COMPARISON OF PRETRAINED TRANSFER LEARNING MODELS ON FACE-MASK DETECTION

2023
MASTER THESIS
COMPUTER ENGINEERING

Salma Abdalbagi ELSIDDIG

Thesis Advisor
Assist. Prof. Dr. Yasin ORTAKCI

COMPARISON OF PRETRAINED TRANSFER LEARNING MODELS ON FACE-MASK DETECTION

Salma Abdalbagi ELSIDDIG

Thesis Advisor
Assist. Prof. Dr. Yasin ORTAKCI

T.C.
Karabuk University
Institute of Graduate Programs
Department of Computer Engineering
Prepared as
Master Thesis

KARABUK
January 2023

I certify that in my opinion the thesis submitted by Salma Abdalbai ELSIDDIG titled "COMPARISON OF PRETRAINED TRANSFER LEARNING MODELS ON FACE-MASK DETECTION" is fully adequate in scope and in quality as a thesis for the degree of Master of Computer Engineering.

Assist. Prof. Dr. Yasin ORTAKCI .....
Thesis Advisor, Department of Computer Engineering

This thesis is accepted by the examining committee with a unanimous vote in the Department of Computer Engineering as a Master of Science thesis. January 30, 2023.

Examining Committee Members (Institutions)	Signature
Member : Assist. Prof. Dr. Yasin ORTAKCI (KBU)	.....
Member : Assist. Prof. Dr. Sait DEMİR(KBU)	.....
Member : Assoc. Prof. Dr. Rafet DURGUT (BOEU)	.....

The degree of Master of Science by the thesis submitted is approved by the Administrative Board of the Institute of Graduate Programs, Karabuk University.

Prof. Dr. Müslüm KUZU .....
Director of the Institute of Graduate Programs

"I declare that all the information within this thesis has been gathered and presented in accordance with academic regulations and ethical principles and I have according to the requirements of these regulations and principles cited all those which do not originate in this work as well."

Salma Abdalbagi ELSIDDIG

ABSTRACT

M. Sc. Thesis

COMPARISON OF PRETRAINED TRANSFER LEARNING MODELS ON FACE-MASK DETECTION

Salma Abdalbagi ELSIDDIG

Karabük University
Institute of Graduate Programs
The Department of Computer Engineering

Thesis Advisor:
Assist. Prof. Dr. Yasin ORTAKCI
January 2023, 38 pages

The global pandemic known as COVID-19 puts huge pressure on researchers to use technological solutions to provide further protection mechanisms. Face masks are one of the most important protection mechanisms among other health protocols. This thesis aims to detect the mask wearing problem by utilizing four CNN models: VGG16, ResNet50V2, InceptionV3 and MobileNetV2 based on Transfer Learning, in addition the provides a comparison based on their performances. The proposed model enhances the classification of mask wearing into three classes; without the mask, the correct wearing of the mask, and not the correct wearing of the mask. The previously mentioned four transfer learning models of CNN architectures were used to train, test, and validate based on the image dataset. The results reveal that the proposed model has performed the classification task successfully. VGG-19 was highlighted as the best pre-trained model.

Key Words : Covid-19, Pre-Trained Models, Mask detection, Transfer learning.
Science Code : 92431

ÖZET

Yüksek Lisans Tezi

YÜZ MASKESI TESPITINDE ÖNCEDEN EĞITILMIŞ TRANSFER ÖĞRENME MODELLERININ KARŞILAŞTIRILMASI

Salma Abdalbagi ELSIDDIG

Karabük Üniversitesi
Lisansüstü Eğitim Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı:
Dr. Öğr. Üyesi Yasin ORTAKCI
Ocak 2023, 38 sayfa

COVID-19 olarak bilinen küresel salgın, bu salgına karşı daha fazla koruma mekanizmaları geliştirme ve teknolojik çözümler sunma konusunda araştırmacılara önemli görevler yüklemektedir. Yüz maskeleri, diğer sağlık protokolleri arasında en önemli koruma mekanizmalarından biridir. Bu tez, CNN tabanlı eğitilmiş modellerden 4 tanesini: VGG16, ResNet50V2, InceptionV3 ve MobileNetV2 kullanarak maske takma sorununu tespit etmeyi ve bunların performanslarına kıyaslamasını yapmayı amaçlamaktadır. Önerilen model kullanıcıları maske takma işlemine göre maskesiz, yanlış makes takmış, doğru maske takmış olarak üç sınıfa ayırır. Bahsedilen CNN tabanlı dört transfer öğrenme modelini eğitmek, test etmek ve doğrulamak için örnek bir veri seti kullanıldı. Sonuçlar, önerilen modelin sınıflandırma görevini başarıyla yerine getirdiğini ortaya koymaktadır. VGG-19, önceden eğitilmiş en iyi model olarak vurgulandı.

Anahtar Kelimeler: Covid-19, Ön Eğitilmiş Modeller, Maske tespiti, Transfer öğrenme.

Bilim Kodu **:** 92431

ACKNOWLEDGMENT

First and foremost, praise and thanks to Allah, Lord of the world before and after, for granting me success in my study and in my life.

Second, I would like to thank my adviser, Assist. Prof. Dr. Yasin ORTAKCI, for his invaluable assistance and support in the preparation of my thesis.

I would like to thank Dr. Omer ABDELMAJEED for his support.

Thank you to my father and mother, my husband and son, and my brothers and sisters.

CONTENTS

Page

Page

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

ABBREVIATIONS INDEX

ABBREVITIONS

CNN	: Convolutional Neural Networks
AI	: Artificial intelligence
WHO	: World Health Organization
GMP	: Global max pooling
AMP	: Average max pooling
ResNet	: Residual network
MTCNN	: Multi-task Cascaded Convolutional Networks
SSD	: Single Shot Detector
ML	: Machine learning
DNN	: Deep Neural Network)
DL	: Deep learning
YOLO	: You Only Look Once
RGB	: Red green blue
TN	: True Negative
T	: Is the total True (True Positive and True Negative together)
F	: Total False (False Positive and False Negative
TP	: True Positive

PART 1

INTRODUCTION

The Coronavirus pandemic severely impacted most of the world's population as it emerged. The COVID-19 virus caused thousands of people's deaths daily worldwide. According to the World Health Organization's (WHO) most recent report on coronaviruses (COVID-19), the illness has infected over 616 million individuals in 213 countries and claimed over 6 million lives [1].

Following the rules established by the medical community can prevent the spread or transmission of the coronavirus. The best method to stop or at least limit the spread is to take precautions against being sick in the first place, such as routinely washing your hands, using disinfectants like 70% alcohol solutions, and avoiding touching your face, especially your eyes, nose, and mouth. Limiting the virus's transmission is possible by using social distancing and strict hygiene practices, such as the mandatory use of facemasks, hand gloves, face shields, and sanitizer [2].

While the benefits of wearing a face mask have been widely publicized thanks to recommendations from the (WHO) and other scientific research, it has been observed that many people do not wear their masks properly. Consequently, nurses and other concerned people have begun public health education campaigns emphasizing the need for mask use. Specifically, these initiatives involve circulating preventative posters and illustrations to educate people on the proper and improper ways to use face protection masks.

Assigning humans to enforce and follow up on persons wearing a mask has an extremely excessive cost. It is feasible to develop a technology that will allow this activity to be done at a lower cost while maintaining a higher quality and accuracy.

Because it can process a substantial quantity of clinical data, artificial intelligence (AI) and Deep learning as a sub-field of AI play an increasingly important part in medicine

In the context of Covid-19, many studies have been conducted to introduce deep learning-based solutions to automate the process of face mask detection. Reliable results have been achieved so far, where companies and governments can now depend on technology to force the World Health Organization recommendations regarding face masking. However, there can be variable approaches to develop a face mask classifier, where each approach has its own advantages and disadvantages. This research is an attempt to emphasize the approach of transfer learning based on Pre-Trained Convolutional Neural Networks (CNN) models.

1.1. BACKGROUND OF STUDY

The Corona pandemic began in December 2019. This was a global epidemic that quickly spread throughout the world, causing a huge amount of illness and death.

Following this the Health Organization has announced the precautions that need to be followed to limit the possibility of the disease being passed on from one person to another.

The two main important soft protocols are mask wearing and social distance, their importance was based on the way the disease is spared which is by the air mostly.

Artificial Intelligence was used during the pandemic in various fields, including measuring temperature, social distancing, and wearing a mask. It had a significant impact on limiting the spread of the disease.

The advancement of neural networks and deep learning in recent years has made significant contributions to the efficiency with which objects may be categorized and located. Two-stage detectors and single-stage detectors are the basis of the current deep learning object identification paradigm. Regarding detection accuracy, the R-CNN family's main two-stage detectors do quite well [3].

1.2. PROBLEM STATEMENT

Covid-19 is a respiratory disease that is like SARS which is transmitted directly through breathing. Because of this, the WHO recommended the use of masks to limit their spread, especially in public places. Fortunately, the technology and specifically the deep learning has proven a great capability in handling the process of face mask detection. Although, developing a face mask classifier using deep learning technology is not a straightforward task, specially building a CNN model from scratch, as it is a tedious task and consumes too much time and resources. Thus, this research aims to set a roadmap for developing such a classifier by employing the transfer learning approach consuming the best-known CNN models that exist in the literature. Therefore, the best pre-trained model is proven for the researchers to start from in future projects.

1.3. RESEARCH OBJECTIVE

The main objective of this research is to investigate and highlight the best pre-trained CNN model among the well-known image classification models, to be used to facilitate the process of building a face mask classifier with the best performance and least cost. This research objective is achieved by accomplishing each of the following sub-objectives:

- Develop and train a face mask classifier on top of each one of the selected pre-trained models.
- Tuning the classifier to get the best performance of each pre-trained model.
- Highlight the best pre-trained model for building the face mask classifier.

1.4. SIGNIFICANCE OF THE STUDY

Much research has been conducted in the context of covid-19 to improve the process of face mask monitoring. However, deciding the proper way to train the detection and classification model was one of the biggest challenges. In this research, we studied the

best pre-trained image classification models and compared their accuracy after building the face mask classifier on top of each of them.

PART 2

LITERATURE REVIEW

Developing a face mask classifier using deep learning is a task that can be achieved in many ways and methods, different previous studies have followed different paths to come up with an acceptable classifier, some of them have modeled the classifier from scratch, and others have employed a pre-trained model. Building a model from scratch is a time-consuming and expensive task as also its performance is still questionable, the opposite of the pre-trained models, which are a reliable and inexpensive approach to building a classifier. This research is interested in the second approach by studying the previous works that employ pre-trained models, based on the literature there are four outstanding pre-trained models that achieved high performance while employed in classifiers, are MobileNet, ResNet, VGG, and Inception.

In 2021 Ms. R. Suganthalakshmi et al used a MobileNet based detector that makes a sound alarm when detecting someone passing without a mask. This detection is conducted on a live video streamed from the camera by feeding images frame by frame to the model [5].

In 2022 Benjaphan Sommana et al, developed a system that included two main modules: face detection and alignment and face mask classification. Firstly, it recognizes all faces in the input image and then for each recognized face it determines if there is a facemask or not. They employed the ResNet50 detection model and trained the face mask classification model by MobileNetV21 A. They exploit several existing face detectors achieving superior performance [6].

Preeti Nagrath et al. uses the two-steps classification approach, in the first step uses SSDMNV2 for face detection task and in the second step uses the MobileNet for mask classification, they achieving an accuracy of 92.64% and F1-Score of 87.7% [7].

In a scientific paper [8] by author Benjaphan, Taya presented MobileNet with global pooling for two classes (mask, no mask) using different metrics to evaluate the model performance and achieved 99.48% accuracy with global average pooling (GAP), 99.48% accuracy of global max pooling (GMP) for the first dataset (Dataset 1), and 100% accuracy with global average pooling (GAP), 99.39% accuracy of global max pooling (GMP) for the second dataset (Dataset 2).

Bosheng et al. presented MobileNet for mask classification to classifying three classes (no face mask-wearing, incorrect face mask-wearing, correct face mask-wearing) and MTCNN (Multi Task Convolutional Neural Network) for face detection with 98.7% accuracy [9].

Farady et al.presented RetinaNet for face detection and ResNet50 for mask classification, three classes are classified (good, bad, none) with a confidence score of 99.84% for the "good" class, and 78.69% for the "bad" class and 65.41% for the "none" class and 81.31% average score [3].

The authors Snyder , Husar in [10] use a combination of resnet-50 and MTCNN to classify three classes (Without Mask, With Mask) with 99.2% Recall and 87.7% F1-score.

Wenxuan et al. used VGG16, which is the base of the SSD (Single Shot Detector) SDD method used for face mask detection, detects two types of classes (wear a mask, did not wear a mask) in real-time environment with 90.9% mAP [11].

Sammy et al. used VGG-16 to classify two face mask classes (face mask wearing, no-face mask wearing) in real-time with 96% detection rate [1]

Li et al. used InceptionV3, as employing two datasets to measure the accuracy of the model, achieving accuracy of 97.11% with the two classes dataset 2, and the accuracy of 94.52% with the three classes dataset 1. Cropping images has improved the accuracy [12].

Manoj et al. used InceptionV3, the model detects two classes with a mask, and without it achieving an accuracy of 99.9% during training and 100% during testing [13].

PART 3

THEORETICAL BACKGROUND

Utilizing technology to achieve our daily life tasks has become a common behavior nowadays, and with technological advancement many new programming languages, frameworks, libraries, and applications have emerged. In this chapter, I provide a gentle introduction to each of the related technologies to this research, starting with artificial intelligence, through machine learning to deep learning. Moreover, deep learning concepts, techniques, and applications are introduced later at this chapter as well.

3.1. ARTIFICIAL INTELLIGENCE (AI)

AI can be defined as a computer system that can handle a large amount of data to perform tasks such as analyzing images and text and making decisions. In today's digital age, AI plays an important part in the field of medicine due to its capacity to deal with a large amount of clinical data [14].

3.2. MACHINE LEARNING (ML)

As a branch of AI, ML utilized an old set of information to obtain new knowledge and information. We can divide ML to two main approaches: unsupervised technique and supervised technique, the ladder is the commonly used technique in ML. While there are several ways in which a computer can learn from already-existing data, the supervised approach is the most effective. Either the raw data was used as the basis for this categorization, or a subject matter expert was consulted for it [15].

3.3. DEEP LEARNING

The word “Deep” is commonly used to refer to the hidden layers in the neural network, sometimes deep learning models are referred to as deep neural networks. In the network, each layer processes the data it receives in a different and unique way, and then informs the following layer. Traditional neural networks have 2-3 hidden layers while deep neural networks can have more hidden layers. Unlike machine learning which requires experts to manually extract the features from the dataset, deep learning does not require an expert because the layers extract the features during the training [16][17].

3.3.1. Deep learning vs Machine Learning

When compared to classic ML-based algorithms, DL-based algorithms excel in feature extraction. During the training phase, an algorithm must be used to accurately extract the image's characteristics, such as its edges, corners, and textures. Figure 3.1 shows the main differences between ML and DL [17].

When using a DNN (Deep Neural Network), these high-level characteristics are automatically extracted from a picture, but when using an ML-based approach, human supervision is required, and the algorithms are often constructed [3].

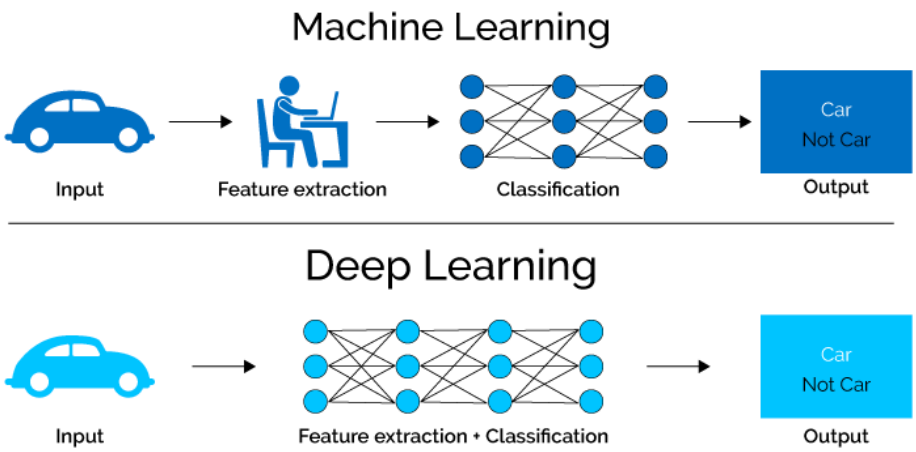


Figure 3.1. Deep Learning vs Machine Learning [17].

3.3.2. Deep Learning Applications in Computer Vision

Computer vision is a branch of AI concerned with understanding and manipulating images or video. Recently, with the development of technology, computer vision has become one of the most important and widespread fields [18].

In the following some examples of computer vision applications that deemed for their high result and reliable efficiency:

Image Classification also known as object classification has played an important role in many different fields in recent years, and there are many applications in our world based on image classification [19] such as:

- Classification of numbers handwritten
- X-ray Classification
- Face Recognition.

Object Detection is a task related to image classification as people, animals, and object detection algorithms use a wide range of image processing techniques to extract the object's features [20].

3.4. BASIC CONCEPTS IN CNN

3.4.1. CNN

CNN is the content of numerous components, such as convolutional layer, pooling layer, and the fully connected layer. It is used in various applications, such as differentiating and categorizing images. The primary method utilized for model training to category face masks is convolutional neural networks, which is part of the Deep Learning Algorithm [7].

3.4.2. CNN Architecture

CNN architectures are among the most widely used deep learning frameworks. CNN has many uses, from object recognition to NLP. The grid-based structure of the data they analyze lends themselves well to this form of deep learning algorithm [7].

CNN is a subset of deep learning algorithms useful for handling data with a spatial or temporal Ral component. Like other neural networks, CNNs use a sequence of convolutional layers to increase their complexity. An important part of CNN is "convolutional layers". A typical CNN structure is illustrated in Figure 3.2.

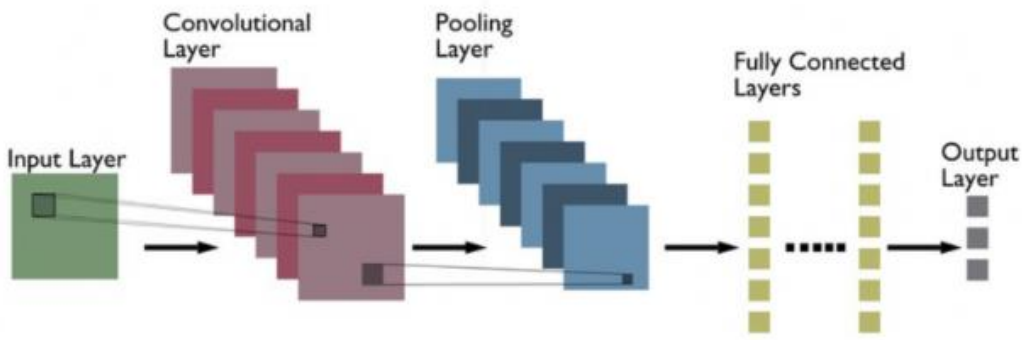


Figure 3.2. The Following are Explanations of The CNN Architectural [7].

3.4.3. Convolution Layer

Convolutional layers are a collection of filters that are used to process an image's data. A feature map, a depiction of the input picture modified by the filtering employed, is what the convolutional layer generates as its output. Stacking convolutional layers allows for the creation of more complicated models that can pick up on subtler aspects in pictures [7]. The convolution layer architecture is illustrated in Figure 3.3.

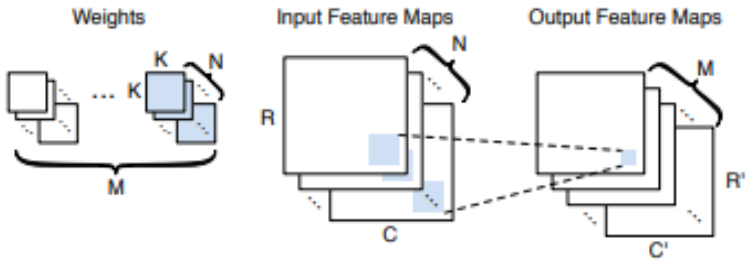


Figure 3.3. Convolution layer Architecture Illustration [15].

3.4.4. Pooling Layer

The pooling process uses a filter on the entire input, the pooling player's role is to reduce the input factors. Pooling has two categories: Max pooling where the filter sends the highest valued pixels from the input to the output array. And average pooling where a down sampled feature map is built by calculating the average value [21]. The pooling layer is illustrated in Figure in Figure 3.4.

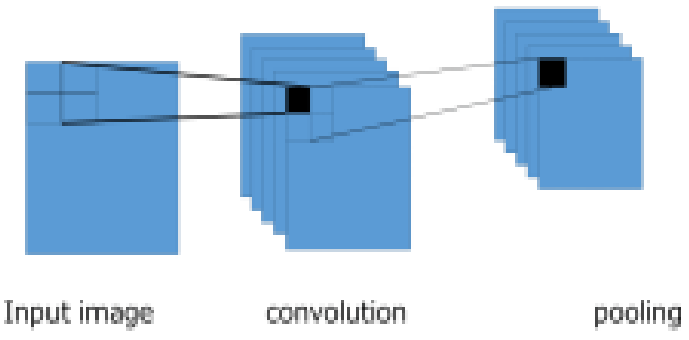


Figure 3.4. Pooling layer illustration [21].

3.4.5. Fully Connected Layer

In fully connected layers each neuron contributes to the activation of each neuron of the next layer where each neuron is connected to all neurons in the next layer. Unlike the partially connected layers where the input is not directly connected to the output layers. The fully connected layer takes the extracted features by the preceding layers and the applied filters as input and produces its output typically using the SoftMax activation function to produce a probability ranging from 0 to 1 [7].

3.5. TRANSFER LEARNING

Transfer Learning is a machine learning method that involves training and producing a model for one task and then reusing it on another. It refers to a situation in which what was learned in one context is used to enhance optimization in another [4].

3.6. PRE-TRAINED MODEL

The term pre-trained model refers to models that have previously been trained to solve a specific problem and then used to solve similar problems instead of training new models from scratch Figure 3.5. show how we use pre- trained model.

There are specific standards for choosing the pre-trained models which are the dataset that the model was trained on, and the targeted problem. For example, to solve the image classification problem the models were trained on the ImageNet dataset.

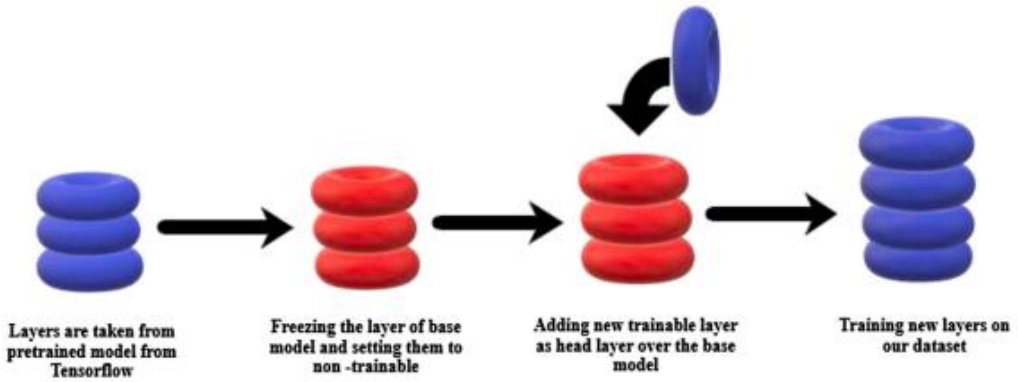


Figure 3.5. Pipeline of Using a Pretrained Model.

3.6.1. MobileNetV2

If you install MobileNetV2 Keras, you will have access to various fully convolutional systems that you can use immediately. MobileNetV2 has been optimized for picture classification. Creating an example of MobileNetV2 is as easy as importing it from Keras. MobileNetV2() returns the network's architecture but not the weights if you do not specify any keywords ,in other words, you will have to train the network on your own. When 'ImageNet' is used, it indicates that the network should be in demand using the ImageNet database [8]. The MobileNetV2 Architecture is shown in Figure 3.6.

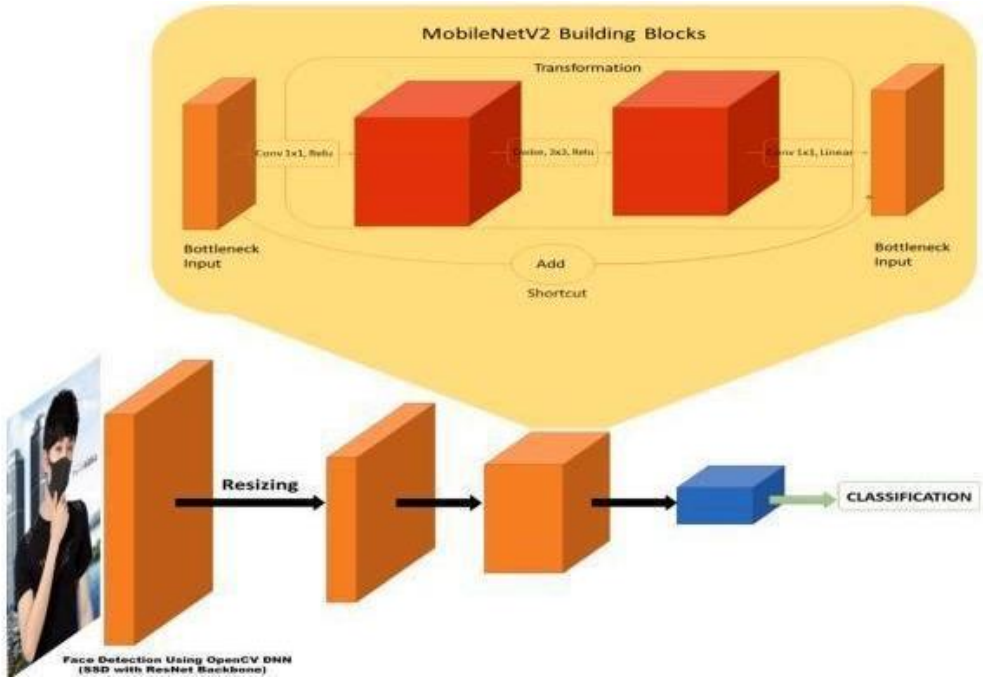


Figure 3.6. MobileNetV2 network's architecture.

3.6.2. ResNet 50

The ResNet CNN architecture, built by Kaiming He et al., achieved a top-five error of just 15.43% to win the ILSVRC 2015 categorization competition. With over a million parameters and 152 layers, this network is complex, even for CNNs; training it on the ILSVRC 2015 dataset would have taken more than 40 days using 32 GPUs. ResNet demonstrates that CNNs may be effectively utilized to address natural languages processing issues like phrase finalization or computer understanding, which the

Microsoft Research Asia team applied in 2016 and 2017 [22]. CNNs are often used for picture classification techniques with 1000 classes. Microsoft's machine comprehension system, which uses CNNs to create responses for more than 100k questions across over 20 categories, is one real-world application/example of ResNet CNN architecture. The ResNet CNN design is scalable to the processing capacity of GPUs and is efficient and robust [22]. The following: Figure 3.7 shows The ResNet architecture.

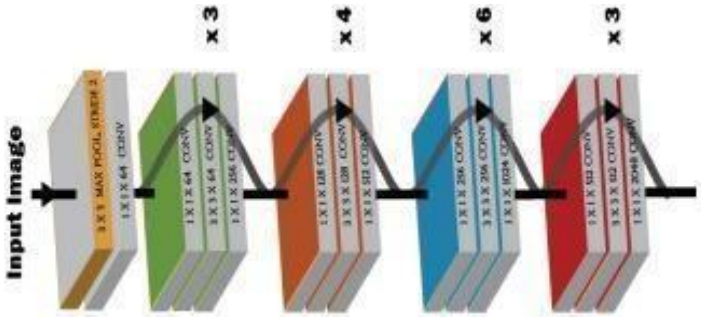


Figure 3.7. ResNet etwork's architecture.

3.6.3. VGG-19

Oxford researchers Karen Simonyan and Andrew Zisserman, together with others, created the CNN architecture known as VGG. Using more than a billion photos for training, the 16-layer CNN known as VGG can have as many as 95 million parameters (1000 classes). Large 224x224 input pictures are no problem for its 4096 convolutional features. Big-filter CNNs are difficult to train and require a large amount of data, resulting in Google Net (Alex Net architecture) and similar CNNs performing better than VGGNet for most image classification tasks with input pictures ranging in size from 100x100 to 350x350. The ILSVRC 2014 classification challenge is one real-world application, an example of the VGGtCNN architecture, which was also won by the GoogleNet CNN design. The VGG CNN model offers a good foundation for many applications in computer vision and is computationally efficient. It may be used for a wide variety of tasks, including object detection. Various neural network designs, including YOLO (You Only Look Once), SSD, etc., take advantage of its deep image features [3].

PART 4

METHODOLOGY

According to the literature review, there are variant methods and models used to achieve the facemask classification. As also many metrics were used to evaluate each of those models. Thus, this chapter justifies the method followed in this thesis to achieve the desired result by highlighting the best pre-trained model that can be used to build a high performing facemask classifier with the lowest cost in time and resources. This chapter starts by stating the research design, describing the selected dataset, showcasing the selected models, determining the evaluation metrics, specifying the experimental setup, and revealing the limitations.

4.1. THESIS DESIGN

This thesis uses the transfer learning technique to develop four facemask classifiers based on different pre-trained models ResNet50-2, MobileNet-2, VGG-19, and Inception-V3. To find the best classifier this research follows the following steps:

- Select the suitable dataset for the facemask classification task. This dataset should include a diverse range of faces, facial expressions, and lighting conditions.
- Pre-process the dataset by resizing, normalizing, and augmenting the images as needed.
- Select the recent and best performing pre-trained deep learning models (based on the literature review) that have been trained on a large dataset of images (such as VGG, ResNet, or Inception) and use them as feature extractors.
- Develop a unified classification Neural Network to carry on the classification task, which will be attached to each pre-trained model to form the facemask classifier.

- Evaluate and compare the performance of the classifiers on a held-out test set.
- Highlight and select the best pre-trained model to be used in the facemask classification tasks based on the selected performance measurement metrics.

4.1.1. Dataset Preparation

In this step, I selected a suitable dataset for facemask classification problems to use with the different pre-trained models under study in this research. Hence, the same dataset is used with all selected pre-trained models in the training phase.

4.1.2. Model Training

Transfer learning technique is used to build a classifier upon each of the selected pre-trained models, by replacing the last classification layer with a unified classification architecture, so that the performance of each classifier is affected by the architecture of the pre-trained model only. The unified classification architecture is designed and fine-tuned to achieve the best performance possible out of all classifiers obtained by each pre-trained model. Figure 4.8 illustrates these steps.

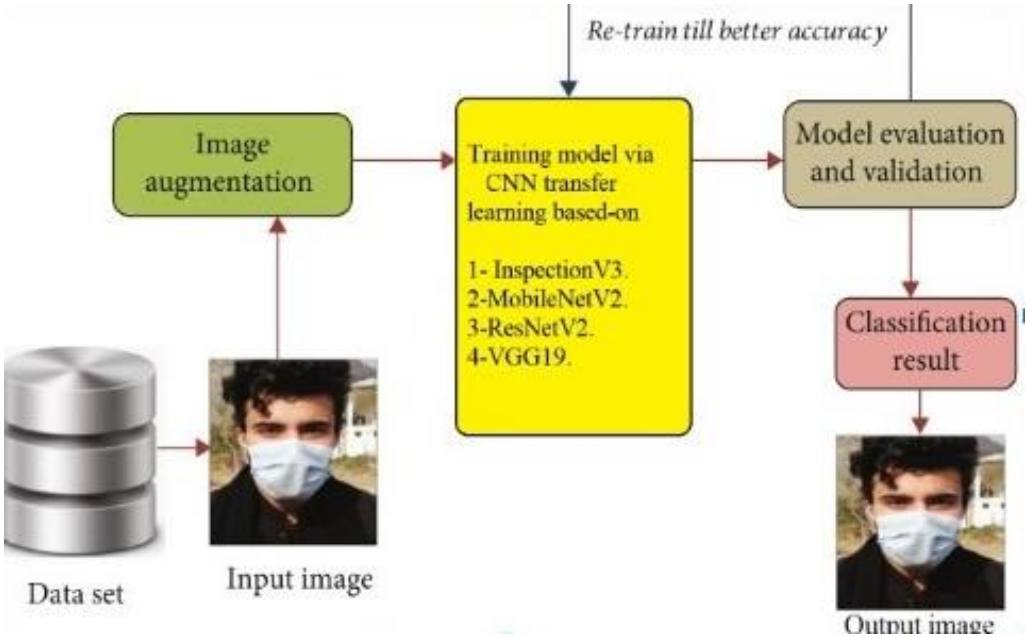


Figure 4.1. Illustrate the Diagram of This Study.

4.1.3. Performance Evaluation

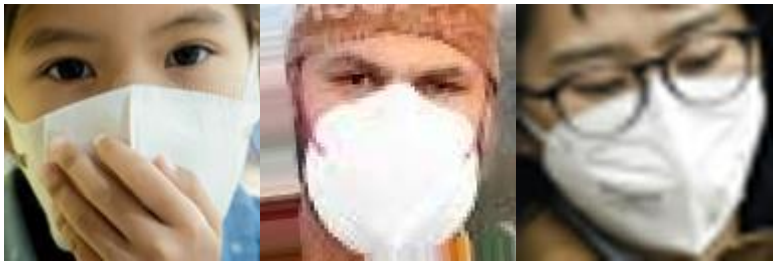
The accuracy, recall and precision metrics are used in combination with the risk of classification errors in the context of COVID-19 to compare the performance of pre-trained model-based classifiers.

4.1.4. Best Pre-Trained Model Selection

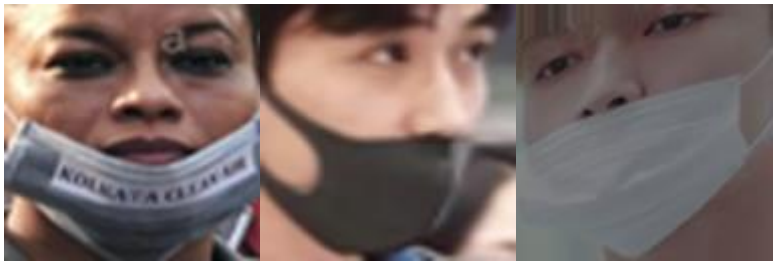
Based on the classifiers testing result, the accuracy, recall, and precision are calculated for each class correct, incorrect, and without-mask. The pre-trained model of the best classifier with the highest score in all the selected metrics is considered the best performing model in the facemask classification.

4.2. DATASET

We collected the dataset from Kaggle [15]. In the final blend, the images were classified into three subfolders: correct, incorrect, and without a mask, total of the images is 8982, and each folder contains 2994 classified images. The format of all images is PNG and they are in RGB colors with assorted sizes (higher or lower than (256Px*256Px)). Figure 4.2 shows samples of the selected dataset.



a. samples of the “correct” class images



b. samples of the “incorrect” class images

c. samples of the “without-mask” class images

Figure 4.2. samples of the classes of the dataset

4.2.1. Dataset Pre-Processing

The pre-processing is used to increase the accuracy of the classifier. The data was preprocessed before feeding it to the model. We reshaped the dimensions of the images to be (128Px*128Px*3) which reduced the images' sizes smoothly without affecting the features and lowered the computations and complexity. On the other hand, while 15% of the dataset remained original, 20% was rotated, 15% was zoomed, 20% was shifted to the left, and 20% was shifted to the right.

After Dataset Pre-Processing, for the data used in training, a number of parameters must be specified for how to pass the data to the neural network.

- Epochs: It is the number of training iterations, where in each iteration the network is trained on the entire training set, and this number varies according to the used algorithm.
- Batch size: It is a hyperparameter that defines the required number of samples of the training dataset that the network should propagate to update its parameters.

The setting of the previously mentioned parameters is shown the following Table 4.1.

Table 4.1. Hyperparameter settings of the Training

HYPERPARAMETER	VALUE
Number of Epochs	30
Batch Size	32
Optimizer	Adam

4.2.2. Dataset Augmentation

The best way to make the machine learning model learn better is by training it with more data. This method lets us produce more training samples, the following are the types of augmentations those applied on the dataset:

- Zooming = 20%
- Rotated = 15%
- Shifted to the left by 20%
- Shifted to the right by 20%

4.3. SELECTED PRE-TRAINED MODELS

According to the literature review, the most relevant models were the CNN based models, and according to the performance, availability and trustworthiness the selected pre-trained models in this study are as follows:

4.3.1. Inception V3

This model consists of groups that hold countless convolution layers, MaxPooling layers, dropout layers, fully connected layers, and the output that passes through the SoftMax without changing the structure as illustrated in Figure 4.1. The weight has been trained in the TL InspectionV3 using the ImageNet dataset. A cover of transferred learned InspectionV3 consists of the model structure, average pooling, flattened layer, and two dense layers.

To tune the weights of the total model the cover ran once with the first epoch then the head of the constructed model was placed on top [14].

To reduce the high dimensions of extracted features the convolution layers were used for extracting features from the image used by the MaxPooling layers during the extraction operation. And to prevent overfitting we used the dropout technique, where 5% of the network weights were dropped for each iteration. For classification issues, a Dense layer with the SoftMax regression equation and three output classes was used.

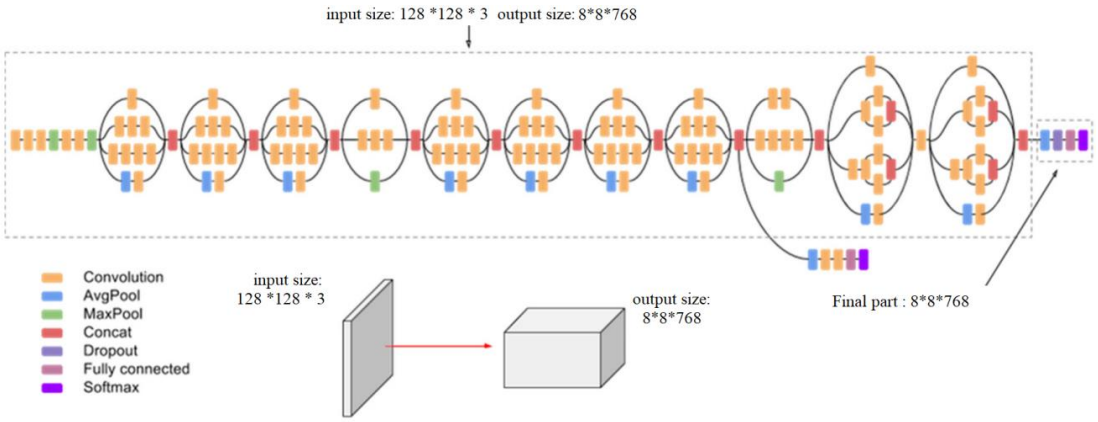


Figure 4.3. Explain The InceptionV3 Design [14].

4.3.2. MobileNet V2

MobileNetV2 pre-trained model was used by the ImageNet dataset as a base network then replaced with a convolution layer and a SoftMax classifier. to reduce overfitting by applying the Adam Optimizer with a very low Learning Rate (LR), 0.0001. We

extracted the features and tuned the value of weight using the pre-trained MobileNetV2. And the features were classified with the SoftMax classifier. The new model was trained using Adam Optimizer of different LR in two stages.

4.3.3. ResNet50 V2

The ResNetV2 model trained on the ImageNet dataset was used as a base network. In the first iteration the top layer of the base network was run once and replaced with the constructed model layers and a SoftMax classifier. A dropout of 0.5 averages was applied to lessen overfitting. While an Adam Optimizers with a small LR, 0.0001. The entire framework is illustrated in Table 4.2.

Table 4.2. The constructed model of ResNetV2.

Iterations No.	**Used layers**
For the first iteration	ResNetV2 network pre-trained by ImageNet
For all iterations	Conv2D
	AveragePooling2D
	Flatten
	Denes (128)
	Dropout (0.5)
	Denes (3)

4.3.4. VGG-19

As our foundation network, a VGG19 trained on the ImageNet dataset was used. We added a common convolution layer and a SoftMax classifier to the base network's top layer to replace the original structure; the model structure is listed in Table 4.3. We added dropout simultaneously to the recently introduced conv2d for reducing overfitting. For feature extraction, we used the pre-trained VGG-19 and the SoftMax classifier by setting a training rate as 0.0001.

Table 4.3. The Constructed Model of ResNet50V2.

Iterations No.	Used layers
For the first iteration	VGG-19 network pre-trained by ImageNet
For all iterations	Conv2D
	AveragePooling2D
	Flatten
	Denes (128)
	Dropout (0.5)
	Denes (3)

4.4. PROPOSED UNIFIED CLASSIFICATION MODEL

This research uses the classification architecture described in Table 4.4. as a unified classification model and uses the pre-trained model as feature extractor to produce the facemask classifier.

Table 4.4. The unified classification model.

Layer	Additional properties
AveragePooling2D	pool_size=(5, 5)
Flatten	
Denes (128)	activation="relu"
Dropout (0.5)	
Denes (3)	

4.5. EVALUATION METRICS

There are several metrics extracted through the confusion matrix to evaluate the machine learning models' performance. In this study, we evaluate the presented models by the most common evaluation metrics: Accuracy, Recall, Precision, and F1-score.

4.5.1. Accuracy

The most parameter that is used to classify model performance is accuracy, which computes the proportion of properly identified samples and is calculated as:

$$ACCURACY = \frac{TN+TP}{T+F} \text{ …….....} \quad (4.1)$$

TP = True Positive

TN = True Negative

T = is the total True (True Positive and True Negative together)

F = total False (False Positive and False Negative

4.5.2. Recall

Recall is the rate of true positive to the positive classes' samples that are appropriately labelled. Equation 2 shows the Recall mathematical representation:

$$recall = \frac{TP}{TP + FN} \quad \text{.....} \quad (4.2)$$

FN = False Negative

4.5.3. Precision

Precision could be the opposite of Recall, where it deals with negative values, i.e., refer to the rate of the true negative to the negative classes samples that are appropriately labelled or how the model can detect the negative instances.

$$precision = \frac{TN}{TN + FP} \text{.....} \quad (4.3)$$

FP = False Positive

4.6. EXPERIMENTAL SETUP

This section describes the implementation requirements in software and hardware as used to gain the reported results.

4.6.1. Software

Python is the programming language used in this research to implement the classifiers. Furthermore, python-based libraries and frameworks such as TensorFlow, Sklearn, and numpy are utilized to ease the process of implementation, the following are the most used functions and classes during the implementation phase

4.6.1.1. TensorFlow

TensorFlow is a machine learning platform that offers a number of libraries, classes, and functions that can be used to speed up machine learning development. The classes and functions utilized in this study are mentioned below:

ImageDataGenerator: python function used to implement data augmentation such as rotation, shifts, flips, and zoom.

preprocess_input: this function is used after loading images to adequate its format.
img_to_array: used to convert image data type to array.
load_img: this function used to load the dataset from URL (Uniform Resource Locators)
to_categorical: used to convert the classes to the binary class matrix.
MobileNetV2: this is the class used to load the MobileNetV2 model.
ResNet50V2: this is the class used to load the ResNet50V2 model.
VGG19: this is the class used to load the VGG19 model.
InceptionV3: this is the class used to load the InceptionV3 model.
AveragePooling2D: is the class that used to add a pooling layer to the model architecture.
Dropout: is the class used to add a dropout layer to the model architecture, to prevent overfitting.
Flatten: is the class used to add a flattening layer to the model architecture.
Dense: is the class used to add a dense layer to the model architecture.
Input is used to build the model by the input and output of the model
Model: is the class that is used to combine the defined layers into the network.
Adam: is the optimization algorithm that uses an adaptive learning rate during the training.

4.6.1.2. Scikit-Learn

The following functions from the open-source machine learning library scikit-learn were used in this work:

- LabelBinarizer: this function is used to convert (correct, incorrect, without mask) to binary labels
- train_test_split: used to split the dataset into 20/80
- classification report: we used this functions to show the confusion matrix result
- confusion matrix: is used to compute a confusion matrix to evaluate the accuracy of classification.

4.6.1.3. Imutils

This library content series of functions used to implement image preprocessing such as translation, rotation, resizing.

4.6.1.4. Numpy

Used to convert image data to array

4.6.2. Hardware

Successful execution of the implantation of the classifiers is a demanding process and requires a high-performing device to finish the execution errors-free, as also finishes in a reasonable time, the following are the specification of the computer in which all the experiments were carried out:

- CPU: Intel Core i5 10^{th} generation.
- GPU: NVIDIA GEFORCE RTX 3060 with 6 GB memory.
- RAM: DDR4 16 GB.

4.6.3. Training Strategy

The pre-trained model layers are frozen, so they are not trained anymore. Hence, the pre-trained model participates as a feature extractor. Therefore, the categorical-cross-entropy loss function was used during the training of the classifier part of the model in

all 30 epochs, where each epoch uses 80% of the dataset for training and the rest (20%) for validation. Furthermore, an augmentation process such as flipping, rotating, and zooming is carried out on the training dataset at the beginning of each epoch to prevent the overfitting problem.

PART 5

RESULTS AND DISCUSSION

To make the comparison and highlight the best pre-trained model for the task of the facemask classification, this chapter at the performance analysis section, it reports analyzes the performance scores for each pre-trained based classifier after the training and validation. Furthermore, it compares (using the selected metrics) the pre-trained models' performance and highlights the best model in the discussion section.

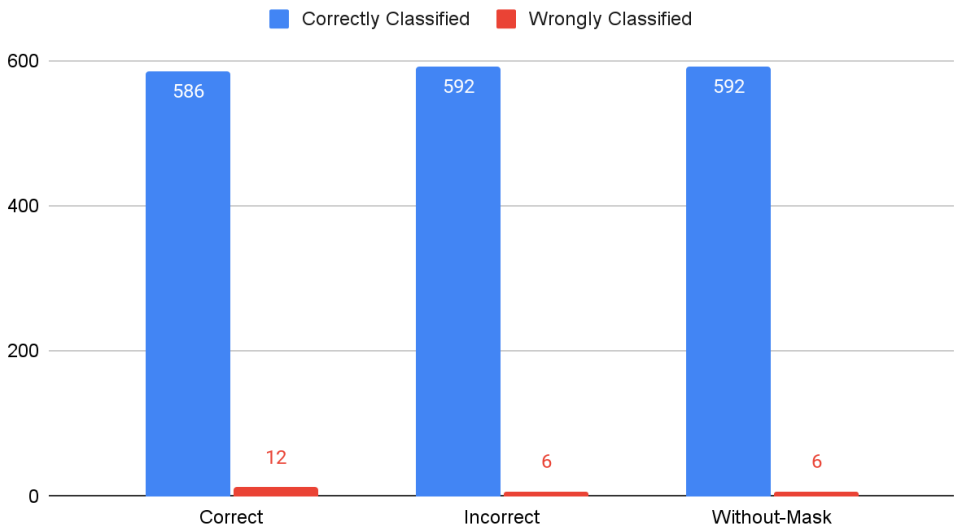
5.1. PERFORMANCE ANALYSIS

This section analyzes the recall and precision for each classifier to report the score of the positive classification, and negative classification for each class. The positive classification is the classification of the image as an instance of a specific class (correct, incorrect, or without-mask), while the negative classification is the classification of the image as NOT an instance of the specific class under analysis.

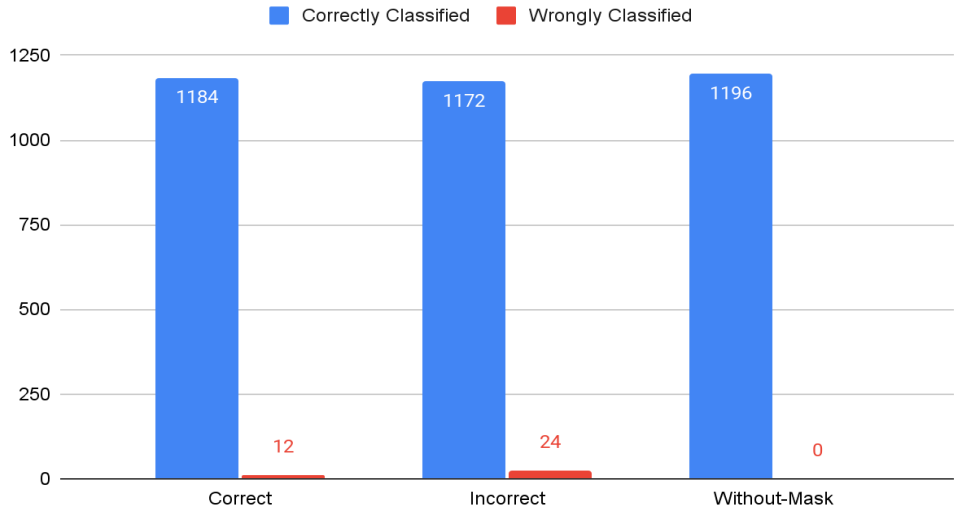
5.1.1. MobileNetV2

The classifier has higher errors with identifying the images of the “Correct” and the “Incorrect” classes according to its positive and negative classification as follows:

- In the positive classification, the classifier was wrong at 12 images from the “Correct” class, 6 images from the “Incorrect” class, and 6 images from the “Without-mask” class as shown in Figure 5.1.a.
- In the negative classification , the classifier was wrong at 12 images from the “Correct” class, and 24 images from the “Incorrect” class, and was correct about all images from the “Without-mask” class as shown in Figure 5.1.b.



(a)



(b)

Figure 5.1. Recall and Precision Analysis for MobileNetV2 Based Classifier.

5.1.2. ResNet50V2

The classifier has higher errors with identifying the images of the “Correct” and the “Incorrect” classes according to its positive and negative classification as follows:

- In the positive classification, the classifier was wrong at 12 images from the “Correct” class, 6 images from the “Incorrect” class, and 6 images from the “Without-mask” class as shown in Figure 5.2.a.

- In the negative classification, the classifier was wrong at 12 images from the “Correct” class, and 24 images from the “Incorrect” class, and 12 images from the “Without-mask” class as shown in Figure 5.2.b.

(a)

(b)

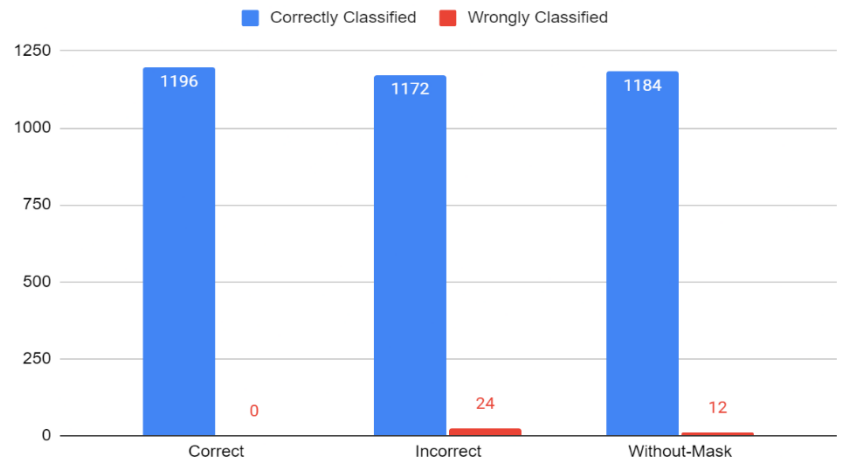
Figure 5.2. Recall and Precision Analysis for ResNet50V2 Based Classifier.

5.1.3. VGG-19

The classifier has higher errors with identifying the images of the “Icorrect” and the “Without-mask” classes according to its positive and negative classification as follows:

- In the positive classification, the classifier was wrong at 12 images from the “Correct” class, 6 images from the “Without-mask”, and was correct about all images from the “Incorrect” class as shown in Figure 5.3.a.
- In the negative classification, the classifier was wrong at 24 images from the “Incorrect” class, 12 images from the “Without-mask”, and was correct about all images from the “Correct” class as shown in Figure 5.3.b.

a)



b)

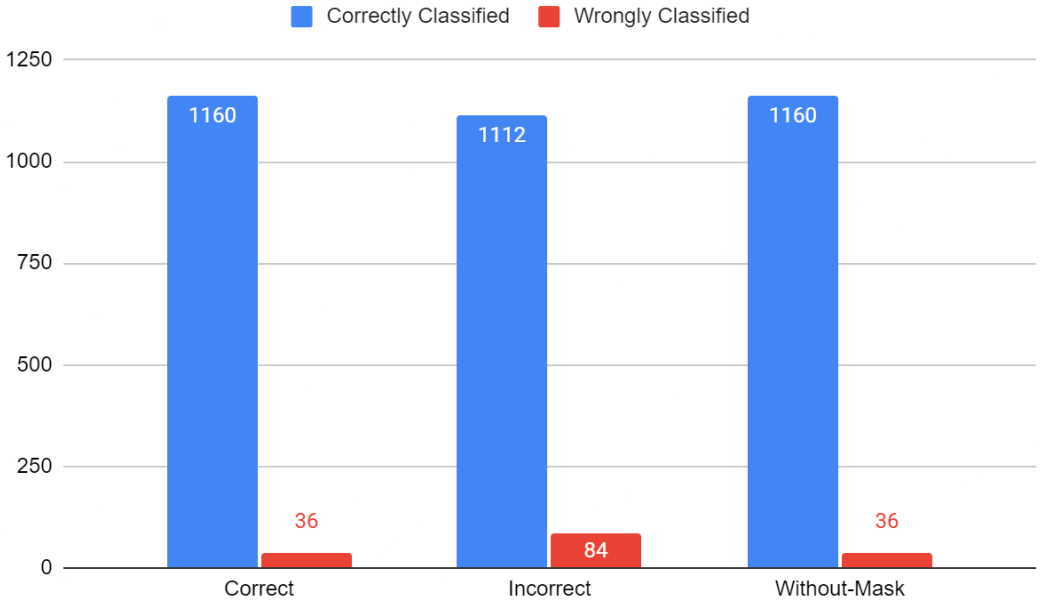
Figure 5.3. Recall and precision analysis for VGG-19 based classifier.

5.1.4. InceptionV3

The classifier has high errors with identifying the images of all the three classes according to its positive and negative classification as follows:

- In the positive classification, the classifier was wrong at 71 images from the “Correct” class, 18 images from the “Incorrect”, and 6 images from the “Without-mask” class as shown in Figure 5.4.a.
- In the negative classification, the classifier was wrong at 36 images from the “Correct” class, 84 images from the “Incorrect”, and 36 images from the “Without-mask” class as shown in Figure 5.4.b.

(a)



b

Figure 5.4. Recall and precision analysis for InceptionV3 based classifier.

5.2. DISCUSSION

The basic metric to compare deep learning models is the accuracy, but in the context of this research the accuracy alone is not enough as we can see in Table 5.1. The MobileNet, ResNet50, and VGG-19 based classifiers have scored the same accuracy of 99%, which is not enough to select and highlight the best model. However, the InceptionV3 is excluded due to its low accuracy. Although, to overcome the equality of the accuracy we use the analysis from the previous section combined with the context of COVID-19 precaution procedure.

Table 5.1. Experiment results of the four deep learning models.

Pre-trained model	Acc.	Correct-Mask		Incorrect-Mask		without-Mask	
		Precision	Recall	Precision	Recall	Precision	Recall
InceptionV3	0.97	0.97	0.91	0.93	0.97	0.97	0.99
MobileNetV2	0.99	0.99	0.98	0.98	0.99	1.00	0.99
ResNetV2	0.99	0.99	0.98	0.98	0.99	0.99	0.99
VGG-16	0.99	1.00	0.98	0.98	1.00	0.99	0.99

As a contribution to the COVID-19 precaution procedure this thesis is interested in preventing people without facemask or with incorrectly placed facemask from entering the protected areas. Therefore, in this context this thesis focuses on the negative classification of the class “correct” (ignoring the risk of the error in the positive classification of this class). Moreover, the positive classification of the classes “incorrect” and “without-mask” can be considered for more protection (ignoring the risk of the error in the negative classification of these two classes)

According to the COVID-19 context and the performance analysis in the previous section, we can find that VGG-19 based classifier over performs the MobileNet and ResNet50 by scoring no errors in the negative classification of the “correct” class as well as no errors in the positive classification the “incorrect” class, as they all scored 6 wrong-classified classifications on the positive classification of the “without-mask” class as illustrated in Figure 5.5.

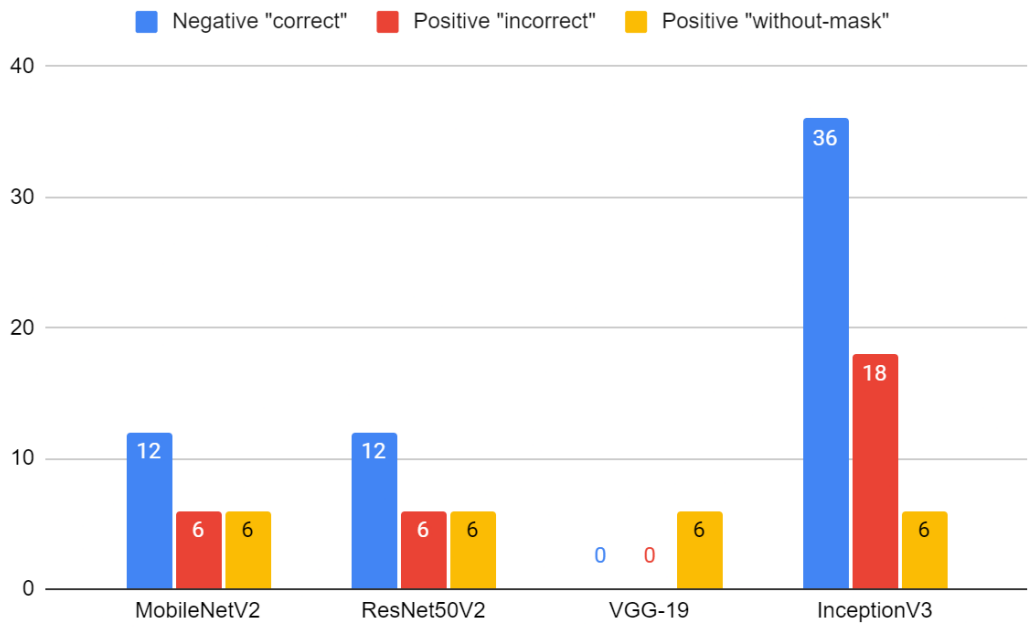


Figure 5.5. The error of the classifications of interest.

5.3. CONCLUSION

This research has investigated the capability of deep learning in the context of the COVID-19 as many of the previous studies have developed different deep learning-based solutions to help in the task of facemask detection and classification. Thus, and in sake of minimizing the cost and optimizing the process of building a high-performing facemask classifier, this thesis has studied the CNN models as pre-trained for the facemask classification task using the transfer learning technique to come up with best facemask classifier combined with final classification layers trained and tuned on top of the base model. Using accuracy, recall, and precision metrics, MobileNetV2, ResNet50V2, VGG-19, and InceptionV3 based classifiers performance results were compared, and based on the accuracy the Inception was excluded, and VGG was highlighted as the best to build a facemask classifier.

Since this research only measures the validation metrics, we recommend that future researchers use those classifiers in real life applications and analyze the performance in the runtime environment.

REFERENCES

1. 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).

2. K. Hammoudi, A. Cabani, H. Benhabiles, and M. Melkemi, "Validating the correct wearing of protection mask by taking a selfie: design of a mobile application 'CheckYourMask' to limit the spread of COVID-19." [Online]. **Available: https://hal.archives-ouvertes.fr/hal-02614790** (2020).

3. I. Farady, C. Y. Lin, A. Rojanasarit, K. Prompol, and F. Akhyar, "Mask Classification and Head Temperature Detection Combined with Deep Learning Networks," ***in 2020 2nd International Conference on Broadband Communications***, Wireless Sensors and Powering, BCWSP (2020).

4. M. Hussain, J. J. Bird, and D. R. Faria, "A study on CNN transfer learning for image classification," ***in Advances in Intelligent Systems and Computing***," ***(***2019).

5. A. Sinha, M. R., A. Hafeeza, P. Abinaya, and A. G. Devi, "Covid-19 Facemask Detection with Deep Learning and Computer Vision, Available: **www.ijert.org ,** (2021).

6. B. Sommana et al., "Development of a face mask detection pipeline for mask-wearing monitoring in the era of the COVID-19 pandemic: A modular approach***," ,19th International Joint Conference on Computer Science and Software Engineering (JCSSE), (***2022***).***

7. P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria, and J. Hemanth, "SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2," (2021).

8. I. B. Venkateswarlu, J., "Face mask detection using MobileNet and global pooling block," in ***4th IEEE Conference on Information and Communication Technology***, CICT 2020, (2020).

9. B. Qin and D. Li, "Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19," ***Sensors (Switzerland)***, vol. 20, no. 18, pp. 1–23, Sep. 2020, doi: 10.3390/s20185236.

10. S. E. Snyder and G. Husari, “Thor: A deep learning approach for face mask detection to prevent the COVID-19 pandemic,” ***in Conference Proceedings - IEEE SOUTHEASTCON***, (2021).

11. W. Han, Z. Huang, A. Kuerban, M. Yan, and H. Fu, “A Mask Detection Method for Shoppers under the Threat of COVID-19 Coronavirus,” ***in Proceedings - 2020 International Conference on Computer Vision, Image and Deep Learning***, (2020).

12. Y. Li, “Facemask detection using inception V3 model and effect on accuracy of data preprocessing methods,” ***in Journal of Physics: Conference Series***, (2021).

13. G. J. Chowdary, N. S. Punn, S. K. Sonbhadra, and S. Agarwal, “Face Mask Detection using Transfer Learning of InceptionV3,” (2020).

14. S. Yeasmin, “Benefits of Artificial Intelligence in Medicine,” ***2nd International Conference on Computer Applications and Information Security***, ICCAIS 2019, pp. 1–6,(2019).

15. E. Horvitz and D. Mulligan, “Data, privacy, and the greater good,” ***Science (1979)***, vol. 349, no. 6245, pp. 253–255, (2015).

16. Y. Xu, X. Wang, Z. Xu and Y. Zhang, ,”Analysis and Application of Control Strategy for Switched Reluctance Drive with Position Sensor,” ***2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)***, Chongqing, China, (2019).

17. M. R. Bhuiyan, S. A. Khushbu, and M. S. Islam, “A Deep Learning Based Assistive System to Classify COVID-19 Face Mask for Human Safety with YOLOv3,” (2020).

18. G. T. S. Draughon, P. Sun, and J. P. Lynch, “Implementation of a Computer Vision Framework for Tracking and Visualizing Face Mask Usage in Urban Environments,” ***in 2020 IEEE International Smart Cities Conference***, ISC2 (2020).

19. Xiangrong Zhang,Kai Jiang, “SpatiallyConstrained Bag-of-Visual-Wordsfor Hyperspectral Image Classification”, ***2016 IEEE International Geoscience & Remote Sensing Symposium : proceedings***, July 10-15, Beijing, China (2016).

20. Apoorva Raghunandan, Mohana, “Object Detection Algorithms for Video Surveillance Applications”, ***Proceedings of the 2018 IEEE International Conference on Communication and Signal Processing (ICCSP) : 3rd - 5th April 2018***, Melmaruvathur, India (2018).

21. Internet: K. O’Shea and R. Nash, “An Introduction to Convolutional Neural Networks”, **http://arxiv.org/abs/1511.08458** (2015)

22. A. Srinivas Joshi, S. Srinivas Joshi, G. Kanahasabai, R. Kapil, and S. Gupta, “Deep Learning Framework to Detect Face Masks from Video Footage; Deep Learning Framework to Detect Face Masks from Video Footage,” ***2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)***, (2020).

23. W. Shen, INESC TEC (Organization), Universidade de Trás-os-Montes e Alto Douro, M. IEEE Systems, International Working Group on Computer Supported Cooperative Work in Design, and Institute of Electrical and Electronics Engineers, **Proceedings of the 2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD)** : May 6-8, Porto, Portu (2019)

RESUME

Salma Abdabagi ElSIDDIG finished her elementary education in Sudan. She completed high school education in Alhilalya Girls Private High School, after that, she started undergraduate program in Omdurman Islamic University, Department of computer science in 2012. Then in 2020, To complete M. Sc. education, she moved to Karabuk University.