A Deep Learning Approach For Real-Time Face Mask Detection

A Thesis
Submitted in Partial Fulfillment of the Requirements for the Degree of Bachelor in Computer Science and Engineering

by

Hafsa
Class Roll: 16CSE-002

Supervised by
Md. Mostafijur Rahman
Assistant Professor

Department of Computer Science and Engineering
University of Barishal
BARISHAL, BANGLADESH

Approval

The Thesis Report “”**A Deep Learning Approach For Real-Time Face Mask Detection**” submitted by **Hafsa**, Roll: 16CSE-002, Session: 2015-16, to the Department of Computer Science Engineering, University of Barishal, Bangladesh, has been accepted as satisfactory for the partial fulfilment of the requirements for the degree of Bachelor of Science (Hons) in Computer Science and Engineering and approved as to its style and contents.

Supervisor's Signature and Date

....
Md. Mostafijur Rahman
Assistant Professor
Department of Computer Science And Engineering,
University of Barishal

Declaration

I, hereby, declare that the work presented in this Thesis is the outcome of the investigation performed by me under the supervision of **Md. Mostafijur Rahman** Sir, Assistant Professor, Department of Computer Science and Engineering, University of Barishal, Bangladesh. I also declare that it has not been previously or concurrently submitted for any other degree at the University of Barishal or any other institution.

Student's Signature and Date

....
Student Name
Date:

Abstract

The COVID-19 pandemic is a big threat to humanity irrespective of gender, caste, creed, and religion. Though several vaccines have been discovered, still the whole world is struggling to reduce the spread of COVID-19. According to the World Health Organization(WHO), wearing face masks properly is one of the practices that help to control the spread of this terrible virus. However, ensuring all people wear face masks is not an easy task. In this context, a deep learning model is proposed to detect whether a person wears a face mask properly or not in a public gathering. In this paper, a simple and effective model has been proposed for real-time monitoring using the deep neural network, MobileNet, OpenCV, Keras and TensorFlow technologies. The model has been trained, validated, and tested upon 2 large datasets contained with mask and without mask photos. Corresponding to the datasets, it is observed that the proposed method produced an accuracy of around 98% for with mask, and 99% for without mask. The aim of this proposed model is to ensure everyone wears a face mask in public places like schools, colleges, universities, hospitals, airports, shopping malls and so on. By monitoring the placement of the face mask on the face, it can be made sure that an individual wears it the right way and helps to reduce the spreading of the virus.

Keywords: Deep Learning, Face mask detector, MobileNet Models, Open Computer Vision (OpenCV), COVID-19

Acknowledgement

First and foremost, praise and appreciation to Allah, the Almighty, for showering His blessings on me throughout my study effort, allowing me to successfully complete the research. I'd like to express my heartfelt appreciation to **Md. Mostafijur Rahman** Sir, Assistant Professor, Department of Computer Science Engineering, University of Barishal, for allowing me to do research and providing me with vital guidance during this process. His dynamism, vision, genuineness, and motivation have all left an indelible impression on me. He showed me how to conduct research and present my findings most clearly and concisely possible. Working and studying under his direction was a wonderful honor and privilege. I'd also like to express my gratitude for his guidance, sensitivity, and good humor.

In addition, I am grateful to **Md. Samsuddoha** Sir, Assistant Professor, Department of Computer Science Engineering, University of Barishal, for mentoring me through this thesis journey. My co-supervisor deserves special thanks for his tremendous advice, unwavering support, and patience. His vast expertise and wealth of experience have aided me throughout my academic career as well as my daily life.

I owe my parents a debt of gratitude for their love, prayers, care, and sacrifices in teaching and preparing me for the future. I am grateful for their love, understanding, prayers, and continued support in helping me finish this research project.

Finally, I would like to express my gratitude to everyone who has helped me complete the research work, whether directly or indirectly.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Face masks have become a vital part of virus-prevention procedures, particularly during the COVID-19 virus pandemic. Face masks protect people not only from COVID-19 infection but also from regular dirt and dust. For everybody now it appears to be a daily necessity. Many organizations and institutions must be able to monitor whether or not people wear face masks in a specific location or at a specific time. The requirements for this data should be very real-time and automated. This research paper proposes a deep learning strategy for real-time face mask detection in this situation. Deep learning is a subset of machine learning, which is a subset of AI, that is an umbrella term for using computer programs and algorithms to accomplish things cleverly and efficiently. Face mask detection is a method of determining whether or not a person is wearing a face mask by comparing stored models of each human face in a group of people. This is essential for ensuring the safety and security of any industry or organization.
This proposed strategy can be used in a variety of sectors, such as the military, defense, schools, colleges, and universities, airplanes, banking, and other public places. The inherent variability of the mask, such as shape, texture, and color, is a hard challenge in face mask recognition. According to our tests, our suggested DNN model and Python method are very efficient and accurate at determining whether or not a single or numerous people in any gathering or institution are wearing a mask. It can notify if someone isn't wearing a face mask properly, in addition to demonstrating accuracy. If the model can be implanted into university CCTV in front of the main gate, lift, office room, and classroom or lab room, the overall safety of university students, teachers, and staff from COVID-19 can be ensured.

1.1 Problem Statement

Though being fully vaccinated, the risk of contracting or transmitting the virus does not completely eradicate it. Every day, a large number of students, teachers, and staff gather at the university to connect with one another. A mask is particularly effective at keeping respiratory particles and droplets from infecting others if a student, teacher, or staff member is affected by COVID-19 and is unaware of it. Wearing a mask can also help prevent germs from other people's respiratory droplets from entering into another person's nose and mouth if a student has not yet received the COVID-19 vaccine. As a result, building an effective surveillance system at a university can save everyone. According to a World Health Organization (WHO) report dated July 12, 2020, the current COVID-19 outbreak has infected over 13,039,853 people and caused more than 571,659 deaths in more than 200 countries and territories, with a mortality rate of approximately 37% compared to less than 1% for influenza [1]. A novel coronavirus has caused person-to-person transmission, but as far as we know, this harmful virus-producing illness can also be transmitted by an asymptomatic carrier who does not have COVID-19 symptoms. Masks are still an effective solution in crowded indoor public locations with a mix of vaccinated and unvaccinated people because COVID-19 can spread by droplets and particles discharged into the air by speaking, singing, coughing, or sneezing. To avoid the spread of COVID-19, people should keep a safe distance from one another, wash their hands often, and wear a mask. (COVID-19) has infected approximately 20 million individuals worldwide, resulting in over 0.7 million deaths, according to the World Health Organization's (WHO) official Situation Report–205 [2]. Viruses can't spread over the world if there's a good monitoring mechanism in place.

1.2 Motivation of the Research

In healthcare centres and other areas where people may be at risk for serious COVID-19 effects, wearing a mask is still suggested. COVID-19 patients have reported a wide range of symptoms, including shortness of breath and difficulty breathing. Lung illness in the elderly puts them at a higher risk [3]. As a result, detecting face masks has become a critical responsibility in the current situation. Face mask detection entails detecting the position of a person's face and then determining whether or not they are wearing a mask.

Hence the motivation for the thesis is to design an effective model for detecting people's face masks, ensuring safety in all aspects of life. As a result, the suggested deep learning model can determine whether or not a person is wearing a mask appropriately. This approach was initially proposed for a university monitoring system, but it can be used efficiently in any public setting. If this concept is used in schools, colleges, universities, shopping malls, medical centres, offices, streets, and other local areas, the rate of COVID-19 infection will be reduced.. And, over time, a secure living environment will emerge.

1.3 Research Question

The goal of this study is to wear a mask appropriately to maintain social distancing and reduce the risk of virus infection. The objective is to create a deep-learning model that can detect whether or not people are wearing masks while interacting with others. As a result, the research questions are:

- **RQ1. How to detect face masks from real-time images or video streaming?**

To recognize the face mask successfully, an improved deep neural network (DNN) model is constructed. At the same time, Keras, Tensorflow, MobileNet, and OpenCV may be used to develop a real-time Face Mask Detector. In addition, this paper will demonstrate how to use this on a live video camera. this model could be connected with CCTV cameras of public places to detect and identify people wearing and not wearing masks.

- **RQ2. Does the proposed model provide better accuracy than the existing face mask detection models?**

For face mask detection, multiple techniques are used in existing design systems. Various models are created with different parameters, and the results of these design systems are quite efficient. Following the model's implementation, the following step is to justify these solutions and make a choice on their flexibility and accuracy in comparison to the proposed method.

1.4 Unique Contribution

Our suggested system can be employed in real-time applications that require face-mask detection for safety purposes due to the COVID-19 outbreak. This thesis can be integrated efficiently and effectively with embedded systems for a variety of applications, including autonomous driving, education, surveillance, and so on. It can also be used to ensure that public safety guidelines are followed in railway stations, offices, schools, and public places. This model will show more accurate results than the other systems currently in use. For the proposed model, a large dataset containing 3837 images is used for training, testing, and validating the model. The dataset has 2 parts - with mask and without the mask. After collecting the images from different sources, all of the images are processed to set a common standard for all. Then they proceed using several functions based on various parameters. Eventually, a unique and large dataset for real-time face mask detection is created. Again, the proposed method has been developed via Keras, Tensorflow, MobileNet, and OpenCV. For training purposes, directly colored images are converted into binary images. Many of the existing models have converted the colored image into a grayscale image and then converted it into a binary image. In this paper, this extra task is effectively eliminated and directly used RGB to RGB conversion process. As a result, the model can provide more accuracy in less time and less complexity. Compared to the existing model, this one is more efficient and easy to implement in the institution's monitoring system. This research primarily focused on the university monitoring systems. That will create a healthy environment and ensure the overall safety of all students, teachers, and staff as well.

1.5 Structure of the Thesis

This section identifies the chapters and presents an overview of each of them. This thesis is organized into six sections. The current chapter serves as a synopsis of the entire thesis work. The rest of the thesis is structured as follows:

Chapter 2: This chapter provides an overview of the background research on the topic of face mask detection. Here are some of the related terminology and technologies that have been discussed.

Chapter 3: This chapter focuses on the relevant work that has already been published. This chapter briefly describes several face mask detection techniques, datasets, limitations, and accuracy.

Chapter 4: This chapter describes the evaluation of the proposed method. The approaches for analyzing results on the custom dataset are also presented in this chapter. Besides, this chapter depicts the scenario of the real-time mask identification process by describing the proposed approach step by step.

Chapter 5: This chapter contains the numerical results and analyses. This chapter presents a comparative analysis and judgment.

Chapter 6: Finally, this chapter wraps up the thesis with a summary discussion and future research direction.

Chapter 2

Background

Deep learning is a machine learning and artificial intelligence (AI) technique that mimics how humans acquire knowledge. Deep learning is highly useful for data scientists who are responsible for gathering, analyzing, and interpreting massive amounts of data; it speeds up and simplifies the process. Deep learning can be regarded as a means to automate predictive analytics at its most basic level. So far, deep learning algorithms are built in a hierarchy of increasing complexity and abstraction, unlike typical machine learning algorithms, which are linear. It is a field that is focused on computer algorithm learning and development on its own. At the same time, deep learning uses artificial neural networks, which are supposed to mimic how humans think and learn [4]. Several investigations must be conducted in order to comprehend and execute the proposed approach for detecting real-time face masks. The most important terms are illustrated below-

2.1 MobileNet

MobileNet is a convolutional neural network built for mobile and embedded vision. They are built on a simplified design that builds lightweight deep neural networks with minimal latency for mobile and embedded devices using depth-wise separable convolutions [5]. MobileNet-v2, for example, is a convolutional neural network with 53 layers. A network trained by an ImageNet database of more than a million images can be loaded using a pre-trained version. The network can classify photos into 1000 different object categories, including keyboards, mice, pencils, and a variety of animals.

2.2 Tensorflow

Google researchers created TensorFlow, an open-source framework. Machine learning, deep learning, and other statistical and predictive analytics tasks are all run using Tensor flow [6]. Google recommends using high-level ones wherever possible to make application programming and data pipeline building easier.

2.3 Keras

Keras is an open-source Artificial Neural Networks library. It gives a Python interface to ANN. It also serves as a TensorFlow library interface. Keras was created with the goal of allowing fast experimentation with deep neural networks. Keras also includes a number of Neural Organization building blocks, such as layers, targets, actuation, capacities, analyzers, and a huge number of apparatuses with image and text data for the execution of deep neural organization codes. Furthermore, it supports Convolutional and repeating neural networks. It is also used to productize deep learning models for mobile devices. Besides, Keras supported a variety of backends up until version 2.3, including TensorFlow, Microsoft Cognitive Toolkit, Theano, and PlaidML. Only TensorFlow is supported as of version 2.4. It is user-friendly, modular, and expandable, with the goal of allowing quick experimentation with deep neural networks.

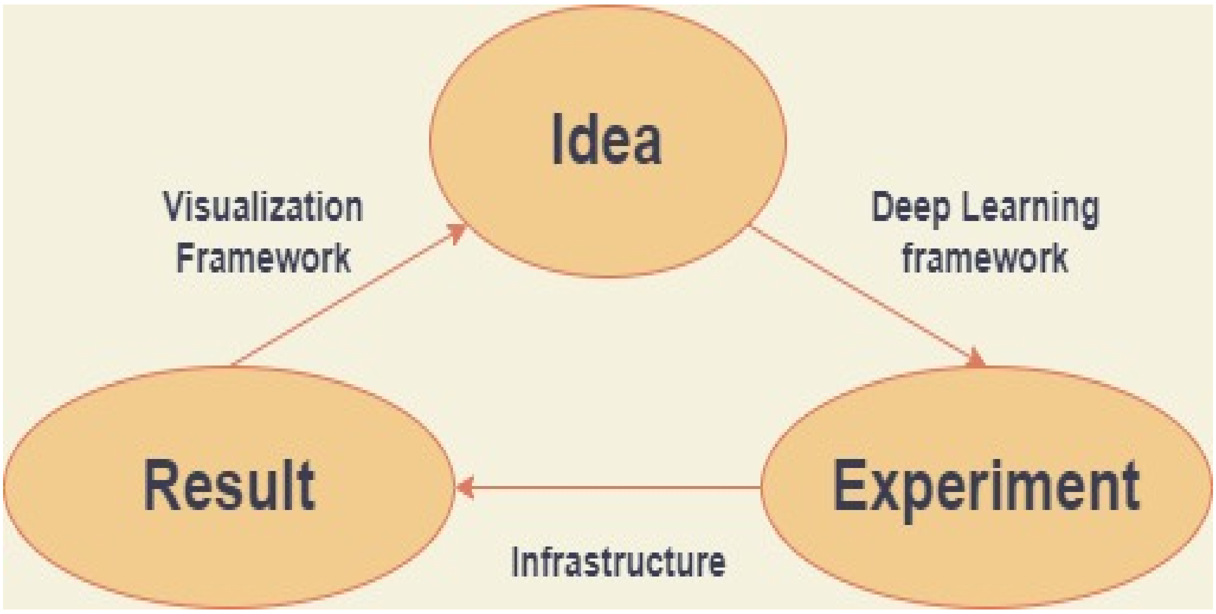

Figure 2.1: Working flow of Keras

Figure 2.1 depicts the structure of Keras, which takes advantage of the TensorFlow platform's full deployment features. Keras models can also be exported as JavaScript for use in the browser, as well as TF Lite for use on iOS, Android, and embedded devices. So far, serving Keras models over a web API has been simple [7].

2.4 OpenCV

OpenCV is a programming library intended primarily for real-time computer vision, machine learning, and image processing [8]. OpenCV was originally written in C++ and allows multi-core processing. These algorithms are also available in Python, Java, and MATLAB bindings. Image processing, video capture, and face detection are the key areas of concentration. OpenCV takes advantage of the underlying heterogeneous to compute platform's hardware acceleration. Nearly 2,500 algorithms for various computer vision approaches are available in OpenCV. OpenCV is a cross-platform library, which means it can run on any platform.

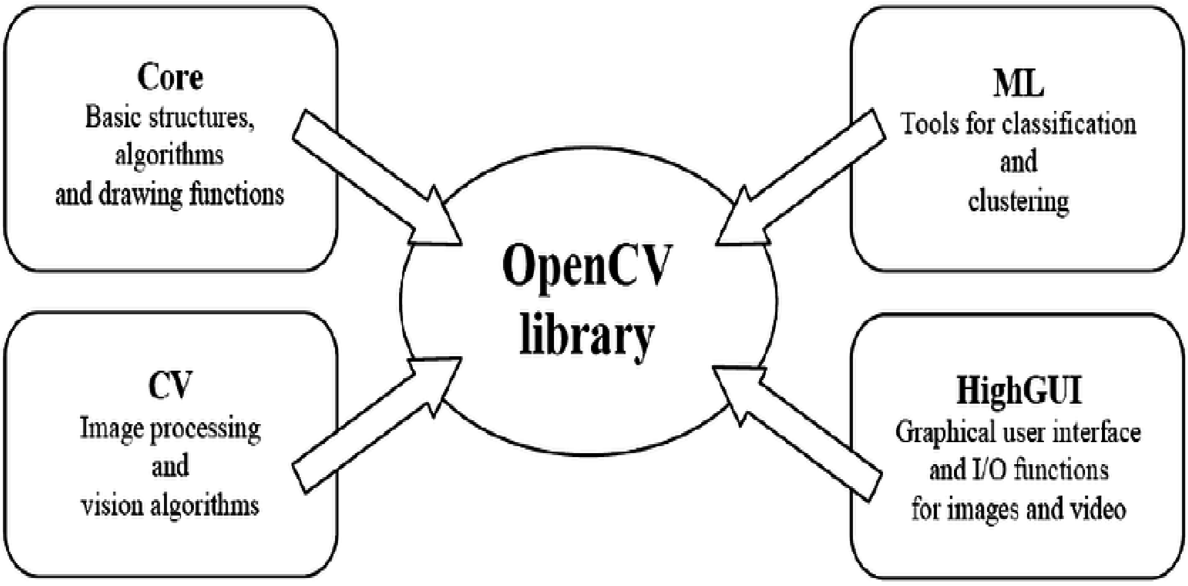

Figure 2.2: Operations of OpenCV

Here fig 2.2 shows the basic operations of OpenCV. OpenCV plays a vital role in detecting real-time face masks in this proposed method.

2.5 Deep Neural Network (DNN)

A deep neural network (DNN), or deep net for short, is a neural network with a certain amount of complexity, usually at least two layers. Deep networks use advanced math modelling to analyze data in complex ways [10]. To properly grasp deep neural networks, though, it's best to think of them as a process. Before deep nets, a few things had to be built. Artificial neural networks arose from the learning component of the modelling process. The hidden layer is used by ANNs to store and evaluate how important each of the inputs is to the output. The hidden layer retains information about the relevance of input and creates correlations between the importance of input combinations.

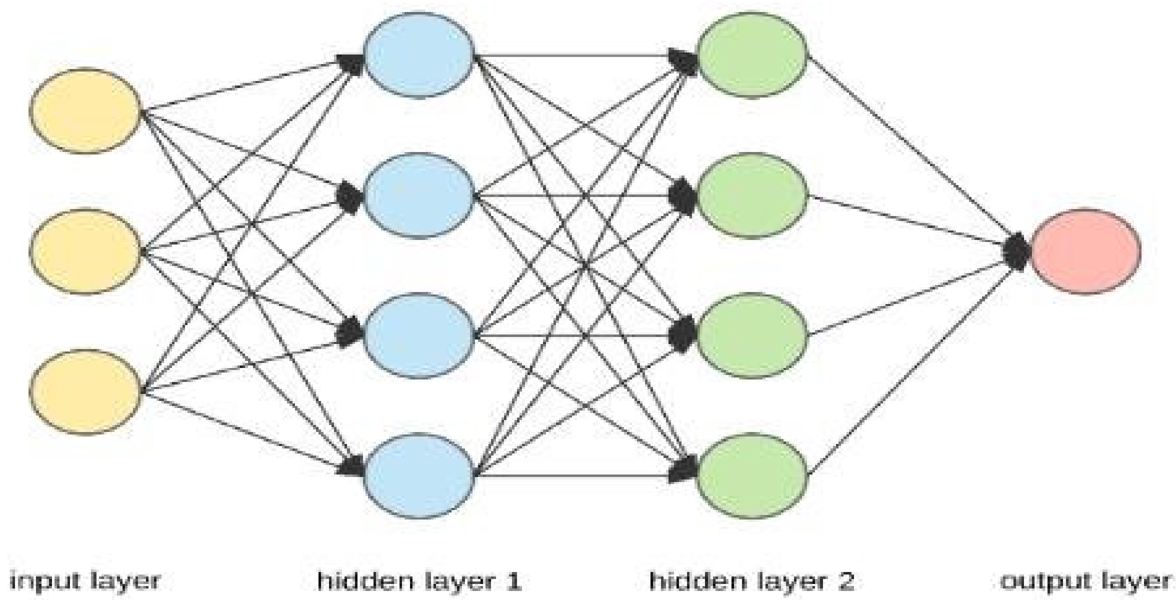

Figure 2.3: 2 Layer Deep Neural Network

Deep neural networks, on the other hand, make use of the ANN component. They argue that this improves a model so well—because each node in the hidden layer creates both associations and grades the relevance of the input to decide the output. As a result, there are several hidden levels in the deep net. Fig 2.3 depicts a simple 2 layers Deep Neural Network. A model's layers are said to be 'deep' if there are numerous deep layers.

2.6 Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a type of neural network that specializes in processing data with a grid-like architecture, such as an image. Each neuron has its receptive field and is coupled to other neurons in such a way that the full visual field is covered. Multiple layers of artificial neurons make up convolutional neural networks. Artificial neurons are mathematical functions that calculate the weighted sum of several inputs and output an activation value, which is an approximate imitation of their biological counterparts [9]. Each neuron's action is determined by its weight. The artificial neurons in a CNN pick out numerous visual properties when fed with pixel values.

Fig 2.4 shows the working method of CNN. Each layer of a ConvNet

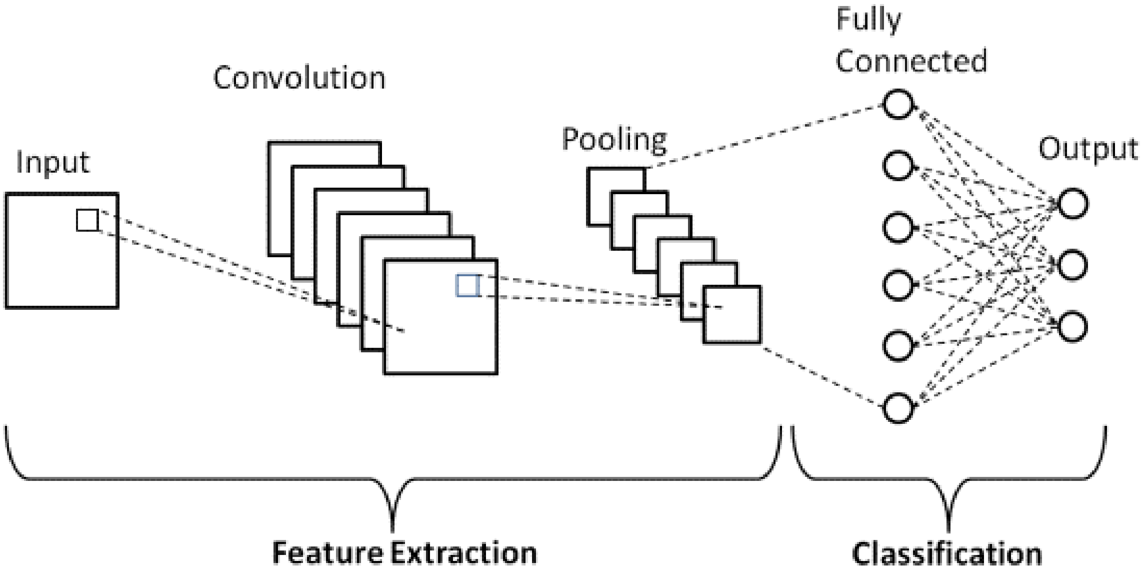

Figure 2.4: Working Method of CNN

generates numerous activation maps when an input image is fed into it. Activation maps highlight the image's most important characteristics. Each neuron takes a pixel patch as input, multiplies its color values by its weights, adds them together, and runs them through the activation function.

2.7 Stages of Creating Deep Learning Model

The study and building of algorithms that can learn from and make predictions on data is a typical task in deep learning. These algorithms work by creating a mathematical model from input data and producing data-driven predictions or judgments. In most cases, the input data needed to develop the model is split into different data sets. Three data sets, in particular, are often employed at distinct stages of the model's development: training, validation, and test sets.

2.7.1 Training Data Set

The model is first fitted using a training data set, which is a collection of instances used to learn the model's parameters (for example, the weights of connections between neurons in artificial neural networks). A supervised learning method, such as gradient descent or stochastic gradient descent, is used to train the model using the training data set. In practice, the training data set is typically made up of pairs of input vectors (or scalars) and output vectors (or scalars), with the response key referred to as the target [11]. For each input vector in the training data set, the current model is run with the training data set and provides a result, which is then compared to the target. The parameters of the model are adjusted based on the results of the comparison and the specific learning technique utilized. Variable selection and parameter estimation can both be part of the model-fitting process.

2.7.2 Validation Data Set

The fitted model is then used to predict responses for observations in a second data set known as the validation data set[11]. While modifying the model's hyperparameters, the validation data set provides an unbiased evaluation of a model fit on the training data set. By terminating early, validation datasets can be used for regularization stopping training when the error on the validation data set increases, as this is a sign of overfitting to the training data set. The fact that the error in the validation dataset may fluctuate during training complicates this basic approach in practice, resulting in several local minima. As a result of this intricacy, various ad hoc rules for determining when over-fitting has genuinely begun have been developed.

2.7.3 Test Data Set

Finally, the test data set is a set of data used to provide an objective assessment of a final model fit on the training data set. The test data set is sometimes known as a holdout data set if the data in it has never been used in training for example, in cross-validation. In certain publications, the phrase "validation set" is used instead of "test set" e.g., if the original data set was partitioned into only two subsets, the test set might be referred to as the validation set. The problem and data available play a big role in determining the sizes and procedures for data set division in training, test, and validation sets.

2.8 Generalization in Deep learning

The generalization of a deep learning model refers to how effectively the concepts acquired by the model apply to specific examples not seen by the model while learning. A successful deep learning model should be able to generalize well from training data to any data in the problem domain. This enables us to make future predictions based on data that the model has never seen before. When we talk about how well a machine learning model learns and generalizes to new data, we use the terms overfitting and underfitting in machine learning. The two most common causes of poor deep-learning algorithm performance are overfitting and underfitting [12].

2.8.1 Statistical Fit

A fit refers to how well a function approximates a target function in statistics. Because supervised machine learning algorithms strive to estimate the unknown underlying mapping function for the output variables given the input variables, this is an appropriate vocabulary to use in machine learning. The goodness of fit is a term used in statistics to describe metrics used to estimate how well a function's approximation matches the target function. Some of these strategies are valuable in machine learning (for example, calculating residual errors), but others imply we know the shape of the target function we're approximating, which isn't always the case.

2.8.2 Overfitting

A model that overfits the training data is referred to as overfitting. When a model learns the information and noise in the training data to the point where it degrades the model's performance on fresh data, this is known as overfitting. This means that the model picks up on noise or random fluctuations in the training data and learns them as concepts [12]. The issue is that these notions do not apply to fresh data, limiting the models' ability to generalize. Nonparametric and nonlinear models, which have more flexibility when learning a target function, are more prone to overfitting. As a result, many nonparametric machine learning algorithms incorporate parameters or strategies that limit and constrain the amount of detail learned by the model.

2.8.3 Underfitting

Underfitting is defined as a model that cannot both model and generalize to new data. A deep learning model is underfitting as evidenced by its poor performance on the training data [12]. Underfitting is rarely considered since, given a decent performance metric, it is simple to discover. The solution is to move on and experiment with different machine-learning techniques. Nonetheless, it serves as a good counterpoint to the issue of overfitting.

Chapter 3

Literature Review

Deep learning is a technology that can be used to analyze large amounts of data and has applications in computer vision, pattern recognition, and speech recognition, among other things. Again, MobileNet has fewer parameters and higher classification accuracy than MovileCV, which is considered a lightweight deep neural network. Dense blocks, as proposed in DenseNets, are integrated into MobileNet to reduce the number of network parameters and increase classification accuracy. Several studies have been conducted to efficiently detect face masks from real-time video streams. Some of the notable research work on face mask identification utilizing various models and approaches are summarized in this part.

3.1 Related Works

Qin et al. [13] developed a classification system for face mask wearers that included image superresolution and used a classification network (SRCNet). Based on 2D facial images, it quantified masks, no masks, and poorly worn masks. The algorithm's backbone consisted of image pre-processing, face detection and cropping, image super-resolution, and identification of face mask-wearing circumstances. The training dataset includes 3835 photos, containing 671 images without a face mask, 134 images of wrong face mask use, and 3030 images of correct face mask use. SRCNet reported a 98.70 percent accuracy rate. A unique data augmentation strategy for mask detection from speech is developed by Ristea et al. [14]. The method might be used for communication among surgeons, forensic sectors, or infectious disorders like COVID-19. Two approaches are proposed in this research. The first is to train cycle-consistent GANs to translate unpaired utterances between two classes with masks and without masks, and the second is to create fresh training utterances using cycle-consistent GANs, giving opposing labels to each translated utterance. Original and translated utterances are transformed into spectrograms and fed into a set of ResNet neural networks of varying levels. A Support Vector Machines (SVM) classifier is used to merge the networks into an ensemble. They will concentrate on multiclass challenges in the future. Execution face mask-related projects have been proposed by Wang et al. [15]. That was an obvious task by providing three samples of masked face datasets, which comprise of Real-world Masked Face Recognition Dataset (RMFRD), Masked Face Detection Dataset (MFDD), and Simulated Masked Face Recognition Dataset (SMFRD). These datasets are publicly available to academia and industry and can be used to construct a variety of masked face applications. In many circumstances, such as face access control, community access control, facial attendance, facial security, checks at train stations, and so on, conventional facial recognition technology is nearly useless. The multi-granularity masked face recognition model we built has a 95%accuracy rate, which is higher than the industry average

Khandelwal et al. [16] described a deep learning model that binarizes a picture based on whether or not a mask is utilized. In this research, the method shows how to use a combination of modern deep learning and classic projective geometry approaches to create a robust social distancing assessment system. At the same time, concentrate on deploying the solution across the Aditya Birla Group's production units (ABG). In addition, the model was trained using 380 photos, while no mask was applied to 460 images, and these images were used to train the MobileNetV2 model. A Deep Dense Face Detector (DDFD) for face detection system was proposed by Z. Wang et al. [17] which can be used for multi-view face detection. The proposed method is the simplest, as it requires no segmentation, bounding-box regression, or SVM classifiers, and it can distinguish faces from a variety of angles. Some commonly used deep learning architectures, as well as their practical applications, are addressed in this study. Four deep learning architectures, including autoencoder, convolutional neural network, deep belief network, and limited Boltzmann machine, are given an up-to-date overview. For the aim of face detection, many types of deep neural networks are explored, and current progress is described. Deep learning approaches are also featured in some specific domains, such as speech recognition, pattern recognition, and computer vision.

A Retina face mask has been proposed by Jiang et al. [18] which is a high-accuracy and efficient face mask detector. ResNet and MobileNet are the models used. The project in question distinguishes between wearing a face mask and not wearing one. The RetinaFaceMask model was proposed as a high-performance one-stage face mask detector in this paper. A fresh dataset with various annotations has been created. A context attention module was also developed, with the goal of learning discriminating features linked with wearing a face mask. Finally, the knowledge gained from the face identification test was transmitted, based on how individuals develop their abilities by repeating comparable activities. The proposed model's benefits were demonstrated by ablation research. The state-of-the-art performance of this RetinaFaceMask model was proved in experiments using both public and novel datasets.

Inamdar et al. [19] suggested a facemask net deep learning network for effective real-time face mask detection. As an input to the Facemask Net model, the suggested methodology can recognize an image or a live video stream. An image is classified as wearing a mask, not wearing a mask, or not wearing a mask. The model was divided into two parts: face mask detector training and face mask detector implementation. This dataset was used to train our FaceMaskNet model, which identified face-masked and non-face-masked photos with an accuracy of 98.6%.Matthias et al. [20] have worked on a face mask identification project that focuses on obtaining real-time photos that indicate whether or not a person is wearing a mask. In this context, investigates the development of CCTV–enabled facial mask recognition software that will make the protocol's monitoring and enforcement easier. The classification process was carried out using a pre-trained deep convolutional neural network (CNN) and the most probable limit (MPL), according to a constructive research technique. To recognize essential facial traits and apply a decision-making algorithm, the suggested technique uses two datasets to train. The Real-World-Masked-Face-Dataset results show that the Real-World-Masked-Face-Dataset has a high success rate in face recognition.

The dataset was utilized for both training and implementing the decision-making algorithm to recognize the primary face features such as eyes, mouth, and nose. By allowing individuals to validate the face mask via their webcam, a proof of concept, as well as a development basis, are provided to help reduce the spread of COVID-19. A model was proposed by Upendra et al. [21] using OpenCV and Deep Learning to develop real-time face mask detection. This suggested model uses the Tensor flow and OpenCV libraries to develop a CNN for facial image detection, and then Deep learning techniques for face detection with and without masks. The dataset consists of 1916 photos of people wearing the mask and 1919 photos of people who aren't wearing it. This proposed model has 98% accuracy in face mask detection after multiple testing with a batch size of 32 and 20 Epoch repeats.

Chapter 4

Proposed Methodology

The suggested system is a deep learning solution that trains the model using OpenCV and TensorFlow. The proposed model will aid in the safety of people in public places by automatically identifying whether or not they wear a face mask. If a person is not wearing a mask or wear it improperly, identify him or her and notify the authorities so that they can take appropriate action. This section briefly covers the solution architecture and how the suggested system will work in an automated manner to prevent virus propagation.

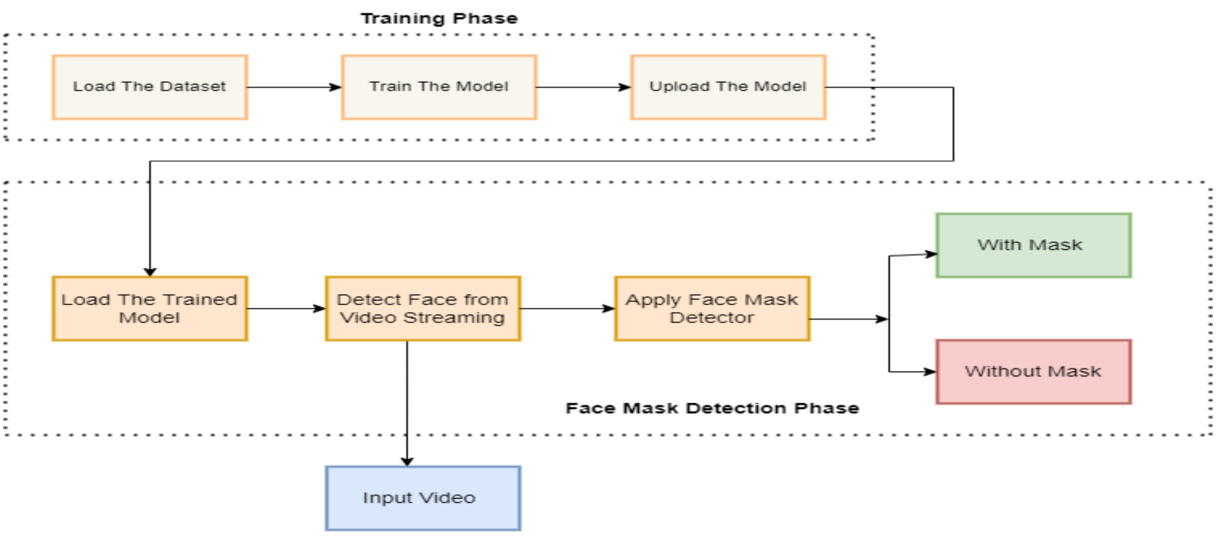

Figure 4.1: Solution architecture of proposed system

The proposed model's whole working technique is depicted in Figure 4.1. With a camera like CCTV or paired with a Raspberry Pi4, the suggested system combines a MobileNet approach to performance optimization with a deep learning algorithm and computer vision to automatically monitor individuals in public places and distinguish persons with masks or no masks.

4.1 Data Collection

A large custom dataset was collected to train, test, and develop the proposed model. The dataset is divided into two subsets: mask and no mask photos. All of the images were gathered from a variety of reliable sources, including Kaggle, Google Photos, and other royalty-free websites. The photos of persons wearing masks and without masks are preprocessed after they are collected. Here, an image data generator is utilized to create a large dataset from the gathered photos, allowing the model to be trained with additional data.

4.2 Data Preprocessing

To develop the Mobilenet model using this proposed method, convert all dataset photos into arrays. After that, masks and non-mask categories are generated. Following that, two empty lists are created in which the data contains all of the photos. The equivalent photos with and without face masks are put to these empty lists, and then a loop is run for both of them to load the preprocessed images from Keras. The size of the target image is 224*224 pixels to set a standard for all images of both dataset.

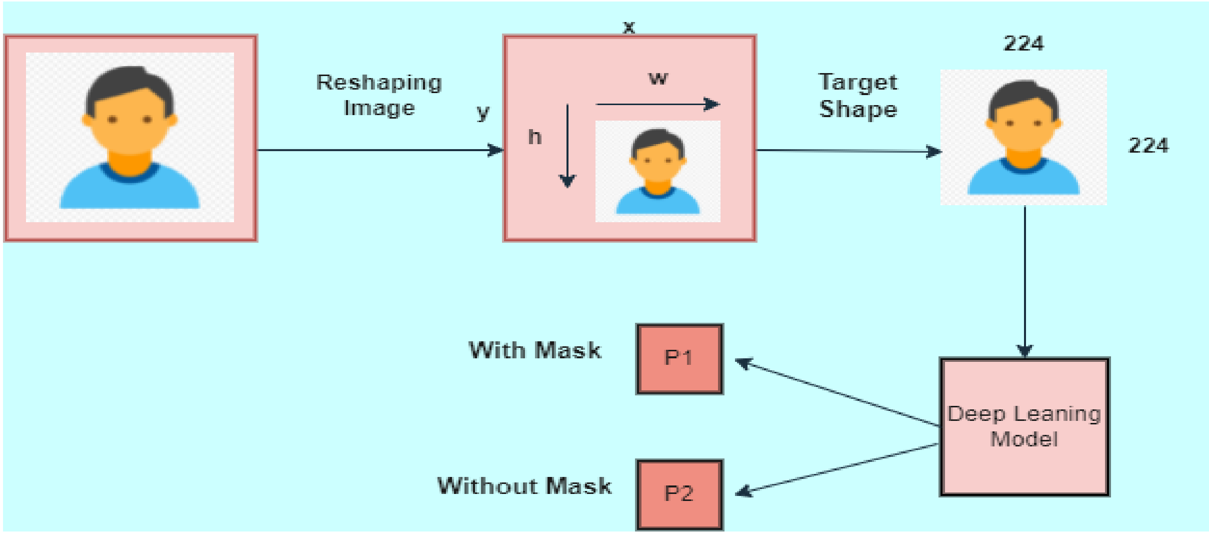

Figure 4.2: Image Reshaping

Here figure 4.2 shows the reshaping process of the images. This technique yields a standard size for subsequent processing for all photos in the collection. All of the photos in the dataset are converted to NumPy array format, and the associated labels are added to the images. The image-to-array method is used to transform the images into arrays in this case. Then, for MobileNet, preprocessed input images are used. The photos are appended to the data list as numerical numbers. Then, within the empty level list of alphabetic values, 2 categories of dataset are appended. All photos are converted to binary numbers in the following step of Mobilenet model preprocessing. Then make a NumPy array out of it. After that, the train-test split function is used to split the data in this scenario. For training, about 80% of the complete dataset is used. The remaining 20% is for testing reasons, with the Random state set to 42. The images in the training data set are also divided into two groups: mask and no-mask images.

4.3 Model Building & Training

The CNN is utilized first to construct the proposed model. Instead of using classic CNN, the suggested methodology uses the Mobilenet model for implementation to improve accuracy. All photos are delivered to the MobileNet model once they have been reshaped and processed. To get the output, they went through the max pooling and completely linked layers. Until now, MobileNet has been the Faster CNN model with fewer parameters. The number of epochs is a hyperparameter that controls how many times the learning algorithm runs over the whole training dataset [22]. The model was trained using four distinct Epochs, with a learning rate of 1e-4. One epoch signifies that each sample in the training dataset got a chance to update the internal model parameters once. The learning rate is 0.0001 for each epoch, and BS is 32. To improve accuracy, the learning rate was kept low for the model. The next stage is to create two models.

- Head Model
- Base model

The head model is for the output and the base model is for the input values. The base model has passed through the head model training step as a parameter of it. The Image data generator functions are used to enhance the dataset photos. By adding features such as rotating, shifting, and flipping, it is possible to make many additional images from a single image. This adds value and provides a huge custom dataset for model development and training. The trained model is the Base model, which is utilized to improve the results. After that, the input tensor defines the shape of the image (224*224*3) where it defines the x and y coordinates and 3 RGB channels. For training the model, all bytes of images are converted into RGBA images. Finally, the generated model is saved in h5 format. The plot of training accuracy and loss is depicted in the plot.png file. For the detection process, set the value of the bounding box as Green and red for with mask and without mask people. The value is defined by the RGB color codes too. The bounding box for mask-wearing people has been set as (0, 1, 0) and for without mask person as (1, 0, 0). The model is trained for four different epoch values, each loop traversing the training dataset. Another nested for-loop within this for-loop iterates over each batch of samples, each batch having the provided "batch size" number of samples. The number of epochs allows the learning procedure to run until the model error has been reduced to a reasonable level[22]. The epochs of 5,10,15, and 20 are used to train the model in order to achieve the best results.

Figure 4.3: 5 epochs vs Training loss and accuracy during model training

Figure 4.3 shows the training loss and accuracy during the model training. After training the model for 5 epochs, the resulting training loss is very small and accuracy is higher. To get more precision, the models were trained again for 10 epochs. The training loss, validation loss, train accuracy, and validation accuracy curve are depicted in Figure 4.4.

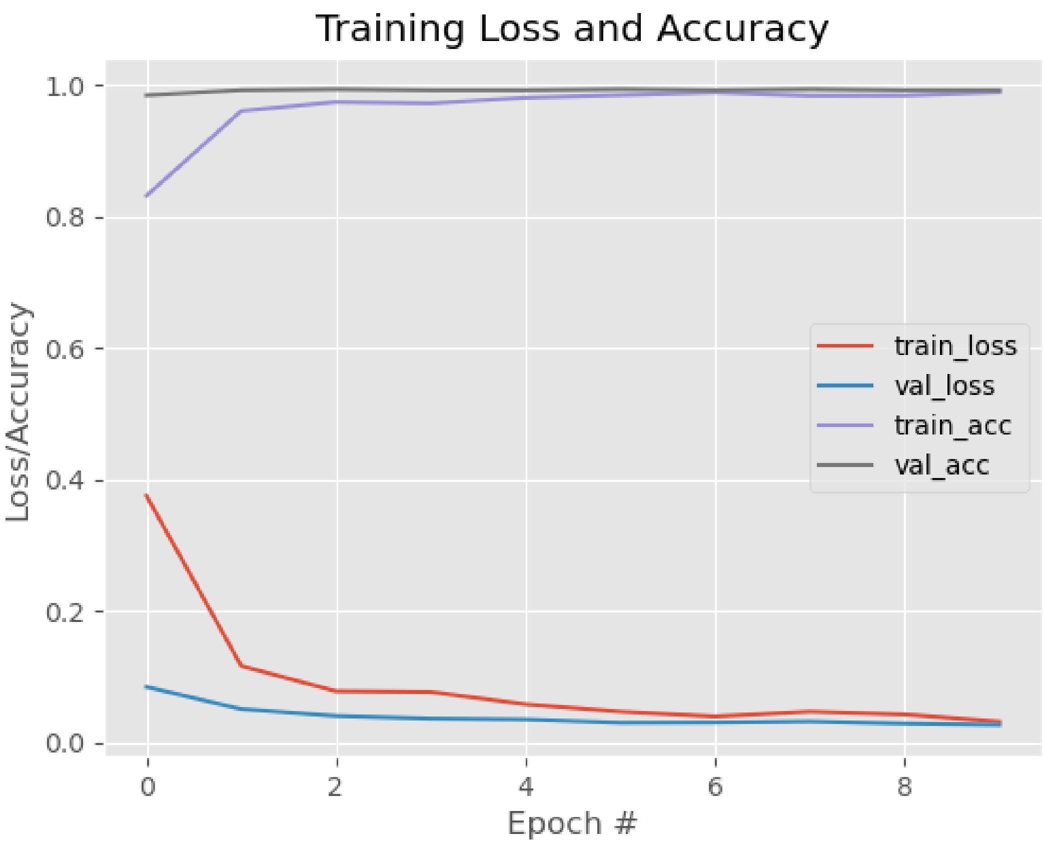

Figure 4.4: 10 epochs vs Training loss and accuracy during model training

This accuracy level is considered more efficient for the proposed model. Figure 4.4 shows the maximum accuracy and lowest train and validation loss without overfitting problem, the proposed model will be implemented for epochs 10 efficiently and effectively.

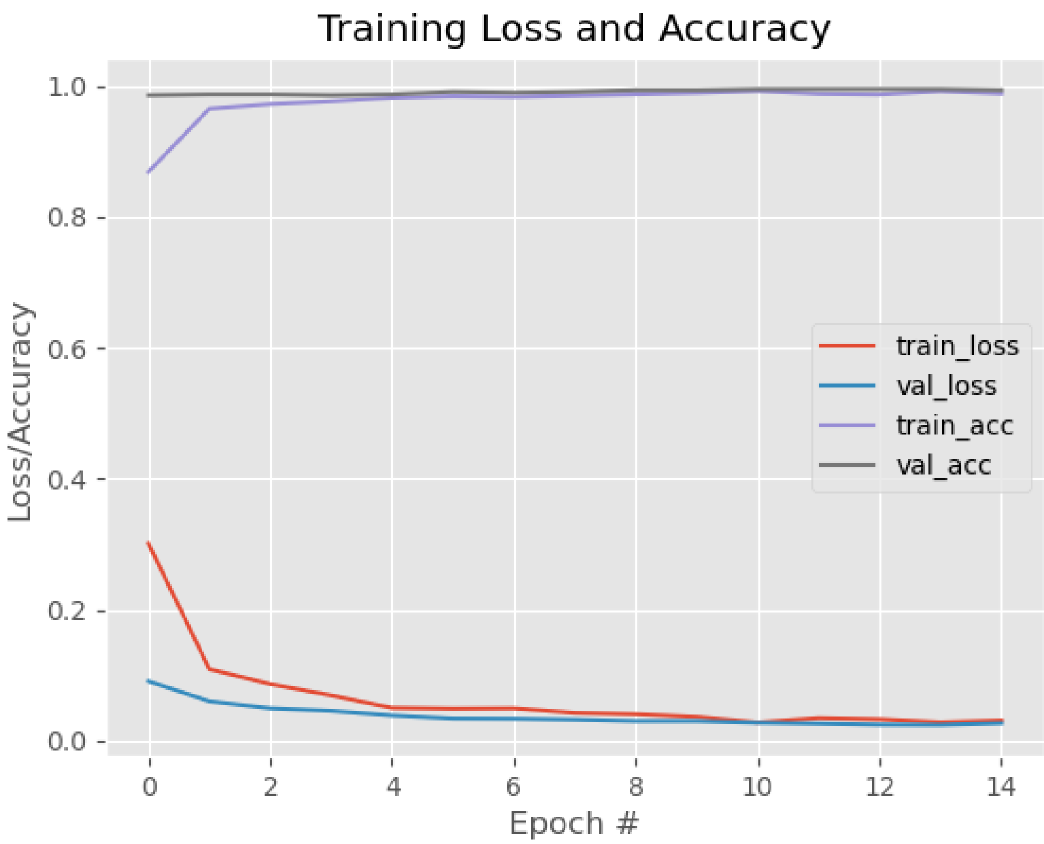

Figure 4.5: 15 epochs vs Training loss and accuracy during model training

As more epochs can generate more accurate results, this model was trained for 15 epochs as well. The resulting curves are shown in Fig 4.5. Though the validation and train loss is the lowest here the overfitting problem is found.

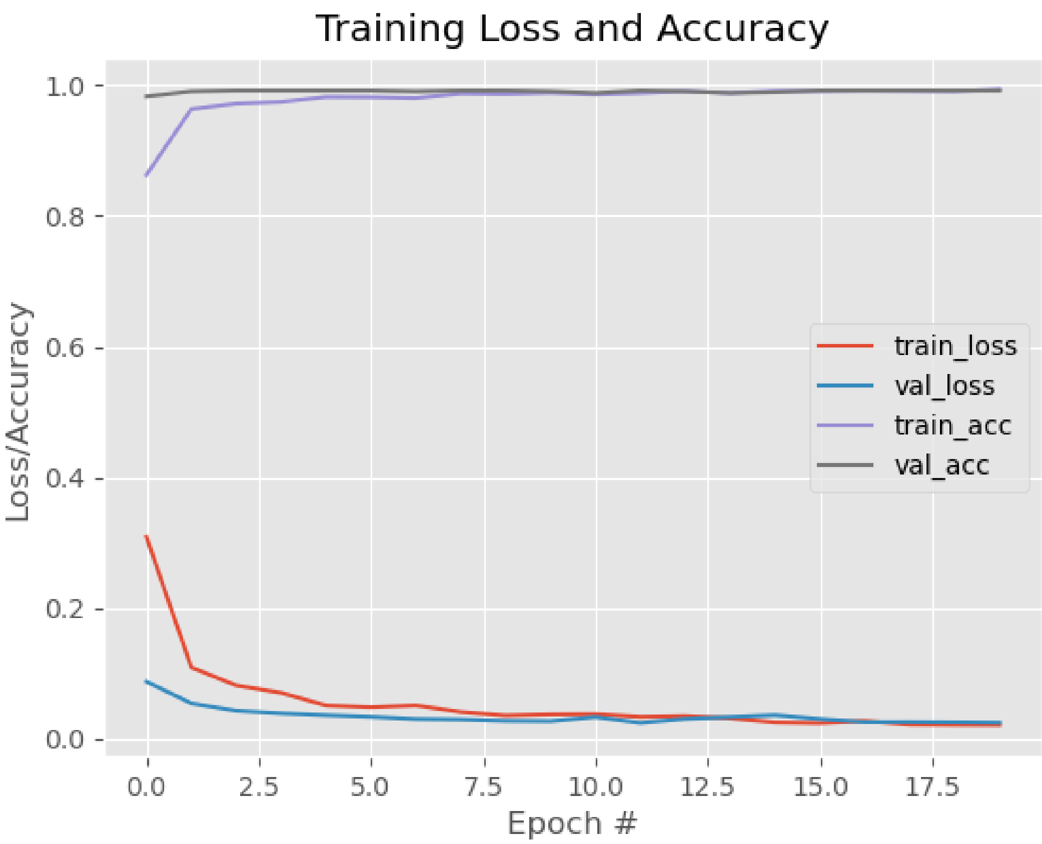

Figure 4.6: 20 epochs vs Training loss and accuracy during model training

For the further experiment, the model was trained for 20 epochs. Fig 4.6 depicted the result for 20 epochs. That also generated the overfitting of data. So far, a deep learning model's generalizability outside of the original dataset is harmed by overfitting [23]. Here figure 4.6 shows the training loss, val_loss, train_acc, Val_acc. The train_loss and val_loss are less for a maximum number of epochs. For 20 epochs, the train_acc and val_acc will increase more. After the model birding and training the model for several epochs, the efficient training stage is found for epochs 10. The accuracy is almost the same for larger epochs, also training the model successfully without any constraint.

4.4 Model Testing

The readNet and OpenCV with the developed Deep Neural Network model are used for testing the propped model. The dataset is loaded using the load model. This model can load the camera like CCTV for video streams. The source is assumed 0 as the model is testing using the primary camera. The number can be set depending on the number of cameras. The model catches the video frame as an individual image. The trained model can detect the human face individually at first. Then identify the position of wearing a mask and detect whether a person wears a face mask or not. If a person wears a face mask, the model has created the green bounding box and shows the accuracy of how appropriately the person wears the mask. If the mask covers the mouth, and nose properly, the model will show more accuracy and be masked as safe. Similarly, the proposed model can identify if a person doesn't wear a mask or wear it improperly. The model will mark the person in a red bounding box and show the accuracy level as well. Not for a single person, this model can be applied for face mask detection of multiple people. The working methodology is the same for this situation as well. The developed model will work simultaneously in multiple-person face mask detection.

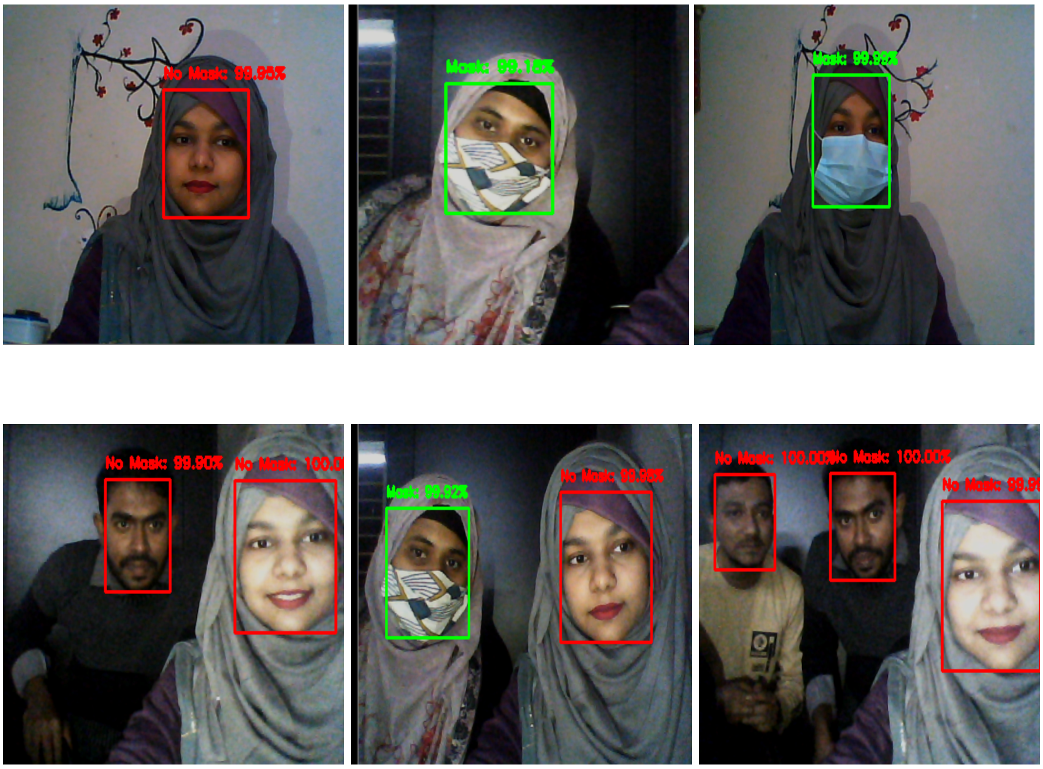

Figure 4.7: Testing the Model

Here the proposed model is validated by testing on a live camera with the accuracy shown in Figure 4.7. The input is loaded from the webcam and read as frame by frame. That can be used in real-time cameras developing the Deep neural network model.

4.5 Model Implementation

The proposed solution is tested and verified using a webcam. It can be implemented with a Raspberry Pi 4 of a CCTV camera. This will aid in the real-time tracking of public locations to prevent the spread of COVID-19. The trained model with the custom data set is installed on the Raspberry Pi 4, and the camera is connected to it. The camera feeds real-time video from public places such as schools, colleges, universities, hospitals, and other locations to the Raspberry Pi4 model, which continuously and automatically monitors public places and detects whether people maintain safe social distances and whether they are wearing masks. As this model is especially proposed for university monitoring systems, if a student, teacher, or staff member doesn't wear a mask properly, the implemented system sends an alert to the university control office to take necessary steps. The system is also the same for other public places. That will ensure a healthy living environment for all of us.

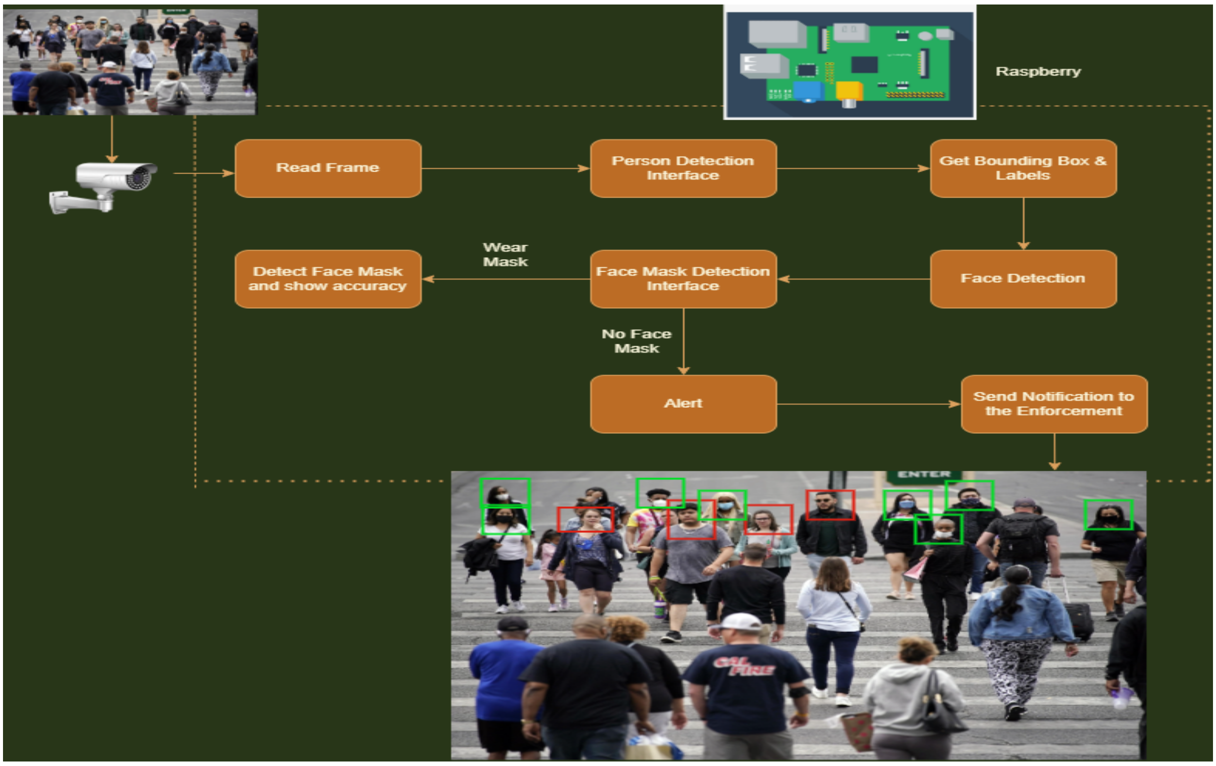

Fig 4.8 - Model Implementation

Here figure 4.8 shows the scenario of how the proposed model will work after implementation. From the public surveillance camera's video input, it can read the frame to detect the human face. After detecting the human face, identify whether the person is wearing a mask or not. The mask may be surgical, KN95, or customized. The face mask detector interface can detect any kind of mask. If the person wears a mask, identify them with the green bounding bow and show the accuracy as well. On the other hand, if the person does wear a face mask or wears it improperly, mark them in a red bounding box and send an alert to the associated control centre to ensure full public safety.

Chapter 5

Experimental Results

After training and validating the model, the expected result is achieved. The evaluation process and result will be shown in this chapter.

5.1 Install Dependencies

To train and implement the proposed model, several dependencies need to be installed to set up the appropriate environment. The following dependencies are required to train the proposed model properly.

- tensorflow==1.15.2
- keras==2.3.1
- imutils==0.5.3
- numpy==1.18.2
- opencv-python==4.2.0.*
- matplotlib==3.2.1
- scipy==1.4.1

5.2 Dataset

The proposed face mask detector model didn't use any morphed masked images dataset. The model is accurate, and because the MobileNet architecture is employed, it is also computationally efficient, making it easier to deploy to embedded systems like Raspberry Pi and Google Coral. The proposed model has 2 classes in the dataset. One is a mask and another is no mask. For this model, a total of 3837 images are used where 1919 are with masks and 1918 images without masks. Some of the samples of the custom database are shown in Fig 5.1.

Figure 5.1: Sample of the Dataset

For the proposed model, the large dataset will help to achieve maximum accuracy to detect real-time face mask detection. The objective of creating a large dataset to reduce the training and validation loss at the same time. The dataset contains different types and coloured masks. So a person either wears surgical, KN95 or any customized mask, the model can be identified easily. For a large amount of with-mask and without-mask photos, the model testing and validation can be done with the lowest validation loss. Whatever the face structure, skin tone and height are, the model can identify the face as well as the face mask effectively.

5.3 Result

The model is trained, validated, and tested four times on two datasets. For both datasets, the model can attain a reasonable level of accuracy for epochs 5. The model has been trained for epochs 10 and up in order to achieve optimum accuracy. Without overfitting, epoch 10 delivered the best results for this suggested model. It has detected faces in the frame and different types of masks with various colors, according to the mask's dataset. The proposed approach achieves a 98% accuracy, which is the optimal accuracy and reduces the cost of error. On dataset 2, which does not have a mask, the model achieves around 99% accuracy.

Parameter	Precision	Recall	f1-score	Support
with-mask	0.98	0.99	0.99	383
without-mask	0.99	0.98	0.99	384
accuracy			0.99	767
macro avg.	0.99	0.99	0.99	767
weighted avg.	0.99	0.99	0.99	767

Table 5.1: Experimental Result of the Proposed Model

The experimental outcome of the proposed model is shown in Table 5.1. MaxPooling is one of the most important factors in reaching this level of accuracy. It gives the internal representation rudimentary translation invariance while also reducing the number of parameters the model must learn. This sample-based discretization procedure reduces the dimensionality of the input representation, which is an image. The use of a large number of neurons and filters can result in poor performance. The improved filter parameters and pool size aid in filtering out the face so that the mask may be detected correctly without over-fitting.

Chapter 6

Conclusion

Wearing a face mask all the time is a demanding and exhausting duty, but it has been mandatory since the COVID-19 crisis since face masks can aid in virus management. In order to perform their services, several public service providers need clients to wear masks. At the outset of this paper, an effective real-time face mask detection model is briefly discussed. The model's learning and performance challenges were then depicted. The method has reached a maximum level of accuracy using basic deep-learning tools and techniques. After training and validation, this model shows an efficient result to detect whether a person wears a face mask or not. In public gathering, people who wear a mask, he/she will marked safe and who doesn't wear a mask or wear it improperly, she/he will marked unsafe without showing accuracy. After that, the implemented system based on the proposed solution will send an alarm and notification to the associated authorities to take action. By applying proper law, safety can be ensured in everywhere. As a result, the infection rate of COVID-19 will be reduced gradually.

6.1 Discussion

The proposed model distinguishes between being an annotated mask and or not on the degree of occlusion in four places (nose, mouth, chin, and eye). As a result, the model will only treat a mask that covers the entire face, including the nose and chin, as "with a mask." The method's key limitations include shifting perspectives and a lack of clarity. It's made more difficult by the movement of unclear faces in the video stream. So far, the model is effective in terms of showing higher accuracy of detecting face masks and sending notifications to the authorities in order to ensure public safety.

6.2 Limitation & Future Work

The model has a few flaws that were discovered. The model is unable to detect faces if the camera height is greater than 10 feet, and it cannot correctly classify partially hidden faces. There are only a few constraints that can easily be overcome in the future.

The use cases stated above are just a few of the numerous features that were included in this solution. We believe this solution can be applied to several other scenarios to provide a more comprehensive sense of security. The following are a few of the features that are currently in development:

- **Distant Measurement**

After detecting a person who doesn't wear a mask or wears it improperly, it's also a threat to spread the virus among other people. So in future, to ensure social distance, this model will be implemented for measuring whether a person stays 6 feet away from others. If not, the model will mark them unsafe and immediately take action.

- **Temperature Screening**

The elevated temperature of the human body is another key symptom of COVID-19 infection. Currently, thermal screening is done using handheld contactless IR thermometers, which requires health workers to be close to the person being screened, putting them at risk of infection. It is also virtually hard to capture the temperature of each and every person in public spots. The proposed use case can be equipped with thermal cameras based on the proposed model.

- **Detection of Coughing and Sneezing**

According to WHO standards, chronic coughing and sneezing are vital signs of COVID-19 infection and one of the main routes of disease transmission to the general public. By improving our suggested methodology with body gesture analysis is to understand if a person is coughing and sneezing in public areas while violating face masks and social distancing guidelines. The deep learning-based approach can be useful in detecting and limiting disease spread, and enforcement agencies can be alerted as a result.

- **Helmet Detection to Enhance Road Safety**

In future, the application of a deep learning approach initially designed for real-time face mask detection can be extended to helmet detection, with the goal of mitigating accidents. The motivation stems from alarming statistics indicating that approximately 41% of motorcycle drivers who succumb to accidents are not wearing helmets, and around 18% of bicyclists neglect helmet usage, contributing to accident rates. To do so, employing a diverse dataset of images portraying individuals with and without helmets, this efficient deep-learning model can undergo fine-tuning through transfer learning. This adaptation anticipates enhancing real-time systems, enabling the identification of individuals without helmets in various scenarios. By addressing helmet non-compliance through advanced technology, this future research aims to contribute to enhancing law enforcement capabilities to a safer environment on roads and reduce accident-related fatalities.

6.3 Conclusion

This paper presents an experimental study on real-time face mask identification using OpenCV and deep learning algorithms. Based on the results, it is a remarkable method for easy face mask recognition from real-time video streaming. Nevertheless, there are a few drawbacks that may easily be overcome in future work. It will help to maintain a secure environment and to ensure additional safety by automatically monitoring public places to prevent the spread of the COVID-19 virus, as well as assisting police by reducing their external surveillance work in containment areas and public places where surveillance is required, using real-time camera feeds from the Raspberry Pi4. As a result, in the current circumstances, when the lock-out is being lifted, this developed scheme will be effective in tracking public locations automatically. The world has gone over the tracking of social separation and the identification of face masks that aid human health in great detail. By deploying a model on a Raspberry Pi 4, the implementation of this approach was successfully tested in real time. This approach can be employed in a variety of public areas, including schools, colleges, universities, temples, shopping malls, metro stations, and airports. Eventually, these real-time actions will achieve the potential to drastically minimize violations, the suggested model would promote public safety by saving time and assisting in the reduction of COVID-19 spread.

References

[**1**] World Health Organization. Accessed on: 21-Jan2021. [Online]

[**2**] W.H.O., “Coronavirus disease 2019 (COVID-19): situation report, 205”. 2020

[**3**] “Coronavirus Disease 2019 (COVID-19) – Symptoms”, Centers for Disease Control and Prevention, 2020. [Online]

[**4**] In-Depth guide to machine learning in the enterprise [Access date - Nov, 21]

[**5**]Cornell University - MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications [Access date - Nov,21]

[**6**] Wikipedia - TensorFlow [Accessed on - 23-oct-21]

[**7**] Keras- Details About Keras [Accessed on - 23-oct-21]

[**8**]Wikipedia - OpenCV [Accessed on - 29-oct-21]

[**9**]DdTechTalks - Convolutional Neural Network CNN [Accessed on - Nov 2021]

[**10**]Jubin Dipakkumar Kothari (2018). A Case Study of Image Classification Based on Deep Learning Using Tensorflow. 6(7), 3888-3892.

[**11**]Wikipedia:Training, validation, and test sets[Access date - Nov,21]

[**12**]Machine Learning Mastery Overfitting and Underfitting With Machine Learning Algorithms [Access date - Dec,21]

[**13**] B. QIN, D. LI, Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19 (2020)

[**14**]N.C. Ristea, R.T. Ionescu, Are you wearing a mask? improving mask detection from speech using augmentation by cycle-consistent gans, arXiv preprint arXiv:2006.10147 (2020)

[**16**] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei, et al., Masked face recognition dataset and application, arXiv preprint arXiv:2003.09093 (2020)

[**17**] P. Khandelwal, A. Khandelwal, S. Agarwal, Using computer vision to enhance safety of workforce in manufacturing in a post COVID world, arXiv preprint arXiv:2005.05287 (2020)

[**17**] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, F.E. Alsaadi:A survey of deep neural network architectures and their applications, Neurocomputing 234, 11 (2017)

[**18** M. Jiang, X. Fan, Retinamask: A face mask detector,arXiv preprint arXiv:2005.03950 (2020)

[**19**] Madhura Inamdar, Ninad Mehendale :Real-time face mask identification using Facemasknet deep learning network, SSRN 3663305 (2020)

[**20**] Matthias, Daniel and Managwu, Chidozie and Olumide:Face mask detection application and dataset(2020)

[**21**] Upendra, Hruthik S and Suman, Shruti and Vishnu, Sai S and Dharani, Jaya: Real-Time Face Mask Detection using OpenCV and Deep Learning(2021)

[**22**] Jason Brownlee:Difference Between a Batch and an Epoch in a Neural Network (2018)

[**23**] Jim Frost: Overfitting Regression Models: Problems, Detection, and Avoidance(2017)