



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# Deep transfer learning for the recognition of types of face masks as a core measure to prevent the transmission of COVID-19

Ricardo Mar-Cupido<sup>a</sup>, Vicente García<sup>a,\*</sup>, Gilberto Rivera<sup>a</sup>, J. Salvador Sánchez<sup>b</sup>

<sup>a</sup> División Multidisciplinaria en Ciudad Universitaria, Universidad Autónoma de Ciudad Juárez, Av. José de Jesús Delgado 18100, Ciudad Juárez, Chihuahua, Mexico

<sup>b</sup> Institute of New Imaging Technologies, Department of Computer Languages and Systems, Universitat Jaume I, Av. de Vicent Sos Baynat, s/n 12071 Castelló de la Plana, Spain

## ARTICLE INFO

### Article history:

Received 15 October 2021  
Received in revised form 31 May 2022  
Accepted 19 June 2022  
Available online 23 June 2022

### Keywords:

Recognition  
Face mask  
Deep learning  
Transfer learning  
COVID-19

## ABSTRACT

The use of face masks in public places has emerged as one of the most effective non-pharmaceutical measures to lower the spread of COVID-19 infection. This has led to the development of several detection systems for identifying people who do not wear a face mask. However, not all face masks or coverings are equally effective in preventing virus transmission or illness caused by viruses and therefore, it appears important for those systems to incorporate the ability to distinguish between the different types of face masks. This paper implements four pre-trained deep transfer learning models (NasNetMobile, MobileNetv2, ResNet101v2, and ResNet152v2) to classify images based on the type of face mask (KN95, N95, surgical and cloth) worn by people. Experimental results indicate that the deep residual networks (ResNet101v2 and ResNet152v2) provide the best performance with the highest accuracy and the lowest loss.

© 2022 Elsevier B.V. All rights reserved.

## 1. Introduction

The COVID-19 outbreak caused by a new strain of coronavirus named SARS-CoV-2 was declared as a global pandemic by the World Health Organization (WHO) on March 11, 2020, because of its high viral transmission and mortality rates. As of 14 October 2021, more than 239.01 million confirmed cases and 4.87 million deaths (<https://covid19.who.int/>) have globally been registered since the first patient was detected in Wuhan, China, in December 2019. The main transmission pathway of COVID-19 is airborne, that is, human-to-human transmission via droplets nuclei or aerosols produced by infected individuals during all expiratory activities, such as coughing, sneezing, talking, singing, shouting or breathing, especially at a distance of less than 1.5 to 2 metres [1].

To mitigate the risk of viral infection and control the spread of COVID-19, several prevention measures such as hand hygiene, social distancing and wearing face masks have been highly recommended by health agencies worldwide and applied in many countries [2–4]. It has been found that the use of face masks in public areas, such as in supermarkets and schools or on public transport, is an effective and easy-to-use barrier against the

spread of COVID-19 infection [5,6]. Although mask fit plays a critical role, the effectiveness of each type of face mask in reducing infection risk depends on the mask material, source, structure and particulate efficacies (particle filtration efficiency and filter quality factor) [7,8]. Examples of some of the face masks most widely used during the COVID-19 pandemic are shown in Fig. 1, whose main features can be summarized as follows [9,10]:

- Cloth masks: Particle filtration efficiency of cloth masks made of a different fabric (typically 100% cotton or 50% cotton/50% polyester blend) varies from 25% to 38%.
- Surgical masks: Commonly, these are comprised of three layers made from non-woven fabric; the middle filter layer removes particulates, whereas the inner layer absorbs droplets from the mouth, and the outer layer is water-repellent. The surgical masks can capture 45%–55% of 0.1  $\mu\text{m}$  particles in the air.
- N95 and KN95 face respirator masks: Both are made from multiple layers of synthetic material (typically a polypropylene plastic polymer) to filter out 95% of 0.3  $\mu\text{m}$  particles.

Cloth coverings and surgical masks do not provide the wearer with the same level of protection from inhaling smaller airborne particles as an N95 (or KN95) mask [11]. Oberg & Brosseau [12] demonstrated that surgical masks do not exhibit adequate filter performance against aerosols measuring 0.9, 2.0, and 3.1  $\mu\text{m}$  in diameter. Lee et al. [13] showed that particles 0.04 to  $\mu\text{m}$  can

\* Corresponding author.

E-mail addresses: [al158883@alumnos.uacj.mx](mailto:al158883@alumnos.uacj.mx) (R. Mar-Cupido), [vicente.jimenez@uacj.mx](mailto:vicente.jimenez@uacj.mx) (V. García), [gilberto.rivera@uacj.mx](mailto:gilberto.rivera@uacj.mx) (G. Rivera), [sanchez@uji.es](mailto:sanchez@uji.es) (J.S. Sánchez).



**Fig. 1.** Types of face masks commonly worn during the COVID-19 pandemic (from left to right, they are KN95 mask, N95 mask, cloth mask, surgical mask, and without mask).

penetrate surgical masks. Although surgical masks are not ideal for containing airborne transmission in a high-risk environment because the SARS-CoV-2 particle has a size similar to that of SARS-CoV (estimated as 0.08 to 0.14  $\mu\text{m}$ ), there is good evidence that they are very useful at a population level [14].

Regarding N95 and KN95 face respirator masks, their main difference refers to the fact that N95 is the United States standard for respirator masks, whereas KN95 is the China standard (and FFP2 is the Europe standard). Besides, N95 masks have slightly stricter requirements for pressure drop while inhaling or exhaling, which means that they have stronger breathability than KN95 masks. On the other hand, KN95 masks usually have earloops, while N95 masks have head straps.

Due to the importance of wearing face masks to control the spreading and transmission of COVID-19, the governments of many countries have enacted laws and rules that mandate their use in public places. This has led to the development of various security and surveillance systems for face mask detection with the aim to identify people who do not wear a face mask.

At the beginning of the pandemic, there was no explicit recommendation on the type of mask to wear. However, with the arrival of new variants of SARS-CoV-2 of increased contagiousness, such as the quick-spreading Omicron, Governments and healthcare agencies changed the guidelines to minimize infections. On January 14, 2022, the US Centers for Disease Control and Prevention (CDC) released the report “Type of Masks and Respirators” [15], which lists a series of recommendations on what type of face mask to use depending on specific risk and vulnerability situations. Thus, some hospitals joined these recommendations by establishing their policies on masks and respirators. For example, the Mayo Clinic Health System website [16] enumerates the types of preferred and acceptable masks that are required for staff, patients and visitors (N95, KN95, and surgical/procedural) to slow the spread of COVID-19, and those that cannot be used because of their reduced effectiveness (masks with vents, bandanas, neck gaiters, face shields, and cloth masks). An article entitled “Why Cloth Masks Might Not Be Enough as Omicron Spreads” [17] in the Wall Street Journal collects the different recommendations given by doctors on the use of the types of face masks, as well as the time it takes to transmit COVID-19 between an infected person and an uninfected person depending on whether they are wearing an N95 mask, surgical mask, cloth mask or none. In this sense, clinicians and infectious disease specialists suggest the importance to educate the public on the different quality of face masks [17].

In this paper, we introduce a model based on deep transfer learning to recognize/classify four different types of face masks (plus faces without a mask). The proposed model could be easily integrated as a software upgrade into existing surveillance systems developed to detect people not wearing a mask. The integration of a classification module may enhance the functionality of surveillance systems by generating alerts to prevent access to people wearing an inappropriate type of face mask depending on where the system is installed. The correct use of the face

mask according to the characteristics of the place can limit the spread and significantly reduce the number of deaths [18]. In some hospitals, cloth masks are allowed as long as it is inside or outside work. However, it is mandatory to replace it with a surgical or procedure mask when at the work unit [16].

Henceforward, this paper is organized as follows. Section 2 reviews some papers that addressed face mask detection for fighting against COVID-19. Section 3 describes the data sets used in the experiments. Section 4 introduces the model adopted to identify the types of face masks. Section 5 presents the experimental set-up and discusses the results. Finally, Section 6 remarks on the main findings and outlines possible directions for further research.

## 2. Some works on face mask detection

This section offers a non-exhaustive sample of works that have been published on the face mask detection problem. We have consciously omitted any study on facial recognition or detection applied to people wearing masks that has not been explicitly related to the COVID-19 pandemic.

A transfer learning approach built by fine-tuning the pre-trained InceptionV3 model was proposed to identify people who are not wearing a mask in public places and crowded areas [19]. Experimental results with the transfer learning model trained for 80 epochs on a testing data set with 479 images achieved a 100% in terms of accuracy, precision, and specificity.

Qin & Li [20] developed a face mask-wearing condition identification model by combining image super-resolution and classification networks to classify images into three categories: no face mask-wearing, incorrect face mask-wearing, and correct face mask-wearing. The authors reported that after 200 epochs, the proposed model achieved a 98.70% of accuracy for the face mask-wearing condition identification. In contrast, when it is used to identify two different face masks, the accuracy in folded face masks vs. N95 was 98.02%, and in medical-surgical vs. basic cloth face masks was 99.19%. A third experiment for colour face mask detection suggests that the proposed model could detect blue, white, black, and other face mask colours with an accuracy of 98.84%, 98.68%, 97.81%, and 98.25%, respectively.

Loey et al. [21] employed the ResNet-50 deep transfer learning algorithm for the feature extraction process and YOLOv2 for real-time detection of medical face masks. The proposed detector optimized with Adam and 60 epochs gave an average precision of 81% on a new combined data set, obtained from two public medical masks data sets.

Singh et al. [22] used the YOLOv3 and faster R-CNN convolutional networks to monitor people wearing face masks in public places. The data set used was composed of two data sets MAFA and WIDER FACE, and images taken from several web sources. Experimental results on the validation data set showed an average precision of 62% after testing 50 epochs.

Loey et al. [23] proposed a hybrid method for face mask detection that comprises two stages: one for feature extraction

using ResNet-50 and a second part to detect face masks using traditional machine learning methods (support vector machine, decision tree and ensemble of classifiers). The authors used four data sets as a two-class classification problem: (1) real face masks (RMFD), (2) fake face masks (SMFD), (3) a combined data set from fake and real face masks, and (4) simulated face masks (LWF). From all experimental results, the best performance was obtained using the SVM, where the accuracy testing on RMFD was 99.64%, SMFD was 99.49% on SMFD, and 100% on LFW.

Wang et al. [24] designed a two-stage method to detect wearing masks based on the transfer model of Faster\_RCNN and InceptionV2 structure. Moreover, they also generated a data set for wearing mask detection that includes 7804 realistic images with 26403 wearing masks and multiple scenes (available at <https://github.com/BingshuCV/WMD>). The effectiveness of the proposal on the whole data set, which contained images of different numbers of people wearing masks (1, 2 to 4, and 5 or more persons), was measured in terms of precision, recall, and the F-measure (F1-score). The experimental results indicate that the model could get competitive testing effectiveness of 93.54% in recall, 94.84% in precision, and 94.19% in F1-score.

Chen, Liu & Zhang [25] devised a transfer learning algorithm combined with a skip-connected structure to improve the classification accuracy of eight masked face poses. Besides, a semi-synthetic masked face pose data set was constructed by superimposing the face pose images taken from the CAS-PEAL-R1 data set [26], and a diversity of the mask images. The aim was to replace the ImageNet data set as a source domain to improve the model's accuracy and generalization ability. Experiments using AlexNet and VGG16 architectures yielded an accuracy testing on each model of 96.44% and 99.29%, respectively.

Jiang et al. [27] introduced an algorithmic improvement to the original YOLOv3 by adding the squeeze and excitation block between the convolutional layer of Darknet53 and replacing the mean squared error loss function with GIoU loss. The authors also built a large data set with 9205 images of mask-wearing faces belonging to three predefined classes: with mask, incorrect mask, and without mask. The final average precision of the proposed model over the three classes was 75%, representing an increase of 8.6% compared to YOLOv3.

Yu & Zhang [28] proposed an improved algorithm based on YOLOv4 for face mask-wearing detection to reduce the computing cost of the network and improve the learning ability of the model. The experiments were carried out on a data set constructed from two sources: (1) the published RMFD and (2) the MaskedFace-Net. The detection task aimed to classify three images: without mask, face mask, and incorrect mask. The model's performance was measured using precision, recall, F1-score, and average precision over all classes. The experimental results indicated that the proposal reached accurate rate recognition of 95.1% for precision, 98.2% for recall, 96.7% for F1-score, and 98.3% for mean average precision.

Sethi, Kathuria & Kaushik [29] presented a highly accurate and real-time technique based on bounding box affine transformation and transfer learning to detect mask faces in public places. The authors reported that their proposal implemented with ResNet50 achieved an accuracy of 98.2%, while the recall in each class was 99.0% (face detection) and 98.24% (mask detection).

Prusty, Tripathi & Dubey [30] implemented a two-class mask detection model based on YOLOv3 and DarkNet53 that consists of a data augmentation block using image filtering techniques such as greyscale and Gaussian blur. The model was trained in two parts according to the data set used. First, using the data set taken from Kaggle, and in the second stage, using the augmented data set. The proposal was evaluated considering three detection problems: (1) individual detection on images, (2) group detection

on images, and (3) group detection in video. In all cases, the confidence level was between 0.98 and 0.99, while in the video analysis of a group of people, the average precision over the two classes was 99.8%.

The MobileNetv2 convolutional neural network to detect people wearing and not wearing a face mask in public places has been used in several studies [31–35]. The experimental results found in the literature suggest that with the use of MobileNetv2 in classification problems of two classes (mask vs. unmask), the accuracy rates are above 98%.

Nagrath et al. [36] integrated the single shot detector (SSD) as a face detector and the SSD MobileNetv2 to determine whether a mask is worn or not. The proposal methodology involves two steps; first, a data set obtained from different sources is pre-processed and augmented; second, a MobileNetv2 is trained using the Adam optimizer. The authors state that their model achieved an accuracy of 92.64% and F1-score of 93%. Gola et al. [37] investigated the performance of MobileNetv2, YOLOv3 and YOLOv4 in the context of real-time face mask detection in the Indian community on various types of masks such as home-made, handkerchiefs, and dupattas, concluding that YOLOv4 performs the best in terms of sensitivity and precision.

Joshi et al. [38] proposed an approach for detecting face masks in videos using the multi-task cascaded convolutional network and the MobileNetv2 model. The experimental evaluation was carried out on a data set compiled from YouTube videos of multiple geographical locations. The proposal model achieved in the facial mask prediction task a precision of 84.39% and an accuracy of 81.74%. Alguzo, Alzubi & Albalas [39] implemented a multi-graph convolutional network to detect people wearing a face mask. The proposed methodology is characterized by using several convolutional filters that transform the extracted facial relief and generalizes image frequencies. The model was evaluated on the publicity data set RMFD. A comparative study suggests that the proposal obtains a high accuracy testing of 97.9%, outperforming six state-of-the-art models.

A methodology that combines a pre-trained RetinaFace model for face detection and the NASNetMobile neural network for classifying the detected faces as masked or non-masked was proposed by [40]. To get cultural diversity, the authors combined several data sets to avoid a model biased toward Asian faces. The proposed method obtained identical results to the DenseNet121 and MobileNetv2 models with an F1-score of 99.13%. However, it performed significantly better in terms of inference speed and also resulted in a smaller size than the DenseNet121 and MobileNetv2 models.

Three different convolutional neural networks (VGG-19, Xception and MobileNetv2) were compared to extract features from images of faces that were further processed using a support vector machine and a  $K$ -nearest neighbours classifier [41]. The best results were obtained when the SVM was combined with MobileNetv2 on a small data set augmented with different techniques such as cropping, padding, and horizontal flipping. The proposal reached an accuracy of 97.11%.

Similarly, Hussain et al. [42] classified individuals who wear the face mask properly, improperly, and without a face mask using VGG-16, MobileNetv2, InceptionV3 and ResNet-50 using a transfer learning approach, obtaining the highest performance with VGG-16 and MobileNetv2 models. This detection system was a module of a complete system called IoT-based Smart Screening and Disinfection Walkthrough Gate, which included a non-contact temperature measurement system and could be installed at the entrance of any place. The authors created a database from different available sources, such as MAFA, MaskedFace-Net and Bing images, to evaluate the proposal. The experimental results showed an accuracy of 99.8% in the detection of three types



Fig. 2. Sample images from the databases for faces with and without a mask.

of situations: without mask, proper mask, and improper mask. Finally, an experiment to evaluate the effectiveness of the method in detecting two types of masks (N95 and surgical) yielded a 98.17% of accuracy.

### 3. Data sets

The present study was carried out using images taken from four publicly available databases:

- The Face Mask Classification (FMC) database [43] contains a total of 440 images of different resolutions: 220 for faces with a mask and 220 for faces without a mask.
- The Face Mask Detection (FMD) database [44] consists of 853 images of faces to detect mask, no mask and incorrectly worn masks.
- The Masked Faces in Real-World for Face Recognition (MFR2) database [45] is a small data set of real-world masked faces of 53 identities and 269 images aligned and preprocessed.
- The Masked Faces (MAFA) database [46] contains 30,811 images of unmasked faces and 35,806 images of masked faces. Faces in the data set have various orientations and occlusion degrees, while a mask occludes at least one part of each face.

Apart from these databases, we also picked up images from Bing Images and Google Images to compensate for the lack of images with N95 and KN95 face respirator masks. Fig. 2 illustrates some example images of faces with and without masks taken from the different databases and repositories used in this work.

After collecting images from the public databases and repositories just mentioned, the distribution of images per class was as follows: 1450 images with N95 respirator mask, 1848 images with KN95 respirator mask, 2685 images with surgical mask, 2114 images with fabric mask, and 2716 images without a face mask. As the distribution of classes was moderately skewed due to the limited number of images with N95 and KN95 face masks, only 1450 images from each category were randomly selected to form a class-balanced data set and thus to avoid the possible bias towards the most represented classes during the recognition process.

### 4. The proposed methodology

This section introduces the methodology designed to classify images based on the different types of face masks. In summary, it consists of a first component to prepare/preprocess the image set and a second one for developing the deep transfer learning model. The flowchart of the proposed methodology is illustrated in Fig. 3.

#### 4.1. Data preprocessing

As each database had a different image format, the first step was to convert all the images into a single format (png) using ImageMagick. Next, we extracted out the region of interest (ROI) from the raw images to keep only the face region and remove the unnecessary parts that could hinder the performance of the recognition process (Fig. 4). The images (ROIs) were then resized to  $224 \times 224$  pixels in 3 channels because this is the default size accepted by the MobileNetv2, ResNet and DenseNet deep transfer learning models used in this work.

Finally, all the images were normalized through scaling down by factor  $1/255$  before feeding to the model. Thus the colour value of every pixel was transformed from range  $[0, 255]$  to range  $[0, 1]$  so that images contribute more evenly to the total loss of the model.

#### 4.2. Data augmentation

An exploitable deep learning model must show a low generalization performance, also called test error, when performing well on unobserved examples [47]. When this does not occur, the model is said to be overfitting, which means that the difference between the training error and the test error is too significant. This problem can be originated when the training data is under-sized [48]. As a rule of thumb, it has been suggested that to get deep learning models with satisfactory performance, each class in the training set should have at least 5000 examples [47].

Two major solutions can be found in the literature to overcome the overfitting [49]. First, the architecture or learning algorithm level includes a variety of strategies, such as dropout, batch normalization, transfer learning, pre-training, early stopping, and limiting the number of parameters by filter size [49,50]. Second, at the data level, data augmentation is considered the natural way to deal with the insufficient amount of data, which consists on creating new training examples from the existing ones either by basic image manipulation, deep learning techniques, or both [50].

Basic image manipulation is the most straightforward strategy for enlarging the training set size via geometric transformations of existing data (e.g., rotation, reflection, flipping, shifting, cropping, elastic distortion, and shearing). Shijie et al. [51] showed that cropping, flipping, and rotation generally yields better results than other data augmentation techniques. In addition, several face mask detection works suggest that this strategy improves the performance and robustness of the deep learning methods [20,23,27,36,42].

For this study, the data set was split into two parts, 80% for training the model and the remaining 20% for testing the recognition performance of the models. Then, we applied some transformations to the training images using Keras ImageDataGenerator: rotation (the rotation parameter was set to 20 degrees), width and height shifting (in both cases, the parameter was set to 0.2), shearing (the parameter was set to 0.15), horizontal flipping, and filling (we filled the empty pixels with the value of their nearest pixel).

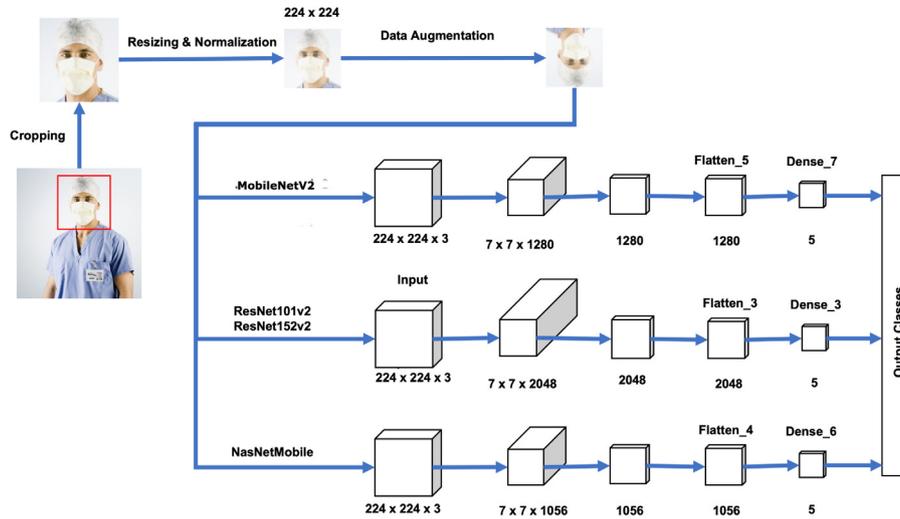


Fig. 3. Framework of the proposed methodology.

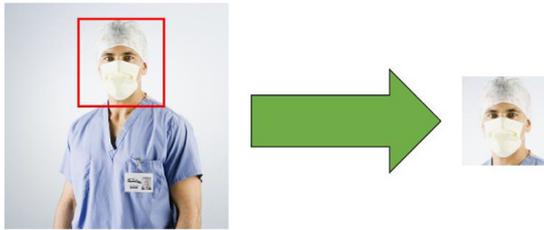


Fig. 4. Region of interest (ROI).

### 4.3. Recognition models

In this paper, we developed some transfer learning based models pre-trained on the ImageNet database to solve the requirement of a huge amount of labelled examples:

- MobileNetv2 is based on an inverted residual structure where the shortcut connections of the residual block are between the thin bottleneck layers [52]. The intermediate expansion layer uses lightweight depth-wise convolutions to filter the features as a source of non-linearity. This architecture consists of the primary full convolution layer through 32 filters and 19 residual bottleneck layers.
- ResNet101v2 and ResNet152v2 are examples of the so-called residual networks, which contain a lot of layers (101 and 152 in this paper) with high performance [53]. The main difference between (v2) and the original (v1) is that (v2) uses batch normalization before each weight layer.
- NasNetMobile is an architecture that aims at searching for an optimal convolutional architecture using reinforcement learning [54]. Two types of convolutional cells are used in the architecture: the normal cells return a feature map of the same input, whereas the reduction cells return a feature map where the height and width of the feature map are reduced by a factor of two.

### 4.4. Performance evaluation in classification tasks

Scalar metrics are the most frequently used alternative to evaluate the quality of a predictive classification/recognition system. The scalar performance metrics can be obtained from a confusion matrix representing the number of successes and errors

Table 1

Confusion matrix for multi-class problems.

Actual	Predicted				Total
	Pred_C <sub>1</sub>	Pred_C <sub>2</sub>	...	Pred_C <sub>m</sub>	
Act_C <sub>1</sub>	a <sub>11</sub>	a <sub>12</sub>	...	a <sub>1m</sub>	n <sub>C<sub>1</sub></sub> = ∑ <sub>j=1</sub> <sup>m</sup> a <sub>1j</sub>
Act_C <sub>2</sub>	a <sub>21</sub>	a <sub>22</sub>	...	a <sub>2m</sub>	n <sub>C<sub>2</sub></sub> = ∑ <sub>j=1</sub> <sup>m</sup> a <sub>2j</sub>
...	...	...	...	...	...
Act_C <sub>m</sub>	a <sub>m1</sub>	a <sub>m2</sub>	...	a <sub>mm</sub>	n <sub>C<sub>m</sub></sub> = ∑ <sub>j=1</sub> <sup>m</sup> a <sub>mj</sub>
Total	∑ <sub>i=1</sub> <sup>m</sup> a <sub>i1</sub>	∑ <sub>i=1</sub> <sup>m</sup> a <sub>i2</sub>	...	∑ <sub>i=1</sub> <sup>m</sup> a <sub>im</sub>	n

distributed across  $m$  classes. In a multi-class problem, it is a square  $m \times m$  matrix, where each  $a_{ij}$  element denotes the number of images or examples whose true class is  $C_i$ ,  $i = 1, \dots, m$ , and was assigned to class  $C_j$ ,  $j = 1, \dots, m$ , by the classifier.

The confusion matrix can be written as in Table 1, where the diagonal entries  $a_{ij}$  ( $i = j$ ) correspond to the correctly classified examples for the class  $C_i$ , and the non-diagonal elements  $a_{ij}$  ( $i \neq j$ ) represent the misclassification examples that were assigned to the class  $C_j$  when the actual class is  $C_i$ .

The most popular scalar performance metrics are accuracy and its complement the error rate:

$$Accuracy = \frac{\sum_{i=1}^m a_{ii}}{n}, \quad (1)$$

and

$$Error = 1 - Accuracy. \quad (2)$$

Accuracy and error rate have been shown to be inappropriate in those cases where the classes do not have an equal number of examples (the class imbalance problem). Therefore, it could not consider the unbalance in the classification results [55].

Straightforward metrics on a single  $C_i$  class can be computed. Considering that  $C_i$  is the class of interest (also called the positive class), then Recall is the true positive rate (TPR) of a classifier on the class  $C_i$ . It is also called the sensitivity of the classifier. For a specific  $C_i$  class, it can be computed as follows:

$$Recall_{C_i} = Sensitivity_{C_i} = TPR_{C_i} = \frac{a_{ii}}{n_{C_i}}. \quad (3)$$

If class  $C_i$  is considered as the positive class, then the rest of the classes can be regarded as the negative class, that is, the negative class is a subset  $NC_{k \neq i} = \{C_k, \dots, C_m\}$  for  $k = 1, \dots, m$  and  $k \neq i$ . Therefore, we can calculate the proportion of all correctly labeled

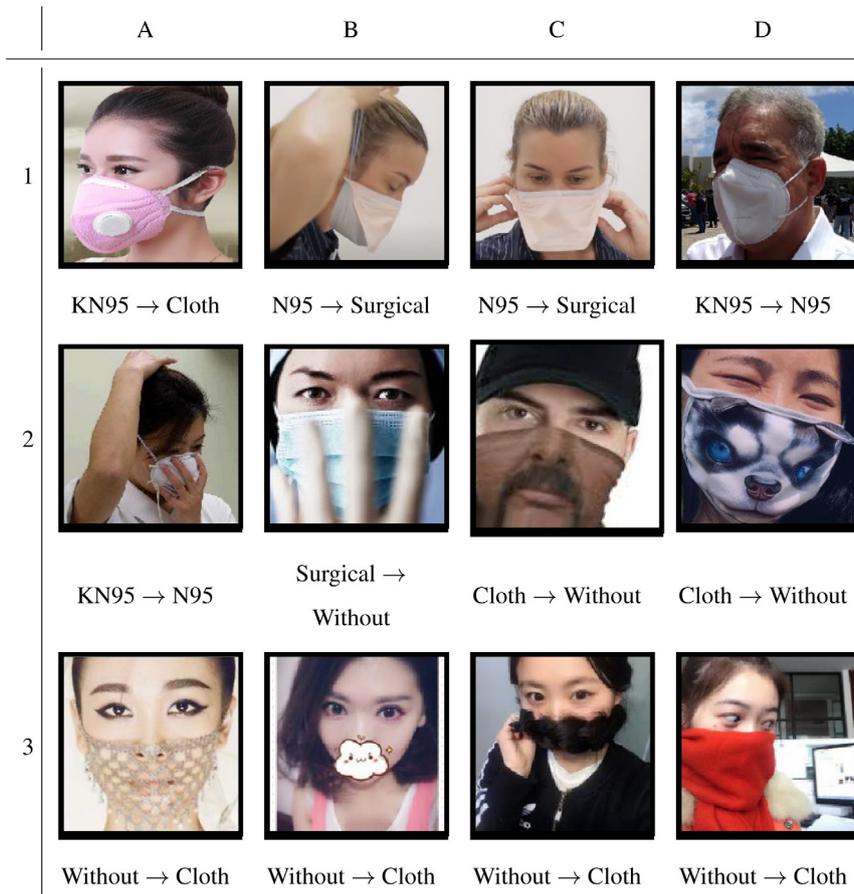


Fig. 5. Examples of images misclassified by the ResNet101v2 model (text means 'true class → predicted class').

**Table 2**  
Accuracy and loss for the recognition models.

	Accuracy	Loss
NasNetMobile	0.9324	0.2160
MobileNetv2	0.9324	0.2131
ResNet101v2	0.9800	0.0734
ResNet152v2	0.9737	0.0808

examples that do not belong to the class  $C_i$  named as the true negative rate (TNR) or specificity,

$$Specificity_{NC_k} = TNR_{NC_k} = \frac{\sum_{k=1}^m a_{kk}}{\sum_{k=1}^m n_{C_k}}, k = 1, \dots, m, k \neq i. \quad (4)$$

Precision or Positive Predictive Value (PPV) indicates the rate of truly samples that belong to  $C_i$  from among all the examples predicted as  $C_i$ . Consequently, it represents how accurate was the algorithm on a given  $C_i$  class. and can be defined as:

$$Precision_{C_i} = PPV_{C_i} = \frac{a_{ii}}{\sum_{j=1}^m a_{ji}}. \quad (5)$$

When the number of samples in each class is unequal, some authors suggest that measures that combine *Recall* and *Precision* may be more appropriate [56,57]. From the text retrieval area, the  $F_\beta$ -score for a class  $C_i$  corresponds to the harmonic mean of *Recall* and *Precision*:

$$F_\beta - score_{C_i} = \frac{(1 + \beta^2) \cdot Recall_{C_i} \cdot Precision_{C_i}}{\beta^2 \cdot Precision_{C_i} + Recall_{C_i}}, \quad (6)$$

where  $\beta$  is a weight to adjust the importance of *Recall* concerning *Precision*. Commonly to give the same importance to both metrics

$$\beta = 1,$$

$$F_1 - score_{C_i} = 2 \cdot \frac{Recall_{C_i} \cdot Precision_{C_i}}{Precision_{C_i} + Recall_{C_i}}. \quad (7)$$

As can be seen, the previous metrics were defined for a two-class problem, where each time, a different class  $C_i$  was selected as the class of interest (positive) while the rest ( $m - 1$ ) was the negative class. However, two extensions to multi-class problems have been proposed in the literature: macro-average and micro-average. It has been demonstrated that the macro-average is the best strategy when the classes are not well-balanced [57]. From  $Recall_{C_i}$ ,  $Specificity_{NC_k}$ , and  $Precision_{C_i}$ , the following macro-average measures can be obtained [57,58]:

$$Recall_{macro-avg} = \frac{1}{m} \sum_{i=1}^m Recall_{C_i}, \quad (8)$$

$$Specificity_{macro-avg} = \frac{1}{m} \sum_{k=1}^m Specificity_{NC_k}, \quad (9)$$

$$Precision_{macro-avg} = \frac{1}{m} \sum_{i=1}^m Precision_{C_i}, \quad (10)$$

$$F_1 - score_{macro-avg} = 2 \cdot \frac{Recall_{macro-avg} \cdot Precision_{macro-avg}}{Precision_{macro-avg} + Recall_{macro-avg}}. \quad (11)$$

## 5. Experiments

The performance of the deep neural networks was analysed over the data set generated as described in Section 3 for the

**Table 3**

Confusion matrix of the pre-trained deep transfer learning models (true class labels on the rows and predicted class labels on the columns).

NasNetMobile						MobileNetv2					
	KN95	N95	Cloth	Surgical	Without		KN95	N95	Cloth	Surgical	Without
KN95	286	3	0	1	0	KN95	275	14	0	1	0
N95	9	279	0	2	0	N95	6	280	2	2	0
Cloth	19	23	225	18	5	Cloth	7	7	246	14	16
Surgical	4	7	1	277	1	Surgical	1	6	1	265	17
Without	1	1	1	2	285	Without	0	0	2	2	286

ResNet101v2						ResNet152v2					
	KN95	N95	Cloth	Surgical	Without		KN95	N95	Cloth	Surgical	Without
KN95	284	3	1	1	1	KN95	282	3	4	0	1
N95	2	285	0	3	0	N95	5	282	2	1	0
Cloth	0	2	280	4	4	Cloth	0	3	272	4	11
Surgical	1	1	0	287	1	Surgical	0	0	2	288	0
Without	0	0	4	1	285	Without	0	0	2	0	288

classification of face images into five different classes: face with a KN95 mask, face with an N95 mask, face with a cloth mask, face with a surgical mask, and face with no mask. To mitigate the problem of long training times, all architectures were trained for 10 epochs with 362 passes over the training samples.

All experiments were performed on the Google Colab computing service using Python programming language and two open-source software libraries, Keras and TensorFlow. The source code is available at <https://github.com/ricardomcupido/Google-Colab-Face-Mask-Classification>. The performance of the models was evaluated using five widely-used criteria: accuracy, recall, precision, specificity, and F1-score.

### 5.1. Results and discussion

Table 2 reports the values of accuracy and loss for the pre-trained deep transfer learning models. As can be observed, the highest training accuracy was achieved with the residual networks ResNet101v2 and ResNet152v2, which were as high as 0.9800 and 0.9737 respectively. Both these pre-trained models based on transfer learning also exhibited the lowest loss values during the training step. Regarding to the results of NasNetMobile and MobileNetv2, one can see that these models obtained an accuracy of 0.9324, whereas the loss of NasNetMobile was the worst.

For a more detailed review of the performance results, Table 3 shows the confusion matrix of the deep transfer learning models. With the help of a confusion matrix, the impact of misclassifications on the model performance can be better analysed. Thus, values in the confusion matrices indicate that all the models mainly failed to identify the images of faces wearing a cloth mask; it has to be noted that most errors were due to mistaking those for images of faces with a surgical mask or images of faces without a mask (for instance, the ResNet152v2 model misclassified 11 images of faces with a cloth mask as being of faces without a mask). Another common error found in these confusion matrices refers to the recognition of images with faces wearing a KN95 that were identified as an image with an N95 face mask and vice versa, which can be explained because these masks are alike and provide much the same structural and functional characteristics.

Hereafter we concentrate on the results of ResNet101v2 as the best performing pre-trained model. From the values in the confusion matrix of ResNet101v2, it is possible to calculate recall, precision, specificity and F1-score. Table 4 shows these performance evaluation measures of each class. As can be viewed, the highest value of recall was for the images of faces with a surgical

**Table 4**

Recall, precision, specificity and F1-score for the ResNet101v2 model.

	Recall	Precision	Specificity	F1-score
KN95	0.9793	0.9895	0.9974	0.9844
N95	0.9828	0.9794	0.9948	0.9811
Cloth	0.9655	0.9825	0.9957	0.9739
Surgical	0.9897	0.9696	0.9922	0.9795
Without	0.9828	0.9794	0.9948	0.9811
Macro-avg	0.9802	0.9801	0.9950	0.9802

mask, whereas the highest values of precision, specificity and F1-score were obtained for the class of KN95 face respirator masks. However, the most important picture of this table is that all classes (types of face masks) yielded high performance.

Fig. 5 collects some examples of images that were misclassified with the ResNet101v2 model. This can help one better understand the rationale of the errors produced during image recognition. For instance, the KN95 face mask in image labelled as A1 was recognized as a cloth mask, which can be due to the profile view of the image and also to the material and colour of the face mask. Faces partially occluded can also be a source of error in the recognition of a face mask as is the case of images A2 and B2. The model did not detect a face mask in images labelled as C2 and D2 because of their resemblance to a face. Errors on images in the third row of Fig. 5 result from misidentifying some article (jewel, sticker, hair, scarf) in the face as a face mask.

### 5.2. Comparison with related works

As mentioned in Section 2, similar proposals for mask detection can be found in the literature. However, they present several variations in the classification model, transfer learning approach, data sets, number of classes, and performance evaluation metrics. Therefore, making a fair and informative comparison under different protocols is not straightforward. Despite this, we compared our experimental results to show that the method proposed in this paper yields competitive recognition rates against state-of-the-art methods.

Table 5 summarizes the performance results of related works and those achieved by our proposal. For comparison purposes, we included the best results from all the experiments carried out in each paper. As can be observed, most of the works dealt with two-class problems to recognize mask vs. unmask, generally giving results higher than 98% in terms of accuracy and F1-score.

The only work focused on identifying two different types of masks (N95 and surgical) by using VGG-16 was the one by

**Table 5**  
Comparison of our proposal against related works.

Reference	Techniques	Classification problem	# Classes	Results
This paper	ResNet101v2	Multi-class	5	Accuracy = 98.00%
[19]	InceptionV3	Two-class	2	Accuracy = 100%
[20]	SRCNet	Multi-class	3	Accuracy = 98.70%
[21]	ResNet50 and YOLOv2	Two-class	2	Average Precision = 81%
[22]	YOLOv3 and Faster R-CNN (ResNet101-FPN)	Two-class	2	Average Precision = 62%
[23]	ResNet50 and SVM	Two-class	2	Accuracy = 100%
[24]	Faster R-CNN and InceptionV2	Two-class	2	F1-score = 94.19%
[25]	VGG16	Multi-class	7	Accuracy = 99.29%
[27]	SE-YOLOv3	Multi-class	3	Average Precision = 73.7%
[28]	CSPDarkNet53, PANet and YOLOv4	Multi-class	3	Average Precision = 98.30%, F1-score = 96.7%
[29]	ResNet50	Two-class	2	Accuracy = 98.20%
[30]	YOLOv3	Two-class	2	Average confidence = 97.00%
[31]	MobileNetv2	Two-class	2	Accuracy = 96.85%
[32]	MobileNetv2 and Single Shot Detector	Two-class	2	Accuracy = 91.70%
[33]	MobileNetv2	Two-class	2	Accuracy = 98.00%
[34]	MobileNetv2	Two-class	2	Accuracy = 98.00%
[35]	InceptionV3	Two-class	2	Accuracy = 98.00%
[36]	Single Shot Multibox Detector and MobileNetv2	Two-class	2	Accuracy = 92.64%, F1-score = 93.00%
[37]	YOLOv4	Two-class	2	Average Precision = 88%, F1-score = 99.54%
[38]	MobileNetv2	Two-class	2	Accuracy = 81.74%
[39]	Multigraph CN-VGG16	Two-class	2	Accuracy = 97.9%
[40]	MobileNetv2, DenseNet121, NASNet	Two-class	2	F1-score = 99.40%
[41]	MobileNetv2-SVM	Two-class	2	Accuracy = 97.11%
[42]	VGG-16	Multi-class	3	Accuracy = 99.81%

Hussain et al. [42], reporting a recall of 97% and 99% for N95 and surgical masks, respectively. As can be seen in Table 4, the ResNet101v2 model applied to our multi-class problem (KN95, N95, cloth, surgical, and without) performed similar or even better than the VGG-16 approach ran over an easier two-class problem.

## 6. Conclusions

This paper has developed a deep transfer learning model for the recognition of different types of face masks. The proposed method consists of two main stages. The first part includes data preprocessing (conversion of image format, ROI detection, and normalization) and data augmentation (rotation, width and height shifting, shearing, horizontal flipping, and pixel filling), whereas the second stage is for the recognition process using a deep transfer learning neural network.

Experiments have investigated the performance of four deep transfer learning models (NasNetMobile, MobileNetv2, ResNet101v2, and ResNet152v2) pre-trained on ImageNet database. The experimental results have shown that both residual networks perform better than the other models, yielding very high accuracy and low loss. In particular, the ResNet101v2 appears as the best deep transfer learning model for the task of distinguishing between four different types of face masks, with an accuracy of 98% and a loss of 0.0734.

Despite its contributions, the results of this paper should not be interpreted without accounting for some limitations that could be addressed in future works. The emphasis of this paper has been on four different types of face masks, but other types should be included for a more real scenario. Although assessing mask fitting is way beyond the scope of this work, this seems to be a critical thing to be watched as a loose-fit can significantly impair even high-filtration masks. On the other hand, the models have been trained with images from public databases, but a further step will consist of elucidating the performance with a set of images of much lower resolution, such as CCTV footage. Finally, the research has analysed four deep transfer learning models, but it could be extended to other models such as Inception and DenseNet.

## CRedit authorship contribution statement

**Ricardo Mar-Cupido:** Methodology, Software, Formal analysis, Investigation. **Vicente García:** Conceptualization, Methodology, Validation, Supervision. **Gilberto Rivera:** Methodology, Validation, Supervision. **J. Salvador Sánchez:** Supervision, Writing – original draft, Writing – review & editing, Funding acquisition.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Funding

This work was supported by Universitat Jaume I, Spain [grant number UJI-B2018-49].

## References

- [1] C.C. Wang, K.A. Prather, J. Sznitman, J.L. Jimenez, S.S. Lakdawala, Z. Tufekci, L.C. Marr, Airborne transmission of respiratory viruses, *Science* 373 (6558) (2021) eabd9149.
- [2] D.K. Chu, E.A. Akl, S. Duda, K. Solo, S. Yaacoub, H.J. Schünemann, et al., Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: A systematic review and meta-analysis, *Lancet* 395 (2020) 1973–1987.
- [3] M.A. Colomer, A. Margalida, F. Alòs, P. Oliva-Vidal, A. Vilella, L. Fraile, Modelling the SARS-CoV-2 outbreak: Assessing the usefulness of protective measures to reduce the pandemic at population level, *Sci. Total Environ.* 789 (2021) 147816.
- [4] M. Abboah-Offei, Y. Salifu, B. Adewale, J. Bayuo, R. Ofosu-Poku, E.B.A. Opore-Lokko, A rapid review of the use of face mask in preventing the spread of COVID-19, *Int. J. Nurs. Stud. Adv.* 3 (2021) 100013.
- [5] S.W. Sim, K.S. Moey, N.C. Tan, The use of facemasks to prevent respiratory infection: A literature review in the context of the Health Belief Model, *Singapore Med. J.* 55 (3) (2014) 160–167.
- [6] E.P. Fischer, M.C. Fischer, D. Grass, I. Henrion, W.S. Warren, E. Westman, Low-cost measurement of face mask efficacy for filtering expelled droplets during speech, *Sci. Adv.* 6 (36) (2020) eabd3083.
- [7] S. Anjum, T. Islam, Types of face coverings (masks) and coronavirus disease 2019 (COVID-19), *Eur. J. Biol. Res.* 11 (3) (2021) 325–331.
- [8] L. Li, M. Niu, Y. Zhu, Assessing the effectiveness of using various face coverings to mitigate the transport of airborne particles produced by coughing indoors, *Aerosol Sci. Technol.* 55 (3) (2021) 332–339.

- [9] M.-C. Kim, S. Bae, J.Y. Kim, S.Y. Park, J.S. Lim, M. Sung, S.-H. Kim, Effectiveness of surgical, KF94, and N95 respirator masks in blocking SARS-CoV-2: A controlled comparison in 7 patients, *Infect. Dis.* 52 (12) (2020) 908–912.
- [10] E.-H. Lee, S.-W. Lee, S.Y. Moon, J. Son, Performance evaluation of commercially available masks in Korea for filtering airborne droplets containing bacteria, *Int. J. Environ. Res. Public Health* 18 (15) (2021) 7909.
- [11] E. O'Kelly, A. Arora, S. Pirog, J. Ward, P.J. Clarkson, Comparing the fit of N95, KN95, surgical, and cloth face masks and assessing the accuracy of fit checking, *PLoS One* 16 (1) (2021) 1–14.
- [12] T. Oberg, L.M. Brosseau, Surgical mask filter and fit performance, *Am. J. Infect. Control* 36 (4) (2008) 276–282.
- [13] S.-A. Lee, S.A. Grinshpun, T. Reponen, Respiratory performance offered by N95 respirators and surgical masks: Human subject evaluation with NaCl aerosol representing bacterial and viral particle size range, *Ann. Occup. Hyg.* 52 (3) (2008) 177–185.
- [14] J. Abaluck, L.H. Kwong, A. Styczynski, A. Haque, M.A. Kabir, E. Bates-Jefferys, E. Crawford, J. Benjamin-Chung, S. Benhachmi, S. Raihan, S. Rahman, N. Zaman, P.J. Winch, M.M. Hossain, H.M. Reza, S.P. Luby, A.M. Mobarak, Normalizing Community Mask-Wearing: A Cluster Randomized Trial in Bangladesh, in: Working Paper Series, (28734) National Bureau of Economic Research, 2021.
- [15] Types of Masks and Respirators - CDC, <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/types-of-masks.html>. (Accessed 28 May 2022).
- [16] COVID-19: Information and latest updates - Mayo Clinic Health System, <https://www.mayoclinichealthsystem.org/covid19>. (Accessed 28 May 2022).
- [17] C. Ansberry, N. Subbaraman, J.R. Brinson, Why cloth masks might not be enough as omicron spreads, 2022, <https://www.wsj.com/articles/cloth-face-mask-omicron-11640984082>. (Accessed 28 May 2022).
- [18] B.M. Gurbaxani, A.N. Hill, P. Paul, P.V. Prasad, R.B. Slayton, Evaluation of different types of face masks to limit the spread of SARS-CoV-2: A modeling study, *Sci. Rep.* 12 (1) (2022) 8630.
- [19] G.J. Chowdary, N.S. Punni, S.K. Sonbhadra, S. Agarwal, Face mask detection using transfer learning of InceptionV3, in: L. Bellatreche, V. Goyal, H. Fujita, A. Mondal, P.K. Reddy (Eds.), *Big Data Analytics*, Springer International Publishing, Cham, 2020, pp. 81–90.
- [20] B. Qin, D. Li, Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19, *Sensors* 20 (18) (2020) 5236.
- [21] M. Loey, G. Manogaran, M.H.N. Taha, N.E.M. Khalifa, Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection, *Sustainable Cities Soc.* 65 (2021) 102600.
- [22] S. Singh, U. Ahuja, M. Kumar, K. Kumar, M. Sachdeva, Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment, *Multimedia Tools Appl.* 80 (2021) 19753–19768.
- [23] M. Loey, G. Manogaran, M.H.N. Taha, N.E.M. Khalifa, A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic, *Measurement* 167 (2021) 108288.
- [24] B. Wang, Y. Zhao, C.L.P. Chen, Hybrid transfer learning and broad learning system for wearing mask detection in the COVID-19 era, *IEEE Trans. Instrum. Meas.* 70 (2021) 1–12.
- [25] S. Chen, W. Liu, G. Zhang, Efficient transfer learning combined skip-connected structure for masked face poses classification, *IEEE Access* 8 (2020) 209688–209698.
- [26] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, D. Zhao, The CAS-PEAL large-scale Chinese face database and baseline evaluations, *IEEE Trans. Syst. Man Cybern. A* 38 (1) (2008) 149–161.
- [27] X. Jiang, T. Gao, Z. Zhu, Y. Zhao, Real-time face mask detection method based on YOLOv3, *Electronics* 10 (7) (2021) 837.
- [28] J. Yu, W. Zhang, Face mask wearing detection algorithm based on improved YOLO-v4, *Sensors* 21 (9) (2021) 3263.
- [29] S. Sethi, M. Kathuria, T. Kaushik, Face mask detection using deep learning: An approach to reduce risk of coronavirus spread, *J. Biomed. Inform.* 120 (2021) 103848.
- [30] M.R. Prusty, V. Tripathi, A. Dubey, A novel data augmentation approach for mask detection using deep transfer learning, *Intell.-Based Med.* 5 (2021) 100037.
- [31] S.A. Sanjaya, S. Adi Rakhmawan, Face mask detection using MobileNetV2 in the era of COVID-19 pandemic, in: Proc. of International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Econom, Sakheer, Bahrain, 2020, pp. 1–5.
- [32] S. Yadav, Deep learning based safe social distancing and face mask detection in public areas for COVID-19 safety guidelines adherence, *Int. J. Res. Appl. Sci. Eng. Technol.* 8 (VII) (2020) 1368–1374.
- [33] F. Mercaldo, A. Santone, Transfer learning for mobile real-time face mask detection and localization, *J. Am. Med. Inform. Assoc.* 28 (7) (2021) 1548–1554.
- [34] S.S. Palani, M. Dev, G. Mogili, D. Relan, R. Dey, Face mask detector using deep transfer learning and fine-tuning, in: Proc. of 8th International Conference on Computing for Sustainable Global Development, New Delhi, India, 2021, pp. 695–698.
- [35] K. Suresh, M.B. Palangappa, S. Bhuvan, Face mask detection by using optimistic convolutional neural network, in: Proc. of 6th International Conference on Inventive Computation Technologies, Coimbatore, India, 2021, pp. 1084–1089.
- [36] P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria, J. Hemanth, SSDMNv2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2, *Sustainable Cities Soc.* 66 (2021) 102692.
- [37] A. Gola, S. Panesar, A. Sharma, G. Ananthkrishnan, G. Singal, D. Mukhopadhyay, MaskNet: Detecting different kinds of face mask for Indian ethnicity, in: D. Garg, K. Wong, J. Sarangapani, S.K. Gupta (Eds.), *Communications in Computer and Information Science*, Vol. 1367, Springer, Singapore, 2020, pp. 492–503.
- [38] A.S. Joshi, S.S. Joshi, G. Kanahasabai, R. Kapil, S. Gupta, Deep learning framework to detect face masks from video footage, in: Proc. of 12th International Conference on Computational Intelligence and Communication Networks, Bhimtal, India, 2020, pp. 435–440.
- [39] A. Alguzo, A. Alzu'bi, F. Albalas, Masked face detection using multi-graph convolutional networks, in: Proc. of 12th International Conference on Information and Communication Systems, Valencia, Spain, 2021, pp. 385–391.
- [40] A. Chavda, J. Dsouza, S. Badgajar, A. Damani, Multi-stage CNN architecture for face mask detection, in: Proc. of 6th International Conference for Convergence in Technology, Maharashtra, India, 2021, pp. 1–8.
- [41] A. Oumina, N. El Makhfi, M. Hamdi, Control the COVID-19 pandemic: Face mask detection using transfer learning, in: Proc. of IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science, Kenitra, Morocco, 2020, pp. 1–5.
- [42] S. Hussain, Y. Yu, M. Ayoub, A. Khan, R. Rehman, J.A. Wahid, W. Hou, IoT and deep learning based approach for rapid screening and face mask detection for infection spread control of COVID-19, *Appl. Sci.* 11 (8) (2021) 3495.
- [43] D. Makwana, B. Wadhvani, Face mask classification: Classify between face with mask and no mask, 2020, <https://www.kaggle.com/dhruvmak/face-mask-detection>. (Accessed 10 June 2021).
- [44] A. Larxel, Face mask detection, 2020, <https://www.kaggle.com/andrewmvd/face-maskdetection>. (Accessed 19 June 2021).
- [45] M.A. Anwar, A. Raychowdhury, Masked face recognition for secure authentication, 2020, <https://github.com/aqeelanwar/MaskTheFace>. (Accessed 17 June 2021).
- [46] S. Ge, J. Li, Q. Ye, Z. Luo, Detecting masked faces in the wild with LLE-CNNs, in: Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, 2017, pp. 426–434.
- [47] I.J. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, Cambridge, MA, USA, 2016.
- [48] J. Wang, L. Perez, The Effectiveness of Data Augmentation in Image Classification using Deep Learning, Tech. rep., Stanford University, 2017.
- [49] S. Gu, M. Pednekar, R. Slater, Improve image classification using data augmentation and neural networks, *SMU Data Sci. Rev.* 2 (2) (2019) 1–43.
- [50] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, *J. Big Data* 6 (1) (2019) 1–48.
- [51] J. Shijie, W. Ping, J. Peiyi, H. Siping, Research on data augmentation for image classification based on convolution neural networks, in: 2017 Chinese Automation Congress, CAC, 2017, pp. 4165–4170.
- [52] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, MobileNets: Efficient convolutional neural networks for mobile vision applications, 2017, ArXiv, [arXiv:1704.04861](https://arxiv.org/abs/1704.04861).
- [53] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, 2016, pp. 770–778.
- [54] B. Zoph, V. Vasudevan, J. Shlens, Q.V. Le, Learning transferable architectures for scalable image recognition, in: Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 8697–8710.
- [55] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, *Inf. Process. Manage.* 45 (4) (2009) 427–437.

- [56] A.P. Bradley, R.P.W. Duin, P. Paclik, T.C.W. Landgrebe, Precision-recall operating characteristic (P-ROC) curves in imprecise environments, in: 18th International Conference on Pattern Recognition, ICPR'06, Vol. 4, 2006, pp. 123–127.
- [57] P. Branco, L. Torgo, R.P. Ribeiro, A survey of predictive modeling on imbalanced domains, *ACM Comput. Surv.* 49 (2) (2016).
- [58] I. Markoulidakis, I. Rallis, I. Georgoulas, G. Kopsiaftis, A. Doulamis, N. Doulamis, Multiclass confusion matrix reduction method and its application on net promoter score classification problem, *Technologies* 9 (4) (2021).