Enhancing Facemask Detection using Deep learning Models

Abdullahi Ahmed Abdirahman[1]*, Abdirahman Osman Hashi[2]*, Ubaid Mohamed Dahir[3]*,
Mohamed Abdirahman Elmi[4]*, Octavio Ernest Romo Rodriguez[5]
Faculty Member, SIMAD University, Department of Computing, Mogadishu Somalia[1, 2, 3, 4]
Department of Computer Science-Faculty of Informatics, İstanbul Teknik Üniversitesi, İstanbul, Turkey[5]

***Abstract*—Face detection and mask detection are critical tasks in the context of public safety and compliance with mask-wearing protocols. Hence, it is important to track down whoever violated rules and regulations. Therefore, this paper aims to implement four deep learning models for face detection and face with mask detection: MobileNet, ResNet50, Inceptionv3, and VGG19. The models are evaluated based on precision and recall metrics for both face detection and face with mask detection tasks. The results indicate that the proposed model based on ResNet50 achieves superior performance in face detection, demonstrating high precision (99.4%) and recall (98.6%) values. Additionally, the proposed model shows commendable accuracy in mask detection. MobileNet and Inceptionv3 provide satisfactory results, while the proposed model based on VGG19 excels in face detection but shows slightly lower performance in mask detection. The findings contribute to the development of effective face mask detection systems, with implications for public safety.**

***Keywords*—Object detection; deep learning; detection; face detection; mask detection; convolutional neural network**

I. INTRODUCTION

Computer vision is a rapidly advancing field that encompasses a wide range of technologies aimed at enabling machines to perceive and interpret visual information, similar to how humans do. One crucial task within computer vision is face detection, which involves locating and identifying human faces in digital images or video streams. Face detection has gained significant attention and importance due to its wide-ranging applications in various domains, including surveillance systems, biometric authentication, facial recognition, human-computer interaction, and social media analysis [1]. Over the years, researchers have made remarkable progress in developing sophisticated face detection algorithms that exhibit high accuracy and robustness. Despite the progress made in general target detection algorithms across various domains, the efficacy of face mask detection techniques remains constrained [2]. In response, researchers have directed their efforts towards this area, employing the "you only look once v2" (YOLOv2) algorithm to devise detection models. Furthermore, advancements have been made by leveraging the YOLOv3 algorithm, which facilitates enhanced feature extraction through an optimized way [3].

However, these challenges arise due to variations in lighting conditions, occlusions, pose variations, complex backgrounds, and scale variations. Lighting variations can lead to significant changes in facial appearance, making it challenging to detect faces consistently. Occlusions, such as glasses, facial hair, or partial face obstructions, further complicate the task by hiding crucial facial features. Additionally, face detection algorithms must handle pose variations, where faces may be rotated, tilted, or viewed from different angles. Complex backgrounds with cluttered scenes pose another challenge, as it becomes difficult to differentiate faces from the surrounding environment[4]. Similarly, scale variations, caused by the varying distances between the camera and the subjects, necessitate robust face detection algorithms that can handle faces of different sizes are required [5].

Over the years, researchers have proposed various face detection techniques, each aiming to address the challenges mentioned above and improve the accuracy and efficiency of face detection algorithms. Early approaches utilized handcrafted features and traditional machine-learning-algorithms, such as Haar cascades and Histogram of Oriented-Gradients (HOG), to detect faces. These methods achieved reasonable results but had limitations in handling pose variations and complex backgrounds. In recent years, the advent of deep learning, particularly convolutional neural networks (CNNs), has revolutionized the field of face detection. CNN-based architectures, such as the Viola-Jones framework, Single Shot MultiBox Detector (SSD), and Faster R-CNN, have demonstrated superior performance in face detection tasks. These models leverage the power of deep learning to automatically learn discriminative features from large-scale datasets, enabling them to handle various challenges faced in face detection. Notably, the use of region-based convolutional neural networks (R-CNN) has greatly improved accuracy by combining region proposals and convolutional networks, allowing for more precise localization of faces [6].

Other researcher improved the YOLOva and made that the YOLO-network generates predictions for bounding boxes in each grid of an image with a size of G×G pixels. However, the network encounters challenges in detecting smaller objects since each bounding box can only be assigned a single class during prediction. The primary issue with YOLO arises from its limitations in accurately localizing objects, particularly when dealing with bounding boxes of unusual ratios [7]. On the other hand, in the realm of face mask detection, various transfer learning approaches have been employed to address the challenges encountered in real-world scenarios. One such method involves utilizing a pre-trained InceptionV3 model as a transfer learning technique to discern individuals wearing or

not wearing masks [3]. For instance, the author in [1] employed a cascading approach using Convolutional Neural Networks (CNNs) to detect faces that are covered with masks. In recent advancements, a dedicated framework called the Retina Face Mask network has been developed to enable accurate and efficient recognition of face masks. Various experiments have been conducted to devise an automated technique for determining whether an individual is wearing a face mask or not. Real-time detection of facial masks has been achieved using the YOLOv3-technique and the Haar-cascading-classifier, but the challenges are remains constrained [3].

Although face detection and recognition, particularly in the presence of masks, pose significant challenges and have been the subject of extensive research in recent years, this research aims to improve the detection performance of masked and unmasked faces, with a specific focus on face masks. The problem of face mask-detection in the fields of image-processing and computer vision is exceptionally complex. The primary objective of this study is to enhance public safety by employing deep learning techniques to identify individuals wearing or not wearing masks in public areas. The developed mask detector can play a crucial role in ensuring our protection. Additionally, witnessing the global impact of the COVID-19 pandemic further motivates the exploration of machine learning techniques to address the real-world problem of habitual mask-wearing when venturing outdoors. Hence, the proposed approach in this study utilizes transfer learning, specifically applying the pre-trained MobileNetV2, RestNet50, InceptionV3 and VGG19 model for fine-tuning the face mask detection task.

II. Related Work

Face detection is a fundamental task in computer vision that involves locating and identifying human faces in digital images or video streams. Over the years, researchers have made significant advancements in developing accurate and robust face detection algorithms [8]. Early approaches in face detection primarily relied on handcrafted features and traditional machine learning algorithms. Author [9] proposed one of the seminal methods, known as the Viola-Jones framework, which utilized Haar-like features and a cascade classifier for rapid face detection. This approach laid the foundation for subsequent advancements in face detection. However, these methods had limitations in handling pose variations, occlusions, and complex backgrounds. The advent of deep learning, particularly convolutional neural networks (CNNs), revolutionized face detection. Researchers explored various CNN architectures for accurate and robust face detection. For instance, author [10] proposed the Single Shot MultiBox Detector (SSD), which combined a deep CNN with a set of anchor boxes for efficient face detection. This approach achieved excellent performance in Region-based convolutional neural networks (R-CNNs) further improved the accuracy of face detection. Author [11] introduced the Faster R-CNN framework, which integrated a region proposal network with a CNN-based object detection network. This method enabled precise localization of faces and achieved state-of-the-art performance in face detection tasks. Subsequent research efforts focused on enhancing the speed and efficiency of R-CNN-based methods, leading to variants like Fast R-CNN. Despite the advancements in face detection techniques, several challenges persist. Variations in lighting conditions pose a significant challenge as they can affect the appearance of faces [12].

To address this, author [13] proposed an illumination-robust face detection method that utilized color normalization and multiple thresholds to handle lighting variations. Occlusions, such as glasses, facial hair, or partial face obstructions, present another challenge. Author [14] introduced a method that employed a deformable part model to handle occlusions and achieve accurate face detection. Pose variations also pose challenges, as faces may be rotated, tilted, or viewed from different angles. Author [15] proposed a pose-aware face detection approach that utilized pose estimations to improve detection accuracy.

Meanwhile, complex backgrounds with cluttered scenes make it difficult to differentiate faces from the surrounding environment. Author [8] addressed this challenge by proposing a context-aware face detection method that leveraged contextual cues to enhance the accuracy of face detection in complex scenes. Additionally, scale variations, caused by the varying distances between the camera and the subjects, require robust face detection algorithms. Similarly, author [16] proposed a scale-aware face detection method that utilized a multi-scale convolutional network to handle faces at different sizes and face localization. In face localization, the goal is to figure out where and how big a certain number of faces are (usually one). In general, there are two ways to find facial parts in a digital image: the feature-based approach and the image-based approach. The feature-based approach tries to pull out parts of the image and compare them to what is known about the face. While image-based methods try to find the best match between the images used for training and the ones used for testing. People often use the following ways to find faces in a still image or a video sequence:

A. Feature based Approaches

Feature-based approaches for face mask detection leverage the distinctive visual characteristics and patterns associated with the presence or absence of masks. These methods primarily rely on handcrafted features and machine learning algorithms to classify faces as masked or unmasked. Some commonly used for features approaches include active shape model and Low level model [11]. And it can be classified by active shape and low level analysis.

The Active Shape Model (ASM) focuses on intricate, flexible aspects, including how features appear and behave. Finding landmark points in a picture that determine the form of any statistically modelled object is the primary objective of ASM. For instance, the eyes, lips, nose, mouth, and eyebrows were removed from a photograph of a person's face. An ASM's statistical face model is created using photos that include hand-marked landmarks during the training process. Three categories of ASMs are distinguished: templates that may alter form, point distribution models (PDMs), and snakes. The first form to use an active contour is known as a snake. To indicate the margins of the head, snakes are employed. A snake must first be positioned near to a head barrier in order to finish the

mission. It then changes into the form of a head after scanning the surrounding edges. Snakes develop by reducing the "esnake" energy function, which is similar to how physical systems operate [17].

Internal energy is the component that derives from the snake's own characteristics and demonstrates how it has evolved organically. Snakes often undergo change by either contracting or growing. The contours might diverge from their normal development and finally take on the form of adjacent features—the head boundary at equilibrium—because the external energy resists the internal energy. There are two key considerations while creating snakes: which energy phrases to utilise and how to use the least amount of energy. Elastic energy is often referred to as "internal energy". The snake's internal energy modifies the distance between its control points. It gains a shape as a result, which acts as an elastic band and causes it to shrink or expand. Image characteristics, on the other hand, rely on external energy. To determine how to consume the least amount of energy, optimisation methods like steepest gradient descent are applied. For quick iteration, there are other greedy algorithms. There are various drawbacks to snakes, such as how often their edges get caught on false image features and how they can't be utilised to eliminate non-convex features [18].

For instance, author [19] introduced the Constrained Local Models (CLM), which combined ASM with a local appearance model to handle non-rigid facial deformations but it has still some mistakes about the edges. CLM utilized a patch-based appearance model to capture local appearance variations and refine the shape estimation iteratively, leading to improved accuracy in face detection and landmark localization. Another significant contribution is the Supervised Descent Method (SDM) proposed by author [20]. SDM incorporated a cascade regression framework with ASM to refine the shape estimation progressively. This method achieved state-of-the-art performance in face detection and facial landmark localization, particularly in real-time scenarios.

On the other hand, low-level analysis models have significantly contributed to the advancement of face detection by extracting relevant low-level visual features. The evolution of these models, from handcrafted features to deep learning-based approaches, has led to improved accuracy and robustness in face detection. Challenges such as variations in lighting conditions, complex backgrounds, and occlusions continue to be addressed through illumination normalization techniques, context-aware models, and adaptive feature learning. Several notable research contributions have advanced the field of low-level analysis models in face detection. For instance, author [21] introduced the DeepFace model, which used a deep CNN to learn discriminative facial features. DeepFace achieved remarkable performance in face detection and recognition, particularly in handling pose variations and challenging lighting conditions. Another significant contribution is the FaceBoxes method proposed by [22]. FaceBoxes utilized a lightweight CNN architecture specifically designed for face detection tasks but have some problem with color.

Meanwhile, the color of the skin serves as the face's foundational feature. There are many benefits to using skin tone as a face-tracking feature. Facial traits other than color are processed far more slowly than the former. Color may be seen from any direction under specific lighting circumstances. Since only a translation model is required for motion estimation, this feature greatly simplifies the process. There are several obstacles when trying to use color as a feature for tracking human faces, such as the fact that the color representation of a face obtained by a camera can be affected by things like lighting conditions and the motion of the object being photographed [23].

B. Image based Approaches

Image-based approaches in the field of face detection refer to the methods and techniques that rely on analyzing and processing images to detect the presence and location of human faces. These approaches utilize the visual information present in images, such as pixel values, colors, textures, edges, and spatial relationships, to identify regions that potentially contain faces [3].

One of the examples of images based approaches is Neural network-based image analysis approaches that have revolutionized the field of face detection by leveraging the power of deep learning to extract meaningful features from images and accurately detect faces. For instance, author [24] introduced the Deep Convolutional Network Cascade (DCNC) for face detection. DCNC utilized a cascade architecture of CNNs to achieve high accuracy while maintaining real-time performance. Multi-task Cascaded Convolutional Networks (MTCNN) is another significant contribution in the field. Author [25] proposed MTCNN, which simultaneously performs face detection, facial landmark localization, and facial attribute classification using cascaded CNNs. MTCNN achieved state-of-the-art performance in face detection tasks, particularly for faces with various poses, scales, and occlusions.

Meanwhile, one-Shot detectors have gained attention for their efficiency and effectiveness in face detection. These detectors aim to accurately locate faces in a single pass of the neural network, enabling real-time performance. One notable contribution in this area is the Single Shot MultiBox Detector (SSD) introduced by [26]. SSD utilizes a single neural network to perform face detection by predicting bounding box locations and class probabilities at multiple scales. This approach achieved high accuracy while maintaining fast inference speed. Another notable contribution is the RetinaFace model proposed by [27]. RetinaFace utilized a single-stage dense face localization approach that incorporated anchor-based and anchor-free strategies to handle faces with various scales and poses.

Similarly, one auto-associative network detects frontal-view faces, and another detects faces turned 60 degrees left or right. Then, a face detection system employing PDBNN was introduced by [28]. PDBNN is like an RBF network with probabilistic learning rules. The system consists of two stages: pre-processing and processing. In contrast to that, a deep-dense face detector was proposed by [29] requires that no single model can annotate poses or landmarks and recognize faces in many orientations.

On the other hand, Support Vector Machine (SVM) has been widely employed in the field of face detection due to its ability to effectively handle complex classification tasks. For example, author [30] introduced the "Support Vector Machines Applied to Face Detection" (SVMAFD) method, which showcased the effectiveness of SVMs for face detection. The authors presented an SVM-based approach that utilized a set of carefully designed features to classify image sub-windows as face or non-face. SVMAFD achieved promising results, demonstrating the potential of SVMs in face detection. Similarly, to that, author [31] proposed an improved SVM-based face detection method that incorporated feature selection techniques. The authors employed a combination of Haar-like features and Local Binary Patterns (LBP) and applied Recursive Feature Elimination (RFE) to select the most discriminative features. Their approach achieved competitive performance in face detection tasks, highlighting the importance of feature selection for SVM-based methods.

III. Proposed Model

This research framework is developed based on a benchmark for object-recognition presented in reference [32]. As it can be seen from Figure 1, it shows that how this benchmark divides object-recognition tasks into training, classification, and detection tasks. Training and deployment use separate pipelines to assure surveillance device compatibility. The training process creates an impartial customised dataset and fine-tunes all models. Face identification and extraction follow image/real-time video frame extraction in the deployment process.

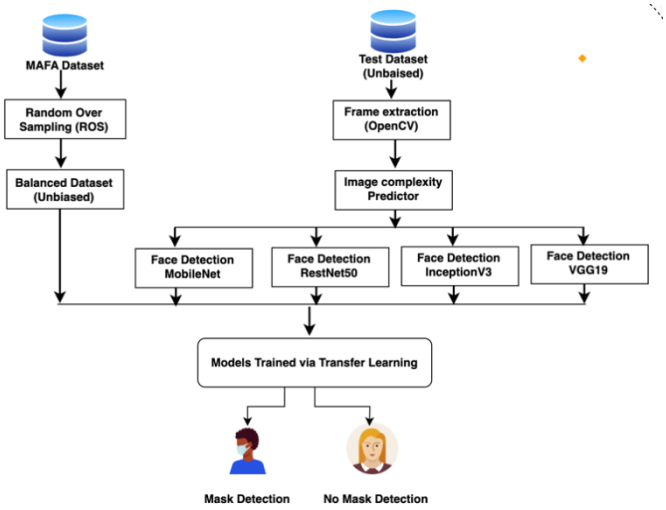

Fig. 1. Proposed methodology.

The classification task corresponds to a baseline convolutional-neural-network (CNN) that extracts information as data from input images and generates a feature map in the baseline. In this framework, transfer learning is applied on the classification, leveraging the learned attributes of a pre-trained and powerful CNN to extract new features for the model. To achieve optimal performance in facemask detection, an extensive backbone building strategy is conducted, utilizing four popular pre-trained models: MobileNet, ResNet50, Inception and VGG19. The novelty of the proposed work lies in the training task, an intermediate module that performs various preprocessing tasks before the actual image classification.

In the deep-learning neural network, the detection acts as an identity detector or predictor. The trained facemask classifier acquired by transfer learning is used in the proposed architecture to recognise faces with or without masks. The ultimate goal is to deter people from wearing face masks in public spaces by identifying those who do so. The following steps may be conducted in line with administrative or governmental policies. Using OpenCV 0.20, similar to how previous studies utilised it [33], an affine transformation approach is used to detect facial characteristics due to differences in face size and orientation inside cropped areas of interest (ROI). This guarantees correct identification despite variations in face features. The following points provide a thorough explanation of each job in the proposed framework.

A pre-trained under supervision is the first step. On the initial biassed MAFA dataset, the CNN model underwent discriminative pre-training. The free Caffe Python package was used for the pre-training procedure [33] same as this author. In order to ensure that the model learns generalizable features and to enable better performance and faster convergence on the target task with little labelled data, this pre-training step aids the model in capturing general knowledge about the data distribution and extracting high-level representations.

A finite-turning of pre-trained is the second step. Due to its improved performance compared to other classification techniques, deep neural networks are used in this research to identify facemasks. Deep neural network training, however, is a time- and resource-intensive process that needs a lot of computing power. Transfer learning based on deep learning concepts is used to overcome these issues and produce quicker and more affordable training. Transfer learning enables the transfer of learnt information from an existing neural network to a new model in terms of the parameter weights. Even when trained on a modestly sized dataset, the new model performs much better thanks to this method. ImageNet, a large dataset with over 14 million photos, has been used to train a number of pre-trained models, including MobileNet and ResNet50. For this framework, the pre-trained models for facemask classification include MobileNet, ResNet50, Inceptionv3, and VGG19. Each of them has five more layers added to the final layer to refine it. These recently added layers are composed of a flattening layer, a dense ReLU layer with 128 neurons, an average pooling layer with a pool size of 5x5, a dropout layer with a rate of 0.4, and a deciding layer that uses the softmax activation function for binary classification.

Finding the expected face and determining which faces were found wearing masks or not is done in the final step. Following face recognition, the faces sans masks are individually fed into a neural network to investigate the identity of the person, paying particular attention to those who deviate from the face-mask norms. However, a fixed-sized input is needed for this stage. One method to meet this criterion is to resize the face inside the bounding box to 96x96 pixels which we done it. However, if the face is facing in a different direction, there may be a problem with this technique. A

simple solution is provided by the use of an affine transformation approach to address this problem. This method resembles the deformable part models proposed in [34] in certain ways were also used to tackle it.

IV. Results and Discussions

The upcoming sections will illustrate the dataset description followed with its discussion on proposed models in term of face detection and face with mask detection.

A. Dataset Description

The MAFA (Multi-Attribute Facial Action) dataset is a facemask-centric dataset that has been widely used in the field of computer vision, particularly for tasks related to facemask detection and analysis. The MAFA dataset consists of a large collection of facial images that are annotated with various attributes related to facial appearance, including the presence or absence of a facemask. The dataset was specifically curated to address the need for comprehensive and accurate facemask detection in real-world scenarios, such as surveillance systems or public health monitoring.

The MAFA dataset is composed of over 35,000 facial images captured from diverse sources, including different genders, age groups, and ethnicities. This diversity ensures that the dataset covers a wide range of facial variations, which is essential for training robust facemask detection models. Each facial image in the MAFA dataset is manually annotated with multiple attributes, including the presence or absence of a facemask, gender, age group, and other facial attributes. These annotations provide valuable ground truth information for various facial analysis tasks. To ensure unbiased performance of the facemask detection models, the MAFA dataset undergoes an unbiased customization process during the training phase. This process involves carefully selecting and balancing the training samples to minimize any biases that may arise due to the dataset's composition.

B. Identify Comparing MobileNet, ResNet50, Inceptionv3 and VGG19

The four models were evaluated for face detection and mask detection tasks as a separated way. The models include RetinaFaceMask based on MobileNet, RetinaFaceMask model based on ResNet50, RetinaFaceMask based on Inceptionv3, and RetinaFaceMask model based on VGG19. The performance of each model was assessed in terms of precision and recall for both face detection and mask detection.

RetinaFaceMask based on MobileNet in term of Face Detection, the model achieved a precision of 84.0% and a recall of 96.0%. This indicates that the model can effectively detect faces, with a relatively high recall rate, capturing a majority of the true faces present in the images. In term of Mask Detection, the model achieved a precision of 81.3% and a recall of 88.2%. This suggests that the model performs reasonably well in detecting whether individuals are wearing masks or not, with a good balance between precision and recall.

On the other hand, the proposed model based on ResNet50 in term of Face Detection has demonstrated excellent performance in face detection, achieving a high precision of 99.4% and a recall of 98.6%. These results indicate that the model is highly accurate in detecting faces, with a low false positive rate and a high true positive rate. Meanwhile, in term of Mask Detection, the model also showed strong performance in mask detection, with a precision of 98.83% and a recall of 98.5%. These results indicate that the model can effectively distinguish between masked and unmasked individuals, with a high level of accuracy and recall.

Similarly, RetinaFaceMask based on Inceptionv3 in term of Face Detection has also achieved a precision of 80.0% and a recall of 91.4% in face detection. Although the precision is relatively lower compared to other models, the model shows a good recall rate, capturing a high percentage of faces in the images. However, in term of Mask Detection, the model achieved a precision of 92.1% and a recall of 86.3%. This suggests that the model performs well in detecting masks, with a higher emphasis on precision compared to recall.

Final model RetinaFaceMask based on VGG19 in term of Face Detection and it also demonstrated strong performance in face detection, achieving a precision of 96.4% and a recall of 98.2%. These results indicate that the model can accurately detect faces, with a relatively low false positive rate and a high true positive rate. Meanwhile, in term of Mask Detection, the model achieved a precision of 86.7% and a recall of 90.2%. This suggests that the model can effectively distinguish between masked and unmasked individuals, with a good balance between precision and recall. It can be seen from Table 1 for all the precision and recalls of the four models.

TABLE I. The Performance of Four Models

Models	Face Detection		Mask Detection	
	Precision	Recall	Precision	Recall
	(%)	(%)	(%)	(%)
MobileNet	84.0	96.0	81.3	88.2
ResNet50	99.4	98.6	98.83	98.5
Inceptionv3	80.0	91.4	92.1	86.3
VGG19	96.4	98.2	86.7	90.2

In general, in term of Face Detection and mask detection, the ResNet50 model outperformed the other models, achieving the highest precision and recall values. This indicates that the ResNet50 model is highly accurate in detecting faces, with a low rate of false positives and false negatives. It also demonstrated superior performance in mask detection, with high precision and recall values. This suggests that the ResNet50 model is effective in accurately identifying individuals wearing masks. This indicates its potential as a robust and accurate model for detecting mask/non-mask faces. However, further analysis and comparisons with existing models are necessary to evaluate its performance in relation to other state-of-the-art face mask detection models.

C. Output of Faces Detetction Result

Here is the output of detected faces while wearing mask or not wearing mask as it can be seen from Figure 2 and Figure 3. In order to determine the best model for detecting mask/non-mask faces using transfer learning, we compare the

performance of MobileNet, ResNet50, Inceptionv3, and VGG19 models based on their given precision and recall from the provided results, it can be observed that the ResNet50-based proposed model achieves the highest accuracy in both face detection and mask detection tasks as already mentioned. With a precision of 99.4% and a recall of 98.6% for face detection, and a precision of 98.83% and a recall of 98.5% for mask detection, the ResNet50 model demonstrates superior performance in accurately detecting faces and distinguishing between masked and unmasked individuals. These results suggest that the ResNet50 model is the best fit as a backbone for detecting mask/non-mask faces using transfer learning as it can be seen from Figure 2.

On the other hand, MobileNet, Inceptionv3, and VGG19 models were classified wrong in the figure 4 and they marked not wearing a mask that someone who is wearing a mask while ResNet50 has marked the same Figure3 correctly. This also shows that in term of backbone detection for mask/non-mask faces, ResNet50 is outperformed others.

Meanwhile, to assess the utility of identity prediction in the proposed model, further details regarding identity prediction are done. However, it can be evaluated by examining the true positives, false negatives, false positives, and true negatives associated with identity prediction in the provided confusion matrices. Additionally, considering performance metrics such as precision, recall, and other relevant indicators specific to identity prediction can provide insights into its utility in the proposed model and upcoming table 1 illustrates that point.

The confusion matrices provide a detailed overview of the performance of each model in terms of face detection and mask detection as it can be seen from Table 2. These matrices reveal the true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) for each model. Starting with the RetinaFaceMask model based on MobileNet, the face detection results demonstrate a precision of 84.0% and recall of 96.0%. This indicates that the model accurately detects 84.0% of the faces present in the dataset, while 16.0% of the faces are missed. The mask detection performance shows a precision of 81.3% and recall of 88.2%, implying that the model correctly identifies 81.3% of the masked faces, but misclassifies 18.7% of the faces as masked.

Fig. 2. Captured with mask.

Fig. 3. Captured without mask.

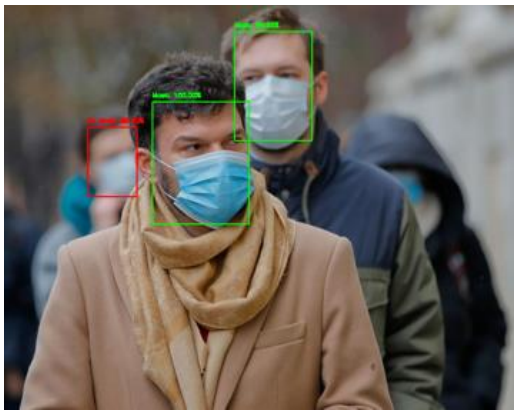

Fig. 4. Captured with / without mask.

TABLE II. CONFUSION MATRICS OF FOUR MODELS

Models		Face Detection		Mask Detection	
		Predicted Positive	Predicted Negative	Predicted Positive	Predicted Negative
MobileNet	Actual Positive	84.0% (4420)	16.0% (234)	81.3% (4420)	18.7% (234)
	Actual Negative	2.2% (108)	97.8% (4713)	2.2% (108)	97.8% (4713)
ResNet50	Actual Positive	99.4% (4680)	0.6% (243)	98.83% (4680)	1.17% (243)
	Actual Negative	2.8% (132)	97.2% (4609)	2.8% (132)	97.2% (4609)
Inceptionv3	Actual Positive	80.0% (4798)	20.0% (203)	92.1% (4798)	7.9% (203)
	Actual Negative	3.4% (127)	96.6% (4530)	13.7% (127)	86.3% (4530)
VGG19	Actual Positive	96.4% (4374)	3.6% (214)	86.7% (4374)	13.3% (214)
	Actual Negative	3.8% (187)	96.2% (4760)	3.9% (187)	96.1% (4760)

Moving on to the proposed model based on ResNet50, the face detection outcomes exhibit an exceptional precision of 99.4% and recall of 98.6%. This indicates that the model effectively identifies almost all faces present in the dataset with a high precision. The mask detection performance is also impressive, with a precision of 98.83% and recall of 98.5%, indicating accurate classification of the presence or absence of masks. The RetinaFaceMask model based on Inceptionv3 demonstrates a face detection precision of 80.0% and recall of 91.4%. Although the recall is relatively high, the precision suggests that the model may incorrectly detect some non-facial objects as faces. For mask detection, the precision is 92.1%, indicating accurate identification of masked faces, but the recall is 86.3%, suggesting some misclassification of masked faces as non-masked.

Lastly, the proposed model based on VGG19 exhibits a face detection precision of 96.4% and recall of 98.2%. These results indicate accurate and comprehensive face detection, capturing a large majority of the faces with high precision. In terms of mask detection, the precision is 86.7%, suggesting a relatively high accuracy in identifying masked faces. The recall of 90.2% implies that some masked faces may be misclassified as non-masked. Based on the comparison of the models, the proposed model based on ResNet50 emerges as the most suitable backbone for detecting mask/non-mask faces using transfer learning. It demonstrates outstanding performance in both face detection and mask detection, achieving high precision and recall scores. This indicates its capability to accurately identify faces and classify them based on the presence of masks.

V. Conclusion

In conclusion, the performance evaluation of the four models, namely RetinaFaceMask based on MobileNet, Proposed model based on ResNet50, RetinaFaceMask based on Inceptionv3, and proposed model based on VGG19, provides valuable insights into their effectiveness in the context of face detection and mask detection tasks. Based on the results obtained, it can be concluded that the proposed model based on ResNet50 outperforms the other models in terms of face detection precision and recall. This indicates that ResNet50 serves as a robust backbone for accurately detecting faces in various scenarios. Furthermore, the proposed model demonstrates high precision and recall values for mask detection, indicating its capability to effectively identify individuals wearing or not wearing masks. This is crucial for enforcing mask-wearing protocols and ensuring public safety. Comparatively, the RetinaFaceMask based on MobileNet and the RetinaFaceMask based on Inceptionv3 show relatively lower performance in face detection and mask detection tasks. Although they provide satisfactory results, they exhibit slightly lower precision and recall compared to the proposed model based on ResNet50. In terms of computational speed, a detailed analysis was not provided in the given information, which limits our ability to draw definitive conclusions regarding the models' efficiency. Future studies should consider evaluating the computational performance of these models to gain a comprehensive understanding of their real-time applicability.

References

[1] S. N. Yahya, A. F. Ramli, M. N. Nordin, H. Basarudin, and M. A. Abu, “Comparison of Convolutional Neural Network Architectures for Face Mask Detection,” International Journal of Advanced Computer Science and Applications, vol. 12, no. 12, 2021, doi: 10.14569/IJACSA.2021.0121283.

[2] A. A. Puzi et al., “Machine Learning Facemask Detection Models for COVID-19,” in IEEE International Conference on Semiconductor Electronics, Proceedings, ICSE, 2022. doi: 10.1109/ICSE56004.2022.9862951.

[3] S. Kumar, D. Yadav, H. Gupta, M. Kumar, and O. P. Verma, “Towards smart surveillance as an aftereffect of COVID-19 outbreak for recognition of face masked individuals using YOLOv3 algorithm,” Multimed Tools Appl, vol. 82, no. 6, 2023, doi: 10.1007/s11042-021-11560-1.

[4] A. Sharma, A. Miran, and Z. R. Ahmed, “The 3D Facemask Recognition: Minimization for Spreading COVID-19 and Enhance Security,” in Lecture Notes in Networks and Systems, 2022. doi: 10.1007/978-981-16-5655-2_60.

[5] J. -, M. Husna, and A. R. Lubis, “OpenCV Using on a Single Board Computer for Incorrect Facemask-Wearing Detection and Capturing,” JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING, vol. 5, no. 2, 2022, doi: 10.31289/jite.v5i2.6118.

[6] H. Nguyen, A. Nguyen, A. Mai, and N. T. Dang, “AI-app development for Yolov5-based face mask wearing detection,” in Proceedings - 2022 9th NAFOSTED Conference on Information and Computer Science, NICS 2022, 2022. doi: 10.1109/NICS56915.2022.10013442.

[7] S. Lee, D. Ko, J. Park, S. Shin, D. Hong, and S. S. Woo, “Deepfake Detection for Fake Images with Facemasks,” in WDC 2022 - Proceedings of the 1st Workshop on Security Implications of Deepfakes and Cheapfakes, 2022. doi: 10.1145/3494109.3527189.

[8] P. Sertic, A. Alahmar, T. Akilan, M. Javorac, and Y. Gupta, “Intelligent Real-Time Face-Mask Detection System with Hardware Acceleration for COVID-19 Mitigation,” Healthcare (Switzerland), vol. 10, no. 5, 2022, doi: 10.3390/healthcare10050873.

[9] J. Waleed, T. Abbas, and T. M. Hasan, “Facemask Wearing Detection Based on Deep CNN to Control COVID-19 Transmission,” in Al-Muthanna 2nd International Conference on Engineering Science and Technology, MICEST 2022 - Proceedings, 2022. doi: 10.1109/MICEST54286.2022.9790197.

[10] L. Kesa, “Chatbot with Facemask Detection Technique,” Int J Res Appl Sci Eng Technol, vol. 9, no. VI, 2021, doi: 10.22214/ijraset.2021.35511.

[11] P. A. Malave, S. M. Wagde, I. S. Vacche, M. S. Gaikwad, and B. Arkas, “Automated Contactless Temperature and Facemask Detection Using Deep Learning,” Int J Sci Res Sci Technol, 2021, doi: 10.32628/ijsrst2183116.

[12] S. Mustafa and M. S. Haruna, “Applying Convolution Neural Network to Facemask Detection from Varied Images,” in 2021 1st International Conference on Multidisciplinary Engineering and Applied Science, ICMEAS 2021, 2021. doi: 10.1109/ICMEAS52683.2021.9692417.

[13] F. H. Almukhtar, “A robust facemask forgery detection system in video,” Periodicals of Engineering and Natural Sciences, vol. 10, no. 3, 2022, doi: 10.21533/pen.v10i3.3072.

[14] M. L. Mokeddem, M. Belahcene, and S. Bourennane, “COVID-19 risk reduce based YOLOv4-P6-FaceMask detector and DeepSORT tracker,” Multimed Tools Appl, 2022, doi: 10.1007/s11042-022-14251-7.

[15] T. Abiodun, E. Ogbuju, and F. Oladipo, “Access Control System for Covid19 Using Computer Vision and Deep Learning Techniques: A Aystematic Review,” Journal of Applied Artificial Intelligence, vol. 3, no. 1, 2022, doi: 10.48185/jaai.v3i1.458.

[16] J. Tomás, A. Rego, S. Viciano-Tudela, and J. Lloret, “Incorrect facemask-wearing detection using convolutional neural networks with transfer learning,” Healthcare (Switzerland), vol. 9, no. 8, 2021, doi: 10.3390/healthcare9081050.

[17] G. Furqan, N. Z. Naqvi, and A. Jaiswal, “Comparative Analysis of Deep Learning Techniques for Facemask Detection,” in Communications in Computer and Information Science, 2022. doi: 10.1007/978-3-031-05767-0_10.

[18] K. Suresh, M. B. Palangappa, and S. Bhuvan, "Face Mask Detection by using Optimistic Convolutional Neural Network," in Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021, 2021. doi: 10.1109/ICICT50816.2021.9358653.

[19] I. B. A. Ouahab, L. Elaachak, M. Bouhorma, and Y. A. Alluhaidan, "Real-time Facemask Detector using Deep Learning and Raspberry Pi," in Proceedings - 2021 International Conference on Digital Age and Technological Advances for Sustainable Development, ICDATA 2021, 2021. doi: 10.1109/ICDATA52997.2021.00014.

[20] P. Prasad, A. Chawla, and Mohana, "Facemask Detection to Prevent COVID-19 Using YOLOv4 Deep Learning Model," in Proceedings of the 2nd International Conference on Artificial Intelligence and Smart Energy, ICAIS 2022, 2022. doi: 10.1109/ICAIS53314.2022.9742863.

[21] I. Journal, "Comprehensive research on Facemask Detection using CNN for the recent COVID-19 outbreak," INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT, vol. 06, no. 05, 2022, doi: 10.55041/ijsrem15683.

[22] S. Saranyan, S. Seshadri, and R. Boothalingam, "Real-time facemask detection and analytics," in 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0, ACMI 2021, 2021. doi: 10.1109/ACMI53878.2021.9528130.

[23] G. P. Bhargav, K. S. Reddy, A. Viswanath, Ba. A. Teja, and A. P. Byju, "An Integrated Facemask Detection with Face Recognition and Alert System Using MobileNetV2," in Smart Innovation, Systems and Technologies, 2022. doi: 10.1007/978-981-16-9873-6_7.

[24] V. Vinitha and V. Velantina, "Covid-19 Facemask Detection With Deep Learning and Computer Vision," International Research Journal of Engineering and Technology (IRJET), vol. 07, no. 08, 2020.

[25] S. V. Militante and N. V. Dionisio, "Deep Learning Implementation of Facemask and Physical Distancing Detection with Alarm Systems," in Proceeding - 2020 3rd International Conference on Vocational Education and Electrical Engineering: Strengthening the framework of Society 5.0 through Innovations in Education, Electrical, Engineering and Informatics Engineering, ICVEE 2020, 2020. doi: 10.1109/ICVEE50212.2020.9243183.

[26] M. S. M. Suhaimin, M. H. A. Hijazi, C. S. Kheau, and C. K. On, "Real-time mask detection and face recognition using eigenfaces and local binary pattern histogram for attendance system," Bulletin of Electrical Engineering and Informatics, vol. 10, no. 2, 2021, doi: 10.11591/EEI.V10I2.2859.

[27] H. Nagoriya and M. Parekh, "Live Facemask Detection System," International Journal of Imaging and Robotics, vol. 21, no. 1, 2021.

[28] G. Chen, B. Bai, H. Zhou, M. Liu, and H. Yi, "Facemask Detection Based on Double Convolutional Neural Networks," Recent Patents on Engineering, vol. 16, no. 3, 2021, doi: 10.2174/1872212115666210827100258.

[29] V. Balasubramaniam, "Facemask Detection Algorithm on COVID Community Spread Control using EfficientNet Algorithm," Journal of Soft Computing Paradigm, vol. 3, no. 2, 2021, doi: 10.36548/jscp.2021.2.005.

[30] W. Boulila, A. Alzahem, A. Almoudi, M. Afifi, I. Alturki, and M. Driss, "A Deep Learning-based Approach for Real-time Facemask Detection," in Proceedings - 20th IEEE International Conference on Machine Learning and Applications, ICMLA 2021, 2021. doi: 10.1109/ICMLA52953.2021.00238.

[31] A. Nowrin, S. Afroz, M. S. Rahman, I. Mahmud, and Y. Z. Cho, "Comprehensive Review on Facemask Detection Techniques in the Context of Covid-19," IEEE Access, vol. 9. 2021. doi: 10.1109/ACCESS.2021.3100070.

[32] M. A. S. Ai et al., "Real-Time Facemask Detection for Preventing COVID-19 Spread Using Transfer Learning Based Deep Neural Network," Electronics (Switzerland), vol. 11, no. 14, 2022, doi: 10.3390/electronics11142250.

[33] B. A. Kumar and M. Bansal, "Face Mask Detection on Photo and Real-Time Video Images Using Caffe-MobileNetV2 Transfer Learning," Applied Sciences (Switzerland), vol. 13, no. 2, 2023, doi: 10.3390/app13020935.

[34] F. Özyurt, A. Mira, and A. Çoban, "Face Mask Detection Using Lightweight Deep Learning Architecture and Raspberry Pi Hardware: An Approach to Reduce Risk of Coronavirus Spread While Entrance to Indoor Spaces," Traitement du Signal, vol. 39, no. 2, 2022, doi: 10.18280/ts.390227.