection techniques [12].

Monitoring, managing and preventing the COVID-19 pandemic and other infectious disease outbreaks require path-breaking tools and innovative solutions. To that end, AI and ML have played a crucial role in our fight against the Corona pandemic [13]. For instance, computer vision (CV) has significantly contributed to teaching computers to comprehend and analyze visual scenes thanks to ML and deep learning (DL) models [14]. Specifically, intelligent machines can be deployed for (i) identifying and tracking objects, (ii) measuring the distance between them, and (iii) responding to ascertained scenes using smartphones, cameras, and other sensory devices [15]. In this regard, CV augmented with DL has considerably contributed to monitoring human crowds, capturing human activity, monitoring social distancing behaviors, and detecting the violation of face mask wearing in public, such as Paris Metro system's surveillance cameras [15].

### 1.2. Contributions

This survey aims to shed light on the scientific community's progress since the start of the pandemic on the development of DL and Deep Transfer Learning (DTL) tools for FMD. More specifically, a well-designed taxonomy of frameworks is created to organize existing approaches better and guide future researchers in the correct branch of work. The taxonomy covers two principal AI tasks related to controlling COVID-19 spreading: social distancing analysis and face-mask-wearing detection. The comparative study of existing AI solutions used to perform the aforementioned tasks, emphasizing state-of-the-art DL models, provides a sturdy basis for future research and applications. Additionally, insightful observations are made to identify both solved challenges and those that remain unresolved, such as the lack of publicly available and annotated real-world datasets, computational cost requirements and the demand for real-time solutions, lack of storage space on edge devices to manage image/video datasets, etc. Accordingly, the main contributions of this paper can be summarized as follows:

- Identifying and systematically reviewing existing DL and TL models used to monitor social distancing in indoor and outdoor environments.
- Describing the state-of-the-art ML- and DL-based methods applied to detect mask-covered faces in the wild.
- Analyzing and discussing the performance of ML and DL models in detecting social distancing respect and face mask usage and identifying their pros and cons.
- Highlighting the open issues for the ongoing research in the field and providing insights about the research directions and applications that can attract considerable interest in the near future.

### 1.3. Review Methodology

The bibliometric research is performed in the context of a narrative review. The recent works related to the FMD based on DL and DTL have been searched. We searched the relevant keywords, like, “face mask detection”, “deep learning”, “transfer learning”, “deep transfer learning”, “domain adaptation”, and with different combinations on different databases (including IEEEXplore, Elsevier, ACM Digital Library, Scopus, etc.) in titles, abstracts, and keywords. The adopted search procedure is explained in Figures 1 and 2.

A total of 196 articles have been returned in response to the relevant queries, out of which 179 are non-duplicates. These articles are further investigated for their relevancy to the theme of the present study, as explained in Figure 2. Finally, 130 papers are considered for this study.

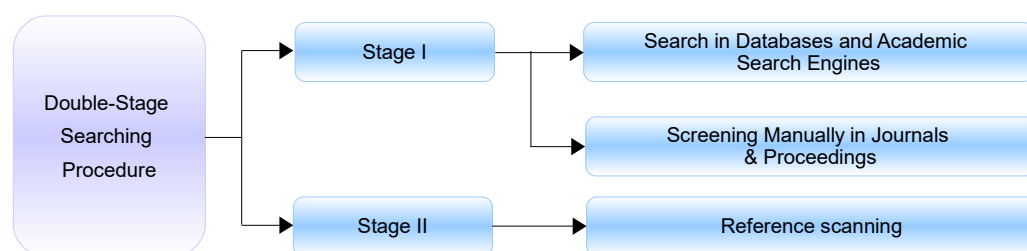


Figure 1. Adopted search procedure for review.

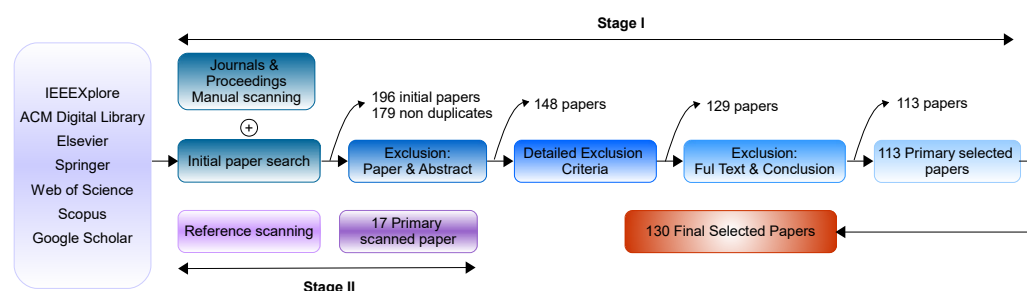


Figure 2. Selection criteria of the research article for the presented review.

## 2. Background

### 2.1. FMD Related Tasks

The detection of face masks in the related literature has evolved into a multi-faceted problem that defines different tasks for each aspect. In the following, we examine each task separately, explain the differences from other tasks and the challenges it raises, and highlight the leading works in the field.

#### 2.1.1. Mask Occlusion Detection

The occluded mask detection task refers to the automatic inference that a person has covered his/her face entirely or partially. It is a binary classification task in which the main concepts of face detection are employed but in reverse principle. The main aim of the ML algorithms is to detect whether the face is covered by a mask, glasses, or any other object. Such applications are used for security [16] (e.g., of automated teller machines (ATMs)), authentication [17] (e.g., for fraud prevention), and public surveillance.

#### 2.1.2. Incorrect Face Mask Wearing Detection

Detecting whether a face mask is worn correctly is of utmost importance since incorrect wearing can remarkably drop the mask's efficiency. This is strengthened by the fact that many individuals still refuse to wear protective masks correctly. To that end, numerous studies have been proposed to investigate this issue by detecting the placement of masks in the faces [18–20]. In [21], the authors propose a dataset for evaluating incorrect face mask-wearing detection techniques. The dataset includes images of incorrectly and correctly

masked faces (CMFD), which are artificially created by applying a mask on many different faces with different levels of face coverage. Also, the authors in [18] introduce a real-time CNN-based approach to identify incorrect utilization of face masks. Authors in [19] combined super-resolution (SR) networks with simpler neural networks trained on the Medical Masks Dataset (MedMasks) for classifying faces as correctly covered by a mask, improperly covered or not wearing a mask in images with multiple persons and achieved an accuracy of 98%.

### 2.1.3. Masked Face Recognition (MFR)

Wearing a mask obstructs conventional face recognition methods, which focus on detecting facial components such as the eye, nose, mouth, and ears to identify individuals. In this detection task [22], which falls under the generic task of occluded face detection, the focus is on whether the detected mask covers the mouth and nose. The recognition task focuses on matching masked faces with either unmasked or masked faces. This challenge has gained increasing interest since the Corona pandemic, when mask-wearing has become part of people's daily life to slow down the spread of the virus.

Three main issues differentiate the task from the conventional occlusion face recognition and make it much harder to solve: (i) the lack of large-scale face datasets with masks, (ii) the fact that the characteristics of the nose and mouth are seriously damaged, and the number of pertinent features is considerably decreased; and (iii) the community interest for detecting and recognizing faces with masks, especially in the case of pandemics. In the same direction, re-using existing DL models can face a significant challenge related to the discrepancy between the source data (SD) and target data (TD). For instance, when unmasked faces are used for training, masked faces are considered in the test or the opposite. Also, when different types of masks are applied, apart from the medical masks, the standard face recognition systems may fail to adapt to new tasks (i.e., new mask types), and further training is needed to efficiently and effectively recognize masked faces [23].

### 2.1.4. Partial Face Recognition

Another task related to FMD is handling distortion that may occur in unconstrained environments, when the photo shoot may have an angle or other objects and faces may partially cover the target face [24]. The resulting faces are of different sizes; some of their features are hidden or out of view, and consequently, conventional algorithms will fail. DL techniques can automatically learn visual features that can be used for face recognition [25] and can thus be beneficial for this task. When the task has to be combined with mask detection, the difficulty increases exponentially. The detection of masks can then be based on advanced DL techniques that combine low-level features with attention mechanisms to use all the visible face features and the potential mask features if a mask is employed [26].

## 2.2. Datasets

Since the pandemic's start, a great effort has been devoted to launching different FMD datasets. However, many of these repositories are either artificially created, so they don't appropriately represent real-world scenarios, or they have noisy samples and wrong annotations, further hindering the detection task. Choosing the suitable dataset for training a mask detection model requires effort.

The SSDMNv2 model [27] was trained using a dataset that included a variety of open-source datasets and images, such as the Medical Mask Dataset from Kaggle (KMMD), the PyImageSearch dataset (PBD), and the masked face recognition dataset (MFRD) [28]. The KMMD dataset contains 678 pictures of people wearing medical masks, along with XML files that describe the masks and provide other details. The PBD dataset includes 1,376 images divided into two classes: people wearing masks (690 images) and people not wearing masks (686 images). This dataset was created using standard images of faces with facial landmarks applied. The use of facial landmarks in the training dataset allowed for the artificial creation of mask-wearing images by positioning masks on faces based on the location of specific facial features, such as the eyes, eyebrows, nose, mouth, and jawline.

To prevent introducing bias into the task, the artificially masked images were not used as non-masked samples. Instead, a larger dataset called the Face Mask Dataset (FMD) was created by combining images from the Wider Face dataset [29] and the MAsked FAcEs (MAFA) dataset [30]. The resulting FMD dataset includes 7959 images from two classes labeled as “with a mask” and “without a mask”. The MAFA dataset, created before the COVID-19 pandemic in 2017, includes a total of 30,811 internet images and 35,806 images of masked faces, some of which are masked using hands or other objects rather than physical masks. The Wider Face dataset is much larger, containing 393,703 faces annotated in 32,203 images and featuring a wide range of scales, poses, and occlusions. Moving on, A much smaller dataset is the masked-faced dataset (MFDS) [31] that includes 200 images. This small number does not allow the dataset to be used for training DL-based algorithms. In addition, different objects have been used to mask faces, reducing the samples that can be used for mask detection tasks. The Face Mask Dataset (FMDs) is designed and released in [32], consisting of 853 images.

In [27], a real-time medical mask detection (RTMMD) is proposed, which includes 5521 images that have two classes annotated as “with\_mask” and “without\_mask”. It is also well adapted to detect assailants covering their faces while performing unlawful deeds. In [33], two FMD datasets are presented for masked face verification (MFV) and masked face identification. The first one includes 800 images of 200 different persons (identities), whereas the second comprises 4916 images of 669 identities. In [34], the medical masks dataset (MMD) is launched by Mikolaj Witkowski in terms of a Kaggle data contest<sup>1</sup>. The MMD dataset contains 682 images, including more than 3000 medical masked faces wearing masks.

In [35], the face mask 12k images dataset (FMID-12k) is released, which includes 11,792 images gathered from various backgrounds. Moving forward, the face mask classification (FMC) data is presented in [36], which has 440 images equally divided into two classes and collected on noisy backgrounds.

In [37], the real-world masked face dataset (RMFD) has been proposed, which contains 90,000 normal faces and 5000 masked faces of 525 individuals. In [38], two face image datasets (FIDS<sub>1</sub> and FIDS<sub>2</sub>) have been proposed. The balanced FIDS<sub>1</sub> consists of 3835 images that belong to two distinct categories, i.e., 1919 images of “without\_mask” and 1916 images of “with\_mask”. These images have been collected from different sources, such as the RMFD, Kaggle datasets, and Bing search application programming interface (API). FIDS<sub>2</sub> comprises 1376 images divided into two classes (686 images of “without\_mask”, and 690 images of “with\_mask”). FIDS<sub>2</sub> is designed using images from the simulated masked face dataset (SMFD) [39]. Authors in [40] introduced the thermal mask dataset (TMD) to cover the lack of masked-faced image datasets in the infrared or thermal spectrum. The dataset contains 153,360 images in both visual and thermal spectra of people wearing surgical masks. Finally, because of the lack of unified and diversified datasets for evaluating both FMD and masked facial recognition (MFR), the authors in [41] develop a large-scale mask detection to quantify the performance of algorithms for both tasks.

Table 1 lists the publicly accessible datasets used for face mask-wearing detection and classification and summarizes their main features. As can be seen, most of these datasets are small and can not be used to train DL models, so they have been augmented in most face mask detection frameworks using standard data augmentation techniques. This approach introduces a bias, which can make the comparative analysis of the results collected on the augmented repositories unfair.



**Table 1.** Summary of existing publicly accessible FMD datasets and their characteristics.

Dataset	# Images/Faces	# Classes	# Masked/Unmasked Faces	Environment	Head Pose	Description
MFDS [31]	200	1	200/0	Real	Various	• Real masked faces applied to track and identify criminals or terrorists.
MAFA [30]	30,811/ 35,806	1	35,806 /0	Real	Various	• Images collected from the Internet. Six attributes are manually annotated for each face region
Masked-LFW	13,097	1	13,097/0	Real	Various	• Faces from the original LFW dataset were masked using software.
SMFD [39]	1570	2	785/785	Simulated	Frontal to Profile	• All the images are web scrapped.
RMFD [37]	95,000	2	5000/90,000	Real	Various	• Part of the dataset is collected from other research datasets and another part is crawled from the Internet.
MMD [34]	6000	20	N/A	Real	Various	• Images acquired from the public domain with extreme attention to diversity.
MaskedFace-Net [21]	137,016	2	67,049/66,734	Real	Frontal	• Face images collected from FFHQ.
FMDs [32]	853	3	717/3232 (+123 incorrect)	Real	Various	• Images collected from the Internet, used to train two-class models.
MFV-MFI [33]	400 (verification) 4916 (identification)	10	200/200 (verification), 2458/2458 (identification)	Real	Various	• A dataset for the MFR task.
MFDD [28]	24,771	1	24,771	Real	Various	• Images from the Internet and images of people wearing COVID-19 masks.
SMFRD [28]	500,000	1	500,000/0	Simulated	Various	• Generated using mask-wearing software based on the Dlib library.
Singh's Dataset [42]	7500	2	N/A	Real	Various	• Combination of MAFA, Wider Face and captured images by surfing various sources.
FIDS1 [38]	3835	2	1916/ 1919	Real	Frontal to profile	• Combination of Kaggle datasets, RMFD dataset and Bing search API.
FIDS2 [38]	1376	2	690/686	Simulated	Frontal to profile	• Created based on SMFD.
AIZOO-Tech	7971	2	12,620/4034	Real	Various	• Designed by modifying the wrong annotations from datasets of Wider Face and MAFA.
FMLD [43]	41,934/ 63,072	3	29,532/ 32,012 (+1528 incorrect)	Real	Various	• Combination of MAFA and Wider Face datasets.
Moxa3K [44]	3000/12,176	2	9161/3015	Real	Various	• Combination of Kaggle datasets recorded from Russia, Italy, China and India during the ongoing pandemic.
UFMD [45]	21,316	3	10,698/ 10,618 (+500 incorrect masked faces)	Real	Frontal to Profile	• Combination of FFHQ, CelebA, LFW, YouTube videos, and the Internet.
WMD [46]	7804/26,403	1	26,403/0	Real	Various	• Collected from real scenarios of fighting against CoVID-19 covering many long-distance scenes.
PWMFD [47]	9205/18,532	3	10,471/7695 (+366 incorrect masked faces)	Real	Frontal to Profile	• Combination of images from WIDER Face, MAFA, RWMFD.
Thermal-mask [40]	75,908	2	42,460/33,448	Real	Various	• The images are in both spectra (visual+thermal) with 18 variations of face mask patterns.
Bing dataset [48]	4039 (tr.:3232/test.:807)	2	N/A	Simulated	Various	• The images are collected from Bing using the bing-images library available in Python.
MedMasks [19]	3835	3	3030/671 (+134 incorrect masked faces)	Real	Various	• Images in uncontrolled environments are pre-processed.

### 2.3. Evaluation Metrics

Various metrics have been used to assess the performance of FMD techniques in the literature. Each method's performance is often evaluated using different criteria to ensure its reliability under different scenarios. The metrics most frequently used for assessing FMD solutions are based on the confusion matrix of the binary classification problem (i.e., mask, no mask). Moreover, by using the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) sample predictions, the researchers measure the prediction accuracy, error rate, recall, precision, and  $F_1$  score as given in the following equations.

Researchers also examine the time performance of their solutions, measuring the times needed for training their models and the inference times. They also measure the throughput of their systems, which is critical for developing real-time face mask detection systems. Table 2 summarizes current evaluation metrics widely used in FMD frameworks.

**Table 2.** A summary of the common evaluation metrics used to evaluate FMD frameworks.

Metric	Description	Formula
Relative error (RE)	The ratio of the absolute error of a variable to its value	$ (y - \hat{y})/\hat{y} $
Mean absolute error/difference (MAE or MAD)	Calculate the difference between the predicted and actual values for a given phenomenon.	$(\sum_{i=1}^m  \hat{y}_i - y_i )/m$
Validation loss ( $L_1$ )	It is determined by evaluating the model on a validation set, which is obtained by dividing the data into training, validation, and test sets using cross-validation.	
Kappa coefficient (K)	Assess the model's prediction accuracy on a test dataset by comparing the predicted values to the true values.	$(\text{total acc-random acc})/(1-\text{random acc})$
Mean squared error/difference (MSE or MSD)	Calculate the average of the differences between the predicted and actual values.	$(\sum_{i=1}^m (\hat{y}_i - y_i)^2)/m$
Root mean squared error/difference (RMSE or RMSD)	The square root of the MSE is taken to express the error in the same units as the original variable.	$\sqrt{(\sum_{i=1}^m (\hat{y}_i - y_i)^2)/m}$
Root mean square percentage error (RMSPE)	Represents the RMSE expressed in percentage.	$\text{RMSE} \times 100\%$
Normalized root mean squared error (NRMSE)	A standardized version of the RMSD that allows for comparison between variables that have different scales.	$\text{RMSE}/(y_{\max} - y_{\min})$
R-squared ( $R^2$ )	The fitness of a regression model to the data. It measures the proportion of variance in the dependent variable that can be explained by the independent variable(s).	$1 - (\sum_{i=1}^m (\hat{y}_i - y_i)^2 / \sum_{i=1}^m (\bar{y} - y_i)^2)$
Theil U1 index	Quantify the difference between the observed and predicted values, with a higher value indicating a better fit and more accurate predictions.	$\sqrt{\sum_{i=1}^m (\hat{y}_i - y_i)^2 / \sum_{i=1}^m y_i^2}$
Theil U2 index	Measures the quality of the predicted results.	$\sqrt{\sum_{i=1}^m (\hat{y}_i - y_i)^2 / m} / \sqrt{(\sum_{i=1}^m y_i^2 / m + \sum_{i=1}^m \hat{y}_i^2 / m)}$
Accuracy (ACC)	Measure how closely the predicted values match the target values.	$(TP + TN)/(TP + TN + FP + FN)$
Error rate (ERR)	Calculate the percentage of incorrect predictions made by the model out of the total number of predictions.	$(FP + FN)/(TP + TN + FP + FN)$
Precision (PPV)	The closeness of predicted results to the true values.	$TP/(TP + FP)$
Recall or True positive rate (TPR)	The ratio of true positive (TP) predictions that are correctly identified.	$TP/(TP + FN)$
False-positive rate (FPR)	The ratio of false positive (FP) predictions among all predictions in the true negative (TN) class.	$FP/(FP + TN)$
True-negative rate (TNR)	Determine the percentage of true negative (TN) predictions correctly identified among all predictions in the true negative (TN) class.	$TN/(TN + FP)$
False-negative rate (FNR)	Measures the proportion of false negative (FN) predictions in the true positive class.	$FN/(TP + FN)$
F1-score	Calculate the percentage of false negative (FN) predictions among all predictions in the true positive class.	$2(\text{PPV} \times \text{TPR})/(\text{PPV} + \text{TPR})$
Matthews correlation coefficient (MCC)	Assess the quality of a binary classification.	$\frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$
Average precision		
Mean average precision (mAP)		$\sum_{k=1}^K \text{AP}/m$
IoU	Intersection over union	$\text{area of overlap}/\text{area of union}$
Precision recall curve (PRC)	Illustrate the balance between precision and recall as the threshold for classification varies.	-
Receiver operating characteristic curve (ROC)	Illustrate the balance between FPR and TPR as the threshold for classification is varied.	-
Area under the ROC (AUROC)	Determine the area under the ROC curve, where a higher value indicates a better classification performance.	-
Cross-validation (CV)	Evaluate the performance of an AI model on unseen data by testing its ability to make predictions and generalize based on a sample of training data.	-
Confusion matrix	A summary of the classification results of an algorithm, typically presented in a table format.	-

### 3. FMD Based on Conventional ML

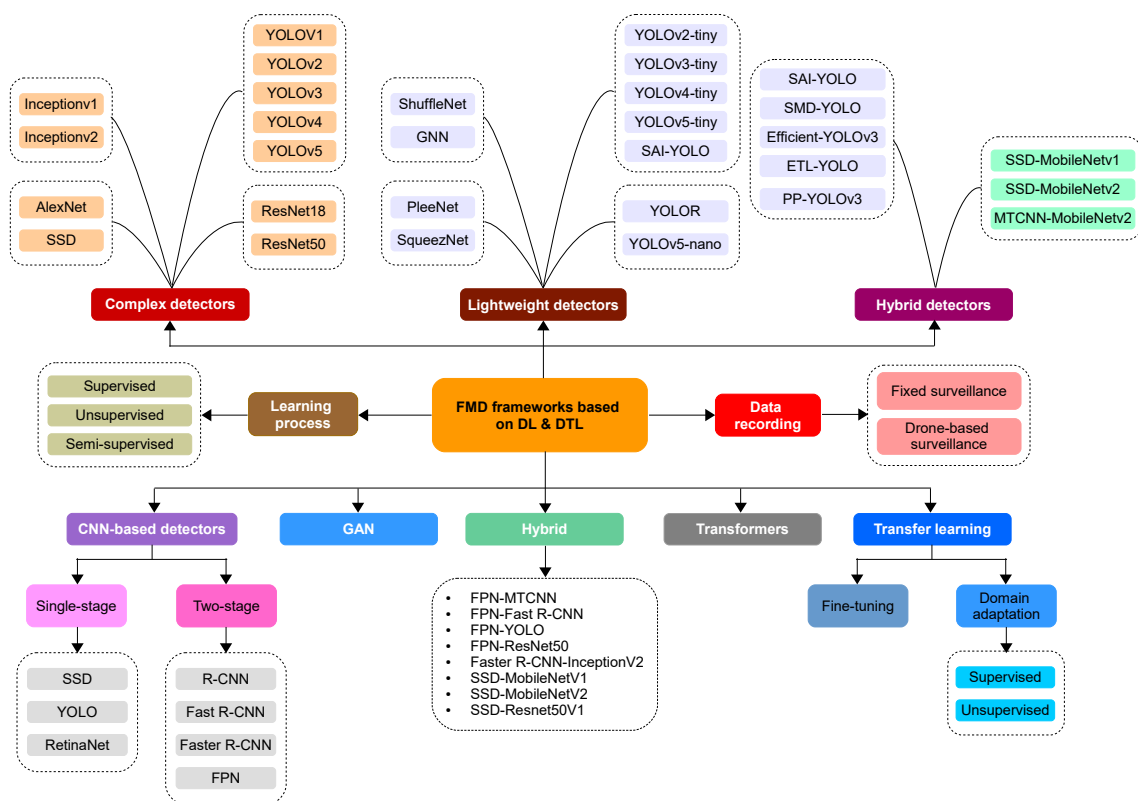
The rapidly growing literature on Face Mask detection (FMD) [13] indicates that FMD approaches are divided into several categories. In this section, we review the most significant works done on FMD. As elucidated in the previous sections, even though the research on the detection of masks that cover people's faces has been going on for decades before the onset and outbreak of the coronavirus pandemic, the methods and algorithms earmarked for the FMD task still need to be improved. Existing FMD techniques can be divided into two categories conventional ML- and DL-based models. This work focuses on the latter group (i.e., deep FMD techniques), demonstrating better prediction performance and more benefits than the ML alternative [49]. In the following paragraphs, we briefly provide the main characteristics of ML-based methods.

As explained in Section 2, a main subtask of FMD is the detection of human faces in the images or video streams. The related ML literature on face detection employs Haar-like features and Haar classifiers to detect faces. This statistical approach builds on the categorization of subsections of an image, which implicitly correspond to facial characteristics (e.g., eyes, eyebrows, nose, nostrils, etc.) and their relative positions [50]. Works that build on face detection and then extract image features to perform FMD employ decision tree classifiers. These give a speedy, simple, and efficient method for detecting masks [51]. Similarly, other techniques build on conventional object detection approaches, such as Haar Cascade [52] and Histogram Of Gradients [53] to detect valuable features, which are then employed by an ML classifier to solve the FMD task. Such techniques have proven to be prominent and influential. Still, they heavily depend on Feature Engineering and demonstrate worse prediction accuracy performance when compared to modern deep learning techniques.

A group of ML methods for detecting partially occluded faces has been presented before 2010 and proposed the use of Support Vector Machines [54], neural networks [55], Gappy PCA that fits around the wrong values [56], Adaboost and other classification techniques on top of handcrafted features extracted from images. The occlusion could be on any part of the face, caused by a mask, the hair, sunglasses, etc. [57]. Since the methods take as input the features detected in each face image, they can be modified to be used for detecting whether the face is clear or has any kind of occlusion, for performing face detection and identification, or for any other face-related classification or regression task (e.g., age estimation, sentiment analysis) [58]. Still, it would take more work to detect the occluded parts and their effect on the accuracy of predictions.

As mentioned above, this work focuses on the DL-based face mask detection (deep FMD) techniques, which are illustrated in Figure 3 using a comprehensive taxonomy. The objective of the taxonomy is to provide a visual summary of the various face mask detection frameworks focusing on (i) the detectors, (ii) the learning processes, and (iii) the data collection alternatives.. In the following section, we survey existing deep FMD methods and the main architectures and models proposed in the literature. We organize their extended presentation according to different criteria, such as the model type, the number of processing steps, and the complexity of the resulting models.





**Figure 3.** A taxonomy of the approaches introduced for solving the Face Mask Detection problem using deep learning.

#### 4. DL-Based FMD

Even though the majority of research works on FMD that employ DL methods have been published during the last three years, the related literature is much richer than that of ML methods. In this section, we list the main approaches in the field, organized by the general architecture employed, the number of processing stages, and the complexity. The works detailed below and their main features and performance are summarized in Table 3.

##### 4.1. Sorted by the Employed Architecture

The detection of face mask wearing is mainly performed on images; thus, image analysis DL models are employed for the task. Although Convolutional Neural Networks (CNNs) dominate the field, a few works still use different DL architectures to solve the task.

##### 4.1.1. Convolutional Neural Networks (CNNs)

CNNs are the most popular architectures in object detection and recognition from images and image classification. Several pre-trained CNN architectures, such as AlexNet, VGG16, VGG19, Inceptionv3 (GoogLeNet), ResNet50, EfficientNet, etc., have been used in a variety of image classification and detection tasks, including medical image detection [59], sentiment analysis and face expression classification [60], etc. However, they have been trained for entirely different tasks. Transfer learning is the primary ML technique that allows these pre-trained models to adapt to new tasks and achieve state-of-the-art performance quickly [61,62]. In [63], transfer learning has been applied to a face mask annotated dataset using AlexNet and VGG16 models to detect if people are using face masks. The authors also combined these architectures with Long short-term memory (LSTM) and Bidirectional LSTM (BiLSTM) layers to further improve the classification process. The dataset used for training comprised a total of 2,000 images organized into four classes: “unmasked”, “masked”, “masked but under the chin”, and “masked but nose open”.

The VGG-16 architecture has been used in [64] as a basis to train a CNN classifier for FMD in real-time. The authors collected over 20,000 images from the web, and the trained image classifier was used as a face mask-wearing alarm system. Authors report an accuracy of 98% in test images and claim that the developed solution is computationally efficient for real-time setups.

A two-stage CNN architecture for FMD is proposed in [65] for detecting instances of improper face mask wearing. The proposed method was trained using 7855 images divided into two classes in the main task. In a side task, authors tried to detect faces in images with multiple persons, using a pre-trained RetinaFace model, and then classify the persons as “mask-wearing”, or “no wearing a mask”. Finally, they employed a centroid tracking technique to track the detected faces between consecutive images and proposed architecture for video analysis.

#### 4.1.2. Generative Adversarial Networks (GANs)

Different models of GANs have been utilized to develop reliable FMD solutions that can preserve individuals’ identities. For instance, a GAN model that automatically removes the masks covering the face areas and regenerates face images by building the missing holes is proposed in [66]. The outcome of this framework is a full-face image looking realistic and natural and realistic. A GAN-based method has been employed in [67] for generating masked faces. A domain-constrained loss is used to train this architecture, bringing the inpainted masked faces as close as possible to their corresponding identified complete faces. Similarly, a GAN-based identity-preserved inpainting solution that alleviates the problem of occluded face recognition is proposed in [68]. Besides, Ding et al. [33] attempted to overcome the lack of large-scale annotated FMD datasets by creating two datasets of synthetic masked face images designed for mask-face detection and recognition. Explicitly, the former includes 400 pairs of 200 identities for verification, while the latter encompasses 4,916 images of 669 identities for identification.

#### 4.2. Sorted by the Number of Processing Stages

DL-based object detection performs well because of its robustness and capability to extract pertinent features. Two popular categories are found in the literature, one-stage, and two-stage object detectors. Examples of one-stage detectors that process images in a single step, are a single shot detector (SSD), you look only once (YOLO), and RetinaNet. Two-stage detectors that select regions and then refine them comprise region-based CNN and its variations and techniques that combine popular CNN architectures for object detection with classifiers for filtering them.

##### 4.2.1. One-Stage FMD

The simpler one-stage approaches divide the image into fixed-sized cells and then apply an object locator to find objects in each cell. The YOLO family of architectures is quite popular for the detection of objects. However, these approaches could perform better in detecting objects of different sizes. A better alternative is to apply multi-scale detection to a single-shot detector and then conduct detection on several feature maps to detect faces of various sizes. Another problem to be handled is the class imbalance problem. Several one-stage techniques include focal loss functions to reduce the loss for easy samples (i.e., cells) and focus on the hard ones.

*Single-Shot Detector (SSD):* It is based on a multi-scale detection to perform detection on different feature maps and detect faces of distinct sizes [69]. In [27], the authors introduce a real-time CNN-based FMD system using SSD and MobileNetV, namely SSDMNv2. In [70], Anithadevi et al. introduce a single-shot multi-box FMD approach based on MobileNetv2 to extract pertinent features and enhance object detection. After considering the significance of developing an accurate model and the limitations of existing models, it proposes a framework based on deep learning, which would automate and simplify the task of monitoring social distancing and face masks through intelligent video analytics. In [19], Qin. et al. combined super-resolution images and SRCNet to develop an FMD approach. They

were able to classify three categories of facemask-wearing conditions (i.e., correct facemask-wearing, incorrect facemask-wearing, and no facemask-wearing), and their proposed method achieved 98.70% accuracy in the face detection phase.

*YOLOv1*: [71] presents a YOLOv1-based FMD, which has been combined with a stack-based virtual machine (WebAssembly) and high-performance neural network inference computing framework (NCNN) to improve performance. This approach enables the privacy preservation of users' data as it is implemented on an edge computing architecture.

*YOLOv2*: It is a single-stage real-time object detection model that has improved the performance of YOLOv1 in different aspects, e.g., using (i) anchor boxes to predict bounding boxes, (ii) a high-resolution classifier, (iii) batch normalization, and (iv) Darknet-19 as a backbone. Despite these improvements, it did not attract significant attention in FMD applications. In [72], YOLOv2 with ResNet50 has been used to develop a medical FMD scheme. Typically, this approach is designed in two stages: (i) extracting deep features using a ResNet50-based DTL architecture and (ii) detecting face masks using YOLOv2.

*YOLOv3*: It is an improved version of YOLOv2, which is built using a DarkNet53 backbone. It includes 107 layers in total, 53 of them being convolutional layers [73]. Ref. [74] discusses the use of DL techniques for visual analytics of the spread of COVID-19 infection in crowded urban environments. The authors describe a CNN model that they developed for visual analytics of the spread of COVID-19 and demonstrate the effectiveness of their model on a dataset of real-world video sequences from crowded urban environments. They show that it is possible to detect individuals potentially infected with COVID-19 based on their physical behavior and interactions with others, with an average detection precision of 69.41%. In [75], the authors use a YOLOv3 model for monitoring social distance in the context of the COVID-19 pandemic. The model is trained to identify individuals in video sequences and to estimate the distance between them, allowing it to identify instances of social distancing violations. The authors evaluate their model on a dataset of real-world video sequences and demonstrate that it can accurately estimate the distance between individuals and identify social distancing violations with a high degree of accuracy. They also discuss the potential applications of their model for real-time monitoring and enforcement of social distancing measures in various settings.

In the same spirit, the authors in [76,77] calibrate videos into bird's eye view before feeding them as inputs to the pre-trained YOLOv3 model. However, both studies need to provide more assessment results. Wu et al. [78] introduce FMD-YOLO, an FMD approach based on Im-Res2Net-101. The latter relies on combining the Res2Net module with deep residual networks (DRN), where non-local mechanisms, deformable convolution, and hierarchical convolutional structure are applied to extract information from the input thoroughly. It is worth noting that many more frameworks that are based on YOLOv3 architecture have been found in the literature [79–99], but they did not provide extensive evaluation results.

The main objective in [100] was the analysis of the effectiveness of deep-learning-based face detection algorithms when applied to thermal images, especially images that depict faces covered by virus-protective face masks. As part of their work, the authors compiled more than 7900 thermal images containing faces with and without masks. Selected raw data pre-processing methods were also investigated, and their influence on the face detection results has been evaluated. It was shown that the use of transfer learning based on features learned from visible light images results in a mean average precision (mAP) greater than 82% for half of the investigated models. The best model was based on the YOLOv3 model (mAP was at least 99.3%, while the precision was at least 66.1%).

*YOLOv4*: A significant number of FMD systems have been built upon YOLOv4 [101–108], but still not all research works present satisfactory experimental results. So in the following, we focus on the studies presenting significant contributions and empirical validations. While YOLOv4 uses the CSPDarknet53 as the original backbone with many parameters with deep layers, the realistic scenes of real-time FMD are generally simple. To detect whether people are wearing masks or not, two kinds of objects are detected: face with a mask or face without a mask. Using the complex CSPDarknet53 as the backbone is inap-

appropriate in this case because the computational cost unnecessarily increases. To overcome this issue, an improved CSPDarknet19, which combines the benefits of CSPDarknet53 and Darknet19, has been proposed in [109]. Cao et al. [110] propose MaskHunter, a real-time FMD based on an improved YOLOv4. MaskHunter uses an improved CSPDarkNet-19 backbone. Concretely, the improved YOLOv4 with CSPDarkNet-19 backbone, neck, and prediction head is developed along with an enhanced mosaic data augmentation approach. The neck architecture of MaskHunter comprises path aggregation network (PAN) [111], FPN, and spatial pyramid pooling (SPP) modules [112]. Figure 4 portrays the flowchart of MaskHunter, which includes the improved CSPDarknet19, the improved neck with BiFPN, the double-head prediction heads, and the mask-guided module that is used to discriminate between people wearing or not masks in night environments.

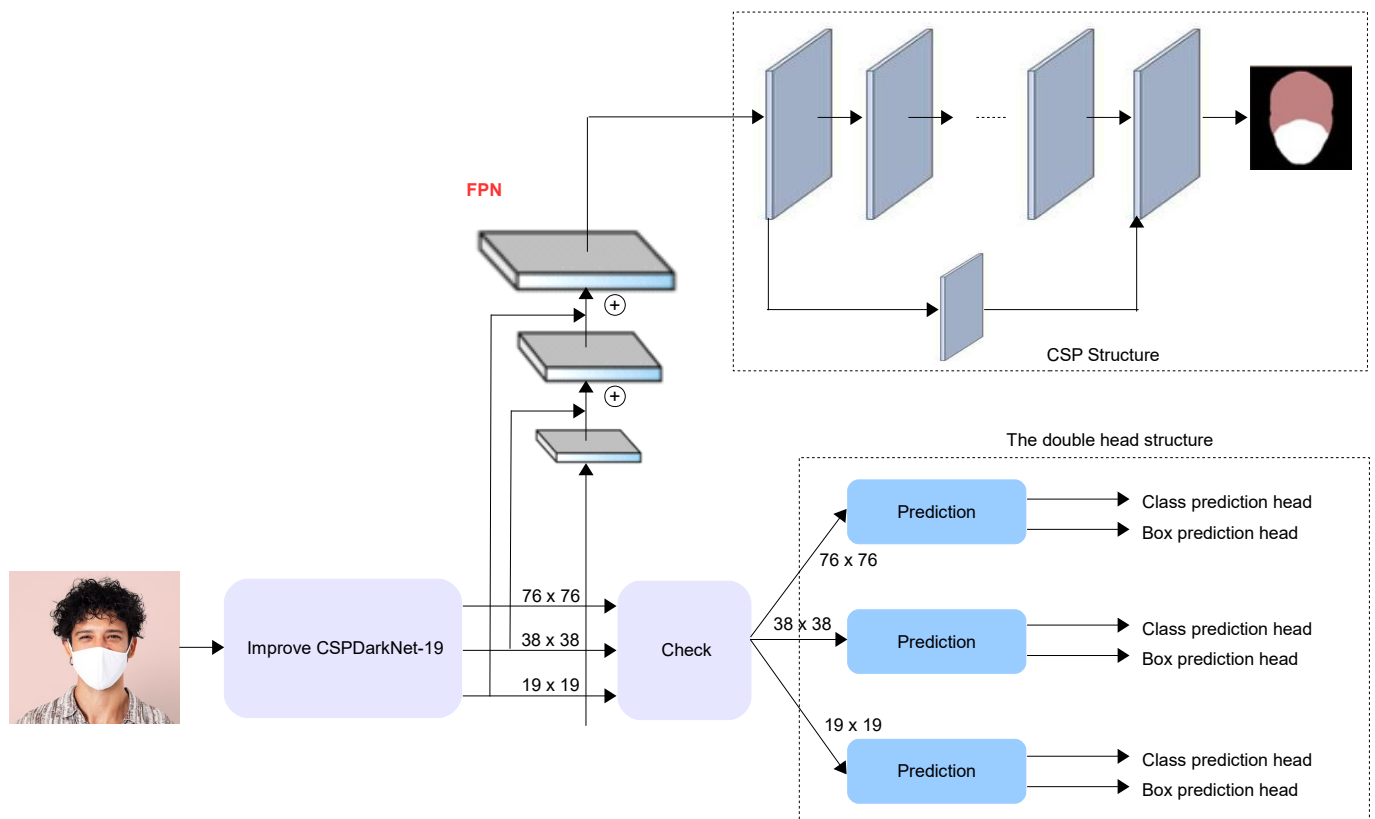
The authors in [113] propose a DL-based VSD detection scheme based on the YOLO v4 model. A fixed and single motionless time of flight (ToF) camera is used to record video data. After detecting people's objects, the Euclidean distance is used to measure the distance between the detected bounding boxes, which is consequently mapped to a real-world distance. An empirical evaluation has shown a mean average precision (mAP) score of 97.84% and a mean absolute error (MAE) between actual and measured social distance values is 1.01 cm. In [114], a DL-based crowd-counting solution is developed for monitoring and controlling the capacity of commercial buildings during the COVID-19 pandemic. It is based on YOLOv4 and has been validated on the COCO dataset. Using route and direction information, the system can determine whether a person leaves or enters the building and count the people still inside it. It can detect violations by comparing the results with a pre-defined threshold. A limitation of this study is the lack of significant assessment. Similarly, in [101], a YOLOv4-based FMD is presented but lacks a thorough empirical evaluation. Authors in [115] developed a YOLOv4-based FMD and thermal scanning kiosk to (i) ease the classification of individuals wearing or not wearing masks and (ii) measure pedestrians' body temperatures using a temperature sensing unit. In [116], Gola et al. have shown the superiority of YOLOv4 versus MobileNetv2, SSD and YOLOv3 in terms of FMD accuracy, whereas in [117] the authors verified the superiority of YOLOv4 over other variations, with a mAP value of 71.69%, and proposed YOLOv4-tiny for limited computational resources environments.

**YOLOv5:** It is a Python implementation of an improved version of YOLOv3 [118] for PyTorch7. It includes changes to activation functions and data augmentation with post-processing to the YOLO architecture, as in YOLOv4. It employs self-adversarial training (SAT) and aggregates images in training, which results in accelerated inference [119]. It has been released with five different models sizes: YOLOv5n (nano), YOLOv5s (small), YOLOv5m (medium), YOLOv5l (large), YOLOv5x (extra large). In [120], Walia et al. proposed an approach using YOLOv5, which involves taking input images from CCTV Cameras. Once a face is detected, it is processed using Stacked ResNet50 to classify between two mask-wearing conditions: correct or not. A robust dataset consisting of 1916 masked and 1930 unmasked images has been collected from various sources, then data augmentation is applied. Experimental Evaluation has shown an accuracy score of 96%, but the method robustness has not been tested with images from other collections and datasets. A promising alternative could be to use cutting-edge facial detection algorithms and transfer learning (hyper-parameter tuning) to detect whether the face has a mask.

**YOLOR:** It is a unified network that uses both implicit and explicit knowledge. The former is used to identify features of the deep layers, while the latter can be attained from annotated data. YOLOR has been employed in FMD tasks and applied in related datasets, such as on VIDMASK [121] where it outperformed YOLOv4 and YOLOv5 with a maximum mAP of 92.1%.

**Others:** the authors in [122] introduce a single-stage FMD approach called RetinaFaceMask, which relies on RetinaFace [123]. It has been built upon (i) using FPN for high-level semantic information fusion and (ii) presenting a context attention module to enable RetinaFaceMask to focus on the characteristics of masks and faces. A cross-class removal approach is also developed to remove the regions with low scores and high IoU values.

These contributions have enabled RetinaFaceMask to outperform RetinaFace [123] in Recall and Precision.



**Figure 4.** The flowchart of MaskHunter, including the improved components. Adapted from: [110].

While most existing FMD methods are developed for simple scenes, monitoring face-mask utilization is often challenging in dense crowds with different occlusions and scales. Additionally, privacy preservation is another issue that impedes the use of public data in centralized training. To overcome these challenges, a cascaded network is introduced in [124] using the Dilation RetinaNet Face Location (DRFL) Network, which helps reduce network parameters and identify faces at different scales. In [125], the authors introduced a new human ear detection pipeline based on the YOLOv3 detector. A well-known face detector named RetinaFace was also added to the detection system to narrow the regions of interest and enhance accuracy. The proposed method has been evaluated on an unconstrained dataset, which shows its effectiveness. The authors in [126] also employed YOLOv3 and FMNMobile using NASNet Mobile and Resnet\_SSD300 algorithms. This achieved an mAP of 91.7% on two datasets of 680 and 1400 images of masked and non-masked faces.

#### 4.2.2. Two-Stage FMD

In the first stage, two-stage FMD techniques are based on generating region proposals, while these proposals are then fine-tuned in the second stage. Overall, two-stage FMD methods can perform better than one-stage FMD solutions but at a higher computational cost.

**R-CNN:** In region-based CNN (R-CNN), candidate regions that may contain objects of interest are proposed using selective search [127]. The proposals are then injected in a CNN module to extract the characteristics, and an SVM is also deployed to recognize object classes. Unfortunately, R-CNN's second stage has a high computation load as the proposals are detected one by one in addition to using an SVM algorithm for final classification.

In [128,129], the authors have chosen R-CNN, Fast R-CNN, and Faster R-CNN algorithms for detecting Mask detection and Social distance. In [130], to address the problems of large inter-class and small intra-class distances and the absence of real FMD datasets,



Zhang et al. first proposed a practical dataset containing 8,635 faces with various mask wear conditions. Next, a context-attention RCNN (CA-RCNN) was introduced to extract distinguishing features and improve the discrimination ability when classifying face masks.

*Fast R-CNN:* This object detector has been used in a few FMD frameworks, mainly for comparison purposes, because it has two main issues: (i) it performs many passes through a single image to extract all the objects required, and (ii) since multiple modules work one after one, the performance of a specific module depends on that of the previous modules. For instance, in [129] Fast R-CNN is utilized along with R-CNN and Faster R-CNN to detect face masks. Typically, face mask regions are identified with CNN based on pixel prediction, mixing pictures, and specific enhancements. Similarly, in [121], Fast R-CNN with a feature Pyramid Network (FPN) and an R101 backbone is deployed for FMD along with other object detection (i.e., YOLOv4, YOLOv4-tiny, YOLOv5, and YOLOR).

*Faster R-CNN:* Sahraoui et al. [131] propose DeepDist, a DL framework for real-time detection of objects and distance violation detection from video sequences, which is based on a Faster R-CNN model. Evaluation of a dataset of real-world video sequences demonstrates the framework's ability to identify objects and estimate their distances accurately. In [42], a transfer learning-based FMD is introduced, which also employs a Faster R-CNN for detecting face masks and counting persons wearing them. Faster R-CNN is shown to have better precision but is less efficient than YOLOv3.

*Mask R-CNN:* An expanded Mask R-CNN (Ex-Mask R-CNN) model, based on ROI wrapping with Resnet-152, is proposed and used for developing an FMD solution in [132]. It is mainly used to reduce the computation complexity by (i) detecting whether a pedestrian is wearing a mask and (ii) using multi-CNN to forecast the suspicious conventional abnormalities in video frames.

*Other methods:* More two-stage methods (also called multi-stage) are found in the FMD literature. For instance, in [133], the authors introduce a DL-based FMD, which localizes faces and detects masks in video frames using MTCNN and MobileNetv2, respectively. However, this approach still needs further improvement and more experiments to evaluate its performance. Loey et al. [134] present a hybrid FMD scheme that combines CNN and conventional ML, which is based on developing two components that (i) extract features using a pre-trained ResNet50 and (ii) classify facemasks using Support Vector Machines (SVM), an ensemble algorithm and decision trees. In [135], the authors propose an ensemble of one-stage and two-stage detectors to reach accurate and real-time results. Firstly, they use a pre-trained ResNet50 as a baseline and then apply the transfer learning concept to improve performance. In [136], a multi-stage FMD approach is proposed based on (i) integrating FPN and ResNet50 (FPN-ResNet50) into a unique DL structure for detecting pedestrians in video frames, (ii) utilizing an MTCNN model for detecting and extracting human faces from these videos.

In [136], a multi-stage FMD approach is proposed based on (i) integrating ResNet50 and FPN into a unique DL structure for detecting pedestrians in video frames, (ii) utilizing an MTCNN model for detecting and extracting human faces from these videos.

Most existing object detectors generally rely on designing CNN-based network architectures to extract discriminative characteristics. However, the difference between faces (with and without masks) is essential when the training dataset size is small. To overcome these issues, Yan et al. [137] propose an FMD scheme, namely CenterFace, which uses a context attention module to activate the adequate attention of the feedforward CNN by adjusting their attention maps' feature refinement. Additionally, an anchor-free detection scheme based on triplet-consistency representation learning is proposed. This helps to integrate the triplet loss and consistency loss to overcome the data scarcity issue and reduce the similarity between occlusions and masks. Figure 5 portrays the flowchart of the CenterFace solution and the CBAM module.



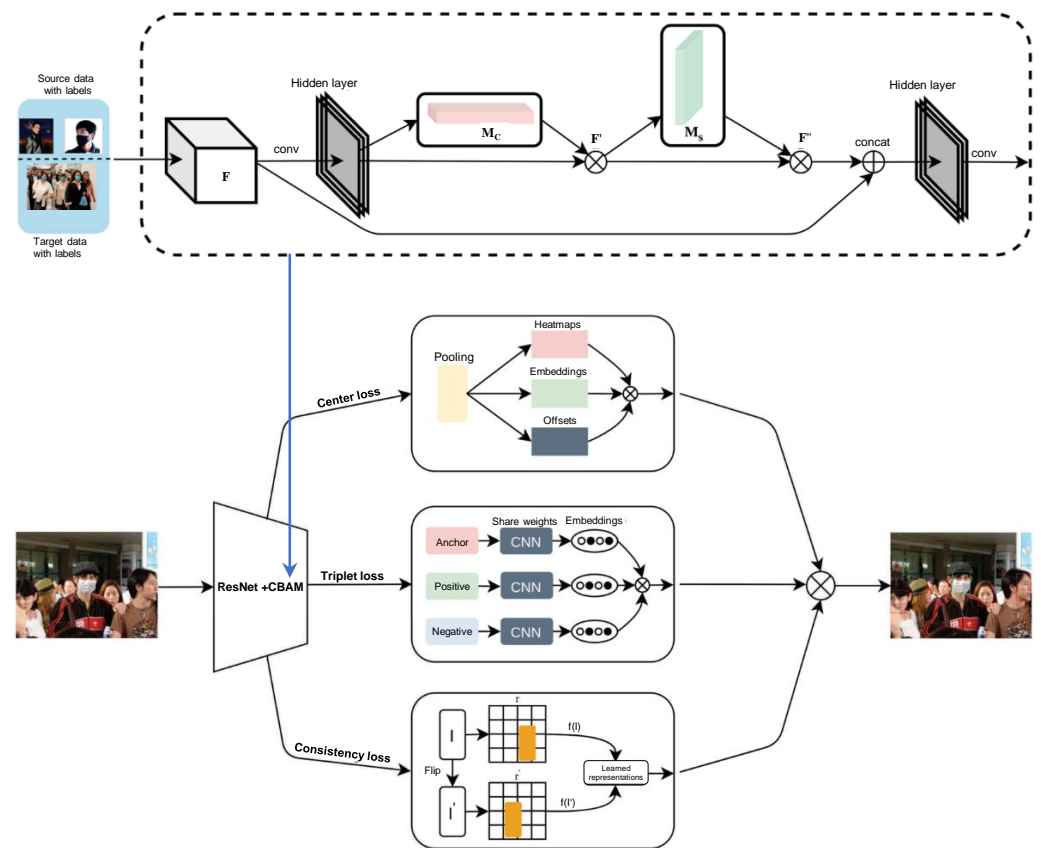


Figure 5. The architecture of CenterFace. Adapted from: [137].

In [138], a two-stage FMD, based on VGG-16 and CNN classifiers, is developed and evaluated on public transportation systems (i.e., buses). The solution has been implemented on a Raspberry Pi and employs an open-source data analytics toolkit. In [139], Bansal et al. propose a CNN-based FMD approach based on (i) detecting face masks using an object detection API, applying different face mask classifiers, including SSD-MobileNetv1, SSD-MobileNetv2, and SSD-Resnet50v1, to classify masks into four classes, i.e., “bare”, “N95”, “surgical”, and “homemade”.

#### 4.2.3. Discussion

Overall, the YOLO series and Faster R-CNN attract increasing attention, especially YOLOv3, YOLOv4, and YOLOv5. Moreover, light-weighted models, e.g., Tiny YOLO-based detectors, are gaining special attention as they can play a crucial role in deploying real-time FMD systems. Improved face detectors like RetinaFaceMask are also promising techniques. By transfer learning strategy, existing object and face detectors can be applied for masked facial detection.

#### 4.3. Sorted by the Complexity of the Models

Researchers have handled the face mask detection task either as an offline or as an online task. In the former case, more complex and resource-demanding architectures are trained and evaluated on offline datasets that comprise unmasked and masked-faced images, with the latter being either synthetically generated (masks are inpainted) or really masked. In the case of FMD as an online task, the driving requirement is that of providing predictions in almost real-time, and in this case, more lightweight and fast models are employed, which can run on edge devices with limited resources.

#### 4.3.1. Complex Object Detectors

In [140], a deep color 2-D PCA (principal component analysis)—CNN (deep C2D CNN) is used to detect face masks. Typically, the attributes of original pixels are mixed with feature representations learned by CNN before performing decision fusion to improve detection. Moving on, the classification of detected face masks is performed using AlexNet. In [141], an MTCNN-based FMD approach is introduced, which is though not appropriate for real-time monitoring due to its computational requirements. In [41], the authors propose DeepmaskNet, a unified model that can be employed in FMD and masked facial recognition (MFR) tasks. The respective framework relies on two CNN scaling techniques, i.e., scaling network depth and input image resolution. The first helps improve the FMD accuracy, while the second allows capturing pertinent fine-grained characteristics with higher-resolution input images. In [142], a near real-time CNN-based FMD approach is proposed. The proposed method first relies on detecting human posture to perform spatial reduction and background filtering in video frames. Openpose [143] is exploited in the first stage to recognize the human body's skeletons and capture facial regions before spatially reducing them. In the second stage, a CNN model that processes images detects the presence of face masks.

#### 4.3.2. Lightweight Object Detectors

Developing lightweight face mask detectors is of utmost importance to enable their implementation on edge devices with limited computation resources, e.g., drones, mobile cameras, etc. Therefore, various studies have focused on developing FMD solutions using lightweight object detectors.

*MobileNetv1*: This lightweight detector is effective for mobile devices, and its performance has been investigated in various studies, such as [144–146]. Typically, it has been used for comparison purposes with other improved variants, e.g., MobileNetv2, MobileNetv3, and MobileNetv4.

*obileNetv2*: Several works employ MobileNetv2 [144,147–151] and similar detectors such as NASNetMobile [152]. In [153], an FMD technique that employs MobileNetv2 as a basis and performs transfer learning has been proposed and tested on SMFD. In [154], a MobilNetv2-based FMD solution is developed and compared with YOLOv3.

In the same direction of lightweight models, after investigating different DL models for FMD, including EfficientNet, ResNet50, ResNet-101, Inceptionv3, VGG19, VGG16, MobileNetv1, and MobileNetv2, Habib et al. [145] introduced a real-time FMD solution which is appropriate for edge devices. This system relies on the MobileNetv2 architecture to extract discriminative features from video frames, which are then fed to an auto-encoder for feature representation abstraction and finally to the classification layer. In [155], an FMD system is developed by employing a classification model based on the MobileNetv2 architecture and the OpenCv's face detector to identify the location of the face and determine whether or not it is wearing a mask. Additionally, the FaceNet model is utilized as a feature extractor and a feedforward multilayer perceptron to perform facial recognition. The model was trained using a set of 13,359 images, including 52.9% with masks and 47.1% without masks.

*YOLOv1-tiny*: YOLOv1-tiny is a real-time object detection model designed based on YOLOv1, where it has been pre-trained on the VOC dataset with 20 classes. Very few FMD studies have used YOLOv1-tiny, such as [156], where the performance of an FMD system based on YOLOv1-tiny is assessed and compared with YOLOv1, YOLOv2, YOLOv2-tiny, YOLOv3-tiny, and YOLOv4-tiny.

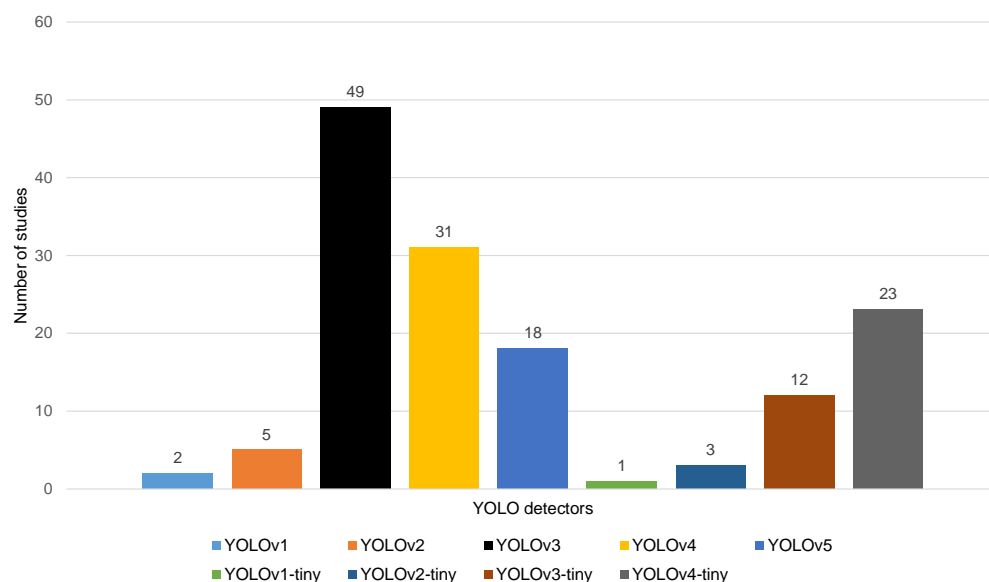
*YOLOv2-tiny*: It is the compressed version of YOLOv2, which has been pre-trained on 80 object classes from the COCO dataset. YOLOv2-tiny has been adopted in some studies to develop real-time FMD systems. For example, in [157], a weight quantization scheme is presented to design a compact CNN model that detects individuals with or without masks based on YOLOv2-tiny.

*YOLOv3-tiny*: It is the real-time compressed version of YOLOv3; it has been pre-trained on the COCO dataset with 80 object classes. Many studies have considered YOLOv3-tiny

as the core of their FMD systems, including [158,159]. For instance, in [159], the accuracy of 95% has been reached on a customized dataset of 135 images. Moving on, the FMD approach proposed in [160] is based on YOLOv3-tiny, which has been improved accordingly to solve the FMD task. In [161], the FMD problem is considered a multi-task object detection problem to detect wrong and correct ways of wearing masks using a YOLOv3-Slim.

**YOLOv4-tiny:** It is the compressed version of YOLOv4 developed for training on low computational resources. It has 16 megabytes of weight, which enables it to be trained using 350 images in less than one hour using a Tesla P100 GPU. Many studies have utilized YOLOv4-tiny to develop real-time DFM solutions, such as [162–165]. Typically, in [163] where a lightweight network FMD approach based on YOLOv4-tiny, namely SAI-YOLO, is proposed to detect drivers wearing masks. Moving on, in [164], the SMD-YOLO approach for FMD is presented using an improved variant of YOLOv4-tiny. In [165], the superiority of YOLOv4-tiny against YOLOv4 has been demonstrated regarding the recall and frame per second (FPS) processing. In [166], ETL-YOLOv4, an improved version of YOLOv4-tiny, is introduced for FMD tasks. In [167], the performance of YOLOv4-Tiny for FMD is investigated and compared with YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. In [168], a lightweight FMD approach is proposed based on an improved YOLOv4-tiny. Moving on, an SPP structure, the backbone network of YOLOv4-tiny, is designed for pooling and fusing the input features at multi-scale. This helps improve the receptive field of the network before combining these multi-scale features with the path aggregation network to repeatedly fuse and enhance them in two paths and then enhance the expressive ability of feature representations. Lastly, label smoothing is utilized to optimize the loss function and overcome over-fitting. In the end, it is worth noting that the YOLO series has been the most widely used detector to build FMD systems. Figure 6 portrays the number of FMD research studies conducted based on the different YOLO series detectors.

**NasNetMobile:** Chavda et al. [65] presented a deep learning-based model to detect people who do not wear face masks properly. In their study, where NasNet-Mobile, DenseNet-121, and MobileNetv2 architectures were used, the highest classification accuracy was achieved via DenseNet-121 architecture, with 99.49%.



**Figure 6.** Statistics on the number of FMD studies built upon the different YOLO detectors.

**Other methods:** In [169], a lightweight-based FMD scheme based on CNN is developed. The proposed model consists of four convolutional, one fully connected, and one output layer. The performance of this approach has been compared against that of MobileNet, NasNetMobile, ResNet101, VGG19, VGG16, and AlexNet, where its superiority has been shown in terms of better balance among accuracy, model size, and time complexity. In [170],

inappropriate mask use is targeted by (i) collecting 2,075 face mask usage images, (ii) labeling them as either “mask”, “no masked”, or “improper mask”, and (iii) studying three scenarios, namely “scenario 1—mask versus no mask versus improper mask”, “scenario 2—mask versus no mask + improper mask”, and “scenario 3—mask versus no mask”. A hybrid deep feature-based face mask detector is then trained and tested. The detector is implemented in three steps: (i) using pre-trained DenseNet201 and ResNet101 as feature generators, (ii) selecting the most discriminative characteristics using an improved ReliefF selector, and (iii) considering the selected characteristics for classification by an SVM.

**Table 3.** Summary of existing DL-based FMD techniques and their characteristics.

Work	Model	Description	Dataset	Best FMD Performance	Advantage/Limitation
[19]	SRCNet	Distinguish face masks using image super-resolution and classification networks.	MedMasks	Acc = 98.70%	Validation on the small test dataset. Not robust to facemask-wearing variation (frontal orientation, posture change, etc.)
[171]	Inceptionv3	FMD using InceptionV3-based TL.	SMFD	Acc = 99%	Validation on a small masked face dataset. FMD assessment in real-life video streaming is missing.
[172]	CNN	Detection of masked, non-masked, and properly-masked faces.	MFDD, RWFCDD, SMFRD.	Acc = 98.6%	Validation on real-world masked face datasets. FMD assessment in real-life video streaming is missing.
[173]	CNN	1,539 faced, and no-faced images were used to train a CNN model.	Private data	Acc = 98.7%	Offline FMD using masked face and non-masked face pictures collected using CCTV cameras. No assessment of real-world masked face datasets.
[126]	FMV3, YOLOv3, ResNet-SSD300	Using a DTL to reduce computational cost.	Private data	Acc = 98%	Offline FMD on a small face image dataset. No assessment of popular real-world masked face datasets.
[128]	R-CNN	FMD using RCNN-based object detection and comparison with SSD-MobileNetv2 and SSD-Inceptionv2 models.	COCO	Acc = 68.72%	Have an overfitting problem and Validation on datasets without assessment in real-world scenarios.
[27]	SSD, MobileNetv2	Real-time FMD using SSD and MobileNetV	KMMD, PBD, RTMMD	Acc = 93%	High computational cost.
[133]	MTCNN, MobileNetv2	FMD using (i) MTCNN-based face detection and (ii) MobileNetv2-based object in the masked region.	Private video data.	Acc = 81%	Low accuracy performance and FMD mainly relies on face detection.
[97]	YOLOv3	FMD using YOLOv3 and Google Colab.	Private image dataset (600 images)	Acc = 96%	• Validation on a small dataset without assessment in real-world scenarios.
[174]	Tiny-CNN (SqueezeNet, Modified SqueezeNet)	Low computational medical FMD and comparison of the performance with SqueezeNet and Modified SqueezeNet.	Combination of FMID-12k, FMC, and private image datasets.	Acc = 99.81%	Validation on face image datasets without assessment in real-world scenarios.
[71]	YOLO-fastest, NCNN	Edge-based FMD using YOLO-fastest, NCNN and WebAssembly.	Combination of Wider Face, MAFA, RWMFD, MMD and SMFD	mAP = 89% (YOLOv3)	Scalable solution where other lightweight DL models can be used. Requires internet connectivity, which makes it computationally expensive.
[44]	YOLOv3, YOLOv3Tiny, SSD, Faster-R-CNN	FD using various CNN architectures.	Moxa3K and MMD	mAP = 63.99% (YOLOv3)	Low detection accuracy; further improvements are required by using effective object detectors.
[142]	CNN	FMD using posture recognition and DL.	Private real-world data.	Acc = 95.8%	Validation on a limited dataset, although they were collected from real-world environments.
[151]	MobileNetv2	Low computational FMD using a lightweight CNN.	MMD, FMD	Acc = 99.75%	Validation on face image datasets without assessment in real-world scenarios.
[12]	CNN	FMD without generating over-fitting.	A balanced masked face dataset of 3832 images.	N/A	The CNN architecture has been trained using the balanced dataset and used in real-world scenarios, though without evaluation.
[150]	CNN	The CNN architecture has been made similar to MobileNetv2 for an efficient computational cost.	FMDD	Acc = 99%	MobileNetv2 has been used to classify pre-processed video frames using OpenCV. There is no comparison with other methods on the training/test splits of the FMDD dataset.
[129]	RCNN, Fast-RCNN, and Faster-RCNN	Real-time analysis of FMD and SDM.	Private dataset	Acc = 93%	Validation on a small dataset without assessment in real-world scenarios.

Table 3. Cont.

Work	Model	Description	Dataset	Best FMD Performance	Advantage/Limitation
[42]	YOLOv3, faster-RCNN	Real-time automated FMD	MAFA, Wider Face	PR = 62%	Real-time implementation is supported with YOLOv3, while its accuracy is lower than faster R-CNN.
[130]	Context-attention RCNN	Detection of faces without masks, with wrong masks, and with correct masks.	A new MAFA-based dataset is created	mAP = 84.1%	Evaluation on a small dataset (4672 images), and the performance needs further improvement.
[63]	TL-AlexNet-LSTM/BiLSTM, TL-VGG16-LSTM/BiLSTM	TL-based AlexNet and VGG16 were combined with LSTM and BiLSTM to detect the manner people use facemasks.	Private data (2000 images)	Acc = 95.66%	Face masks via real-time video recordings were not supported, and the Validation was on a small dataset.
[175]	Improved YOLOv4	FMD and standard wear detection	RMFD+, MaskedFace-Net, private data	mAP = 98.3%	Insufficient feature extraction for difficult detection samples. FMD, when the light is insufficient, was not treated.
[176]	YOLOv3, YOLOv4-tiny	The YOLO network using Darknet applied is a state-of-the-art real-time object detection system.	A novel publicly annotated dataset	mAP = 90.69%	Applied only with surgical mask (helms or shield masks)
[64]	CNN (VGG-16)	FMD in real-time with an alarm system.	A new web dataset.	Acc = 98%	The prevention by creating an alarm stipulation these rules are not observed properly.
[170]	ResNet101	Pre-trained ResNet101 and DenseNet201 are used to generate image features, a Relief selector is used to find discriminative features, and an SVM for classifying images.	MaskedFace-NET, a private dataset with three classes.	Acc = 99.75%	Not compared with existing solutions on the custom dataset.
[138]	VGG-16 and CNN	Automatic FMD system in public transportation using Raspberry Pi.	Private data	Acc = 99.4%	Not validated on public datasets, which makes it difficult to compare the performance with existing solutions.
[72]	TL-based ResNet50, YOLOv2	Using TL-based ResNet50 for feature extraction and YOLOv2 classifier to detect medical masks.	MMD, FMD	Acc = 81%	Do not discriminate between medical and normal masks.
[110]	YOLOv4	Real-time FMD using effective structures of backbone, neck, and prediction head based on YOLOv4	MAFA, WiderFace	AP = 94%	Achieve real-time FMD and suitable for dark/night environments.
[121]	Mask RCNN, YOLOv4, YOLOv5, and YOLOR	FMD using different lightweight CNN models for detecting mask-wearing in videos.	ViDMASK	mAP = 97.1% (YOLOR)	The dataset and code are publicly available. Faces are shot from various angles.
[137]	ResNet-18	FMD using triplet-consistency representation learning.	WiderFace, MAFA	L <sub>1</sub> = 91.5% (MAFA), L <sub>1</sub> = 54.1% (Wider Face)	The performance drops under noisy environments (i.e., the hard set of Wider Face), and the privacy preservation are not addressed.

#### 4.4. FMD Based on Deep Transfer Learning (DTL)

When developing FMD techniques, the authors have faced different challenges. For instance, training DL models is challenging, considering the diversity of mask types, camera angles in video frames, etc. Additionally, another issue is the unavailability of real large-scale FMD datasets (data scarcity) to train DL models. To that end, DTL has recently been adopted in many FMD solutions. DTL consists of two steps: (i) using the knowledge learned on one domain/task and (ii) transferring it to a new, similar domain/task. For example, the knowledge from detecting face masks in annotated datasets could be transferred to another face mask dataset that does not have labels [177,178]. In the context of DL, one popular way to apply transfer learning is by a series of steps that comprise: (i) taking layers from a previously trained network, (ii) freezing them to preserve the information they contain (during future training stages), (iii) adding some new and trainable layers on the top of the frozen ones (this helps in turning the old features into predictions on the target dataset), and (iv) training the new layers on the target dataset [179].

An FMD approach combining TL and DL is proposed in [180], which is based on (i) introducing an Efficient-YOLOv3-based DTL, where EfficientNet is the feature extraction backbone, (ii) using a loss function based on CIoU to improve the accuracy of face mask and decrease the number of network parameters. Moreover, the masks have been categorized into two unqualified masks (scarves, sponge masks, cotton masks, etc.) and qualified masks



(disposable medical masks, N95 masks, etc.). Also, an FMD dataset has been created, and the DTL approach has been combined with MobileNet to improve the overall solution's generalizability, solve the data scarcity problem and tackle the over-fitting problem. In [181], the YOLOv3-based DTL scheme is proposed for creating a mask detection model, where an improved FMD that includes 300 images has been created using data augmentation, i.e., image filtering. In [182], a DTL scheme based on PaddlePaddle-YOLOv3 (PP-YOLOv3) combined with data augmentation and model compression is introduced to solve an FMD task. This resulted in an mAP of 86.69% in public scenes with a speed processing time of 11.84 ms per frame. Experimental results showed that compared with YOLOv3 and FasterRCNN, the model has faster accuracy and detection speed. In [183], DTL has been applied on YOLOv3 using public datasets and donation datasets for training. The resulting models can recognize faces with a 98.7% accuracy rate and identify faces, including those with face masks, with a 92.7% accuracy rate.

*Fine-tuning:* A last, optional, step in DTL is fine-tuning, which consists of unfreezing the entire model (or part of it) and re-training it on the new data with a very low learning rate. This can potentially achieve meaningful improvements by incrementally adapting the pre-trained features to the new data. An increasing number of DTL studies have been developed to tackle different FMD tasks. For instance, in [18], an approach for detecting incorrect face mask-wearing using CNN and DTL is proposed. Accordingly, various CNN models are fine-tuned, including MobileNetv2, Xception, Inceptionv3, ResNet50, NASNet, and VGG19.

In a survey of deep learning models for the FMD task [184], the authors concluded that deep learning models such as the Inceptionv3 CNN achieved almost perfect results and stated the need for sharing real-world FMD images in order to fine-tune DL methods.

In [43], a DTL-based FMD system is proposed by initializing several CNN models with parameters learned from ImageNet dataset. Then, they are fine-tuned on FMD using a standard cross-entropy loss. Jiang et al. [122] propose the RetinaFaceMask, a single-stage FMD approach to assist in the monitoring of the COVID-19 pandemic. The approach helps to resolve some of the problems already encountered by other studies by (i) establishing a new annotated dataset that helps in fine-tuning the DL models to discriminate between correct and incorrect mask-wearing scenarios, (ii) proposing a context attention framework to learn discriminating characteristics corresponding to the states of wearing face masks, (iii) transferring the knowledge from the face detection task to FMD, emulating humans' ability to adapt to similar tasks quickly.

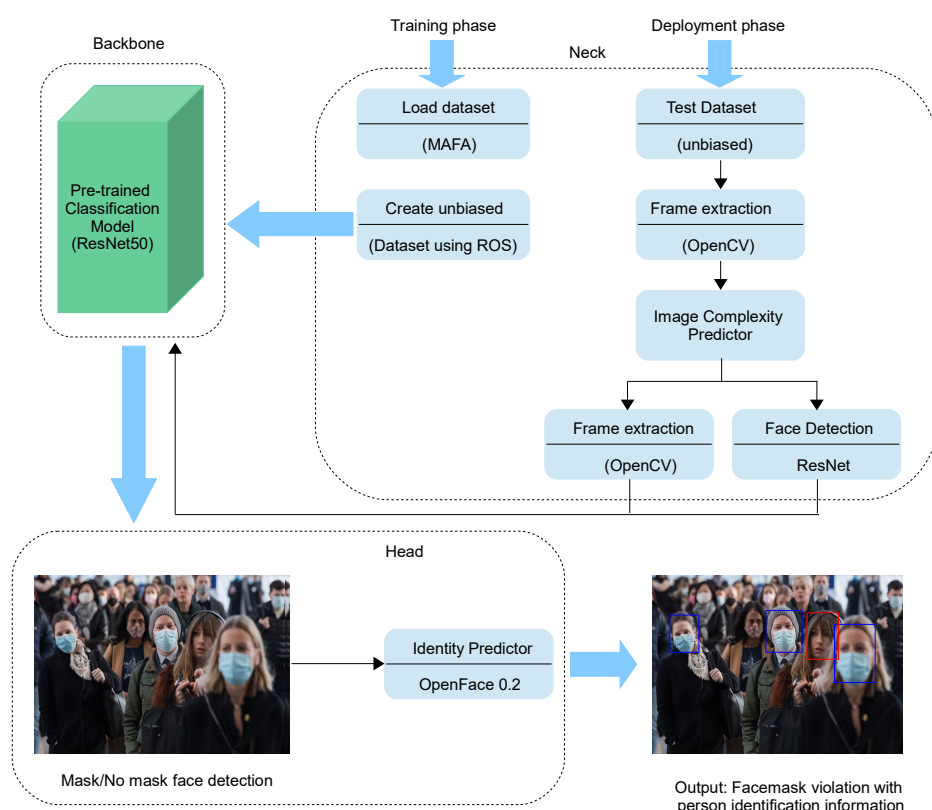
In [171], Chowdary et al. propose a DTL method on inceptionv3. They fine-tune the pre-trained Inceptionv3 and test it on an SMFD task. In [185], the pre-trained MobileNet, ResNet Classifier, and VGG are used to an FMD and SDM framework. A real-time DTL-based FMD approach for mobile devices that employ MobileNetv2 is implemented in [144]. The pre-trained MobileNetv2 network has been utilized to extract discriminative features, while a softmax classifier has been used to classify video frames into a mask or no mask. Additionally, an image augmentation technique has been deployed to avoid overfitting and improve the overall performance [186]. In [46], a hybrid TL and broad learning FMD solution are proposed, where Faster R-CNN and Inceptionv2 are combined and fine-tuned to detect face masks regions. Second, broad learning is used to verify real facial masks.

In [171], a DTL-based FMD framework is proposed by fine-tuning the pre-trained Inceptionv3 model. Moreover, data augmentation is employed to overcome the data shortage problem and improve training and testing performance. In [187], a DTL-based FMD on the Spartan Face Detection and Facial Recognition System is proposed. The approach is based on a stacking ensemble of deep learning models and covers four primary tasks: mask detection, mask type classification, mask position classification, and identity recognition. In [188], the pre-trained Faster R-CNN Inception ResNet v2 model is used to implement a DTL-based FMD, where the pre-trained model weights on the COCO datasets are used as the starting point for DTL on another FMD dataset, which includes 3300 images with different types of face masks. Besides, because the lower parts of human faces are usually occluded and cannot be utilized in the face recognition learning process,



the authors in [189] propose a framework that recognizes human faces using any available facial components. Such components may vary depending on wearing or not wearing masks and on the mask position. Typically, a DTL-based FMD approach based on an improved FaceNet model, is proposed. With DTL, the initial weights have been transferred, and the model has then been fine-tuned on the masked CASIA (M-CASIA) dataset [190]. Similarly, in [191], a Faster R-CNN model is fine-tuned and assessed for the FMD task on the FMD dataset. In [192], a DTL-FMD is introduced by fine-tuning a MobileNetv2 base model with the imageNet weights and adding a new fully-connected head to process input data for analysis. Finally, in [193], a real-time CNN-Based lightweight mobile FMD system is proposed. Three different pre-trained models, including VGG16, MobileNet, ResNet50 are fine-tuned to enhance the detection performance. For instance, the first 23 layers of the MobileNet architecture are frozen, then three more dense layers are added before fine-tuning the new model.

*Domain adaptation (DA):* It refers to the process of adapting models across different domains. This is because training and test data could have a discrepancy or fall from other data distributions. For example, when an FMD model is trained only on unmasked faces data, while it is tested on masked faces data, the performance significantly drops, and DA can be called to solve this issue. DA aims at building ML tools that are able to generalize well into a TD process and deal with the gap across domain distributions. In this regard, some studies have attempted to investigate the roles of DA for better generalization of developed models. For instance, in [135] and the FMD scheme aggregating both one-stage and two-stage object detectors are proposed to achieve accurate and real-time mask detection. Typically, ResNet50 is first adopted as a baseline before applying a DTL technique. Figure 7 illustrates the architecture of the FMD system using ResNet50-based TL.



**Figure 7.** A ResNet50-based TL architecture for FMD. Adapted from: [135].

In [194], Mandal et al. propose a supervised DA to boost the performance of a MFR scheme. In doing so, faces without masks are considered in the source domain (SD), while faces with masks have been reserved for the target domain (TD). Moving on, a ResNet50 model has been trained and validated under two case studies. The model has been trained

only on the SD and the tested on the TD in the first scenario. In the second scenario, the model is trained on the SD and a portion of TD before being tested on the remaining part of TD. Table 4 summarizes most relevant DTL-based FMD frameworks and their characteristics.

**Table 4.** Summary of existing DTL-based FMD frameworks and their characteristics.

Work	DTL Model	Description	Dataset	Best FMD Performance	Advantage/Limitation
[122]	TL-based ResNet50 and MobileNetv1	FMD using ResNet or MobileNet as the backbone, FPN as the neck, and context attention modules as the heads.	MAFA + FMD	Acc = 91.9% (ResNet)	No assessment on real-world masked face datasets.
[195]	Transfer learning	Relies on adopting transfer learning to detect face masks in both images and video streams.	RMFD	Acc = 98%	(i) Works on a variety of devices (e.g., smartphones, etc.) and is also able to process in real-time images and video streams, (ii) the approach is not well interpretable activation since they do not use activation maps.
[171]	Inceptionv3-based DTL	FMD using Inceptionv3-based DTL.	SMFD	Acc = 99%	Validation on a small masked face dataset. FMD assessment in real-life video streaming is missing.
[18]	MobileNetv2, Xception, Inceptionv3, ResNet50, NASNet, VGG19	Detection of incorrect face mask-wearing using CNN and DTL.	Private data	Acc = 83%	(i) Implemented via Android app that works with real scenarios, and the solution can identify mask misuse, (ii) Unable to detect incorrect lateral adjustment and glasses underneath. Moreover, the system was applied with surgical and FP2 masks. Not applied with masks that have sequins and other drawings.
[196]	DTL based on combining SVM and MobileNetv2	MFD using deep feature selection and award-winning pre-trained DL models.	Collected data of 1376 images	Acc = 97.1%	Tested on a small-sized dataset. Not tested on the challenging occluded face.
[197]	YOLOv3 and Darknet53	Data augmentation and DTL for FMD.	data collected from Kaggle.	Prec = 99.8%	The automated system detects masks using an augmented dataset.
[198]	VGG-19 transfer learning DCNN	A software model that could be used in existing surveillance applications.	FMDC	Acc = 98%	(i) Implemented via live feed footage in IP cameras. (ii) Tested on the artificially created dataset. Not suitable with a web server or linkage of multiple IP cameras.
[189]	Improved FaceNet	FMD using residual inception networks.	M-CASIA	Acc = 99.2%	Validated on a simulated dataset; however, any unrealistic part in the simulated images might cause some inaccuracies in the recognition.
[191]	Faster-RCNN	Automated real-time FMD using Faster-RCNN-based DTL.	FMD	AvPrec = 81%, AvRec = 84%	The performance needs further improvement.
[192]	MobileNetv2	Using DTL and fine-tuning to detect face masks.	Private data	Acc = 98.2%	Validation of public dataset is required to compare the performance with other existing FMD solutions.
[193]	VGG16, MobileNetv1, ResNet50	Development of a real-time CNN-based lightweight mobile FMD system.	VGGFace2, MaskedFace-Net	Acc = 99.6%	Less computational power and resources are required using MobilNetv1.
[48]	VGG-16, MobileNetv2, Inceptionv3, ResNet50, and CNN	IoT- and DTL-based FMD scheme for rapid screening.	MAFA, Masked Face-Net, and Bing	Acc = 99.81% (VGG-16)	Validated on public datasets that mostly have artificially created and noisy face mask images.
[171]	Inceptionv3	FMD using TL and image augmentation.	SMFD	Acc = 100%	Evaluation on large-scale real-world datasets is required. The type of mask cannot be detected.
[194]	ResNet50-based DA	MFR using DA.	Private data	F1 = 89.7% (unmasked), F1 = 44.73% (masked)	Have problems in detecting masked faces. The dataset used is unbalanced.

Other approaches to domain adaptation for FMD include using adversarial training or cycle consistency to align the feature distributions between the source and target domains or using domain-invariant features that are less sensitive to domain shift.

## 5. Evaluation, Discussion and Findings

### 5.1. Comparative Analysis

Singh et al. [42] utilize two object detectors, e.g., YOLOv3 and Faster R-CNN FMD. While Faster R-CNN has outperformed YOLOv3 in terms of accuracy, YOLOv3 has pre-

sented can run faster, which makes it a good fit for real-time applications. Overall, selecting the best model for a specific application is mainly related to environmental conditions. Moving on, the work conducted by Alganci et al. [199] has resulted in similar conclusions. Besides, SSD, Faster R-CNN, YOLOv3, and YOLOv3-Tiny are employed in [44] to address the issues of FMD in medical environments. Based on the results derived from the Moxa3K dataset, YOLOv3-Tiny outperforms the other models in terms of accuracy and inference speed, which makes it the most appropriate for real-time applications. Table 5 compares existing FMD frameworks in terms of their performance and characteristics.

**Table 5.** Comparison of existing FMD frameworks in terms of different characteristics, (i) adopted method, (ii) support of real-time implementation, (iii) used dataset, (iv) sample size, (v) mask type, (vi) recognition type, and (vii) validation accuracy.

Work	Method	Real-Time	Dataset	Sample Size	Mask Type	Recognition Type	Validation Accuracy
[200]	PCA	No	ORL Face	400	Real	Masked/Unmasked	95%
[33]	Latent part detection	No	CASIA-WebFace	4,916	Augmented	Masked/Unmasked	97.94%
[65]	CNN	No	Private data	7855	Mix	Masked/Unmasked	99.45%
[19]	CNN	No	KMMD	3835, 134, 3030	Real	Correct/Incorrect/Unmasked	98.70%
[201]	Mixture of Gaussians	Yes	Private data	N/A	Real	Masked/Unmasked	95%
[134]	ResNet50-based DTL	No	RMFD, SMFD, LFW	10,000, 1570, 13,000	Mix	Masked/Unmasked	100%
[18]	CNN	No	Private data	3200	Real	Correct/Incorrect/Unmasked	83%
[20]	MobileNet	No	Private data	770, 500	Augmented	Correct/Incorrect/Unmasked	90%
[21]	Haar feature cascade	No	CMFD, IMFD	137,016	Mix	Correct/Incorrect/Unmasked	-
[171]	Inceptionv3	No	SMFD	1570	Augmented	Masked/Unmasked	99.90%
[193]	MobileNetv2-based DTL	Yes	VGGFace, Tailored dataset	1,022,811, 1849	Mix	Correct/Incorrect/Unmasked	99.96%

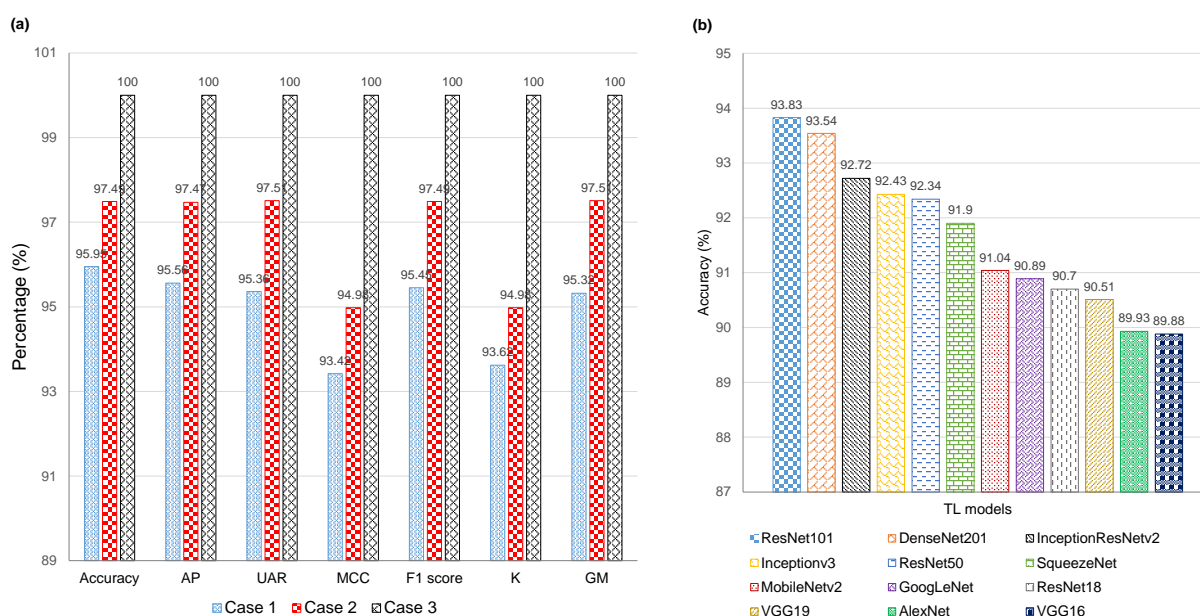
In [170], pre-trained ResNet101 and DenseNet201 have been fine-tuned and aggregated to form an efficient feature generation module tested on three different scenarios. In Scenario 1, 2075 images are used to form a three-class classification task that is investigated by considering three individual classes, i.e., mask, no mask, and improper mask. Moving on, wrong mask and no mask classes are combined in Scenario 2 to form a two-class classification task. Hence, a “non-compliance” set is formed using 2075 images. Similarly, by excluding improper mask sets, another two-class classification problem is developed in Scenario 3 while only 1546 are used. Figure 8a reports obtained results in terms of different evaluation criteria, including the accuracy (ACC), average precision (AP), unweighted average recall (UAR), Mathew correlation coefficient (MCC), F1 score, Cohen’s kappa (CK), and geometric mean (GM) [71,72].

Besides, various pre-trained CNN models have been fin-tuned and applied as feature generators to solve the problem of FMD under the first scenario. Figure 8b portrays the accuracy performance of 12 pre-trained CNN models.

## 5.2. Critical Discussion

The detection of face masks is challenging due to the variations in the appearance of faces with masks, including the types and degrees of obstruction, as well as the diverse types of masks that are used. Face mask detection is essential for facilitating interactions between humans and computers and managing image databases. Despite the successes of existing face detectors, there is a need for more advanced models that can handle event analysis and video surveillance tasks, which can be challenging due to a lack of suitable datasets with correctly masked faces and facial recognition, as well as the presence of noise on the face caused by the mask. While some research has addressed these issues, there is still a need for a large dataset to develop an efficient face mask detection model. Overall, there are several current challenges to the task of FMD:

- Lack of suitable datasets: one major challenge is the Lack of datasets with a sufficient number of images of faces with masks, as well as a diverse range of mask types and wearing conditions. This can make it difficult to train and evaluate face mask detection algorithms, as the performance of these algorithms is often dependent on the quality and diversity of the training data.
- Small intra-class distance and significant inter-class distance: another challenge is the small intra-class distance and large inter-class distance between masked and non-masked faces, making it difficult to accurately distinguish between these two classes. This may require specialized algorithms or techniques that can extract distinguishing features and increase the separation between these classes.
- Noise caused by masks: the presence of masks on the face can also introduce noise that can interfere with the performance of face mask detection algorithms. This may be due to factors such as the mask's texture, reflections or shadows, and the occlusion of facial features.
- Variability in mask appearance: face masks come in a wide variety of shapes, sizes, and colors, and they may also be worn in different ways (e.g., covering the nose and mouth, covering only the nose, or hanging around the neck). This variability can make it challenging for a model to detect masks accurately.
- Occlusions: face masks can occlude parts of the face, making it difficult for the model to identify features such as the eyes, nose, and mouth.
- Lighting and background: the model may have difficulty detecting masks in low light conditions or against cluttered or complex backgrounds.
- False positives and false negatives: the model needs to minimize false positives (incorrectly identifying a mask when none is present) and false negatives (failing to identify a mask when one is present).
- Real-time performance: there is also a need for face mask detection algorithms that are able to perform in real time, as these algorithms may be used in applications such as surveillance or event analysis, where speed is critical.
- Adversarial examples: an attacker can create "adversarial examples" (images specifically designed to fool the model) that could cause the model to make incorrect predictions.



**Figure 8.** Performance comparison of DTL-based CNN models: (a) Performance of combining pre-trained ResNet101 and DenseNet201 under three different scenarios, and (b) accuracy performance of 12 pre-trained CNN models under scenario 1. Data from: [170].

Typically, these challenges highlight the need for developing robust and efficient face mask detection algorithms that can handle a wide range of conditions and variations in the appearance of masked faces. Besides, the use of masks has been shown to reduce the infection rate of COVID-19 by 40% effectively, but detecting the wearing of masks in the real world can be challenging due to factors such as lighting conditions, occlusion, and the presence of multiple objects. This can lead to poor detection performance, and using non-medical masks such as cotton masks, sponge masks, and scarves may also reduce the protective effect of mask-wearing.

## 6. Open Challenges

Over the past two years, researchers have proposed various deep learning-based methods for face mask detection. However, many of these methods have struggled with detecting small or poorly shot masks occluded by other objects, leading to low detection accuracy.

### 6.1. Lack of Annotated Datasets

A major challenge for researchers is the Lack of properly annotated real datasets containing images of faces with and without masks shot under various conditions (e.g., during the day or night, indoors or outdoors, from a distance or close, etc.). Most of the datasets that were presented in Section 2.2 and summarized in Table 1 either contain synthetic images of face masks, where masks have been painted over the faces or contain properly shot and cropped face images, without distortion or noise, which is typical for the images collected by surveillance cameras. The datasets that contain images from such cameras mostly refer to the social distancing detection task, and it is hard to detect mask-wearing on them.

Another issue that is not properly addressed yet is that face masks (including those used for preventing COVID-19 diffusion) come in different shapes and colors. There are still no properly annotated datasets that cover this. Most of the existing datasets assume the typical white or teal medical masks only and ignore all other fabric-based or multi-colored medical masks that became so popular in the last few years [137].

Ignoring all the above issues and creating synthetic datasets that reproduce the researchers' beliefs on how a face mask looks, is shaped or is worn, we introduce a bias to the dataset, which may hinder the trained model's performance in real conditions.

### 6.2. Computational Cost

There are many challenges to using computational platforms for face mask detection (FMD). The hardware used for this purpose must be affordable, compact, and energy-efficient, with enough memory and processing power to quickly analyze images using convolutional neural networks (CNNs) or other models. To protect the privacy of the people being monitored, all processing should be done on the device itself, without any communication with cloud servers. One potential solution to these challenges is to use low-power binary neural network classifiers to detect face masks' presence and proper positioning. These classifiers can be implemented on field-programmable gate arrays (FPGAs), capable of high-throughput binary operations, and can perform FMD classification tasks on an edge device.

Low-power binary NN-based COVID-19 FMD, such as BinaryCoP [202], can be employed to identify correct facial mask wearing and positioning using edge computing and mobile devices. Another solution is to use light-weight CNN models, such that in [203], for performing multiple access control-related tasks, including the detection of face masks, the monitoring of body temperature, and the counting of people entering a building or an area (e.g., a concert hall). Such models can be trained offline and used for inference on single-board computers (SBCs) such as Raspberry Pi, Jetson Nano, or others. The performance of edge-based solutions in FMD tasks in indoor and outdoor environments has been proven better than their cloud-based alternatives [204], with the respective time performance for



image processing slightly worse. However, if we also consider the time needed to transfer images to the cloud, edge-based solutions are almost 50% faster.

Cooperative edge computing [205], task offloading, and federated learning (FL) [124] are a few of the techniques that are expected to gain the interest of researchers in the next few years since they preserve data privacy transfer processing to the edge and minimize transfer times.

### 6.3. Security and Privacy

Large-scale monitoring and FMD of individuals in public areas do not only need well-performing AI tools but also require to include privacy preservation modules. This is because any kind of surveillance can violate citizens and regulations, including the general data protection regulation (GDPR), which sets strict regulations on the use and share of personal [11]. Although FMD developers deny privacy problems that come with FMD systems as they do not identify individuals in most cases, and detect the existence of face masks, adding privacy protection mechanisms to them will increase the users' trust and acceptance of the FMD technology.

To cope with this challenge, the authors in [11], explain a privacy-protection FMD system that might be developed, demonstrating various implementation and performance evaluation options.

### 6.4. Difficulty in Recognizing People's Emotions

Due to the obligation of wearing masks in public areas that have been set in various countries, emotion recognition has become challenging. Despite the promising progress achieved in tackling this problem, the new requirements to wear face masks compromise what has been done in recent years. New studies have been explored to detect persons' moods using DL and ML models when face masks are worn to close this gap. For instance, ref. [206] validates a real-time CNN-based emotion recognition approach on a modified version of the AffectNet dataset. Synthetic face masks have been added to each subject from the AffectNet dataset. Additionally, the number of emotions has been reduced from eight emotions (Happiness, Disgust, Anger, Fear, Surprise, Sadness, Contempt, Neutral) to five (Anger-Disgust, Fear-Surprise, Happiness, Sadness, Neutral).

### 6.5. Masked Face Attacks

With the start of the COVID-19 pandemic, face masks have significantly flattened the COVID-19 curve. However, this opens new challenges for face recognition as wearing a mask hides multiple discriminative features of a face. On the other hand, face presentation attack detection (PAD) is critical for ensuring the security of face recognition systems. Although the increasing number of FMD frameworks, few studies have been proposed to investigate the impact of the masked face on PAD. For instance, to reflect the actual real-world situations, Fang et al. [207] investigate (i) attacks with subjects wearing masks and (ii) attacks with real face masks placed on presentations. Moving on, in [208], a physical universal adversarial perturbation (UAP), namely the adversarial mask, is proposed to simulate attacks against face recognition systems, which are deployed on face masks using carefully crafted patterns.

## 7. Future Directions

### 7.1. Interpretability and Explainability

Explainable Artificial Intelligence (XAI) refers to methods and techniques for creating AI systems that can provide understandable and interpretable explanations for their predictions, decisions, and actions. XAI aims to build AI systems that can be trusted and used safely and effectively by humans. This can be achieved through techniques such as transparency, interpretability, and post-hoc explanations [209,210].

As it is made evident from this study, the accuracy of face mask detection algorithms is the main concern of researchers, along with the complexity of the models and the respective training and inference (time) performance. Although neural network architectures have



significantly boosted the prediction/detection performance over their ML predecessors, the research community still needs to understand deep learning architectures, break down the structure of pre-trained models, and show how predictions are made. As a consequence, the results of FMD models are expected to be interpretable, and explainable [211], and this is in accordance with the general trend for interpretability and XAI. The works in the field of face recognition aim to develop comprehensible facial depictions in which different dimensions represent different face segments or features and propose end-to-end learnable filters, which are locally activated using spatial loss functions, such as the spatial activation diversity loss function in [212]. The generation of expression pattern-maps and their association with expression features is another approach that improves the interpretability of facial expression detection tasks [213]. Such approaches can be further combined with ML or DL classifiers in a two-stage approach and provide interpretable FMD and robustness in the cases of partial face occlusion.

In an attempt to explain the results of the facial matching process, authors in [214] have defined a new evaluation protocol called the “inpainting game”, and introduced an explainable face recognition (XFR) algorithm that explains which facial characteristics (e.g., nose, eyebrows, etc.) are matched in each case. The XFR algorithm generates network attention maps for facial images consistent across the human participants and provides insight into the features that make each face unique. In a similar line of research, saliency maps can be employed to explain face matching by measuring the contribution of different face parts [215]. Examining the effect of perturbation, occlusion, or noise on each of the parts alone or in a combination of parts is a first step towards the explainability of face recognition tasks. Such techniques can be combined with FMD algorithms to provide explainable decisions that link the detection of a mask (or proper mask-wearing) with the coverage of specific facial features (i.e., nose, mouth).

### 7.2. Further Generalization for FMD Techniques

The current study has revealed the large variety of techniques and models for FMD and the plethora of datasets used for their training. Many studies are trained and validated on custom and private datasets, whereas others employ more significant benchmarks. The main characteristic of the datasets is the variety of images they contain (e.g., controlled or uncontrolled poses, real or synthetic images, two or more classes, etc.), and this consequently affects the models being trained and limits their ability to generalize to more, but relevant, tasks.

Despite the many works, datasets, and models trained for the general image classification task, such as VGG-16, ResNet50, Inceptionv3, or EfficientNet, or the works that focus on building generalized models for face detection, such as MTCNN or FaceNet [216] there is still an open need for generalized pre-trained models that detect masked faces and extract their features. Research in this direction will allow deep transfer learning techniques to be applied to multiple FMD-related tasks and boost the performance of lightweight models trained with only a few training samples [217–219]. The first works that apply transfer learning of Inceptionv3 [171], RetinaFace [65], MobileNet [220], and similar networks show the way for future researchers that want to build general models for the Faced Mask Detection task.

Transformers can also be employed in the specific FMD task, following the successful paradigms in the generic image classification task [221,222]. The resulting models can be further trained using only a small number of images for the task and achieve state-of-the-art performance [223].

### 7.3. Federated FMD

FL is an ML technique where multiple decentralized devices, such as smartphones or edge devices, work together to improve a shared model without sharing their raw data. The devices train a local version of the model on their data and then share the updates with a central server, aggregating the updates to improve the global model. This allows for training models on distributed data without the need to centralize and share

the data [224]. The need to continuously monitor the proper use of masks, especially in public areas, and more importantly, without breaching citizens' privacy requests the implementation of lightweight detection methods. The latter can process video streams on edge in real-time without storing any information locally. Solutions, such as WearMask [71], or BinaryCoP [202] apply transfer learning to very fast and lightweight models such as YOLO-Fastest or adopt fast binarized NN inference frameworks such as Finn [225], and operate as mask-wearing sensors. They can be deployed as mobile apps or run on FPGA-powered hardware that allows high-speed face detection and classification in two stages [226].

FL is another approach that can help protect user privacy by design and, simultaneously, by training DL or ML models locally using sensitive data, such as people's faces. By splitting the FMD task into two or more cascade stages, such as the location of faces in images, the extraction of image features, and the classification of the detected faces as masked and non-masked (and/or improperly masked), it would be easier to federate the learning task [124] and take advantage of locally trained models. Similar architectures can be employed for social distancing screening [227], detecting littered face masks [228], etc. Besides, another way to implement FL for face mask detection is to have multiple parties contribute to the model training by providing their datasets of images. Each party could train a local model on their data and then send the model parameters to a central server, aggregating the parameters and using them to update the global model. This process could be repeated iteratively until the worldwide model has learned to detect masks accurately from the combined datasets [229].

Moreover, there are many potential benefits to using FL for face mask detection. For example, it can allow organizations to train a model without requiring access to sensitive or personal data. It can also enable organizations to collaborate on ML projects without sharing data [230]. Additionally, FL can be used to train models on data distributed across multiple devices, such as smartphones or cameras, which can be useful for real-time face mask detection in various settings [231].

Lastly, it is worth noting that all the aforementioned technologies constitute the future of face detection and identification tasks and consequently find their application to face mask detection. An in-depth knowledge of the existing literature as presented in the current study and a careful examination of the open challenges and future solutions is the first step for researchers that wish to contribute to this field. The next step involves the collection of benchmark datasets that can be used for training and fine-tuning the new state-of-the-art models.

## 8. Conclusions

The use of FMD systems has become increasingly important in recent times due to the COVID-19 pandemic. Wearing a face mask is an effective way to reduce the spread of the virus. Many businesses and organizations have implemented policies requiring face masks in public spaces. Face mask detection systems can help enforce these policies and ensure that people follow the guidelines to protect themselves and others from the virus. This paper comprehensively reviewed FMD based on deep and transfer learning models. It was clear that DL and TL have the potential to be practical tools for FMD, with the ability to handle a wide range of variations in the appearance of masked faces and achieve high accuracy rates in many cases. At the same time, it is essential to note that any FMD algorithm's performance is likely to depend on the quality and diversity of the training data, as well as the specific design and implementation of the algorithm. As such, further research is needed to fully understand the limitations and potential of deep learning for face mask detection and to develop robust and efficient algorithms that can handle real-world conditions and variations.

On the other hand, it was demonstrated that the FMD is a multi-faceted task that poses various challenges to computer vision and ML engineers. The challenges related to the proper definition of the task are detecting a covered face or a mask that has been properly worn, which affects the model that has to be trained to handle the respective task properly.

Another challenge relates to the scale of the face and face mask detection, which spans from a single person in controlled conditions to the detection in the wild, usually in combination with the detection of proper social distancing. Different algorithms apply to the massive identification of persons in a video or image stream and the detection of mask-wearing than in the case of face mask-wearing detection in controlled environments and single-person images. However, the biggest challenge relies on the scalability of solutions, which must handle video streams with dozens of persons in the case of public surveillance applications. This, in turn, sets the need for distributed processing using lightweight models and for federated training of the respective models.

The future of face mask detection in intelligent cities resides in using generic models that will be fine-tuned on specific detection tasks, using fewer resources for training and processing. The FL approach will take advantage of the resources available on edge devices and allow lightweight models to better and faster adapt to the emerging needs of each task.

**Author Contributions:** Conceptualization, Y.H. and I.V.; methodology, Y.H. and I.V.; software, Y.H. and I.V.; validation, Y.H. and I.V.; formal analysis, S.A.-M.; investigation, N.A.-M.; resources, K.A.; data curation, Y.H. and I.V.; writing—original draft preparation, Y.H. and I.V.; writing—review and editing, S.A.-M., N.A.-M., K.A. and A.M.; visualization, Y.H. and I.V.; supervision, S.A.-M.; project administration, S.A.-M. and K.A.; funding acquisition, S.A.-M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research work was made possible by research grant support (QUEX-CENG-SCDL-19/20-1) from Supreme Committee for Delivery and Legacy (SC) in Qatar. The statements made herein are solely the responsibility of the authors.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Note

<sup>1</sup> <https://www.kaggle.com/vtech6/medical-masks-dataset>, accessed on 5 January 2022.

## References

- Hyysalo, J.; Dasanayake, S.; Hannu, J.; Schuss, C.; Rajanen, M.; Leppänen, T.; Doermann, D.; Sauvola, J. Smart mask—Wearable IoT solution for improved protection and personal health. *Internet Things* **2022**, *18*, 100511. [CrossRef]
- WHO Coronavirus (COVID-19) Dashboard. Available online: <https://covid19.who.int/> (accessed on 13 January 2022).
- Himeur, Y.; Al-Maadeed, S.; Almaded, N.; Abualsaud, K.; Mohamed, A.; Khattab, T.; Elharrouss, O. Deep visual social distancing monitoring to combat COVID-19: A comprehensive survey. *Sustain. Cities Soc.* **2022**, *85*, 104064. [CrossRef]
- Fanelli, D.; Piazza, F. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos Solitons Fractals* **2020**, *134*, 109761. [CrossRef] [PubMed]
- Galbadage, T.; Peterson, B.M.; Gunasekera, R.S. Does COVID-19 spread through droplets alone? *Front. Public Health* **2020**, *8*, 163. [CrossRef] [PubMed]
- Liao, M.; Liu, H.; Wang, X.; Hu, X.; Huang, Y.; Liu, X.; Brenan, K.; Mecha, J.; Nirmalan, M.; Lu, J.R. A technical review of face mask wearing in preventing respiratory COVID-19 transmission. *Curr. Opin. Colloid Interface Sci.* **2021**, *52*, 101417. [CrossRef] [PubMed]
- Agarwal, C.; Kaur, I.; Yadav, S. Hybrid CNN-SVM Model for Face Mask Detector to Protect from COVID-19. In *Artificial Intelligence on Medical Data*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 419–426.
- Elharrouss, O.; Al-Maadeed, S.; Subramanian, N.; Ottakath, N.; Almaadeed, N.; Himeur, Y. Panoptic segmentation: A review. *arXiv* **2021**, arXiv:2111.10250.
- Liberatori, B.; Mami, C.A.; Santacatterina, G.; Zulich, M.; Pellegrino, F.A. YOLO-Based Face Mask Detection on Low-End Devices Using Pruning and Quantization. In Proceedings of the 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 23–27 May 2022; pp. 900–905.
- Wakchaure, A.; Kanawade, P.; Jawale, M.; William, P.; Pawar, A. Face Mask Detection in Realtime Environment using Machine Learning based Google Cloud. In Proceedings of the 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 9–11 May 2022; pp. 557–561.
- Kühl, N.; Martin, D.; Wolff, C.; Volkamer, M. “Healthy surveillance”: Designing a concept for privacy-preserving mask recognition AI in the age of pandemics. *arXiv* **2020**, arXiv:2010.12026.
- Kaur, G.; Sinha, R.; Tiwari, P.K.; Yadav, S.K.; Pandey, P.; Raj, R.; Vashisth, A.; Rakhra, M. Face Mask Recognition System using CNN Model. *Neurosci. Inform.* **2021**, *2*, 100035. [CrossRef]

13. Mohamed, M.M.; Nessiem, M.A.; Batliner, A.; Bergler, C.; Hantke, S.; Schmitt, M.; Baird, A.; Mallol-Ragolta, A.; Karas, V.; Amiriparian, S.; et al. Face mask recognition from audio: The MASC database and an overview on the mask challenge. *Pattern Recognit.* **2022**, *122*, 108361. [\[CrossRef\]](#)
14. Mohamed, S.K.; Abdel Samee, B.E. Social Distancing Model Utilizing Machine Learning Techniques. In *Advances in Data Science and Intelligent Data Communication Technologies for COVID-19*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 41–53.
15. Selvakarthi, D.; Sivabalaselvamani, D.; Ashwath, S.; Kalaivanan, A.A.; Manikandan, K.; Pradeep, C. Experimental Analysis using Deep Learning Techniques for Safety and Riskless Transport-A Sustainable Mobility Environment for Post Covid-19. In *Proceedings of the 2021 6th International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, 20–22 January 2021; pp. 980–984.
16. Ray, S.; Das, S.; Sen, A. An intelligent vision system for monitoring security and surveillance of ATM. In *Proceedings of the 2015 Annual IEEE India Conference (INDICON)*, New Delhi, India, 17–20 December 2015; pp. 1–5.
17. Chen, Q.; Sang, L. Face-mask recognition for fraud prevention using Gaussian mixture model. *J. Vis. Commun. Image Represent.* **2018**, *55*, 795–801. [\[CrossRef\]](#)
18. Tomás, J.; Rego, A.; Viciano-Tudela, S.; Lloret, J. Incorrect facemask-wearing detection using convolutional neural networks with transfer learning. *Healthcare* **2021**, *9*, 1050. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Qin, B.; Li, D. Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19. *Sensors* **2020**, *20*, 5236. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Rudraraju, S.R.; Suryadevara, N.K.; Negi, A. Face mask detection at the fog computing gateway. In *Proceedings of the 2020 15th Conference on Computer Science and Information Systems (FedCSIS)*, Sofia, Bulgaria, 6–9 September 2020; pp. 521–524.
21. Cabani, A.; Hammoudi, K.; Benhabiles, H.; Melkemi, M. MaskedFace-Net—A dataset of correctly/incorrectly masked face images in the context of COVID-19. *Smart Health* **2021**, *19*, 100144. [\[CrossRef\]](#)
22. Asghar, M.Z.; Albogamy, F.R.; Al-Rakhami, M.S.; Asghar, J.; Rahmat, M.K.; Alam, M.M.; Lajis, A.; Nasir, H.M. Facial Mask Detection Using Depthwise Separable Convolutional Neural Network Model During COVID-19 Pandemic. *Front. Public Health* **2022**, *10*, 855254. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Jeevan, G.; Zacharias, G.C.; Nair, M.S.; Rajan, J. An empirical study of the impact of masks on face recognition. *Pattern Recognit.* **2022**, *122*, 108308. [\[CrossRef\]](#)
24. Azeem, A.; Sharif, M.; Raza, M.; Murtaza, M. A survey: Face recognition techniques under partial occlusion. *Int. Arab J. Inf. Technol.* **2014**, *11*, 1–10.
25. He, L.; Li, H.; Zhang, Q.; Sun, Z. Dynamic feature learning for partial face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7054–7063.
26. Li, Y.; Guo, K.; Lu, Y.; Liu, L. Cropping and attention based approach for masked face recognition. *Appl. Intell.* **2021**, *51*, 3012–3025. [\[CrossRef\]](#)
27. Nagrath, P.; Jain, R.; Madan, A.; Arora, R.; Kataria, P.; Hemanth, J. SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. *Sustain. Cities Soc.* **2021**, *66*, 102692. [\[CrossRef\]](#)
28. Wang, Z.; Wang, G.; Huang, B.; Xiong, Z.; Hong, Q.; Wu, H.; Yi, P.; Jiang, K.; Wang, N.; Pei, Y.; et al. Masked face recognition dataset and application. *arXiv* **2020**, arXiv:2003.09093.
29. Yang, S.; Luo, P.; Loy, C.C.; Tang, X. Wider face: A face detection benchmark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27–30 June 2016; pp. 5525–5533.
30. Ge, S.; Li, J.; Ye, Q.; Luo, Z. Detecting masked faces in the wild with lle-cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21–26 July 2017; pp. 2682–2690.
31. Bu, W.; Xiao, J.; Zhou, C.; Yang, M.; Peng, C. A cascade framework for masked face detection. In *Proceedings of the 2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, Ningbo, China, 19–21 November 2017; pp. 458–462.
32. Bhandary, P. Face Mask Dataset Dataset (FMDs). Available online: <https://www.kaggle.com/andrewmvd/face-mask-detection>. (accessed on 13 January 2022).
33. Ding, F.; Peng, P.; Huang, Y.; Geng, M.; Tian, Y. Masked face recognition with latent part detection. In *Proceedings of the 28th ACM International Conference on Multimedia*, Seattle, WA, USA, 12–16 October 2020; pp. 2281–2289.
34. Witkowski, M. Medical Masks Dataset (MMD). Available online: <https://humansintheloop.org/resources/datasets/medical-mask-dataset/>. (accessed on 13 January 2022).
35. Jangra, A. Face Mask 12k Images Dataset. Available online: <https://www.kaggle.com/datasets/ashishjangra27/face-mask-12k-images-dataset>. (accessed on 13 January 2022).
36. Makwana, D. Face Mask Classification. Available online: <https://www.kaggle.com/datasets/dhruvmak/face-mask-detection>. (accessed on 13 January 2022).
37. Real-World Masked Face Dataset, RMFD. Available online: <https://github.com/X-zhangyang/Real-World-Masked-Face-Dataset>. (accessed on 13 January 2022).
38. Dey, S.K.; Howlader, A.; Deb, C. MobileNet mask: A multi-phase face mask detection model to prevent person-to-person transmission of SARS-CoV-2. In *International Conference on Trends in Computational and Cognitive Engineering*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 603–613.
39. Bhandary, P. Simulated Masked Face Dataset. Available online: <https://github.com/prajnasb/observations>. (accessed on 13 January 2022).



40. Queiroz, L.; Oliveira, H.; Yanushkevich, S. Thermal-mask—a dataset for facial mask detection and breathing rate measurement. In Proceedings of the 2021 International Conference on Information and Digital Technologies (IDT), Zilina, Slovakia, 22–24 June 2021; pp. 142–151.
41. Ullah, N.; Javed, A.; Ghazanfar, M.A.; Alsufyani, A.; Bourouis, S. A novel DeepMaskNet model for face mask detection and masked facial recognition. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 9905–9914. [\[CrossRef\]](#)
42. Singh, S.; Ahuja, U.; Kumar, M.; Kumar, K.; Sachdeva, M. Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment. *Multimed. Tools Appl.* **2021**, *80*, 19753–19768. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Batagelj, B.; Peer, P.; Štruc, V.; Dobrišek, S. How to Correctly Detect Face-Masks for COVID-19 from Visual Information? *Appl. Sci.* **2021**, *11*, 2070. [\[CrossRef\]](#)
44. Roy, B.; Nandy, S.; Ghosh, D.; Dutta, D.; Biswas, P.; Das, T. MOXA: A deep learning based unmanned approach for real-time monitoring of people wearing medical masks. *Trans. Indian Natl. Acad. Eng.* **2020**, *5*, 509–518. [\[CrossRef\]](#)
45. Irem Eyiokur, F.; Kemal Ekenel, H.; Waibel, A. A Computer Vision System to Help Prevent the Transmission of COVID-19. *arXiv* **2021**, arXiv:2103.08773.
46. Wang, B.; Zhao, Y.; Chen, C.P. Hybrid transfer learning and broad learning system for wearing mask detection in the covid-19 era. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [\[CrossRef\]](#)
47. Jiang, X.; Gao, T.; Zhu, Z.; Zhao, Y. Real-time face mask detection method based on YOLOv3. *Electronics* **2021**, *10*, 837. [\[CrossRef\]](#)
48. Hussain, S.; Yu, Y.; Ayoub, M.; Khan, A.; Rehman, R.; Wahid, J.A.; Hou, W. IoT and deep learning based approach for rapid screening and face mask detection for infection spread control of COVID-19. *Appl. Sci.* **2021**, *11*, 3495. [\[CrossRef\]](#)
49. Nowrin, A.; Afroz, S.; Rahman, M.S.; Mahmud, I.; Cho, Y.Z. Comprehensive review on facemask detection techniques in the context of covid-19. *IEEE Access* **2021**, *9*, 106839–106864. [\[CrossRef\]](#)
50. Mita, T.; Kaneko, T.; Hori, O. Joint haar-like features for face detection. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, Beijing, China, 17–21 October 2005; Volume 2, pp. 1619–1626.
51. Jauhari, A.; Anamisa, D.; Negara, Y. Detection system of facial patterns with masks in new normal based on the Viola Jones method. *J. Phys. Conf. Ser.* **2021**, *1836*, 012035. [\[CrossRef\]](#)
52. Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer society Conference on Computer Vision and Pattern Recognition, CVPR 2001, Kauai, HI, USA, 8–14 December 2001; Volume 1, p. I.
53. Chelbi, S.; Mekhmoukh, A. A practical implementation of mask detection for COVID-19 using face detection and histogram of oriented gradients. *Aust. J. Electr. Electron. Eng.* **2022**, *19*, 129–136. [\[CrossRef\]](#)
54. Yoon, S.M.; Kee, S.C. Detection of Partially Occluded Face Using Support Vector Machines. In Proceedings of the MVA, Nara, Japan, 11–13 December 2002; pp. 546–549.
55. Sharifara, A.; Rahim, M.S.M.; Anisi, Y. A general review of human face detection including a study of neural networks and Haar feature-based cascade classifier in face detection. In Proceedings of the 2014 International Symposium on Biometrics and Security Technologies (ISBAST), Kuala Lumpur, Malaysia, 26–27 August 2014; pp. 73–78.
56. Colombo, A.; Cusano, C.; Schettini, R. Gappy PCA classification for occlusion tolerant 3D face detection. *J. Math. Imaging Vis.* **2009**, *35*, 193–207. [\[CrossRef\]](#)
57. Ichikawa, K.; Mita, T.; Hori, O.; Kobayashi, T. Component-based face detection method for various types of occluded faces. In Proceedings of the 2008 3rd International Symposium on Communications, Control and Signal Processing, Saint Julian's, Malta, 12–14 March 2008; pp. 538–543.
58. Thom, N.; Hand, E.M. Facial attribute recognition: A survey. In *Computer Vision: A Reference Guide*; Springer: Cham, Switzerland, 2020; pp. 1–13.
59. Bhandari, M.; Shahi, T.B.; Siku, B.; Neupane, A. Explanatory classification of CXR images into COVID-19, Pneumonia and Tuberculosis using deep learning and XAI. *Comput. Biol. Med.* **2022**, *150*, 106156. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Li, J.; Jin, K.; Zhou, D.; Kubota, N.; Ju, Z. Attention mechanism-based CNN for facial expression recognition. *Neurocomputing* **2020**, *411*, 340–350. [\[CrossRef\]](#)
61. Sayed, A.N.; Himeur, Y.; Bensaali, F. Deep and transfer learning for building occupancy detection: A review and comparative analysis. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105254. [\[CrossRef\]](#)
62. Himeur, Y.; Elnour, M.; Fadli, F.; Meskin, N.; Petri, I.; Rezgui, Y.; Bensaali, F.; Amira, A. Next-generation energy systems for sustainable smart cities: Roles of transfer learning. *Sustain. Cities Soc.* **2022**, *85*, 104059. [\[CrossRef\]](#)
63. Koklu, M.; Cinar, I.; Taspinar, Y.S. CNN-based bi-directional and directional long-short term memory network for determination of face mask. *Biomed. Signal Process. Control* **2022**, *71*, 103216. [\[CrossRef\]](#)
64. Militante, S.V.; Dionisio, N.V. Deep Learning Implementation of Facemask and Physical Distancing Detection with Alarm Systems. In Proceedings of the 2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE), Surabaya, Indonesia, 3–4 October 2020; pp. 1–5.
65. Chavda, A.; Dsouza, J.; Badgujar, S.; Damani, A. Multi-stage cnn architecture for face mask detection. In Proceedings of the 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 2–4 April 2021; pp. 1–8.
66. Din, N.U.; Javed, K.; Bae, S.; Yi, J. A novel GAN-based network for unmasking of masked face. *IEEE Access* **2020**, *8*, 44276–44287. [\[CrossRef\]](#)

67. Geng, M.; Peng, P.; Huang, Y.; Tian, Y. Masked face recognition with generative data augmentation and domain constrained ranking. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 2246–2254.
68. Ge, S.; Li, C.; Zhao, S.; Zeng, D. Occluded face recognition in the wild by identity-diversity inpainting. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 3387–3397. [[CrossRef](#)]
69. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In *Computer Vision—ECCV 2016. ECCV 2016; Lecture Notes in Computer Science*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer: Cham, Switzerland, 2016; Volume 9905.
70. Anithadevi, N.; Abinisha, J.; Akalya, V.; Haripriya, V. An Improved SSD Object Detection Algorithm For Safe Social Distancing and Face Mask Detection In Public Areas Through Intelligent Video Analytics. In Proceedings of the 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 6–8 July 2021; pp. 1–7.
71. Wang, Z.; Wang, P.; Louis, P.C.; Wheless, L.E.; Huo, Y. Wearmask: Fast in-browser face mask detection with serverless edge computing for covid-19. *arXiv* **2021**, arXiv:2101.00784.
72. Loey, M.; Manogaran, G.; Taha, M.H.N.; Khalifa, N.E.M. Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. *Sustain. Cities Soc.* **2021**, *65*, 102600. [[CrossRef](#)]
73. Redmon, J.; Farhadi, A. YoloV3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
74. Pi, Y.; Nath, N.D.; Sampathkumar, S.; Behzadan, A.H. Deep Learning for Visual Analytics of the Spread of COVID-19 Infection in Crowded Urban Environments. *Nat. Hazards Rev.* **2021**, *22*, 04021019. [[CrossRef](#)]
75. Ahmed, I.; Ahmad, M.; Rodrigues, J.J.; Jeon, G.; Din, S. A deep learning-based social distance monitoring framework for COVID-19. *Sustain. Cities Soc.* **2021**, *65*, 102571. [[CrossRef](#)] [[PubMed](#)]
76. Shalini, G.; Margret, M.K.; Niraimathi, M.S.; Subashree, S. Social Distancing Analyzer Using Computer Vision and Deep Learning. *J. Phys. Conf. Ser.* **2021**, *1916*, 012039. [[CrossRef](#)]
77. Widiatmoko, F.; Berchmans, H.J.; Setiawan, W. Computer Vision and Deep Learning Approach for Social Distancing Detection During Covid-19 Pandemic. Ph.D. Thesis, Swiss German University, Kota Tangerang, Indonesia, 2021.
78. Wu, P.; Li, H.; Zeng, N.; Li, F. FMD-Yolo: An efficient face mask detection method for COVID-19 prevention and control in public. *Image Vis. Comput.* **2022**, *117*, 104341. [[CrossRef](#)]
79. Basu, A.; Ali, M.F. COVID-19 Face Mask Recognition with Advanced Face Cut Algorithm for Human Safety Measures. In Proceedings of the 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 6–8 July 2021; pp. 1–5.
80. Nagaraj, P.; Phebe, G.S.; Singh, A. A Novel Technique to Classify Face Mask for Human Safety. In Proceedings of the 2021 Sixth International Conference on Image Information Processing (ICIIP), Shimla, India, 26–28 November 2021; Volume 6; pp. 235–239.
81. Anthoniraj, S. Face Mask Detection with Computer Vision & Deep Learning. In Proceedings of the 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 8–9 October 2021; pp. 1–4.
82. Subhash, S.; Sneha, K.; Ullas, A.; Raj, D. A Covid-19 Safety Web Application to Monitor Social Distancing and Mask Detection. In Proceedings of the 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC), Bangalore, India, 30 September–2 October 2021; pp. 1–6.
83. Setyawan, N.; Putri, T.S.N.P.; Al Fikih, M.; Kasan, N. Comparative Study of CNN and YOLOv3 in Public Health Face Mask Detection. In Proceedings of the 2021 8th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Semarang, Indonesia, 20–21 October 2021; pp. 354–358.
84. Liu, S.; Agaian, S.S. COVID-19 face mask detection in a crowd using multi-model based on YOLOv3 and hand-crafted features. In Proceedings of the Multimodal Image Exploitation and Learning 2021, Online, 12–16 April 2021; Volume 11734, p. 117340M.
85. Gawde, B.B. A Fast, Automatic Risk Detector for COVID-19. In Proceedings of the 2020 IEEE Pune Section International Conference (PuneCon), Pune, India, 16–18 December 2020; pp. 146–151.
86. He, J. Mask detection device based on YOLOv3 framework. In Proceedings of the 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), Harbin, China, 25–27 December 2020; pp. 268–271.
87. Aswal, V.; Tupe, O.; Shaikh, S.; Charniya, N.N. Single camera masked face identification. In Proceedings of the 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 14–17 December 2020; pp. 57–60.
88. Ren, X.; Liu, X. Mask wearing detection based on YOLOv3. *J. Phys. Conf. Ser.* **2020**, *1678*, 012089. [[CrossRef](#)]
89. Darawsheh, A.; Siam, A.A.; Shaar, L.A.; Odeh, A. High-performance Detection and Predication Safety System using HUAWEI Atlas 200 DK AI Developer Kit. In Proceedings of the 2022 2nd International Conference on Computing and Information Technology (ICCIT), Tabuk, Saudi Arabia, 25–27 January 2022; pp. 213–216.
90. Kumar, K.S.; Kumar, G.A.; Rajendra, P.P.; Gatti, R.; Kumar, S.S.; Nataraja, N. Face Mask Detection and Temperature Scanning for the Covid-19 Surveillance System. In Proceedings of the 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), Bangalore, India, 27–28 August 2021; pp. 985–989.
91. Rakhsith, L.; Karthik, B.; Nithish, D.A.; Kumar, V.K.; Anusha, K. Face Mask and Social Distancing Detection for Surveillance Systems. In Proceedings of the 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 3–5 June 2021; pp. 1056–1065.
92. Prabha, P.A.; Karthikeyan, G.; Kuttralanathan, K.; Venkatesun, M.M. Intelligent Mask Detection Using Deep Learning Techniques. *J. Phys. Conf. Ser.* **2021**, *1916*, 012072. [[CrossRef](#)]



93. Zhang, K.; Jia, X.; Wang, Y.; Zhang, H.; Cui, J. Detection System of Wearing Face Masks Normatively Based on Deep Learning. In Proceedings of the 2021 International Conference on Control Science and Electric Power Systems (CSEPS), Shanghai, China, 28–30 May 2021; pp. 35–39.
94. Amin, P.N.; Moghe, S.S.; Prabhakar, S.N.; Nehete, C.M. Deep Learning Based Face Mask Detection and Crowd Counting. In Proceedings of the 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 2–4 April 2021; pp. 1–5.
95. Xiang, Y.; Yang, H.; Hu, R.; Hsu, C.Y. Comparison of the Deep Learning Methods Applied on Human Eye Detection. In Proceedings of the 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA), Shenyang, China, 22–24 January 2021; pp. 314–318.
96. Avanzato, R.; Beritelli, F.; Russo, M.; Russo, S.; Vaccaro, M. YOLOv3-based mask and face recognition algorithm for individual protection applications. Available online: <https://eur-ws.org/Vol-2768/p7.pdf> (accessed on 7 February 2023).
97. Bhuiyan, M.R.; Khushbu, S.A.; Islam, M.S. A deep learning based assistive system to classify COVID-19 face mask for human safety with YOLOv3. In Proceedings of the 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 1–3 July 2020; pp. 1–5.
98. Li, C.; Cao, J.; Zhang, X. Robust deep learning method to detect face masks. In Proceedings of the 2nd International Conference on Artificial Intelligence and Advanced Manufacture, Manchester, UK, 15–17 October 2020; pp. 74–77.
99. Vinh, T.Q.; Anh, N.T.N. Real-Time Face Mask Detector Using YOLOv3 Algorithm and Haar Cascade Classifier. In Proceedings of the 2020 International Conference on Advanced Computing and Applications (ACOMP), Quy Nhon, Vietnam, 25–27 November 2020; pp. 146–149.
100. Głowacka, N.; Rumiński, J. Face with Mask Detection in Thermal Images Using Deep Neural Networks. *Sensors* **2021**, *21*, 6387. [\[CrossRef\]](#)
101. Mahurkar, R.R.; Gadge, N.G. Real-time COVID-19 Face Mask Detection with YOLOv4. In Proceedings of the 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 4–6 August 2021; pp. 1250–1255.
102. Protik, A.A.; Rafi, A.H.; Siddique, S. Real-time Personal Protective Equipment (PPE) Detection Using YOLOv4 and TensorFlow. In Proceedings of the 2021 IEEE Region 10 Symposium (TENSYP), Jeju, Republic of Korea, 23–25 August 2021; pp. 1–6.
103. Prasad, P.; Chawla, A.; Mohana. Facemask Detection to Prevent COVID-19 Using YOLOv4 Deep Learning Model. In Proceedings of the 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India 2022; pp. 382–388. [\[CrossRef\]](#)
104. Qin, Z.; Guo, Z.; Lin, Y. An Implementation of Face Mask Detection System Based on YOLOv4 Architecture. In Proceedings of the 2022 14th International Conference on Computer Research and Development (ICCRD), Shenzhen, China, 7–9 January 2022; pp. 207–213. [\[CrossRef\]](#)
105. Gupta, A.; Thapar, D.; Deb, S. Smart Camera for Enforcing Social Distancing. In Proceedings of the 2021 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNiS), Jaipur, India, 18–22 December 2021; pp. 349–354.
106. Ubaid, M.T.; Khan, M.Z.; Rumaan, M.; Arshed, M.A.; Khan, M.U.G.; Darboe, A. COVID-19 SOP's Violations Detection in Terms of Face Mask Using Deep Learning. In Proceedings of the 2021 International Conference on Innovative Computing (ICIC), Lahore, Pakistan, 9–10 November 2021; pp. 1–8.
107. Vella, S.; Scerri, D. Vision-based Health Protocol Observance System for Small Rooms. In Proceedings of the 2021 IEEE 11th International Conference on Consumer Electronics (ICCE-Berlin), Berlin, Germany, 15–18 November 2021; pp. 1–6.
108. Mokeddem, M.L.; Belahcene, M.; Bourenane, S. Yolov4FaceMask: COVID-19 Mask Detector. In Proceedings of the 2021 1st International Conference On Cyber Management And Engineering (CyMaEn), Virtual, 26–28 May 2021; pp. 1–6.
109. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
110. Cao, Z.; Shao, M.; Xu, L.; Mu, S.; Qu, H. MaskHunter: Real-time object detection of face masks during the COVID-19 pandemic. *IET Image Process.* **2020**, *14*, 4359–4367. [\[CrossRef\]](#)
111. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt, Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768.
112. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [\[CrossRef\]](#)
113. Rahim, A.; Maqbool, A.; Rana, T. Monitoring social distancing under various low light conditions with deep learning and a single motionless time of flight camera. *PLoS ONE* **2021**, *16*, e0247440. [\[CrossRef\]](#) [\[PubMed\]](#)
114. Rodriguez, C.R.; Luque, D.; La Rosa, C.; Esenarro, D.; Pandey, B. Deep learning applied to capacity control in commercial establishments in times of COVID-19. In Proceedings of the 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), Bhimtal, India, 25–26 September 2020; pp. 423–428.
115. Pandya, S.; Sur, A.; Solke, N. COVIDSAVIO: A Novel Sensor-Fusion and Deep Learning Based Framework for Virus Outbreaks. *Front. Public Health* **2021**, *9*, 797808. [\[CrossRef\]](#) [\[PubMed\]](#)
116. Gola, A.; Panesar, S.; Sharma, A.; Ananthakrishnan, G.; Singal, G.; Mukhopadhyay, D. MaskNet: Detecting Different Kinds of Face Mask for Indian Ethnicity. In *Proceedings of the International Advanced Computing Conference*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 492–503.

117. Kumar, A.; Kalia, A.; Verma, K.; Sharma, A.; Kaushal, M. Scaling up face masks detection with YOLO on a novel dataset. *Optik* **2021**, *239*, 166744. [CrossRef]
118. Jocher, G.; Stoken, A.; Borovec, J.; Changyu, L.; Hogan, A.; Diaconu, L.; Ingham, F.; Poznanski, J.; Fang, J.; Yu, L.; et al. ultralytics/yolov5: v3. 1-Bug Fixes and Performance Improvements. Version v3. 2020; Volume 1. Available online: <https://zenodo.org/record/4154370#Y-8cJR9Bw2w> (accessed on 7 February 2023).
119. Yap, M.H.; Hachiuma, R.; Alavi, A.; Brüngel, R.; Cassidy, B.; Goyal, M.; Zhu, H.; Rückert, J.; Olshansky, M.; Huang, X.; et al. Deep learning in diabetic foot ulcers detection: A comprehensive evaluation. *Comput. Biol. Med.* **2021**, *135*, 104596. [CrossRef] [PubMed]
120. Walia, I.S.; Kumar, D.; Sharma, K.; Hemanth, J.D.; Popescu, D.E. An Integrated Approach for Monitoring Social Distancing and Face Mask Detection Using Stacked ResNet-50 and YOLOv5. *Electronics* **2021**, *10*, 2996. [CrossRef]
121. Ottakath, N.; Elharrouss, O.; Almaadeed, N.; Al-Maadeed, S.; Mohamed, A.; Khattab, T.; Abualsaud, K. ViDMASK dataset for face mask detection with social distance measurement. *Displays* **2022**, *73*, 102235. [CrossRef]
122. Jiang, M.; Fan, X.; Yan, H. Retinamask: A face mask detector. *arXiv* **2020**, arXiv:2005.03950.
123. Deng, J.; Guo, J.; Ververas, E.; Kotsia, I.; Zafeiriou, S. Retinaface: Single-shot multi-level face localisation in the wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 5203–5212.
124. Zhu, R.; Yin, K.; Xiong, H.; Tang, H.; Yin, G. Masked Face Detection Algorithm in the Dense Crowd Based on Federated Learning. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 8586016. [CrossRef]
125. Nguyen Quoc, H.; Truong Hoang, V. Real-time human ear detection based on the joint of yolo and retinaface. *Complexity* **2021**, *2021*, 7918165. [CrossRef]
126. Addagarla, S.K.; Chakravarthi, G.K.; Anitha, P. Real time multi-scale facial mask detection and classification using deep transfer learning techniques. *Int. J. Adv. Trends Comput. Sci. Eng.* **2020**, *9*, 4402–4408. [CrossRef]
127. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
128. Gathani, J.; Shah, K. Detecting masked faces using region-based convolutional neural network. In Proceedings of the 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), Rupnagar, India, 26–28 November 2020; pp. 156–161.
129. Meivel, S.; Devi, K.I.; Maheswari, S.U.; Menaka, J.V. Real time data analysis of face mask detection and social distance measurement using Matlab. *Mater. Today Proc.* **2021**. [CrossRef] [PubMed]
130. Zhang, J.; Han, F.; Chun, Y.; Chen, W. A novel detection framework about conditions of wearing face mask for helping control the spread of covid-19. *IEEE Access* **2021**, *9*, 42975–42984. [CrossRef] [PubMed]
131. Sahraoui, Y.; Kerrache, C.A.; Korichi, A.; Nour, B.; Adnane, A.; Hussain, R. DeepDist: A Deep-Learning-Based IoV Framework for Real-Time Objects and Distance Violation Detection. *IEEE Internet Things Mag.* **2020**, *3*, 30–34. [CrossRef]
132. Gupta, P.; Sharma, V.; Varma, S. A novel algorithm for mask detection and recognizing actions of human. *Expert Syst. Appl.* **2022**, *198*, 116823. [CrossRef] [PubMed]
133. Joshi, A.S.; Joshi, S.S.; Kanahasabai, G.; Kapil, R.; Gupta, S. Deep learning framework to detect face masks from video footage. In Proceedings of the 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), Bhimtal, India, 25–26 September 2020; pp. 435–440.
134. Loey, M.; Manogaran, G.; Taha, M.H.N.; Khalifa, N.E.M. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. *Measurement* **2021**, *167*, 108288. [CrossRef] [PubMed]
135. Sethi, S.; Kathuria, M.; Kaushik, T. Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread. *J. Biomed. Inform.* **2021**, *120*, 103848. [CrossRef]
136. Snyder, S.E.; Husari, G. Thor: A Deep Learning Approach for Face Mask Detection to Prevent the COVID-19 Pandemic. In Proceedings of the SoutheastCon 2021, Atlanta, GA, USA, 10–13 March 2021; pp. 1–8.
137. Yang, C.W.; Phung, T.H.; Shuai, H.H.; Cheng, W.H. Mask or Non-Mask? Robust Face Mask Detector via Triplet-Consistency Representation Learning. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2022**, *18*, 1–20. [CrossRef]
138. Kumar, T.A.; Rajmohan, R.; Pavithra, M.; Ajagbe, S.A.; Hodhod, R.; Gaber, T. Automatic face mask detection system in public transportation in smart cities using IoT and deep learning. *Electronics* **2022**, *11*, 904. [CrossRef]
139. Bansal, A.; Dhayal, S.; Mishra, J.; Grover, J. COVID-19 Outbreak: Detecting face mask types in real time. *J. Inf. Optim. Sci.* **2022**, *43*, 357–370. [CrossRef]
140. Gupta, S.; Sreenivasu, S.; Chouhan, K.; Shrivastava, A.; Sahu, B.; Potdar, R.M. Novel face mask detection technique using machine learning to control COVID-19 pandemic. *Mater. Today Proc.* **2021**. [CrossRef]
141. Mundial, I.Q.; Hassan, M.S.U.; Tiwana, M.I.; Qureshi, W.S.; Alanazi, E. Towards facial recognition problem in COVID-19 pandemic. In Proceedings of the 2020 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM), Medan, Indonesia, 3–4 September 2020; pp. 210–214.
142. Lin, H.; Tse, R.; Tang, S.K.; Chen, Y.; Ke, W.; Pau, G. Near-realtime face mask wearing recognition based on deep learning. In Proceedings of the 2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 9–12 January 2021; pp. 1–7.
143. Cao, Z.; Simon, T.; Wei, S.E.; Sheikh, Y. Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7291–7299.

144. Asif, S.; Wenhui, Y.; Tao, Y.; Jinhai, S.; Amjad, K. Real Time Face Mask Detection System using Transfer Learning with Machine Learning Method in the Era of Covid-19 Pandemic. In Proceedings of the 2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 28–31 May 2021; pp. 70–75.
145. Habib, S.; Alsanea, M.; Aloraini, M.; Al-Rawashdeh, H.S.; Islam, M.; Khan, S. An Efficient and Effective Deep Learning-Based Model for Real-Time Face Mask Detection. *Sensors* **2022**, *22*, 2602. [\[CrossRef\]](#) [\[PubMed\]](#)
146. Sen, S.; Sawant, K. Face mask detection for covid\_19 pandemic using pytorch in deep learning. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1070*, 012061. [\[CrossRef\]](#)
147. Sanjaya, S.A.; Rakhmawan, S.A. Face Mask Detection Using MobileNetV2 in The Era of COVID-19 Pandemic. In Proceedings of the 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), Sakheer, Bahrain, 26–27 October 2020; pp. 1–5.
148. Boulila, W.; Alzahem, A.; Almoudi, A.; Afifi, M.; Alturki, I.; Driss, M. A Deep Learning-based Approach for Real-time Facemask Detection. *arXiv* **2021**, arXiv:2110.08732.
149. Lad, A.M.; Mishra, A.; Rajagopalan, A. Comparative Analysis of Convolutional Neural Network Architectures for Real Time COVID-19 Facial Mask Detection. *J. Phys. Conf. Ser.* **2021**, *1969*, 012037. [\[CrossRef\]](#)
150. Nayak, R.; Manohar, N. Computer-Vision based Face Mask Detection using CNN. In Proceedings of the 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatre, India, 8–10 July 2021; pp. 1780–1786.
151. Taneja, S.; Nayyar, A.; Nagrath, P.; et al. Face Mask Detection Using Deep Learning During COVID-19. In *Second International Conference on Computing, Communications, and Cyber-Security*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 39–51.
152. Kayali, D.; Dimililer, K.; Sekeroglu, B. Face Mask Detection and Classification for COVID-19 using Deep Learning. In Proceedings of the 2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Kocaeli, Turkey, 25–27 August 2021; pp. 1–6.
153. Singh, R.; Singh, I.; Kapoor, A.; Chawla, A.; Gupta, A. Co-Yudh: A Convolutional Neural Network (CNN)-Inspired Platform for COVID Handling and Awareness. *Sn Comput. Sci.* **2022**, *3*, 241. [\[CrossRef\]](#) [\[PubMed\]](#)
154. Aadithya, V.; Balakumar, S.; Bavishprasath, M.; Raghul, M.; Malathi, P. Comparative Study Between MobilNet Face-Mask Detector and YOLOv3 Face-Mask Detector. In *Sustainable Communication Networks and Application*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 801–809.
155. Talahua, J.S.; Buele, J.; Calvopiña, P.; Varela-Aldás, J. Facial recognition system for people with and without face mask in times of the covid-19 pandemic. *Sustainability* **2021**, *13*, 6900. [\[CrossRef\]](#)
156. Kumar, A. A cascaded deep-learning-based model for face mask detection. *Data Technol. Appl.* **2022**, 1–24. doi/10.1108/DTA-02-2022-0076/full/html [\[CrossRef\]](#)
157. Al-Hamid, A.A.; Kim, T.; Park, T.; Kim, H. Optimization of Object Detection CNN With Weight Quantization and Scale Factor Consolidation. In Proceedings of the 2021 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia), Gangwon, Republic of Korea, 1–3 November 2021; pp. 1–5.
158. Bharathi, S.; Hari, K.; Senthilarasi, M.; Sudhakar, R. An Automatic Real-Time Face Mask Detection using CNN. In Proceedings of the 2021 Smart Technologies, Communication and Robotics (STCR), Sathyamangalam, India, 9–10 October 2021; pp. 1–5.
159. Jaisharma, K.; et al. A Deep Learning Based Approach for Detection of Face Mask Wearing using YOLO V3-tiny Over CNN with Improved Accuracy. In Proceedings of the 2022 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 16–17 February 2022; pp. 1–5.
160. Liu, G.; Zhang, Q. Mask Wearing Detection Algorithm Based on Improved Tiny YOLOv3. *Int. J. Pattern Recognit. Artif. Intell.* **2021**, *35*, 2155007. [\[CrossRef\]](#)
161. Jiang, X.; Xiang, F.; Lv, M.; Wang, W.; Zhang, Z.; Yu, Y. YOLOv3-Slim for Face Mask Recognition. *J. Phys. Conf. Ser.* **2021**, *1771*, 1–9. [\[CrossRef\]](#)
162. Sathyamurthy, K.V.; Rajmohan, A.S.; Tejaswar, A.R.; Kavitha, V.; Manimala, G. Realtime Face Mask Detection Using TINY-YOLO V4. In Proceedings of the 2021 4th International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 16–17 December 2021; pp. 169–174.
163. Zhao, Z.; Hao, K.; Ma, X.; Liu, X.; Zheng, T.; Xu, J.; Cui, S. SAI-YOLO: A Lightweight Network for Real-Time Detection of Driver Mask-Wearing Specification on Resource-Constrained Devices. *Comput. Intell. Neurosci.* **2021**, *2021*, 4529107. [\[CrossRef\]](#)
164. Han, Z.; Huang, H.; Fan, Q.; Li, Y.; Li, Y.; Chen, X. SMD-YOLO: An efficient and lightweight detection method for mask wearing status during the COVID-19 pandemic. *Comput. Methods Programs Biomed.* **2022**, *221*, 106888. [\[CrossRef\]](#) [\[PubMed\]](#)
165. Anand, R.; Das, J.; Sarkar, P. Comparative Analysis of YOLOv4 and YOLOv4-tiny Techniques towards Face Mask Detection. In Proceedings of the 2021 International Conference on Computational Performance Evaluation (ComPE), Shillong, India, 1–3 December 2021; pp. 803–809.
166. Kumar, A.; Kalia, A.; Kalia, A. ETL-YOLO v4: A face mask detection algorithm in era of COVID-19 pandemic. *Optik* **2022**, *259*, 169051. [\[CrossRef\]](#)
167. Hraybi, S.; Rizk, M. Examining YOLO for Real-Time Face-Mask Detection. In Proceedings of the 4th Smart Cities Symposium (SCS 2021), Online, 21–23 November 2021.
168. Zhu, J.; Wang, J.; Wang, B. Lightweight mask detection algorithm based on improved YOLOv4-tiny. *Chin. J. Liq. Cryst. Disp.* **2021**, *36*, 1525–1534. [\[CrossRef\]](#)
169. Farman, H.; Khan, T.; Khan, Z.; Habib, S.; Islam, M.; Ammar, A. Real-Time Face Mask Detection to Ensure COVID-19 Precautionary Measures in the Developing Countries. *Appl. Sci.* **2022**, *12*, 3879. [\[CrossRef\]](#)



170. Aydemir, E.; Yalcinkaya, M.A.; Barua, P.D.; Baygin, M.; Faust, O.; Dogan, S.; Chakraborty, S.; Tuncer, T.; Acharya, U.R. Hybrid deep feature generation for appropriate face mask use detection. *Int. J. Environ. Res. Public Health* **2022**, *19*, 1939. [CrossRef] [PubMed]
171. Chowdary, G.J.; Pun, N.S.; Sonbhadra, S.K.; Agarwal, S. Face mask detection using transfer learning of inceptionv3. In Proceedings of the International Conference on Big Data Analytics, Sonapat, India, 15–18 December 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 81–90.
172. Inamdar, M.; Mehendale, N. Real-Time Face Mask Identification Using Facemasknet Deep Learning Network. 2020. Available online: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3663305](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3663305) (accessed on 7 February 2023).
173. Rahman, M.M.; Manik, M.M.H.; Islam, M.M.; Mahmud, S.; Kim, J.H. An automated system to limit COVID-19 using facial mask detection in smart city network. In Proceedings of the 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Vancouver, BC, Canada, 9–12 September 2020; pp. 1–5.
174. Mohan, P.; Paul, A.J.; Chirania, A. A tiny CNN architecture for medical face mask detection for resource-constrained endpoints. In *Innovations in Electrical and Electronic Engineering*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 657–670.
175. Yu, J.; Zhang, W. Face mask wearing detection algorithm based on improved YOLO-v4. *Sensors* **2021**, *21*, 3263. [CrossRef] [PubMed]
176. Vrigkas, M.; Kourfalidou, E.A.; Plissiti, M.E.; Nikou, C. FaceMask: A New Image Dataset for the Automated Identification of People Wearing Masks in the Wild. *Sensors* **2022**, *22*, 896. [CrossRef]
177. Kumar, B.A.; Bansal, M. Face Mask Detection on Photo and Real-Time Video Images Using Caffe-MobileNetV2 Transfer Learning. *Appl. Sci.* **2023**, *13*, 935. [CrossRef]
178. Al-Kababji, A.; Bensaali, F.; Dakua, S.P.; Himeur, Y. Automated liver tissues delineation techniques: A systematic survey on machine learning current trends and future orientations. *Eng. Appl. Artif. Intell.* **2023**, *117*, 105532. [CrossRef]
179. Himeur, Y.; Al-Maadeed, S.; Kheddar, H.; Al-Maadeed, N.; Abualsaud, K.; Mohamed, A.; Khattab, T. Video surveillance using deep transfer learning and deep domain adaptation: Towards better generalization. *Eng. Appl. Artif. Intell.* **2023**, *119*, 105698. [CrossRef]
180. Su, X.; Gao, M.; Ren, J.; Li, Y.; Dong, M.; Liu, X. Face mask detection and classification via deep transfer learning. *Multimed. Tools Appl.* **2022**, *81*, 4475–4494. [CrossRef] [PubMed]
181. Sevilla, R.V.; Alon, A.S.; Melegrito, M.P.; Reyes, R.C.; Bastes, B.M.; Cimagala, R.P. Mask-Vision: A Machine Vision-Based Inference System of Face Mask Detection for Monitoring Health Protocol Safety. In Proceedings of the 2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kinabalu, Malaysia, 13–15 September 2021; pp. 1–5.
182. Jian, W.; Lang, L. Face mask detection based on Transfer learning and PP-YOLO. In Proceedings of the 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Nanchang, China, 26–28 March 2021; pp. 106–109.
183. Watcharabutsarakham, S.; Marukat, S.; Suntiwichaya, S.; Junlouchai, C. Partial Facial Identification using Transfer Learning Technique. In Proceedings of the 2021 16th International Joint Symposium on Artificial Intelligence and Natural Language Processing (ISAI-NLP), Phra Nakhon Si Ayutthaya, Thailand, 22–23 December 2021; pp. 1–4.
184. Mbunge, E.; Simelane, S.; Fashoto, S.G.; Akinuwa, B.; Metfula, A.S. Application of deep learning and machine learning models to detect COVID-19 face masks-A review. *Sustain. Oper. Comput.* **2021**, *2*, 235–245. [CrossRef]
185. Teboulbi, S.; Messaoud, S.; Hajjaji, M.A.; Mtibaa, A. Real-Time Implementation of AI-Based Face Mask Detection and Social Distancing Measuring System for COVID-19 Prevention. *Sci. Program.* **2021**, *2021*, 8340779. [CrossRef]
186. Fan, X.; Jiang, M.; Yan, H. A deep learning based light-weight face mask detector with residual context attention and Gaussian heatmap to fight against COVID-19. *IEEE Access* **2021**, *9*, 96964–96974. [CrossRef]
187. Song, Z.; Nguyen, K.; Nguyen, T.; Cho, C.; Gao, J. Spartan Face Mask Detection and Facial Recognition System. *Healthcare* **2022**, *10*, 87. [CrossRef]
188. Razavi, M.; Alikhani, H.; Janfaza, V.; Sadeghi, B.; Alikhani, E. An automatic system to monitor the physical distance and face mask wearing of construction workers in COVID-19 pandemic. *SN Comput. Sci.* **2022**, *3*, 27. [CrossRef] [PubMed]
189. Mounsouy, W.; Tawanbunjerd, T.; Liamsomboon, N.; Kusakunniran, W. Face recognition under mask-wearing based on residual inception networks. *Appl. Comput. Inform.* **2022**. [CrossRef]
190. M-CASIA: CASIA-WebFace+masks. Available online: <https://paperswithcode.com/dataset/casia-webface-masks> (accessed on 22 January 2022).
191. Sabir, M.F.S.; Mehmood, I.; Alsaggaf, W.A.; Khairullah, E.F.; Alhurairi, S.; Alghamdi, A.S.; El-Latif, A.; et al. An automated real-time face mask detection system using transfer learning with faster-rcnn in the era of the covid-19 pandemic. *Comput. Mater. Contin.* **2022**, pp. 4151–4166.
192. Palani, S.S.; Dev, M.; Mogili, G.; Relan, D.; Dey, R. Face Mask Detector using Deep Transfer Learning and Fine-Tuning. In Proceedings of the 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 17–19 March 2021; pp. 695–698.
193. Kocacinar, B.; Tas, B.; Akbulut, F.P.; Catal, C.; Mishra, D. A Real-Time CNN-based Lightweight Mobile Masked Face Recognition System. *IEEE Access* **2022**, *10*, 63496–63507. [CrossRef]
194. Mandal, B.; Okeukwu, A.; Theis, Y. Masked face recognition using resnet-50. *arXiv* **2021**, arXiv:2104.08997.
195. Mercaldo, F.; Santone, A. Transfer learning for mobile real-time face mask detection and localization. *J. Am. Med. Inform. Assoc.* **2021**, *28*, 1548–1554. [CrossRef] [PubMed]

196. Oumina, A.; El Makhfi, N.; Hamdi, M. Control the covid-19 pandemic: Face mask detection using transfer learning. In Proceedings of the 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra, Morocco, 2–3 December 2020; pp. 1–5.
197. Prusty, M.R.; Tripathi, V.; Dubey, A. A novel data augmentation approach for mask detection using deep transfer learning. *Intell.-Based Med.* **2021**, *5*, 100037. [[CrossRef](#)] [[PubMed](#)]
198. Zhang, E. A Real-Time Deep Transfer Learning Model for Facial Mask Detection. In Proceedings of the 2021 Integrated Communications Navigation and Surveillance Conference (ICNS), Dulles, VA, USA, 19–23 April 2021; pp. 1–7.
199. Alganci, U.; Soydas, M.; Sertel, E. Comparative research on deep learning approaches for airplane detection from very high-resolution satellite images. *Remote Sens.* **2020**, *12*, 458. [[CrossRef](#)]
200. Ejaz, M.S.; Islam, M.R.; Sifatullah, M.; Sarker, A. Implementation of principal component analysis on masked and non-masked face recognition. In Proceedings of the 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 3–5 May 2019; pp. 1–5.
201. Nieto-Rodriguez, A.; Mucientes, M.; Brea, V.M. System for medical mask detection in the operating room through facial attributes. In *Proceedings of the Iberian Conference on Pattern Recognition and Image Analysis*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 138–145.
202. Fafous, N.; Vemparala, M.R.; Frickenstein, A.; Frickenstein, L.; Badawy, M.; Stechele, W. Binarycop: Binary neural network-based covid-19 face-mask wear and positioning predictor on edge devices. In Proceedings of the 2021 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), Portland, OR, USA, 17–21 June 2021; pp. 108–115.
203. Petrović, N.; Kocić, Đ. Iot-based system for COVID-19 indoor safety monitoring. In Proceedings of the IcETran, Belgrade, Serbia, 8 October 2020.
204. Quiñonez, F.; Torres, R. Evaluation of AIoT performance in Cloud and Edge computational models for mask detection. *Ingenius* **2022**, *27*, 1–19.
205. Kong, X.; Wang, K.; Wang, S.; Wang, X.; Jiang, X.; Guo, Y.; Shen, G.; Chen, X.; Ni, Q. Real-time mask identification for COVID-19: An edge computing-based deep learning framework. *IEEE Internet Things J.* **2021**, *8*, 15929–15938. [[CrossRef](#)]
206. Magherini, R.; Mussi, E.; Servi, M.; Volpe, Y. Emotion recognition in the times of COVID19: Coping with face masks. *Intell. Syst. Appl.* **2022**, *15*, 200094. [[CrossRef](#)]
207. Fang, M.; Damer, N.; Kirchbuchner, F.; Kuijper, A. Real masks and spoof faces: On the masked face presentation attack detection. *Pattern Recognit.* **2022**, *123*, 108398. [[CrossRef](#)]
208. Zolfi, A.; Avidan, S.; Elovici, Y.; Shabtai, A. Adversarial Mask: Real-World Adversarial Attack Against Face Recognition Models. *arXiv* **2021**, arXiv:2111.10759.
209. Tjoa, E.; Guan, C. A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4793–4813. [[CrossRef](#)]
210. Minh, D.; Wang, H.X.; Li, Y.F.; Nguyen, T.N. Explainable artificial intelligence: A comprehensive review. *Artif. Intell. Rev.* **2022**, *55*, 3503–3568. [[CrossRef](#)]
211. Hossain, M.S.; Muhammad, G.; Guizani, N. Explainable AI and mass surveillance system-based healthcare framework to combat COVID-19 like pandemics. *IEEE Netw.* **2020**, *34*, 126–132. [[CrossRef](#)]
212. Yin, B.; Tran, L.; Li, H.; Shen, X.; Liu, X. Towards interpretable face recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, South Korea, 27 October–2 November 2019; pp. 9348–9357.
213. Zhang, J.; Yu, H. Improving the Facial Expression Recognition and Its Interpretability via Generating Expression Pattern-map. *Pattern Recognit.* **2022**, *129*, 108737. [[CrossRef](#)]
214. Williford, J.R.; May, B.B.; Byrne, J. Explainable face recognition. In *Proceedings of the European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 248–263.
215. Mery, D.; Morris, B. On Black-Box Explanation for Face Verification. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 4–8 January 2022; pp. 3418–3427.
216. Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 815–823.
217. Izhar, F.; Ali, S.; Ponum, M.; Mahmood, M.T.; Ilyas, H.; Iqbal, A. Detection & recognition of veiled and unveiled human face on the basis of eyes using transfer learning. *Multimed. Tools Appl.* **2022**, *82*, 4257–4287. [[PubMed](#)]
218. Holkar, A.; Walambe, R.; Kotecha, K. Few-Shot learning for face recognition in the presence of image discrepancies for limited multi-class datasets. *Image Vis. Comput.* **2022**, *120*, 104420. [[CrossRef](#)]
219. Jin, B.; Cruz, L.; Gonçalves, N. Deep facial diagnosis: Deep transfer learning from face recognition to facial diagnosis. *IEEE Access* **2020**, *8*, 123649–123661. [[CrossRef](#)]
220. Venkateswarlu, I.B.; Kakarla, J.; Prakash, S. Face mask detection using mobilenet and global pooling block. In Proceedings of the 2020 IEEE 4th Conference on Information & Communication Technology (CICT), Chennai, India, 3–5 December 2020; pp. 1–5.
221. Lanchantin, J.; Wang, T.; Ordonez, V.; Qi, Y. General multi-label image classification with transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 16478–16488.
222. Chen, H.; Wang, Y.; Guo, T.; Xu, C.; Deng, Y.; Liu, Z.; Ma, S.; Xu, C.; Xu, C.; Gao, W. Pre-trained image processing transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 12299–12310.



223. Aziz, M.N.A.; Mutalib, S.; Aliman, S. Comparison of Face Coverings Detection Methods using Deep Learning. In Proceedings of the 2021 2nd International Conference on Artificial Intelligence and Data Sciences (AiDAS), IPOH, Malaysia, 8–9 September 2021; pp. 1–6.
224. Nguyen, D.C.; Pham, Q.V.; Pathirana, P.N.; Ding, M.; Seneviratne, A.; Lin, Z.; Dobre, O.; Hwang, W.J. Federated learning for smart healthcare: A survey. *Acm Comput. Surv. (CSUR)* **2022**, *55*, 1–37. [[CrossRef](#)]
225. Umuroglu, Y.; Fraser, N.J.; Gambardella, G.; Blott, M.; Leong, P.; Jahre, M.; Vissers, K. Finn: A framework for fast, scalable binarized neural network inference. In Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays, Monterey, CA, USA, 22–24 February 2017; pp. 65–74.
226. Fang, T.; Huang, X.; Saniie, J. Design flow for real-time face mask detection using PYNQ system-on-chip platform. In Proceedings of the 2021 IEEE International Conference on Electro Information Technology (EIT), Mt. Pleasant, MI, USA, 14–15 May 2021; pp. 1–5.
227. Fourati, L.C.; Samiha, A. Federated learning toward data preprocessing: COVID-19 context. In Proceedings of the 2021 IEEE International Conference on Communications Workshops (ICC Workshops), Montreal, QC, Canada, 14–23 June 2021; pp. 1–6.
228. Peyvandi, A.; Majidi, B.; Peyvandi, S.; Patra, J.C.; Moshiri, B. Location-aware hazardous litter management for smart emergency governance in urban eco-cyber-physical systems. *Multimed. Tools Appl.* **2022**, *81*, 22185–22214. [[CrossRef](#)]
229. Kim, J.; Park, T.; Kim, H.; Kim, S. Federated Learning for Face Recognition. In Proceedings of the 2021 IEEE International Conference on Consumer Electronics (ICCE), Penghu, Taiwan, 16–18 June 2021; pp. 1–2. . [[CrossRef](#)]
230. Niu, Y.; Deng, W. Federated learning for face recognition with gradient correction. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2022; Volume 36, pp. 1999–2007.
231. Meng, Q.; Zhou, F.; Ren, H.; Feng, T.; Liu, G.; Lin, Y. Improving federated learning face recognition via privacy-agnostic clusters. *arXiv* **2022**, arXiv:2201.12467.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.