Contents

[1. Introduction 1](#_Toc113643408)

[1.1 Background 1](#_Toc113643409)

[1.2 Motivation 1](#_Toc113643410)

[1.3 Introduction 3](#_Toc113643411)

[1.3.1 Guidelines for Using the Facial Mask 5](#_Toc113643412)

[1.3.2 WHO's list of approved facial mask types 6](#_Toc113643413)

[1.4 Problem statement 8](#_Toc113643414)

[2. Related works 9](#_Toc113643415)

[2.1 Computer vision-based classification 12](#_Toc113643416)

[2.2 Databases for facial masks 15](#_Toc113643417)

[2.2.1 Available datasets for faces 15](#_Toc113643418)

[2.2.2 Available datasets for masked and non-masked faces 15](#_Toc113643419)

[2.2.3 Facial Mask Classification During Coronavirus 16](#_Toc113643420)

[2.3 Tools and Procedures 17](#_Toc113643421)

[2.3.1 Alex Net 17](#_Toc113643422)

[2.3.2 VGG-16 17](#_Toc113643423)

[2.3.3 Longest short-term memory (LSTM) 18](#_Toc113643424)

[2.3.4 Bidirectional LSTM 18](#_Toc113643425)

[3. Related Knowledge 19](#_Toc113643426)

[3.1 Convolutional Neural Network 19](#_Toc113643427)

[3.2 VGG16 20](#_Toc113643428)

[3.3 VGG19 21](#_Toc113643429)

[3.4 Artificial neural network (ANN) 21](#_Toc113643430)

[3.5 Inception version2 22](#_Toc113643431)

[3.6 Resnet 50 23](#_Toc113643432)

[3.7 Decision tree 23](#_Toc113643433)

[3.8 Support Vector Machines 24](#_Toc113643434)

[3.9 Ensemble 24](#_Toc113643435)

[4. Methodology 24](#_Toc113643436)

[4.1 Dataset 24](#_Toc113643437)

[4.2 Data Preprocessing Exploratory Analysis 25](#_Toc113643438)

[4.3 Train Test Split 26](#_Toc113643439)

[4.4 Model Development and Training 26](#_Toc113643440)

[4.4.1 Transfer Learning Models 26](#_Toc113643441)

[4.4.2 Implementation of Study 1 27](#_Toc113643442)

[4.4.3 Implementation of Study 2 27](#_Toc113643443)

[4.4.4 Implementation of Study 3 28](#_Toc113643444)

[4.5 Model Evaluation 28](#_Toc113643445)

[5. Results and Discussion 29](#_Toc113643446)

[5.1 Proposed Models Results 29](#_Toc113643447)

[5.2 Paper 1 Models Results 32](#_Toc113643448)

[5.3 Paper 2 Models Result 33](#_Toc113643449)

[5.4 Paper 3 Models Result 35](#_Toc113643450)

[5.5 Comparative Study 37](#_Toc113643451)

[References 38](#_Toc113643452)

Table of Figures

[**Figure 1:** Architecture of VGG-16 model. 22](#_Toc113643453)

[**Figure 2:** Class Distribution of Dataset. 27](#_Toc113643454)

[**Figure 3:** Samples Representation of Dataset 28](#_Toc113643455)

[**Figure 4:** Face Mask Classification - Confusion Matrices 32](#_Toc113643456)

[**Figure 5:** Face Mask Classification - Confusion Matrices 34](#_Toc113643457)

[**Figure 6:** Face Mask Classification - ResNet50 Confusion Matrix 35](#_Toc113643458)

[**Figure 7:** Face Mask Classification - Confusion Matrices 37](#_Toc113643459)

[**Figure 8:** Face Mask Classification - ResNet Confusion Matrix 38](#_Toc113643460)

[**Figure 9:** Face Mask Classification - Comparative Bar Chat 40](#_Toc113643461)

Table of Tables

[**Table 1:** Initial Statistics of Compiled Dataset. 28](#_Toc113643476)

[**Table 2:** Tuned Hyper parameters for ResNet50. 31](#_Toc113643477)

[**Table 3:** Face Mask Classification - Classification Reports 32](#_Toc113643478)

[**Table 4:** Face Mask Classification - Classification Reports 34](#_Toc113643479)

[**Table 5:** Face Mask Classification - ResNet50 Reports 35](#_Toc113643480)

[**Table 6:** Face Mask Classification - Classification Reports 36](#_Toc113643481)

[**Table 7:** Face Mask Classification – ResNet50 Report 38](#_Toc113643482)

[**Table 8:** Face Mask Classification - Hyper parameter tuning 39](#_Toc113643483)

[**Table 9:** Face Mask Classification – Comparative Classification Report 40](#_Toc113643484)

# Introduction

## Background

We are all familiar with the destruction that the COVID-19 pandemic and its variant have caused for millions of people. The breakout of this pandemic has immensely affected tens of millions of people throughout the globe since 2019. Since December 12, 2019, the virus has caused around 5,287,534 total deaths worldwide. Luckily, medical experts and scientists have discovered a vaccine against this virus. Vaccinations are utilized to slow the transmission rate of the virus. An estimated 8,324,380,164 people have been vaccinated *(Kurlekar et al., 2021)*. Yet the new COVID variants like Omicron, SARS, BA.5, and BA.4.6 are still a huge threat. These variants are still a threat that remains challenging to control *(Murray et al., 2021)*. Considering the transmissibility rate of these new variants, they are still effective and cause major treatment issues for re-infection and new infections. Rather than finding an effective cure for a disease, prevention is one of the best solutions for slowing down the fast propagation of this virus. WHO, abbreviated as the World Health Organization, has recommended and clinically proven that public healthcare measures are very useful. Public healthcare measures like having well-fitted face masks, getting hands sanitized and hygienic, avoiding crowded places, and getting vaccinated with all COVID doses and boost doses can prevent or decrease the risk of COVID infection *(Gupta et al., 2021)*.

## Motivation

Yet the technique of wearing face masks plays a very important part in slowing down the transmission of this hazardous disease. According to the present scenarios, run time detection of a proper facial masks or utilization of facial masks is essential at this time. People are encouraged to wear veils or face masks appropriately in public spots and crowded places. Every association and workplace is instructed to promote the importance and utilization of masks to their employees and students in educational institutes. There should be some manual checks and interventions to ensure that people follow the instructions to wear a mask. According to previous studies, it is known that wearing face masks is very valuable, even during the spread of previous respiratory viruses and diseases like swine flu. Yet wearing masks has its proportionality rates, the efficiencies of face masks like N95 and other surgical masks can highly be seen as per their performance. According to the research, these masks ultimately block the transmission of the SARS virus by up to 68%–91% (Qin et al., 2020). Wearing a face mask is highly beneficial for various airborne diseases, particles, and viruses. Masks effectively target that these harmful pathogens won't be able to enter the respiratory system of one individual from any other. A mask is a non-pharmaceutical product. Face mask wearing is a comparatively non-invasive and cheap practice that beneficially reduces the spread of respiratory diseases and infections. It is normal to say that, since COVID-19 broke out, the worldwide utilization of facemasks have become routine. It is generally used by people to avoid direct contact with germs in the open environment. Every person or patient exposed to infection with COVID is required to put the facial mask on his or her face to prevent rapid transmission of this virus.

To avoid infection, healthy persons are also recommended to put a facial masks on their faces to save themselves from getting infected. A face mask has a hygienic methodology. If it is worn properly, it can beneficially disturb the motion of particles that are expelled out of the mouth during coughing or sneezing. This disruption helps to prevent various disease transmissions, including COVID-19. If people start to wear them improperly, the importance of the facial masks to avoiding an infectious outbreak in a wider environment or among the general public will decrease. To avoid such issues, it was necessary to develop an automatic detection methodology to judge the real-time mask-wearing conditions. This methodology for predicting and checking masks in different scenarios was very important and contributed a lot to personal protection and public epidemic prevention. For automatic identification, we ultimately design models accordingly to the distinctive facial characteristics in all face mask-wearing conditions. With the innovations in IT and tech, we have many work opportunities for the development of programs using deep learning and digital image processing algorithms. Since many deep learning methods can help in demonstrating quality work in a lot of other domains, this includes the detection of different objects, classifications of image data, segmentation of image data, and distance identification (Chavda et al., 2021). Nowadays, monitoring the usage of a facemask in real-time for a lot of people is getting challenging. Manual monitoring is typically difficult to implement due to the amount of manpower required to effectively save people’s spaces as well as guarantee that people are utilizing facial masks in the right way. To maximize this factor of human labor to this extent, a particular worker’s group would be in daily interaction with a lot of people. This practice would increase the infection points and would risk the lives of these employees, along with additional financial issues and managerial effort (Boulila et al., 2021).

## Introduction

In reaction to the emergence of the latest coronavirus in the city of Wuhan, China, WHO announced the People's Health Related Emergency of International Consequence (PHEIC) dated January 30, 2020. Just after this pandemic brought forward influenza A H1N1 (Baker et al., 2020) throughout 2009, a coronavirus was later declared by the World Health Organization to be a pandemic on March 11, 2020. According to the statement made by WHO dated February 11, 2020, severe acute respiratory syndrome coronavirus 2 or SARS-CoV-2 viruses, which share genetic ancestry with the coronavirus that triggered a SARS epidemic in 2003, is a virus that causes a coronavirus illness like coronavirus (Liu et al., 2020).

Since about August 29, 2020, approximately 188 different countries, as well as territories, have received reports of nearly 24.7 million new cases, which have resulted in about 837,000 fatalities as well as 16.2 million recoveries. In this instance, healthcare facilities have implemented aggressive infection surveillance criteria. The World Health Organization issued a few recommendations but also directions for the wider populace to abide by to prevent further spread of this pandemic respiratory illness, COVID-19 among many other people, including consistent practice of hand washing, establishing good social distance, donning the mask in public, attempting to ignore social occasions, and practicing self-isolation and residential containment (Cheng et al., 2020).

Additionally, following World Health Organization norms, the government put into place certain regulations like screening and quarantining incoming passengers, shutting or controlling metropolitan as well as national or international borders (ChengVcc et al., 2020), and conducting extensive reverse transcription polymerase chain reactions, or RT-PCR testing to find cases. Subsequently, to lessen the possibility of communal transmission, just stay-at-home orders, confinement, house confinement, and cancellations of large meetings, including travel bans, were implemented to varying degrees or at different times in different nations. There were no documented vaccinations or pharmaceutical treatments that were 100% successful against coronavirus, so it is uncertain when the outbreak may terminate. 3,5 Although many vaccinations have been developed and effort has been put into immunization, neither of these can promise 100% effectiveness against SARS (Cheng.V.C et al., 2019). The United States Organization for Disease Identification, Control, and Prevention has advised that people put on facial masks to deal with a continuing coronavirus epidemiological crisis. General populace facial masks are worn substantially more commonly throughout many more Asian nations that have good experience managing novel and emerging coronavirus outbreaks. Facial masks have been recommended as a very productive technique of containing the COVID-19 epidemic in China. 9 Facial masks, according to the World Health Organization, should only be used by people who are frequently coughing and sneezing, treating covid patients who have probably COVID-19, or both (Olliaro et al., 2021).

The adoption of face masks is advised by numerous government organizations throughout the entire world to reduce the COVID-19 outbreak. Internationally, there are differences in compulsory usage as well as enforcement. Although some nations have approved laws mandating the wearing of face masks, various nations, such as China, India, Japan, South Korea, and Taiwan, have issued stricter regulations (Bar et al., 2021).

The governmental departments have encouraged the general public to use face masks, including respirators, due to an exponential expansion as well as the emergence of coronavirus (Abd et al., 2020). The USA has advised that a use of surgical face masks, textile face masks, and fabric face masks, especially for prolonged respiratory uses, should be carefully considered. Health workers should prioritize using professional face masks, such as surgical face masks and N95 respiratory protection, as these seem to defend against infection. After requiring widespread use, nations including Germany and Brazil gave single-use face masks to the wider population (Maclntyre et al., 2017). The rest of the nation can utilize some homemade fabric face masks, which offer lower coverage but have the same effectiveness if professional face masks, surgical face masks, or respiratory protection are not easily accessible.

The Czech Republic, as well as Britain, recommended individuals utilize reusable face masks, while Japan they have provided wool face masks while requiring their usage (Davies et al., 2013). The best face mask policy selection for containing the spread of the infection has not been determined by sufficient evidence. As an alternative to solitary facial masks and surgical facial masks utilized in the healthcare industry, it is advocated for the general population to use disposable facial coverings or facial masks. Reusable face mask use is somewhat successful in preventing the spread of viruses, especially when paired with non-clinical measures like promoting some social separation while maintaining handwashing (Ding et al., 2021). One of the questionnaires revealed.

According to reports, due to the coronavirus pandemic, approximately 28,000 people here between the ages of (Liu et al., 2020) and (Liao et al., 2020) donned facial masks. Approximately 70% of individuals in Vietnam, India, China, and Italy, as well as in other countries, utilized disposable facial masks to guard against coronavirus epidemics. The proportion is 50% or more in the United States and other nations. 9,11 According to perceptions over the previous six months, face mask use was consistently lowest (25% of the total) in the Northern European countries and consistently high (>75%) in several regions, including South Asia and South and Central America. The improvement throughout age has been seen in several countries, which could be useful in assessing the rising instances of COVID-19 with similar mask-related rules and recommendations (Doremalen et al., 2020).

The related dataset can be used to assist in examining whether regulations relate to practice globally as well as to fund public healthcare communications-based campaigns and initiatives that focus on face masks used to assist in minimizing the dispersion of coronavirus. We want to clarify various facets of face masks. In particular, we thoroughly examine many types of face masks, their material makeup, and their efficiency in shielding users from airborne infection, as well as rules for face mask application and face mask maintenance. When it comes to face mask categories as well as what to select if neither the medical nor surgical face mask is accessible, we prefer to inform the general public better (Lauer et al., 2020).

### Guidelines for Using the Facial Mask

There are accounts of people with pulmonary tuberculosis (TB) using medical facial masks, which greatly decreased dissemination as well as provided an additional method for limiting tuberculosis transmission between infected individuals and healthy individuals (Migliori et al., 2019). More research has shown that when you sneeze, this surgical face mask disengages additional individual covid 19 viruses. Additionally, the meta-analysis of the randomized trials revealed that both facial masks and N95 respiratory protection masks were reasonably productive in protecting healthcare professionals against pandemics as well as influenza-like illnesses, which were verified in the lab (Long et al., 2020).

A tried-and-true way of treating respiratory illnesses is to wear a mask. In a pertinent experiment, a fabric-based face mask used twelve inches away from the patient who had a coronavirus infection blocked 96% of the virus quantity that might normally be present (Bae et al., 2020).  It has been demonstrated that for each 10-fold increment in the virus quantity, there is a corresponding 26% increase in patient mortality following severe virus-related diseases (Jiang et al., 2019). According to studies regarding aerosol exposure, most kinds of facial masks are slightly helpful in saving an operator. According to the 2008 study by Van der Sande et al., every mask can reduce aerosol exposures by as much as 90% over time. Notwithstanding its poor fit, scientists discovered that wearing either sort of mask is likely to minimize the danger of this infection through virus transmission in a population.

Nevertheless, because few nanoparticles are released as aerosol, examining particulate filtering is most likely to underestimate the effectiveness of face masks (Van et al., 2008). Either handmade or medical face masks may significantly lower the number of bacteria as well as slow the overall spread of the infection, but surgical face masks are significantly superior to manufactured face masks as well as more effective in preventing dissemination. The significant problem is the lack of these surgical face masks, including N95 respiratory protection. The CDC currently recommends that fit individuals utilize cotton facial masks in public (Vaccinated et al., 2021). The significance of going through this process of masking a face while providing medical care has been observed throughout Chinese health facilities, in which health doctors (especially in the quarantine regions) seemed to have no COVID-19 infectious diseases despite being around Corona patients, whereas other doctors had 9 different lethal infections in healthcare facilities because they did not wear the facial masks on their faces (Vaccinated et al., 2021).

### WHO's list of approved facial mask types

Governmental departments and WHO advise wearing face masks throughout the coronavirus outbreak to prevent the spread of SARS-CoV-2. The use of facial masks has followed different recommendations made by numerous people's health organizations as well as governments. The World Health Organization and some different private and non governmental medical hospitals or universities agree upon wearing facial masks able to minimize the rapid spread of highly contagious lung illnesses, in the particular case of coronavirus (Setti et al., 2020). The World Health Organization advises using a variety of face masks during the coronavirus epidemic. Several categories are listed here.

* Fabric facial masks
* Surgical face masks or medical face masks
* Respirators

We aim to prioritize the safety of human life as much as possible. To make this possible, we are implementing deep learning algorithms and models that have recently been applied to various areas in real life and used to tackle numerous complex problems, yielding important outcomes. Massive amounts of data can be analyzed and examined quickly and accurately thanks to deep learning (DL). Therefore, the report provides a technique that would easily enforce facial-mask wearing as well as easily supervise it in the real-time recordings of CCTV clips. A suggested network makes it simple as well as effective for business settings and government agencies to monitor facemasks. The uniqueness of this research in comparison to other publications is the suggestion of an effective and precise method for real-time videos. The suggested method offers real-time, accurate facemask wear detection, including whether it is worn properly or not. To accomplish this, a comprehensive dataset is gathered from various valid sources like Kaggle, etc (Su et al., 2022). Here we have utilized a dataset containing three classes (mask, no mask, and semi mask) and applied five deep learning models. The classes are customized by Mask, No Mask, and Semi-Mask. And the models included are ANN (artificial neural network), CNN (convolutional neural network), VGG16 (convolutional neural network with 16 layers), VGG19 (convolutional neural network with 19 layers), and MPL-CNN (multilayer perceptron convolutional neural network). Facemask detection consists of these 5 models as part of the DL architecture. These models produce great results for object detection and object classification and have the advantages of speed and edge device compatibility. The suggested techniques can be applied to actual CCTV cameras in open places to verify whether the public is dressing appropriately as well as wearing identification marks. A possible solution is easily implementable with the minimum number of resources.

Our work ensures we attain our goal of by utilizing deep learning, effective facial mask identification in runtime is possible. DL based algorithms as well as models demonstrate good performance in various real-life problems and applications. The strategy suggested in our work is built on four steps. The first stage was to perform and work accordingly towards the goal of developing a DL model that can find and navigate facial masks as well as determine whether these are being worn properly. a phase which uses edge computing to deploy the DL model online and find masks in real time. In this work, we suggest using ANN (artificial neural network), CNN (convolutional neural network), VGG16 (convolutional neural network with 16 layers), VGG19 (convolutional neural network with 19 layers), and MPL-CNN (multilayer perceptron convolutional neural network) to identify facemasks. Several experiments have been carried out, and the results indicate that the suggested approach performs well (above 90% for training and testing accuracy). Yet all the experiments and several comparison tasks suggest that models like ANN, CNN, VGG16, VGG19, and MPL-CNN showed good effectiveness in the training data as well as correctness and precision. The task that has been proposed in this report helps to identify whether a person is wearing a proper facial mask or not. The detection takes place manually and automatically compares the data and footage. By running different models, deep learning models can identify an appropriate face mask. The detection is based on the criteria that the nose, mouth, and chin are covered. We used a dataset from the Kaggle Face Mask detection dataset, which is split into three groups: with the facial mask, without the facial mask, and finally, Semi mask. Using convolution layers, the performance of this model in terms of prediction is good. The system ultimately classifies and identifies the people wearing masks in the wrong manner in real-time throughout a CCTV video. In this way, we can automatically identify the optimal use of face masks. Yet it is all done in the digital image processing domain and deep learning. That is why it is called multidisciplinary. A disciplinary field allows the computer to extract a useful amount of data from all sources of digital images, videos, and other visual inputs. In computer vision, object detection and classification are essential for comprehending the scene and context of the digital image. Computer vision algorithms may identify the object's face mask. Face mask detection in congested environments is crucial, and computer vision paves the way for automatic mask recognition (Gurcan et al., 2009).

## Problem statement

classifying accurately the masked faces non-masked faces and semi-masking human faces is the difficult and hard task for the digital image processing related algorithms the characteristics needed to correctly forecast the corresponding class and then the classification of the individual is limited from naked human face to only the eyes or maybe sometimes a forehead of a human. This study depends on the existing pre-train ResNet-50 framework, which is trained on human masked faces, non-masked faces, and semi-mashed faces to deal with the problem of classifying the person’s class when he or she is wearing the facial mask or not.

# Related works

Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N. Taha, and Nour Eldeen M. Khalifa (Loey et al., 2021) proposed a face mask classification technique that was composed of two components. One component used a ResNet50 to extract the facial features from the images, and the second component included SVM (Support Vector Machines), ensemble techniques, and tree classifiers to classify the scenario. They displayed a testing score of 99.6 and 99.4 on two multiple datasets; the datasets were the Real-World Masked Face Dataset (RMFD) and the Simulated-World Masked Face Dataset (SMFD), respectively.

Mental et al. (Meenpal et al., 2019) established the technique for producing precise face embeddings in the form of vectorized features of any picture of input dimension based on an RGB formatted structure. The approach extracted the function using VGG with Pre-defined weights. VGG-16 to be more specific. To differentiate the faces displayed in the picture semi-semantically, full CNNs are trained. Gradient Descent was employed for the optimization, and Binomial Cross Entropy was selected as the loss function. Experimenting using segmental face masks on the "Multi Parsing Humans Dataset" yielded an accuracy of 94.8 percent.

(Lil et al., 2020) describe a novel, real three-dimensional facial mask assault identification method based on VRA (visual refractive analysis). In light of the suggested method, the facial image was initially evaluated using an intrinsic image decomposition algorithm to determine the reflectance of the image. The pixel intensity histograms are then created from three perpendicular planes to highlight the variations in reflectance intensity pictures for comparison between the real face and the 3D face mask. Following that, given that the reflectance image of a seamless surface was more vulnerable to changes in light, the one-dimensional CNN (convolutional neural network) was utilized to characterize the distinctness of the components that respond individually to highlight fluctuations.

Nieto-Rodríguez A., Mucientes M., and Brea V.M. (Nieto et al., 2015) proposed a method for detecting the presence or absence of a necessary operating room health mask. The eventual aim is to reduce the number of false positive face detection systems while avoiding missing mask detections to trigger alarms only for medical personnel who don't wear a protective mask. The suggested technique was able to achieve an accuracy of 95%.

(Grassi et al., 2007) presented the idea of doing some minor preparation of a somewhat smooth picture using a successively darkened elliptical mask centered on the supports of the ears. When utilized in conjunction with DCT for the sake of extracting functions, RBF Neural Networks, and MPL during classification, it allows the machine's results to be increased unaffected by total compute intensity and reduces the training period of MLP.

In (Wang et al., 2020), the authors proposed a solution to simplify the procedure of recognizing people who had not covered their faces with a mask. A modified model is to be developed. This modification was done by changing the ending layers, and hence it is also called transfer learning. The suggested system was created by fine-tuning InceptionV3, which is a very common deep learning architecture. SMFD was used as the dataset for the learning of the suggested system.

The image augmentation approach was used to overcome most possible scenarios. The main idea of this augmentation was to improve model learning capability. This system achieved an accuracy of 99.9%.

(Nagrath et al., 2021) proposed a system that was deployable on smart AI-based hardware electronic components just like jetsons nano. For the sake of system’s real-time efficiency, the architecture was modified, including the s and, therefore, components. One component contained an SSD (Single Shot Detection) with multi-box detections, and the second component was a MobileNetv2 model. The first part is for detection and the second is for classification. The base concept they used was dealing with facial landmarks to detect face masks as the hidden portion of landmarks.

(Maharani et al., 2020) proposed a system made via the Python Deep Learning Toolkit. The main architecture used by them was MobileNetv2, made by CNN. It was not used as is, but it was fine-tuned with the present weights to make the model simpler and the learning procedure fast. The accuracy obtained was around 97%. They trained the system using a local dataset in India.

(Inamdar et al., 2020) also proposed the modern transfer learning based method that was fine-tuned a VGG16 and used a triple loss FaceNet, whose first publication was made by Google in 2015. It was a training technique based on anchor, positive, and negative data. The picture from the data repository will be chosen at random as the image that will play the role of an anchor, where the positive reference will be the image of the person who was selected as the anchor, and for the negative reference, a different image will be chosen. After this, FaceNet is responsible for adjusting the distance between positive and negative data. This adjustment is made by changing the parameters of the FaceNet. The final shortest distance finding method was done via centroid tracking, which is also done here (Ambata et al., 2019). It achieved up to 82.2% training accuracy and 78% testing accuracy, whereas the fine-tuned VGG16 showed a training accuracy of 100%, so the testing accuracy was also 100%.

(Srinivasan et al., 2020) present complete and efficient method for performing face identification and face masking categorization utilizing detection and recognition, and the binary classification depends on the CNN. Using security footage datasets, YOLOv3 was used for the detection of the person in the crowd, and MobileNetV2-based binary classifiers were used. This report also includes a comparison of several facial identification methods as well as facial mask classifications methods. The purpose of this system was to group the people twhowere not wearing masks in public and warn the authorities about the violation of social distancing during the crucial time of COVID-19. Yolov3 is already well known for detection. The accuracy that was achieved was around 91.2%. The architecture was made while keeping in mind that it had to be in real-time. With Yolo-v3 being light, it took 7.12 seconds to process 78 frames. For the classification part, several experiments were conducted as MobileNetv2 outperformed Xception and ResNet50. Both the other models displayed overall accuracy of 50% and 48.7%, respectively.

(Munjal et al., 2021) proposed A genuine tool for detecting people wearing or not wearing face masks is offered. The overall accuracy of around 98% was attained after training with a dataset of roughly 4000 photos. The system was trained and combined with multiple CNNs in the experiment to determine the optimal benchmarking for the best score for this sort of model. This may be deployed in publicly accessible venues such as airports, railroads, campuses, and institutions, among others.

(Sethi et al., 2021) have developed a unique technique for combining one level with the approach of level detection to achieve a short judgment time while maintaining good precision. It started by using ResNet50 as a foundation, which utilized the core principle of combining better domain knowledge into multiple ideas. The test was carried out on three popular core models: ResNet50, AlexNet, and MobileNet. It should be highlighted that when combined with ResNet50, the suggested method achieves an accuracy of around 98.2%. Similarly, the suggested model achieves a high accuracy of 11.07% and 6.44% in mask recovery when compared to the contemporary public and the model released as Retina-Face-Masks-Detector. The suggested systems’ superb efficiency is particularly useful in the case of surveillance pieces of equipment.

In (Meenpal et al., 2019), the author's goal is to create a 2D face portion that could discover or identify any face in the image, regardless of how consistent it is. which shows how it creates precise facial masks for separating any form of a picture that is not sized correctly. To identify abilities from RGB images of any size, the approach employs certain pre-defined training methods using the weights of the VGG-16 design. The testing accuracy obtained was 93.8%. In (Song et al., 2021), they adopt a flow where, first, there is data augmentation, which employs limited sample traffic lights in datasets to solve the problem of extremely maldistributed diverse samples. Second, the Yolov3 architecture makes it desirable to get a bit more specific. Yolov3's thorough techniques are used. Finally, their technique displayed superior detection skills in comparison to other methods, with a precision of 91%, a mean average precision of 84.76%, and a recall of 90%.

(Sanjaya et al., 2020) presents a face mask classification, which governments may utilize to mitigate, evaluate, avoid, and prepare measures regarding COVID-19. In this work, face mask identification is achieved using a machine learning approach and the picture classification method MobileNetV2. Preparation of the data includes the process of collecting and applying some preemptive measures to make the data good enough for the training. Splitting the data, evaluating the system, and implementing the model are the phases of developing the model. The constructed model can recognize people wearing and not wearing a face mask with an accuracy score of 96.85%. Following the implementation of the system in 25 cities using multiple picture sources, the proportion of people using face masks in the regions has a high association with the alertness index. They (Wang et al., 2020) used RWDF (the Real-World Masked Face dataset) as their training data.

## Computer vision-based classification

Among the more productive fields of machine learning-based research is computer vision. Within the last 20 years, significant advancements in concerning areas have been made. The model to learn deep neural networks to an appropriate or required depth was put forth by (Hinton et al., 2006). When ImageNet became available, the breakthrough was made in 2012 (Krizhevsky et al., 2020), when the 8-layer artificial neural network was developed and surpassed every previous method for identifying photos in an ImageNet. In the architecture, researchers used convolution layers with corrected linear activation functions and a dropout. They find a path for the reason more effective computer vision models. Computer vision includes facial recognition as a subset.

Facial recognition exhibits great accuracy in identifying human faces when symmetric computer vision algorithms are applied. This is crucial because it allows for authentication via the identification of people (Huang et al., 2008). In this chapter, we concentrate on discussing how facial recognition-based research has advanced thanks to modern deep learning-based approaches. FaceNet (Deng et al., 2009) 's overall accuracy on the LFW dataset in 2015 was 99.63%, utilizing GoogleNet-24 architecture. Working with VGGFace, (Parkhi et al., 2022) obtained an overall 98.95% accuracy on LFW data.  VGGNet- 6 frameworks are employed throughout this instance. Following Microsoft's work, which successfully used the same architecture for digital imaging classification in 2015, many prevalent algorithms now rely on the ResNet as well as its derivatives.

The face recognition challenge has a variation that involves recognizing facial images with some occlusion. When individuals wear caps, sunglasses, face masks, and some other things that can cover part of the face while keeping other parts of that image intact, normal facial recognition techniques become limited (Hariri et al., 2022). For example, conventional systems like ResNet, VGG-16, or VGGFace have learned to distinguish between characteristics in full-facial photos. Unfortunately, extracting characteristics from obstructed or covered faces and complete facial images does not provide the required information about face identification. Some of these issues are connected to our proposed work. In this work, we concentrated on finding a solution to the issue of face identification for masked individuals. We came to this conclusion because the eyes, as well as the nose, were crucial elements for recognition; it becomes difficult to identify the correct people when they are not wearing the mask in training images.

The detection of faces with concealment has been the subject of numerous studies. The restorative technique, as well as the discard opacity dependant strategy, are both the main strategies employed in this context (Cao et al., 2018). Depending on the image used for training, restoration techniques attempt to recover portions of images that were obscured. The 3-dimensional facial recognition method is used by Bachi et al. (Bagchi et al., 2014) to restore obscured facial features. The Due To Initial Position Procedure is used by the system to record the 3-dimensional input of the person's front face. Secondly, obstructed portions of the face were found, and then a restorative procedure was employed, employing a principal component analysis (PCA). Comparable to this, (Drira et al., 2013) recognize and estimate obscured regions of the face using a 3-dimensional based statistical technique. The principle component analysis method is employed in the recovery of obstructed facial regions.

To prevent a poor reconstructing procedure, the discarding obstruction-based technique fully excludes using occluded regions. The feature extraction, as well as classification operations, utilized the remaining facial part. The facial digital image was segmented into tiny local areas by (Priya et al., 2014). Then, obstructed facial features were found using a support vector machine, and then they were eliminated. The leftover incomplete facial is then identified using a mean-based set of the weight matrix. Additionally, (Weng et al., 2016) removed obscured facial characteristics, identified regions of interest, as well as extracted characteristics from such feature points. Additionally, they employed the matching technique to compare their retrieved data to images in the gallery by calculating the distance between the characteristics of both faces.

Numerous academics have concentrated considerable research work on the topic of face mask use throughout the coronavirus epidemic (Duan et al., 2018). This differs beyond this work, though, because the issue of whether or not a mask is worn is the mask recognition challenge, whereas this is only the facial detection difficulty with an additional restriction that the mask is covered. In this regard, we are aware that there are significantly fewer studies investigating how to employ the most cutting-edge face-based recognition algorithms to identify the person wearing the facial mask. Since fewer traits may be used to identify someone who is wearing a facial mask, this is a considerably difficult problem.

That using the quantization-based pooled strategy on a VGG-16 which is a pretrained algorithm and merely taking into account the characteristics maps just at the final convolution layer focused on detecting facial characteristics on face mask by cutting eyes as well as forehead from the image. (Geng et al., 2018) employed two techniques to recognize human faces behind masks. To start, a further training dataset was produced using simple faces simulated as covered human faces with masks. They then employed disguised photos of every identification and Dimension Constricted Ranks (DCR) loss, which divides the complete digital image into two centers. The DCR next makes sure which features on the masked human faces are comparable to those on the whole human face. The challenge of incomplete facial recognition can also be used to mask facial recognition.

The alignment-free facial representation technique dependent on Multi-Keypoint Descriptor (MKD) was created by (Liao et al., 2012), in which an image's real information determines the face's local variable dimensional description. utilized the quick filtering technique after that for face recognition. Radial basis function (RBF) structures were suggested by (Mandal et al., 2021) for 100 people's complete facial photos. To solve challenges using incomplete facial images, irrespective of dimension (He et al., 2018), the dynamic features match (DFM) approach with slipping costs was suggested. This work is nearly linked. ResNet is used in place of VGG16 due to its superior effectiveness as well as its strength and robustness to alter illumination, facial expressions, and obstruction. The goal of the work is to solve the issue of mask face identification during the COVID-19 epidemic by demonstrating a solution that incorporates a mask as the second characteristic in addition to the entire face. The essay is broken down into other chapters.

## Databases for facial masks

The computer vision community was informed of the latest research projects about the creation of databases including pictures of facial masks as soon as the epidemic spread over the world. Nevertheless, the identification of facial mask classes and the creation of the related datasets were hot research areas long before the epidemic. Additionally, several articles focused particularly on the faces that were obscured. for example, detection of any facial mask, non-facial mask, or semi-mask.

### Available datasets for faces

These specialized databases are employed to train algorithms that can find human faces in still photos and video frames.  The largest available dataset geared for computer vision tasks on human facial images is WiderFace (Yang et al., 2016). This dataset, which has 32,213 photos and 393,713 facial labels, was released in 2016. It incorporates pictures from a huge Wider collection. The Facial Detection Dataset, as well as the Benchmark (FDDB) (Jain et al., 2010), are also significant data resources. The 2010 release of FDDB included annotations from 5173 faces in 2843 photos. The data collection was produced using actual situations, like features that were obscured. The first data collection project on the subject was FDDB. 468 human faces can be found in 205 photos that make up labeled Faces in Wild (AFW). It was made public in early 2012 as a result of research into stance and facial estimation (Gong et al., 2012).

### Available datasets for masked and non-masked faces

A significant problem with facial detection for obscured human faces and image analytics has already been solved by scientists (jain et al., 2011). Well before the epidemic, computer vision-related research personnel had been unsuccessfully attempting to produce the standard data collection focused on carrying out the current tasks for years. Nevertheless, one important data collection, known as the MAFA, was released in early 2017 and contains 35,805 masked facial annotations on 30,810 photographs that were retrieved from the Internet (Ge et al., 2017). A minimum of one human face on every photograph is hidden by the mask. The COVID-19 outbreak necessitates paying special heed to developing the latest datasets that depict realistic or hypothetical contemporary COVID-19 facial mask circumstances. In this regard, MaskedFace-Net was introduced in 2021 to address the issue of a scarcity of substantial masked datasets (Cabani et al., 2021). 137,116 customizable images are included. Specifically, "properly face masked data collection" as well as "an inaccurately face masked dataset" are two kinds of face masks put on features using picture editing software. The facial mask detection data collection is a Real-World facial mask dataset and a simulated facial mask recognition data collection (SMFRD), with 24,772, 94,000, and 600,000 digital images, respectively, simply presented in (Wang et al., 2020). The online available images have been scraped for MFDD, and only if subjects are hiding behind face masks are tags included. The writer claims that RMFRD is the largest data collection of real-world facial masks in existence. The collection consists of 6,000 photos, 526 of which show people wearing masks, and 80,000 of them show actual individuals. The data collection is accessible for free on GitHub. The collection does, though, contain some incorrectly mask-wearing faces.

### Facial Mask Classification During Coronavirus

Since the start of the COVID-19 epidemic, several deep learning models have been made available with various strategies for addressing issues of facial mask detection and classification. For instance, the authors of (Liu et al., 2019) provide a hybrid approach that combines deep learning and traditional machine learning. Decision trees as well as support vector machines are employed to categorize placements of masks in the first section, which utilizes the ResNet-50 framework as a characteristics extractor. According to (Addagarla et al., 2019), the authors describe the strategy that utilized super-resolution pictures with classification systems like SRCNet to recognize scenarios where people were wearing facial masks. Additionally, it has lately been suggested to employ YOLOv2 for the detection and ResNet-50 for the feature extractor (Worby et al., 2020). The researchers of (Krizhevsky et al., 2017) suggested the so-called FM-Yolo methodology to determine whether individuals are appropriately wearing face masks in public settings. For the phases of characteristics extractor and features fusion, this system uses Im-Res2Net-1o1 as well as path aggregation via the algorithm En-PAN, correspondingly. The researchers use different publicly accessible databases to compare their methods to several cutting-edge detection techniques. On these two databases, their conclusions fared better than the approaches they were matched to. Another detecting technique that makes use of the novel architecture for recognizing masking faces is the Context-Attention R-CNN (Deng et al., 2009). The approach is founded on several contextual characteristics extractor, dissociation branching, and attentiveness elements. A database of the 8635 human faces was also produced by researchers throughout various experimental circumstances. The model-based mean average precision (mAP), which was 6.9% better than Faster R-mAP, of the convolutional neural network was 84.2% for the dataset in question. The detecting model, Facial Mask Recognizing Networks (FMRN), which utilizes the algorithm for the classification of images by pose recognition, is the last one. The photos are then processed by the network following Class (Bastys et al., 2009).

## Tools and Procedures

Recently, we have seen a considerable increase in the performance of digital image processing applications depending on AI and deep learning-based. Because convolutional neural networks, using various deep learning-based techniques, can quickly discern minute characteristics that a human eye cannot even see in image recognition implementations, they are currently strongly desirable in various applications across a variety of disciplines. A very crucial aspect of convolutional neural networks is that they can recognize specific patterns immediately from the image pixels without extreme preprocessing (Aslan et al., 2021). Within the parameters of the investigation, convolutional neural networks (AlexNet, as well as VGG16 frameworks) were utilized. Longest Short-Term Memory, as well as BiLSTM, were used to develop algorithms in combination with such architectures.  MATLAB applications, such as model training as well as testing operations, were carried out.

### Alex Net

In an ImageNet Large Scale Visually Recognized Challenge (ILSVRC) in 2012, they submitted AlexNet, the variation of conventional LeNet, which performed better. It makes use of a sizable ImageNet training dataset as well as GPUs, which offer around a 10-fold boost in computational capabilities. AlexNet has a Softmax outcome layer and three fully connected layers, including five convolution layers. To be able to straighten a rectified linear unit (ReLU), an activation function comes after every convolutional layer (Konstantin et al., 2019). To cut down on model size, every convolution layer has the maximum pooling setting. There are two fully connected layers having 4096 outcomes following the convolution layer at the network's conclusion. To identify the input dataset, there is one more fully linked layer left. A Softmax function is used in the final layer to classify 1000 items (Abd et al., 2018).

### VGG-16

In an ILSVRC 2014 competition, a VGG was represented by the Simonyan with Zisserman (Simonyan et al., 2014), which finished in second place. Every of the VGG's convolution layers' smaller (3 3) filtering performs better. This is because using a large number of modest filters in succession can duplicate the overall effects of larger sizes. Little filters are easy to use throughout the system, which produces outstanding generalization performance. Due to its clarity as well as its great generalization performance, a VGG is also still extensively accessible in a variety of production steps. One of the most widely used variants, VGG16, has the Softmax outcome layer in addition to twelve convolution layers, three fully linked layers, and other layers. The ReLU as an activation function comes after every convolution layer. Maximum pooling is used in every convolutional layer to cut down on particle sizes. There are 2 fully linked layers containing 4096 outcomes following the convolution layer just at the network's conclusion. To identify all input datasets, another completely linked layer is left. The Softmax function is used in the final layer to classify 1000 items (Jain et al., 2019).

### Longest short-term memory (LSTM)

In artificial intelligence, we have the longest short-term memory, which is a different topology from a conventional neural network. Recurrent Neural Networks (RNN) are a specific subclass of the neural network that may train over time how different pieces of input are related. To solve the long-term reliance issue with repeated neural networks, the longest short-term memories were created. RNNs have the shape of a series of repetitive neural network modules. Similar to Tanh as an activation layer, the module, which is repetitive in an ordinary, recurrent neural network, has a straightforward architecture. LSTMs, it is a dual-category module with different architectures. Long-term storage modules (LSTMs) carry out this function by learning. When long-term connections are necessary, the LSTM framework uses hidden units, often known as memory cells. Such memory units are used to retain the inputs, which must be maintained over a certain time. Through doors it contains in its construction, it determines whether this knowledge seems important (Yang et al., 2019).

### Bidirectional LSTM

BiLSTM is the modified version of a bidirectional recurrent neural network using the longest short-term memory framework. Two separate recurrent neural network layers make up the bi-directional recurrent neural network, which processes input data in opposite directions (Bengio et al., 1994). Inputs are represented by X, outcomes are represented by Y, condition duration is represented by t, and a hidden state is represented by h. Additionally, t+1 denotes the following, and t-1 denotes the preceding. The vanishing gradient’s problem, which affects recurrent neural networks since they have a basic neural network topology, frequently results in subpar performance (Hochreiter et al., 1996). Long short-term memory, the unique recurrent neural network cell, is suggested as a solution to this issue (Graves et al., 2005). The one-way forwarded Long short-term memory's hidden state solely preserves past characteristics; it makes no allowances for a future. The Bi-directional recurrent neural networks and long short-term memory work together to efficiently utilize both past and prospective characteristics. The algorithm has two concurrent layers in both transmission orientations, unlike long short-term memory networks. The backward long short-term memory layer captures prospective features, whereas the forward long short-term memory layer extracts past characteristics (Jia et al., 2019). The fundamental concept behind BILSTMs is to offer both discrete hidden layers, one for the past and one for the future, to record prior and prospective knowledge correspondingly. The result is then produced by combining both hidden states (Schmidhuber et al., 2005).

# Related Knowledge

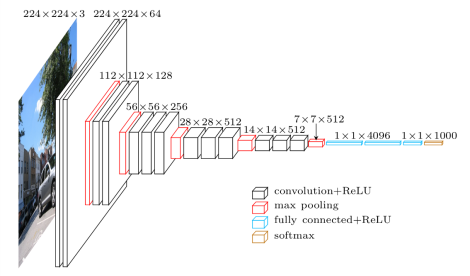
## Convolutional Neural Network

Deep Learning has established itself as a very potent tool over the last few decades due to its capacity for handling massive amounts of data. Hidden layer technology is much more popular than conventional methods, particularly for pattern recognition. Convolutional Neural Networks are among the most widely used deep neural networks. Convolutional neural networks are a type of artificial neural network used most frequently to interpret visual data in deep learning. Normally, matrix multiplications come to mind when we think about neural networks, but that is not the case with Convolutional network. It makes use of a unique method called convolution. Convolution is a mathematical procedure that takes two functions and creates a third function that describes how the appearance of one is changed by the second in mathematics.

Artificial neurons are arranged in numerous layers to form convolutional neural networks. Artificial neurons are based on mathematical functions that compute the weighted sum of several inputs and generate an output value, roughly imitating their biological counterparts. Each layer of a Convolutional network creates a number of activation functions that are forwarded to the subsequent layer when an image is received.  Typically, the very first layer extracts fundamental features like edges that run horizontally or diagonally. The following layer receives this output and detects more intricate features like corners or multiple edges. The network may recognize increasingly more complex elements, including objects, faces, etc., as we go further into it. The classification layer generates a series of confidence ratings (numbers between 0 and 1) that indicate how likely it is for the picture to be a member of a "class," depending on the feature maps of the final output layer. The output of the last layer, for instance, can be the probability that the input image contains any of the class like cats, dogs, or horses detected by the Convolutional network.

## VGG16

In order to win the 2014 ILSVR(ImageNet) competition, the convolution neural net (CNN) architecture VGG16 was used. One of the best vision model architectures to date is thought to be this one. The most distinctive feature of VGG16 is that, rather than focusing on having a lot of hyper-parameters, they concentrated on having convolution layers of 3-by-3 filter with a stride 1 and maintained the same padding and MaxPooling layer of 2x2 filtering with stride 2. The convolution and maximum pool layer arrangements are maintained throughout the whole architecture. The output is provided by a SoftMax after two fully connected layers (FC). VGG16 comprises 16 layers, each of which has a weight, as indicated by the number 16. The architecture of the VGG16 model is shown in below Figure 1.



**Figure 1:** Architecture of VGG-16 model.

VGG16 received a received a 3-dimensional image of the shape of 224 x 224 on the input layer. After the input layer, the first two layers are based on the 64 channels of filter size and padding size. Following, a MaxPooling of 2 x 2 followed by the 2 convolutional layers of 128 channels with 3 x3 filter size. The convolutional layer is again followed by the same MaxPooling layer of 2 x 2 stride. After this, the 2 sets of 3 convolutional layers and MaxPooling layers are added in the network. Each layer is base on the 256 filters of 3-by-3 filter and same padding size. We obtained a (7, 7, 512) feature map after adding a convolutional and max-pooling layer to the stack. This result is flattened to create a (1, 25088) feature vector. There are then 3 fully connected layers; the first layer uses the most recent feature vector as input and produces a vector of shape (1, 4096); the second layer also produces a vector of same size (1, 4096); however, the third layer produces a vector of size (1, 1000), which is used to implement the SoftMax activation function to classify 1000 categories. ReLU is used by every hidden layer as its activation function. As ReLU promotes quicker learning and lessens the chance of vanishing gradient issues, it becomes computationally more efficient.

## VGG19

Over a million photos from the ImageNet dataset were used to train the deep neural network VGG-19. The network, which has 19 layers, can classify photos into 1000 different object classes, including a laptop, mouse, pen, and numerous animals. This has led to the network learning detailed feature representations for a variety of images. Advanced CNN VGG19 has layers that have already been trained and has a strong comprehension of the characteristics of a picture in terms of shape, color, and structure. The extremely deep VGG19 has been trained on a massive variety of images for challenging classification tasks. In order to complete my classification objective of differentiating between photos with and without mask, we didn't train the VGG19 model any further; instead, we simply added a shallow 2 dense layer network on top of it and freezing the weights of VGG19 network. The architecture of the VGG-19 model is following:

* This network received a fixed-size (224 \* 224) RGB picture as input, indicating that the grid was shaped (224,224,3).
* The mean RGB value of each pixel, calculated throughout the whole training set, was the only preprocessing that was carried out.
* They were able to cover the entirety of the image by using kernels that were (3 \* 3) in size with a stride size of 1 pixel.
* To maintain the image's pixel density, spatial padding was applied.
* Stride 2 was used to conduct max pooling over a 2 \* 2-pixel window.
* Rectified linear unit (ReLU) was used after this to add non-linearity to the model in order to enhance classification accuracy and computation time. As opposed to earlier models that used tanh or sigmoid functions, this one performed far better.
* Applied three fully linked layers, the first 2 of which had a size of 4096, followed by a layer with 1000 channels for classification using the 1000-way ILSVRC, and the third layer being a SoftMax activation function.

## Artificial neural network (ANN)

Artificial neural network learning has been successfully used to train real-values, discrete-values, as well as vector-values functions addressing issues like analyzing visual-based situations, voice recognition, and learning robotic control techniques. Artificial neural network learning is impervious to inaccuracies in training-data. The discovery that biologically based learning-systems are composed of extremely intricate webs of interconnected neurons in brains served as inspiration for studying the ANNs. Every connection of a neuron in the human brain is typically linked strongly to l04-105 additional synapses, forming a tightly interconnected system of about 1011–1012 synapses. Therefore, the human brain requires about 10 to 1 second on average to generate remarkably sophisticated judgments. Artificial neural networks  are driven to incorporate this form of  distributed representation-based, massively parallelized processing. Artificial neural networks are typically constructed from a densely linked system of simple units, each of which accepts multiple real-value inputs and outputs a single real-value output. However, Artificial neural networks are somewhat driven by the biological based neural network systems because these networks are much more complicated than what is represented by Artificial neural networks.

## Inception version2

Convolutional layers like 5\*5, which significantly reduce input size, were occasionally used in Inception Version 1. A neural network's performance suffers as a result of this. This is because if an input's dimensionality is reduced too much, overall neural networks are vulnerable to information loss. Additionally, complexity is reduced while using larger convolutional layers, such as 55, as opposed to 33. In consideration of the factorization, we can go even more by splitting the 33 convolutional layers into such an asymmetrical 13 as well as the 31 convolutional layers. This is the same as moving the two-layer infrastructure with the 33 convolutional receptive fields, but 33% less expensive. Whenever an input size is large for every layer, linear decomposition performs best, but only when an input dimension is mxm, where m is just between 12 and 20. The supplementary classification enhances a network's convergence network, per the Inception Version 1 framework. By extending beneficial gradients to previous layers of the network, researchers contend that it may aid in reducing the impact of the vanishing gradient problem in deep neural networks (to minimize the overall loss). However, researchers observed that this classification didn't significantly enhance convergence before training. in this architecture of Inception Version 2. The twin 33 convolutional layers take the place of 55 convolutions. Due to the fact that the 55 convolutions are 2.78 times more expensive than the 33 convolutions, it also reduces calculation time and hence boosts computational speed. Therefore, employing twin 3x3 convolutional layers rather than five 5x5 layers improves architectural performance. Also transformed the 1xn a  aswells nx1 singular value decomposition by this framework are nXn factorizations. As was mentioned previously, the 33 convolutions can be reduced to the 13 convolutions that are followed by the 31 convolutions, which have a 33% lower complexity of computation than the 33 convolutions. Rather than expanding a module,  the characteristic banks of a module were increased to address an issue of representational constraint.

## Resnet 50

The residual network is referred to as ResNet. In 2015, a digital image processing research article titled "Deep Residual Learning for Image Recognition" originally conceived this novel artificial neural network  for additional information. There are other versions of the ResNet that use the same basic idea but have various numbers of network layers. The form that can operate using 50 neural networks' layers is known as Resnet50. A ResNet50 framework depends on the concepts mentioned previously; however, there is one significant distinction. Due to concerns about the length of time required to train network layers, the building elements in this instance were changed into the bottleneck architecture. Rather  then preceding 2 networks layers, this utilised the stack of 3. In order to create a Resnet 50 design, every one of the 2-layer units in the Resnet 34 was changed to the 3-layer constraint unit. Compared to a 34-layers ResNet algorithm, it has substantially high accuracy. The throughput of a 50-layer ResNet is 3.8 billion FLOPS.

## Decision tree

Every internal node of the decision-tree indicates the test on the feature (such as whether the coin will land on its top or bottom), every leaf connection representing the class tag (decision made following calculating relevant characteristics), while branching indicate connectives of characteristics which result in any of these class tags. Standard approach are represented by a routes between roots to branch. The foundation decision-making loop is shown in the picture below with labels "Raining (Yes), No Raining (No)". The algorithmic method for dividing the data collection into segments depends on several criteria used to build the decision-trees. It is among the most popular and useful techniques for supervised, supervised-learning. The non-parametric and supervised learning technique called "decision decision-trees is utilized for both classification and regression applications.

Classification based trees are the tree algorithms where a goal parameter can assume the discrete list of the values. Regression trees are the decision trees when a targeted variable can assume attribute variables (usually rational numbers). This is referred to as a classification or regression tree (CART).

## Support Vector Machines

The supervised machine learning based model called the support vector machine (SVM) employs categorization techniques to solve the two-group classification tasks. After delivering a Support Vector Machine model collections of labelled training set for every category, those who capable to categorize fresh text.

They offer two key benefits over more recent techniques like neural networks: greater speed as well as improved accuracy with fewer samples (in the thousands). As a result, this approach is excellent for textual classification issues, where it is typical to only have exposure to the dataset with a few hundred tags on each example.

## Ensemble

Ensemble technique in the statistical as well as machine learning-based combine several training algorithms to achieve higher predicted performances than the individual learning model could. The machine learning based ensemble, while normally indeed for considerably greater flexibility to the exist amongst possibilities, only contains the list finite collection of different models, apart from the statistical ensemble in the statistical mechanics, that is generally unlimited.

# Methodology

The methodology section will discuss the dataset, preprocessing of the dataset, techniques and models used for the classification of face mask.

## Dataset

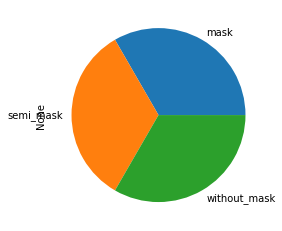
For the classification of face mask images, we proposed the three types of faces. (With mask, without mask, semi mask). Unfortunately, there is no public ally available dataset based on these types of images. Here, we used the two different datasets from Kaggle for the classification of face masks. The first dataset of the face mask images totally contains the 7553 images from two categories (with mask and without mask). The dataset contains the 3725 sample images with face mask and 3828 sample images without mask. From the second dataset of face mask images, we extract the 898 samples images with semi mask. Semi mask images represent the faces that are wearing mask but mask is slip down from nose. The complete statistics of the compiled dataset is available in Table 2.

**Table 1:** Initial Statistics of Compiled Dataset.

|  |  |  |
| --- | --- | --- |
| Face Mask Type | Number of Samples | From |
| With Mask | 3725 | Dataset 1 |
| Without Mask | 3828 | Dataset 1 |
| Sami Mask | 898 | Dataset 2 |

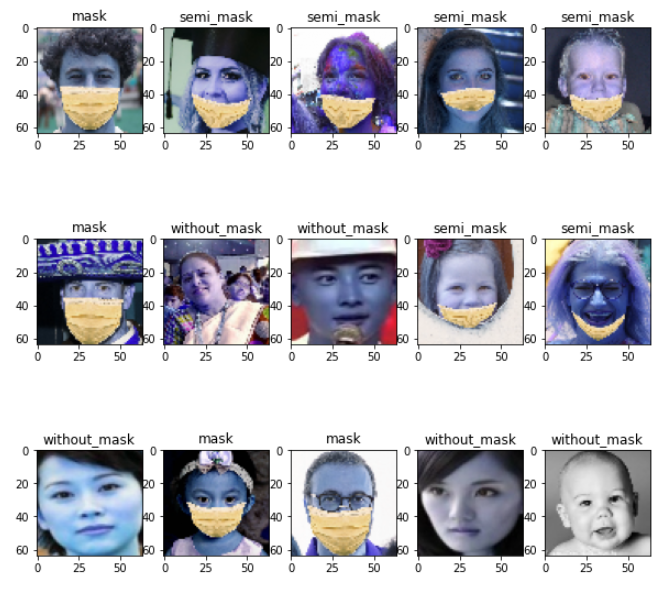
## Data Preprocessing Exploratory Analysis

In the preprocessing of the dataset, we remove the noisy samples. We also keep the images that have the faces only and manually removed the images of the persons with full body. As according to the initial statistics of the dataset, the compiled dataset is very class unbalanced dataset. The dataset was converted into the class balanced dataset by down sampling technique. Down Sampling technique select the samples from each category equal to the count of minimum samples category. Now we have the semi mask that have the lowest count with 989 samples. Down sampling technique will randomly select the 898 samples for each category. We used built-in resample function of scikit-learn library for down sampling. After down sampling our dataset, we got the 898 samples for each category (With mask, Without Mask, Sami Mask). The class distribution of the dataset is also plotted in Figure 2.



**Figure 2:** Class Distribution of Dataset.

We also plot the images with annotation label for displaying purpose. The samples of the dataset with class label are presented in Figure 3.



**Figure 3:** Samples Representation of Dataset

## Train Test Split

The dataset needs to be divided into multiple subsets for the training and testing of the model just after preprocessing. The scikit-learn library's developed train-test-split method was used to partition the collection into three portions. With some ratio for one group and the remaining samples for the second subset, the built-in train test split function chooses a random sample from each class. The observations in the partitioned subset were not overridden by it. We divided the face mask samples into two separate subsets using the reprocessed data. The processed dataset was dividing into training and testing set with the ratio of 80% and 20%. After splitting the dataset, the train and test set contain the 2694 and 90 samples.

## Model Development and Training

The model development section is divided into four section that will discuss the implementation of customized models, implementation of paper 1, paper 2 and paper 3 respectively. Each implementation is further based on the development of several deep learning models. We will discuss the implementation of each model and training of each model in the below subsections.

### Transfer Learning Models

We used different deep learning models with transfer learning approach. Firstly, we used the VGG-16, VGG-19 and Inception ResNet V2 model with the transfer learning technique. For VGG-16, VGG-19 and Inception ResNet V2 model, drop the output layer and extract the features from last hidden layer. After getting the features from last layer, we flatten the weights by using the Flatten layer. Three dense layers were added with the shape of 100, 10 and 3 neurons respectively followed by the flatten layer for each model. The last dense layer of shape 3 was used as the output layer. The ReLU activation function was used on the hidden layer while the SoftMax activation function was used for the output layer of each model.

We also train the convolutional neural network (CNN), artificial neural network (ANN) and CNN+MLP model for the classification of face masks. For CNN, we apply the two convolutional layer of kernel size 3 with 16 and 32 filters respectively followed by the MaxPooling layers with the kernel size of 2. After flatting the weights of convolutional layers one dense layer with 64 neurons and output layer with 3 neurons was added in our proposed CNN model. SoftMax as an activation function was also used for proposed CNN model. For KNN, we just add the input layer with 32 neurons and output layer with 3 neurons. The SoftMax was also used for the ANN model. For CNN+MLP model, we apply the two convolutional layer of kernel size 3 with 16 and 32 filters respectively followed by the MaxPooling layers with the kernel size of 2. After flatting the output of convolutional layers, we added the two dense layers with 8 and 4 neurons. Lastly, the output layer with the neuron equal to the class length was added in the model for classification.

Collectively, we used three deep learning model develop from scratch and 3 deep learning models including VGG-16, VGG-19 and Inception ResNet V2 with transfer learning technique. After the development of all the model, the model was compiled with Adam Optimizer and categorical cross entropy as loss function.

### Implementation of Study 1

Secondly, we implement the methodology proposed by the author of the study (Hariri et al., 2022). By following the methodology of the author, ResNet50 was used for classification of face masks with transfer learning approach. The pretrained ResNet50 was initialized with the training weights. The output layer of the trained RasNet50 model was drop and extract the features from the second last layer. BY the end of the feature extraction layer two dense layer with the shape of 10 and 3 was added in the model. The last layer was used for the classification of face mask images.

### Implementation of Study 2

The study proposed by the (Loey et al., 2021) was implemented to classify the face mask images. The proposed study implements the ResNet50 model for feature extraction purposes. We also implement the ResNet50 model with the pretrained weights and drop the output layer. The features of the face mask images were extracted from the last layer of pretrained ResNet50 model. After extracting the features of the samples from pretrained model, different machine learning models were used for the classification of face mask images according to study (Loey et al., 2021). Decision Tree Classifier, support vector machine and voting classifier from ensemble learning was used for the classification of face mask on extracted features set. All the default values of the hyper parameters were used for the training of the models.

### Implementation of Study 3

Author (Mandal et al., 2021) also proposed a study for the classification of face mask images. We also followed the methodology of the study and implemented it. By following the methodology of the study, once again we used the pretrained ResNet50 model. The published study performed the hyper parameter tuning for the significant results of the pretrained model. After initializing the ResNet50 model with pretrained model different combinations of the hyper parameter were used for the better assessment t of model. The details of the tunned hyper parameters and their different value are available in Table 2.

**Table 2:** Tuned Hyper parameters for ResNet50.

|  |  |
| --- | --- |
| Hyper parameters | Values Set |
| Batch Size | 256, 32 |
| Optimizer | Adam, SGD |
| Dropout | 0.5, 0.4, 0.3 |
| Learning Rate | 0.002, 0.0016, 0.001 |

## Model Evaluation

Researchers used a variety of evaluation metrics, such as accuracy, precision, recall, and f1-score, to assess the classification model. The proportion of actual positive instances that accurately foresee good outcomes to all truly positive situations is known as recall or sensitivity. Precision, on the other hand, or confidence, on the other hand, refers to the proportion of expected positive cases that are actually true positives. So, we can mention the recall means “how many samples of particular class you find over the all samples of that class," and the precision will be “how many are correctly classified among that class." The harmonic means of recall and precision is known as the f1-score. The trained classifiers were assessed using the test set, which contained 90 images. The formula for that metric was used to construct the evaluation metrics. Eq 1-4 gives the equation used to calculate the measurements.

Eq. 1

Eq. 2

Eq. 3

Eq. 4

# Results and Discussion

The result and discussion section of this report will present the evaluation scores of the trained models. Moreover, the comparative analysis of all the tainted models with the help of tables and graphical representation in the last of this section.

## Proposed Models Results

For the classification of face mask images, 3 deep learning models including CNN, ANN and CNN+MLP was trained and 3 pretrained model including the VGG-16, VGG-19 and Inception ResNet V2 with transfer learning approach was used. All the model were trained on the 10 epochs with Adam optimizer and categorical cross entropy as loss function. The training samples images (s=2694) were used for the training of the models. BY following the training of the models, the test sample were used for the evaluation of the model. We got the 0.94%, 0.33% and 0.87% accuracy score for CNN, ANN and CNN+MLP model. Rest of the evaluation measures ware also calculated using the test instances of the dataset. The complete classification report of the models is available in Table 3. The confusion metrices of CNN, ANN and CNN+MLP were also plotted for the better assessment of the models in Figure 4a, 4b, 4c respectively.

**Table 3:** Face Mask Classification - Classification Reports

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classification Report of CNN Model** | | | | | |
|  | precision | recall | f1-score | | support |
| 0 | 0.86 | 1.00 | 0.92 | | 30 |
| 1 | 1.00 | 0.83 | 0.91 | | 30 |
| 2 | 1.00 | 1.00 | 1.00 | | 30 |
|  |  |  |  | |  |
| accuracy |  |  | 0.94 | | 90 |
| macro avg | 0.95 | 0.94 | 0.94 | | 90 |
| weighted avg | 0.95 | 0.94 | 0.94 | | 90 |
| **Classification Report of ANN Model** | | | | | |
|  | precision | recall | f1-score | support | |
| 0 | 1 | 0.93 | 0.97 | 30 | |
| 1 | 0.93 | 0.93 | 0.93 | 30 | |
| 2 | 0.94 | 1 | 0.97 | 30 | |
|  |  |  |  |  | |
| accuracy |  |  | 0.96 | 5870 | |
| macro avg | 0.96 | 0.96 | 0.96 | 5870 | |
| weighted avg | 0.96 | 0.96 | 0.96 | 5870 | |
| **Classification Report of CNN+MLP Model** | | | | | |
|  | precision | recall | f1-score | support | |
| 0 | 1.00 | 0.80 | 0.89 | 30 | |
| 1 | 0.81 | 0.83 | 0.82 | 30 | |
| 2 | 0.86 | 1.00 | 0.92 | 30 | |
|  |  |  |  |  | |
| accuracy |  |  | 0.88 | 5870 | |
| macro avg | 0.89 | 0.88 | 0.88 | 5870 | |
| weighted avg | 0.89 | 0.88 | 0.88 | 5870 | |

|  |  |
| --- | --- |
| (a) | (b) |
| (c) | |

**Figure 4:** Face Mask Classification - Confusion Matrices

After the full training of the transfer learning-based models (VGG-16, VGG19 and Inception ResNet V2), we used the samples images from test set for the evaluation of the models. We got the 0.33% accuracy score VGG-16 and VGG-19 while the 0.99% for the Inception ResNet V2 model on the test samples. We also calculate the precision, recall and f1-score of all models for complete assessment of training process. The complete classification report of all the models is available in Table 4 and confusion matrices in Figure 5.

**Table 4:** Face Mask Classification - Classification Reports

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classification Report of VGG-16 Model** | | | | | |
|  | precision | recall | f1-score | | support |
| 0 | 0.33 | 1 | 0.5 | | 30 |
| 1 | 0 | 0 | 0 | | 30 |
| 2 | 0 | 0 | 0 | | 30 |
|  |  |  |  | |  |
| accuracy |  |  | 0.33 | | 90 |
| macro avg | 0.11 | 0.33 | 0.17 | | 90 |
| weighted avg | 0.11 | 0.33 | 0.17 | | 90 |
| **Classification Report of VGG-19 Model** | | | | | |
|  | precision | recall | f1-score | support | |
| 0 | 0 | 0 | 0 | 30 | |
| 1 | 0.33 | 1 | 0.5 | 30 | |
| 2 | 0 | 0 | 0 | 30 | |
|  |  |  |  |  | |
| accuracy |  |  | 0.33 | 90 | |
| macro avg | 0.11 | 0.33 | 0.17 | 90 | |
| weighted avg | 0.11 | 0.33 | 0.17 | 90 | |
| **Classification Report of Inception-ResNet V2 Model** | | | | | |
|  | precision | recall | f1-score | support | |
| 0 | 1.00 | 1.00 | 1.00 | 30 | |
| 1 | 1.00 | 1.00 | 1.00 | 30 | |
| 2 | 1.00 | 1.00 | 1.00 | 30 | |
|  |  |  |  |  | |
| accuracy |  |  | 1.00 | 5870 | |
| macro avg | 1.00 | 1.00 | 1.00 | 5870 | |
| weighted avg | 1.00 | 1.00 | 1.00 | 5870 | |

|  |  |
| --- | --- |
| VGG-16 | VGG-19 |
| Inception ResNet V2 | |

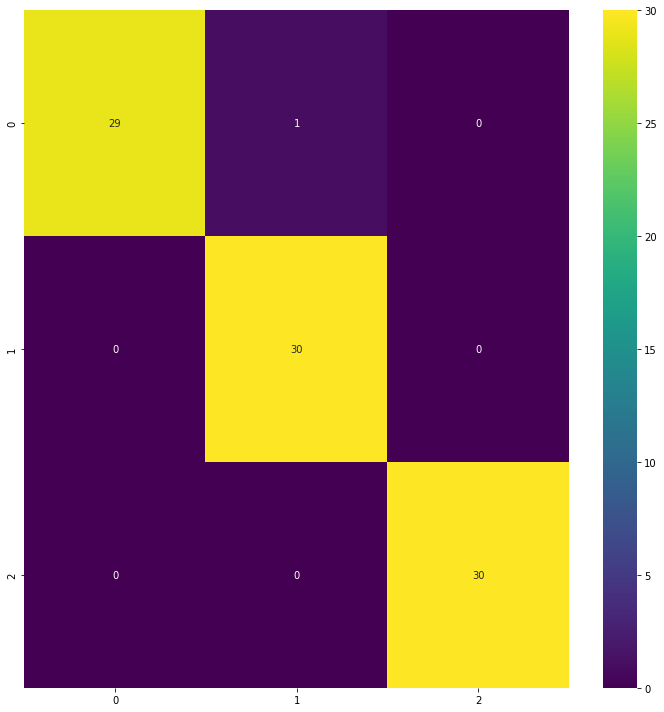
**Figure 5:** Face Mask Classification - Confusion Matrices

## Paper 1 Models Results

As paper 1 proposed the ResNet50 model with transfer learning technique, we also used the pretrained ResNet50 model for the classification of face mask images. After the Initialization of the model, all the training samples of the dataset was used to train the developed model. For the assessment of the trained ResNet50 model, test images were used and we got the 0.98% accuracy score. Other evaluation scores were also calculated using the samples of test set. The complete classification report of the ResNet50 model for face mask classification is presented in Table 5 and confusion matrix of the model in Figure 6.

**Table 5:** Face Mask Classification - ResNet50 Reports

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Classification Report of ResNet Model** | | | | |
|  | precision | recall | f1-score | support |
| 0 | 1.00 | 0.97 | 0.98 | 30 |
| 1 | 0.97 | 1.00 | 0.98 | 30 |
| 2 | 1.00 | 1.00 | 1.00 | 30 |
|  |  |  |  |  |
| accuracy |  |  | 0.99 | 90 |
| macro avg | 0.99 | 0.99 | 0.99 | 90 |
| weighted avg | 0.99 | 0.99 | 0.99 | 90 |



**Figure 6:** Face Mask Classification - ResNet50 Confusion Matrix

## Paper 2 Models Result

The proposed study used the different machine learning models with extracted features for face mask classification. In our study we also initialize the ResNet50 pretrained model with the training weights to extract the features. All the training samples were passed to the model and features were extracted from last layer against each sample. After the extraction of the samples, decision tree, SVM and voting classifiers were used for the classification of face masks. By following the training process of the models, we used the test set to calculate the evaluation measures. We got the 0.98%, 0.97%, and 0.98% accuracy score for DT, SVM and voting classifier respectively. The complete classification report of all the models is presented in Table 6 and confusion matrices in Figure 7.

**Table 6:** Face Mask Classification - Classification Reports

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Classification Report of DT Model** | | | | | |
|  | precision | recall | f1-score | | support |
| 0 | 1.00 | 0.97 | 0.98 | | 30 |
| 1 | 0.97 | 1.00 | 0.98 | | 30 |
| 2 | 1.00 | 1.00 | 1.00 | | 30 |
|  |  |  |  | |  |
| accuracy |  |  | 0.99 | | 90 |
| macro avg | 0.99 | 0.99 | 0.99 | | 90 |
| weighted avg | 0.99 | 0.99 | 0.99 | | 90 |
| **Classification Report of SVM Model** | | | | | |
|  | precision | recall | f1-score | support | |
| 0 | 1.00 | 0.93 | 0.97 | 30 | |
| 1 | 0.94 | 1.00 | 0.97 | 30 | |
| 2 | 1.00 | 1.00 | 1.00 | 30 | |
|  |  |  |  |  | |
| accuracy |  |  | 0.98 | 90 | |
| macro avg | 0.98 | 0.98 | 0.98 | 90 | |
| weighted avg | 0.98 | 0.98 | 0.98 | 90 | |
| **Classification Report of Voting Model** | | | | | |
|  | precision | recall | f1-score | support | |
| 0 | 1.00 | 0.97 | 0.98 | 30 | |
| 1 | 0.97 | 1.00 | 0.98 | 30 | |
| 2 | 1.00 | 1.00 | 1.00 | 30 | |
|  |  |  |  |  | |
| accuracy |  |  | 0.99 | 5870 | |
| macro avg | 0.99 | 0.99 | 0.99 | 5870 | |
| weighted avg | 0.99 | 0.99 | 0.99 | 5870 | |

|  |  |
| --- | --- |
| Decision Tree | SVM |
| Voting Classifier | |

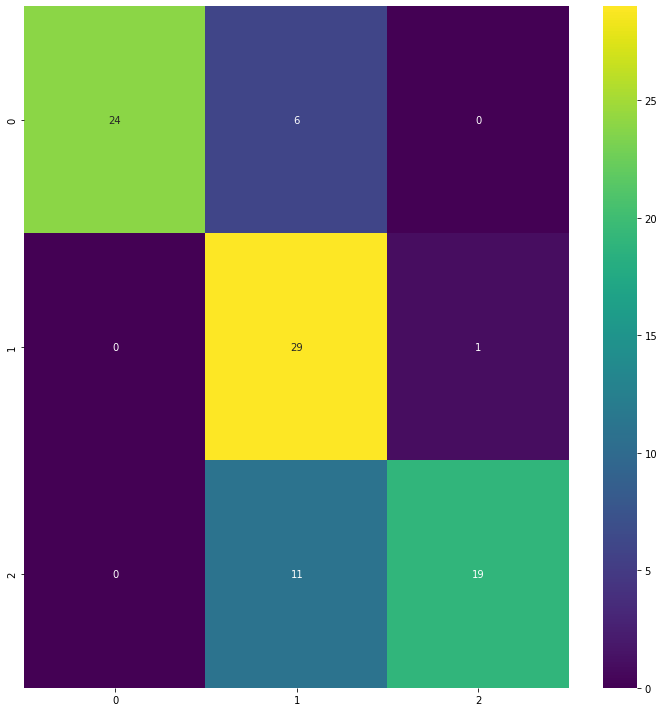
**Figure 7:** Face Mask Classification - Confusion Matrices

## Paper 3 Models Result

The author of the 3rd implemented paper also proposed the ResNet50 model with hyper parameter tuning technique for the classification of face mask images. The combination of different hyper parameters was used to gain the significant score of evaluation measures (methodology section). We also trained the ResNet50 model with each combination of the model using the training samples of the dataset. By following the training process, we used the test images of the dataset to evaluate the performance of model every time. We got the 0.98% accuracy score from ResNet50 model. The complete classification report of the ResNet50 model in shown in Table 7 and confusion matric in Figure 8.

**Table 7:** Face Mask Classification – ResNet50 Report

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Classification Report of ResNet50 Model** | | | | |
|  | precision | recall | f1-score | support |
| 0 | 1.00 | 0.80 | 0.89 | 30 |
| 1 | 0.63 | 0.97 | 0.76 | 30 |
| 2 | 0.95 | 0.63 | 0.76 | 30 |
|  |  |  |  |  |
| accuracy |  |  | 0.80 | 90 |
| macro avg | 0.86 | 0.80 | 0.80 | 90 |
| weighted avg | 0.86 | 0.80 | 0.80 | 90 |



**Figure 8:** Face Mask Classification - ResNet Confusion Matrix

Lastly, we contribute in the hyper parameter tunning of the ResNet50 model. Different combination of hyper parameters was used for the training of the model. Every time we used the training samples for the training of the model and test samples for the assessment of the model. The accuracy score for each combination of hyper parameters is shown in Table 8.

**Table 8:** Face Mask Classification - Hyper parameter tuning

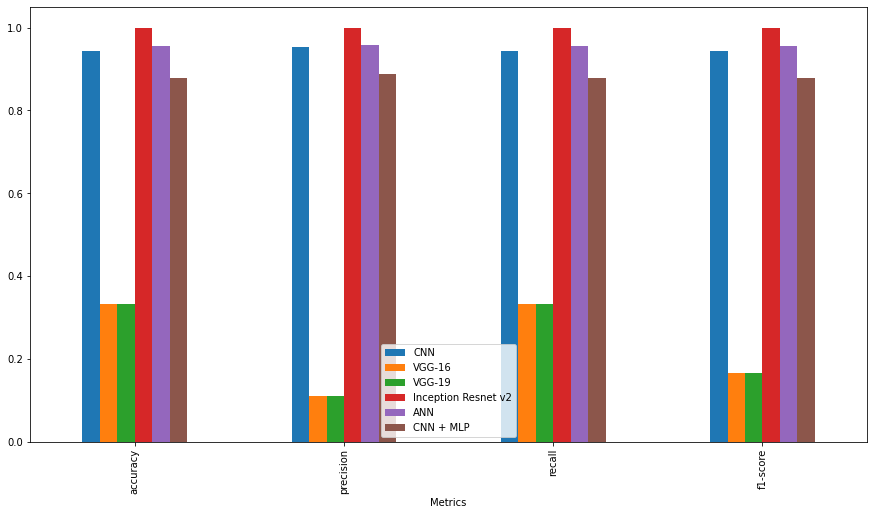
| **Optimizers** | **Learning Rate** | **Epochs** | **Accuracy Score** |
| --- | --- | --- | --- |
| Adam | 0.0010 | 10 | 1.0000 |
| Adam | 0.0010 | 15 | 1.0000 |
| Adam | 0.0010 | 20 | 0.9888 |
| Adam | 0.0001 | 10 | 0.9777 |
| Adam | 0.0001 | 15 | 1.0000 |
| Adam | 0.0001 | 20 | 1.0000 |
| SGD | 0.0010 | 10 | 0.9888 |
| SGD | 0.0010 | 15 | 0.9888 |
| SGD | 0.0010 | 20 | 0.9888 |
| SGD | 0.0001 | 10 | 0.9333 |
| SGD | 0.0001 | 15 | 0.9666 |
| SGD | 0.0001 | 20 | 0.9555 |

## Comparative Study

Finally, we did the comparative analysis of the proposed model. We found the highest accuracy score with the Inception ResNet V2 model for the classification of face mask images. Table 9 displays an analysis of all trained models in comparison. The comparative bar chat of all models is also presented in Figure 9 for better understanding. By completely analyzing the result, we hypothesize that the Inception ResNet V2 model is robust enough to make prediction in real environment.

**Table 9:** Face Mask Classification – Comparative Classification Report

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Metrics** | **CNN** | **VGG-16** | **VGG-19** | **Inception** | **ANN** | **CNN+MLP** |
| **accuracy** | 0.9444 | 0.3333 | 0.3333 | 1.00 | 0.9555 | 0.8777 |
| **precision** | 0.9523 | 0.1111 | 0.1111 | 1.00 | 0.9569 | 0.8878 |
| **recall** | 0.9444 | 0.3333 | 0.3333 | 1.00 | 0.9555 | 0.8777 |
| **f1-score** | 0.9440 | 0.1666 | 0.1666 | 1.00 | 0.9555 | 0.8772 |



**Figure 9:** Face Mask Classification - Comparative Bar Chat

# References

[1].  Kurlekar, M.S., 2021. Face Mask Detection System Using Deep Learning. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(7), pp.1327-1332.

[2]. Murray, C.J. and Piot, P., 2021. The potential future of the COVID-19 pandemic: will SARS-CoV-2 become a recurrent seasonal infection?. *Jama*, *325*(13), pp.1249-1250.

[3]. Gupta, S., Sreenivasu, S.V.N., Chouhan, K., Shrivastava, A., Sahu, B. and Potdar, R.M., 2021. Novel face mask detection technique using machine learning to control COVID’19 pandemic. *Materials Today: Proceedings*.

[4]. Qin, B. and Li, D., 2020. Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19. *Sensors*, *20*(18), p.5236.

[5]. Chavda, A., Dsouza, J., Badgujar, S. and Damani, A., 2021, April. Multi-stage CNN architecture for face mask detection. In *2021 6th International Conference for Convergence in Technology (i2ct)* (pp. 1-8). IEEE.

[6]. Boulila, W., Alzahem, A., Almoudi, A., Afifi, M., Alturki, I. and Driss, M., 2021, December. A deep learning-based approach for real-time facemask detection. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 1478-1481). IEEE.

[7]. Baker, D.M., Bhatia, S., Brown, S., Cambridge, W., Kamarajah, S.K., McLean, K.A., Brindl, N., Lapolla, P., Pérez-Ajates, S., Raubenheimer, K. and Xu, W., 2020. Medical student involvement in the COVID-19 response. *The Lancet*, *395*(10232), p.1254.

[8]. Liu, P., Cai, J., Jia, R., Xia, S., Wang, X., Cao, L., Zeng, M. and Xu, J., 2020. Dynamic surveillance of SARS-CoV-2 shedding and neutralizing antibody in children with COVID-19. *Emerging microbes & infections*, *9*(1), pp.1254-1258.

[9]. Cheng, V.C.C., Wong, S.C., Chuang, V.W.M., So, S.Y.C., Chen, J.H.K., Sridhar, S., To, K.K.W., Chan, J.F.W., Hung, I.F.N., Ho, P.L. and Yuen, K.Y., 2020. The role of community-wide wearing of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV-2. *Journal of Infection*, *81*(1), pp.107-114.

[10]. chengVcc, W.O.N.G.S.C., 2020. Preparedness and proactive infection control measures against the emerging novel coronavirus in china. *JHospInfect*, *104*(3), p.254.

[11] Cheng, V.C., Wong, S.C., Chen, J.H., Yip, C.C., Chuang, V.W., Tsang, O.T., Sridhar, S., Chan, J.F., Ho, P.L. and Yuen, K.Y., 2020. Escalating infection control response to the rapidly evolving epidemiology of the coronavirus disease 2019 (COVID-19) due to SARS-CoV-2 in Hong Kong. *Infection Control & Hospital Epidemiology*, *41*(5), pp.493-498.

[12] Olliaro, P., Torreele, E. and Vaillant, M., 2021. COVID-19 vaccine efficacy and effectiveness—the elephant (not) in the room. *The Lancet Microbe*, *2*(7), pp.e279-e280.

[13] Bar-Zeev, N. and Inglesby, T., 2020. COVID-19 vaccines: early success and remaining challenges. *The Lancet*, *396*(10255), pp.868-869.

[14] Abd-Elsayed, A. and Karri, J., 2020. Utility of substandard face mask options for health care workers during the COVID-19 pandemic. *Anesthesia and analgesia*.

[15] MacIntyre, C.R., Chughtai, A.A., Rahman, B., Peng, Y., Zhang, Y., Seale, H., Wang, X. and Wang, Q., 2017. The efficacy of medical masks and respirators against respiratory infection in healthcare workers. *Influenza and other respiratory viruses*, *11*(6), pp.511-517.

[16] Davies, A., Thompson, K.A., Giri, K., Kafatos, G., Walker, J. and Bennett, A., 2013. Probando la eficacia de las máscaras caseras:¿ protegerían en una pandemia de gripe. *Preparación para la Salud Pública*, *7*, pp.413-418.

[17] Ding, Z., Qian, H., Xu, B., Huang, Y., Miao, T., Yen, H.L., Xiao, S., Cui, L., Wu, X., Shao, W. and Song, Y., 2021. Toilets dominate environmental detection of severe acute respiratory syndrome coronavirus 2 in a hospital. *Science of The Total Environment*, *753*, p.141710.

[18] Liu, J., Liao, X., Qian, S., Yuan, J., Wang, F., Liu, Y., Wang, Z., Wang, F.S., Liu, L. and Zhang, Z., 2020. Community transmission of severe acute respiratory syndrome coronavirus 2, Shenzhen, China, 2020. *Emerging infectious diseases*, *26*(6), p.1320.

[19] Liu, J., Liao, X., Qian, S., Yuan, J., Wang, F., Liu, Y., Wang, Z., Wang, F.S., Liu, L. and Zhang, Z., 2020. Community transmission of severe acute respiratory syndrome coronavirus 2, Shenzhen, China, 2020. *Emerging infectious diseases*, *26*(6), p.1320.

[20] Doremalen, N., Bushmaker, T. and Morris, D.H., MyndiG. Holbrook, Amandine Gamble, Brandi N. Wi lliamson, Azaibi Tamin, Jennif er L. *Harcourt, Natalie J. Thornburg, SusanI. Gerber, JamesO. Lloyd-Smith, Emmied eW it, V incen tJ. Munster. Aerosolan dSurfaceStabilityof SARS-CoV-2asCompared with SARS-CoV-1. NewEngland Journalof Medici ne*, *2*, p.020.

[21] Lauer, S.A. and Grantz, K.H., 2020. Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler. The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*, *172*(9), pp.577-582.

[22] Migliori, G.B., Nardell, E., Yedilbayev, A., D'Ambrosio, L., Centis, R., Tadolini, M., Van Den Boom, M., Ehsani, S., Sotgiu, G. and Dara, M., 2019. Reducing tuberculosis transmission: a consensus document from the World Health Organization Regional Office for Europe. *European Respiratory Journal*, *53*(6).

[23] Long, Y., Hu, T., Liu, L., Chen, R., Guo, Q., Yang, L., Cheng, Y. and Huang, J., Effectiveness of N95 respirators versus surgical masks against influenza: a systematic review and meta-analysis. J Evid Based Med. 2020.

[24] Bae, S., Kim, M.C., Kim, J.Y., Cha, H.H., Lim, J.S., Jung, J., Kim, M.J., Oh, D.K., Lee, M.K., Choi, S.H. and Sung, M., 2020. Effectiveness of surgical and cotton masks in blocking SARS–CoV-2: a controlled comparison in 4 patients. *Annals of internal medicine*, *173*(1), pp.W22-W23.

[25] Jiang, S.C., Zhang, Z.W., Fu, Y.F., Li, Z.L., Zhu, F., Lan, T., Chen, Y.E., Yuan, M. and Yuan, S., 2019. Every 10-fold increase in viral load results in 26% more patient deaths: a correlation analysis. *Int J Clin Exp Med*, *12*(12), pp.13712-22.

[26] Van der Sande, M., Teunis, P. and Sabel, R., 2008. Professional and home-made face masks reduce exposure to respiratory infections among the general population. *PloS one*, *3*(7), p.e2618.

[27] Vaccinated, W.Y.V.B.F., 2021. How to Protect Yourself and Others. *Natl. Cent. Immun. Respir. Dis.(NCIRD), Div. Viral Dis.,(available at https://www. cdc. gov/coronavirus/2019-ncov/vaccines/fully-vaccinated. html)*.

[28] Vaccinated, W.Y.V.B.F., 2021. How to Protect Yourself and Others. *Natl. Cent. Immun. Respir. Dis.(NCIRD), Div. Viral Dis.,(available at https://www. cdc. gov/coronavirus/2019-ncov/vaccines/fully-vaccinated. html)*.

[29] Setti, L., Passarini, F., De Gennaro, G., Barbieri, P., Perrone, M.G., Borelli, M., Palmisani, J., Di Gilio, A., Piscitelli, P. and Miani, A., 2020. Airborne transmission route of COVID-19: Why 2 meters/6 feet of inter-personal distance could not be enough. *International journal of environmental research and public health*, *17*(8), p.2932.

[30]. Su, X., Gao, M., Ren, J., Li, Y., Dong, M. and Liu, X., 2022. Face mask detection and classification via deep transfer learning. *Multimedia Tools and Applications*, *81*(3), pp.4475-4494.

[31]. Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M. and Yener, B., 2009. Histopathological image analysis: A review. *IEEE reviews in biomedical engineering*, *2*, pp.147-171.

[32] Loey, M., Manogaran, G., Taha, M.H.N. and Khalifa, N.E.M., 2021. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. *Measurement*, *167*, p.108288.

[33] Meenpal, T., Balakrishnan, A. and Verma, A., 2019, October. Facial mask detection using semantic segmentation. In *2019 4th International Conference on Computing, Communications and Security (ICCCS)* (pp. 1-5). IEEE.

[34]  Li, L., Xia, Z., Jiang, X., Ma, Y., Roli, F. and Feng, X., 2020. 3D face mask presentation attack detection based on intrinsic image analysis. *Iet Biometrics*, *9*(3), pp.100-108.

[35] Nieto-Rodriguez, A., Mucientes, M. and Brea, V.M., 2015, June. System for medical mask detection in the operating room through facial attributes. In *Iberian Conference on Pattern Recognition and Image Analysis* (pp. 138-145). Springer, Cham.

[36]  Grassi, M. and Faundez-Zanuy, M., 2007, June. Face recognition with facial mask application and neural networks. In *International Work-Conference on Artificial Neural Networks* (pp. 709-716). Springer, Berlin, Heidelberg.

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

[38] Nagrath, P., Jain, R., Madan, A., Arora, R., Kataria, P. and Hemanth, J., 2021. SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. *Sustainable cities and society*, *66*, p.102692.

[39] Maharani, D.A., Machbub, C., Rusmin, P.H. and Yulianti, L., 2020, December. Improving the capability of real-time face masked recognition using cosine distance. In *2020 6th International Conference on Interactive Digital Media (ICIDM)* (pp. 1-6). IEEE.

[40] Inamdar, M. and Mehendale, N., 2020. Real-time face mask identification using facemasknet deep learning network. *Available at SSRN 3663305*.

[41] Ambata, L.U., Del Castillo, I.A.P., Jacinto, J.R.H. and Santos, C.M.T., 2019, November. Public and private vehicle quantification and classification using vehicle detection and recognition. In *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)* (pp. 1-6). IEEE.

[42] Srinivasan S, Singh RR, Biradar RR, Revathi SA. Covid-19 monitoring system using social distancing and face mask detection on surveillance video datasets. In2021 International conference on emerging smart computing and informatics (ESCI) 2021 Mar 5 (pp. 449-455). IEEE.

[43] Munjal, P., Rattan, V., Dua, R. and Malik, V., 2021. Real-Time Face Mask Detection using Deep Learning. *Journal of Technology Management for Growing Economies*, *12*(1), pp.25-31.

[44]  Sethi, S., Kathuria, M. and Kaushik, T., 2021. Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread. *Journal of biomedical informatics*, *120*, p.103848.

[45] Meenpal, T., Balakrishnan, A. and Verma, A., 2019, October. Facial mask detection using semantic segmentation. In *2019 4th International Conference on Computing, Communications and Security (ICCCS)* (pp. 1-5). IEEE.

[46] Song, S., Li, Y., Huang, Q. and Li, G., 2021. A new real-time detection and tracking method in videos for small target traffic signs. *Applied Sciences*, *11*(7), p.3061.

[47] Sanjaya, S.A. and Rakhmawan, S.A., 2020, October. Face mask detection using MobileNetV2 in the era of COVID-19 pandemic. In *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)* (pp. 1-5). IEEE.

[48] Wang, Y., Tian, H., Zhang, L., Zhang, M., Guo, D., Wu, W., Zhang, X., Kan, G.L., Jia, L., Huo, D. and Liu, B., 2020. Reduction of secondary transmission of SARS-CoV-2 in households by face mask use, disinfection and social distancing: a cohort study in Beijing, China. *BMJ global health*, *5*(5), p.e002794.

[49] Hinton, G.E., Osindero, S. and Teh, Y.W., 2006. A fast learning algorithm for deep belief nets. *Neural computation*, *18*(7), pp.1527-1554.

[50]. Krizhevsky, A., Sutskever, I. and Hinton, G.E., ImageNet Classification with Deep Convolutional Neural Networks (AlexNet) ImageNet Classification with Deep Convolutional Neural Networks (AlexNet) ImageNet Classification with Deep Convolutional Neural Networks.

[51]. Huang, G.B., Mattar, M., Berg, T. and Learned-Miller, E., 2008, October. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*.

[52] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., 2009, June. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee.

[53]. Parkhi, O.M., Vedaldi, A. and Zisserman, A., 2015. Deep face recognition.

[54] Hariri, W., 2022. Efficient masked face recognition method during the covid-19 pandemic. *Signal, image and video processing*, *16*(3), pp.605-612.

[55]. Cao, Q., Shen, L., Xie, W., Parkhi, O.M. and Zisserman, A., 2018, May. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)* (pp. 67-74). IEEE.

[56] Bagchi, P., Bhattacharjee, D. and Nasipuri, M., 2014. Robust 3d face recognition in presence of pose and partial occlusions or missing parts. *arXiv preprint arXiv:1408.3709*.

[57] Drira, H., Amor, B.B., Srivastava, A., Daoudi, M. and Slama, R., 2013. 3D face recognition under expressions, occlusions, and pose variations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(9), pp.2270-2283.

[58]. Priya, G.N. and Wahida Banu, R.S.D., 2014. Occlusion invariant face recognition using mean based weight matrix and support vector machine. *Sadhana*, *39*(2), pp.303-315.

[59]. Weng, R., Lu, J. and Tan, Y.P., 2016. Robust point set matching for partial face recognition. *IEEE transactions on image processing*, *25*(3), pp.1163-1176..

[60]. Duan, Y., Lu, J., Feng, J. and Zhou, J., 2018. Topology preserving structural matching for automatic partial face recognition. *IEEE Transactions on Information Forensics and Security*, *13*(7), pp.1823-1837.

[61]. Geng, M., Peng, P., Huang, Y. and Tian, Y., 2020, October. Masked face recognition with generative data augmentation and domain constrained ranking. In *Proceedings of the 28th ACM International Conference on Multimedia* (pp. 2246-2254).

[62]. Liao, S., Jain, A.K. and Li, S.Z., 2012. Partial face recognition: Alignment-free approach. *IEEE Transactions on pattern analysis and machine intelligence*, *35*(5), pp.1193-1205.

[63]. Mandal, B., Okeukwu, A. and Theis, Y., 2021. Masked face recognition using resnet-50. *arXiv preprint arXiv:2104.08997*.

[64]. He, L., Li, H., Zhang, Q. and Sun, Z., 2018. Dynamic feature learning for partial face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7054-7063).

[65]. Yang, S., Luo, P., Loy, C.C. and Tang, X., 2016. Wider face: A face detection benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5525-5533).

[66]. Jain, V. and Learned-Miller, E., 2010. *Fddb: A benchmark for face detection in unconstrained settings* (Vol. 2, No. 6). UMass Amherst technical report.

[67]. Gong, B., Shi, Y., Sha, F. and Grauman, K., 2012. 2012 IEEE Conference on Computer Vision and Pattern Recognition.

[68]. Jain, A.K., Klare, B. and Park, U., 2011, March. Face recognition: Some challenges in forensics. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)* (pp. 726-733). IEEE.

[69]. Ge, S., Li, J., Ye, Q. and Luo, Z., 2017. Detecting masked faces in the wild with lle-cnns. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2682-2690).

[70]. Cabani, A., Hammoudi, K., Benhabiles, H. and Melkemi, M., 2021. MaskedFace-Net–A dataset of correctly/incorrectly masked face images in the context of COVID-19. *Smart Health*, *19*, p.100144.

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

[72] Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X. and Pietikäinen, M., 2019. Deep learning for generic object detection: A survey. arXiv 2018. *arXiv preprint arXiv:1809.02165*.

[73] Addagarla, S.K., Chakravarthi, G.K. and Anitha, P., 2020. Real time multi-scale facial mask detection and classification using deep transfer learning techniques. *International Journal*, *9*(4), pp.4402-4408.

[74] Worby, C.J. and Chang, H.H., 2020. Face mask use in the general population and optimal resource allocation during the COVID-19 pandemic. *Nature communications*, *11*(1), pp.1-9.

[75] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), pp.84-90.

[76] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., 2009. ImageNet: A Large-Scale Hierarchical Image Database, in ‘CVPR09’. *IEEE Computer Society, Miami, Florida, USA*, pp.248-255.

[77] Bastys, A., Kranauskas, J. and Masiulis, R., 2009. Iris recognition by local extremum points of multiscale Taylor expansion. *Pattern recognition*, *42*(9), pp.1869-1877.

[78] Aslan, M.F., Unlersen, M.F., Sabanci, K. and Durdu, A., 2021. CNN-based transfer learning–BiLSTM network: A novel approach for COVID-19 infection detection. *Applied Soft Computing*, *98*, p.106912.

[79] Konstantin, E. and Johannes, S., 2019. A comparison of deep networks with ReLU activation function and linear spline-type methods. *Neural Networks*, *110*, pp.232-242.

[80] Abd Almisreb, A., Jamil, N. and Din, N.M., 2018, March. Utilizing AlexNet deep transfer learning for ear recognition. In *2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)* (pp. 1-5). IEEE.

[81] Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[82] Jain, G., Sharma, M. and Agarwal, B., 2019. Optimizing semantic LSTM for spam detection. *International Journal of Information Technology*, *11*(2), pp.239-250.

[83] Yang, B., Sun, S., Li, J., Lin, X. and Tian, Y., 2019. Traffic flow prediction using LSTM with feature enhancement. *Neurocomputing*, *332*, pp.320-327.

[84] Bengio, Y., Simard, P. and Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, *5*(2), pp.157-166.

[85] Hochreiter, S. and Schmidhuber, J., 1996. LSTM can solve hard long time lag problems. *Advances in neural information processing systems*, *9*.

[86] Graves, A., Fernández, S. and Schmidhuber, J., 2005, September. Bidirectional LSTM networks for improved phoneme classification and recognition. In *International conference on artificial neural networks* (pp. 799-804). Springer, Berlin, Heidelberg.

[87] Jia, M., Huang, J., Pang, L. and Zhao, Q., 2019, July. Analysis and research on stock price of LSTM and bidirectional LSTM neural network. In *3rd International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2019)* (pp. 467-473). Atlantis Press.

[88] Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural networks. 2005 Jul 1;18(5-6):602-10.