Masked Face Detection and Recognition from Images

By

Aroobah Iftikhar

(Registration No: 00000320624)

Thesis Supervisor: Dr. Arslan Shaukat

Department of Computer and Software Engineering

College of Electrical & Mechanical Engineering

National University of Sciences and Technology

Islamabad, Pakistan

September 2023

Annex A

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by **NS Aroobah Iftikhar** Registration No. **00000320624**, of College of E&ME has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the thesis.

Signature : ____________________

Name of Supervisor: **Dr Arsalan Shaukat**

Date: 19-09-2023

Signature of HOD: ____________________
(Dr Usman Qamar)
Date: 19-09-2023

Signature of Dean: ____________________
(Brig Dr Nasir Rashid)
Date: 19 SEP 2023

DEDICATION

Dedicated to my exceptional parents: ***Mr. & Mrs. Iftikhar Ahmed****, my adored siblings, (****Mohsin and Ansha****), and my supervisor, (****Dr. Arslan Shaukat****) whose unwavering support and cooperative efforts have played a significant role in my achievement*

ACKNOWLEDGEMENT

I am deeply grateful to Allah Almighty for bestowing upon me the determination, persistence, and knowledge that are essential for completing tasks with utmost satisfaction. Without His divine blessings, my accomplishments would have been unattainable. Indeed, His benevolent intervention has paved the way for my successes, and without His divine blessings, my achievements would have been impossible. May peace and blessings be upon Muhammad (SAW).

My gratitude extends to my supervisor, Dr. Arslan Shaukat (Tenured Associate Professor, College of Electrical & Mechanical Engineering, NUST), who has been gracious in providing invaluable guidance and suggestions for my thesis from the head start. He is truly worthy of commendation and may Allah the Almighty bestow upon him innumerable favors in this life and the hereafter. Additionally, I would like to express my appreciation to my entire thesis committee, Muhammad Usman Akram, and Dr. Ahsan Shehzad, for their collaboration and valuable recommendations.

I am indebted to my beloved parents for their support and prayers throughout my life, and to my friend Rimsha Tariq, who has been a constant source of strength. Finally, I would like to express my gratitude to all my acquaintances who have provided encouragement and support throughout this entire period.

ABSTRACT

The research field of computer vision has recently taken interest in the active problem of masked face recognition due to the COVID 19 pandemic. The use of face masks as a preventative measure against the spread of COVID-19 has presented a new challenge for the technology of face recognition. Masked Face Recognition (MFR) has emerged as a crucial issue within the field of face recognition after the COVID-19 epidemic. MFR is a specific type of facial occlusion issue that obstructs vital facial features such as the mouth, nose, or chin. The purpose of research on Masked Face Detection and Recognition is to fine-tune a pre-trained model that can more accurately recognize masked faces and detect whether the person is wearing a mask or not, which can be advantageous in various applications such as security and surveillance, healthcare, retail, law enforcement, the workplace, and social media. The objective of this dissertation is to examine the potential of machine learning techniques for enhancing the performance of masked face detection and recognition systems. This thesis proposes an approach to enhance the performance of the single neural network architecture such as pretrained InceptionV3 as unified model capable of both detection and recognition of masked images by achieving 99% and 98% respectively on MFR2 dataset. Pretrained VGG16 with transfer learning and fine tuning is trained and tested on publicly available datasets for the detection of masked faces, which are MDMFR dataset, Kaggle face mask detection dataset, facedatahybrid and for recognition results obtained on MFR2. The findings of this research offer valuable insights into the potential of pretrained networks with transfer learning to improve the performance of masked face detection and recognition systems and pave the way for future research in this area.

Keywords: COVID-19, Masked Face Detection, Masked Face Recognition, VGG16, InceptionV3, Transfer Learning

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

The impact of technology on the world has been profound, particularly in the realm of security, where the emerging technology like artificial intelligence has been applied to enhance the competence of various processes, including the identification and detection of individuals.

Facial recognition is one area where AI has made significant strides. Traditionally, facial recognition systems have been utilized in the process of recognizing people through a comparison of their facial features to a database of known faces. However, the advent of the pandemic COVID19 and the utilization of masks on a large scale have posed a challenge to the efficiency of these systems. as crucial facial features are obscured.

To address this challenge, researchers have proposed various methods for the detection and recognition of masked faces, which could be categorized into two things: Face Detection with mask and without mask and Face recognition with and without mask. Masked face detection involves identifying whether a person is wearing a mask or not. On the other hand, masked face recognition seeks to identify individuals from images of their faces, even when wearing masks.

This thesis will analyze the current advances in masked face detection and recognition, evaluating the different techniques proposed and assessing their performance on a variety of datasets. Furthermore, the thesis will suggest a novel methodology for the recognition of masked faces utilizing both machine learning and deep learning techniques.

The findings of this thesis will provide significant perspectives into the difficulties and prospects that are linked with masked face detection and recognition and will contribute to the development of more reliable and robust facial recognition systems that can be effectively employed across various applications, even in scenarios involving the presence of face masks.

1.1 Motivation

In a world where mask usage is prevalent, the precise identification of individuals with partially covered faces has become of utmost importance. But there are limited unified models with higher accuracy that are also evaluated on diverse datasets that have been proposed. The motivation behind this research is to design a unified model that can more accurately recognize masked faces and also detect if a face is wearing a mask or not.

Although the current face detection and recognition systems perform well in regular scenarios, they face difficulties recognizing masked faces. Hence, this research aims to develop innovative algorithms that utilize distinguishable facial features to enhance security, surveillance, and identity verification. The outcomes of this study could revolutionize human-computer interaction and reinforce overall security measures.

By empowering industries and governments to adapt to ever-changing security needs, this study ultimately contributes to a safer future. In conclusion, the exploration of masked face detection and recognition can potentially redefine the role of technology in security practices and society on a larger scale.

1.2 Problem Statement

The development of effective systems for detecting and recognizing masked faces poses a significant challenge due to several factors. These factors include the wide variation in face mask types, each with its own unique features, which can impede the identification of individuals by face recognition algorithms. Additionally, masks make identification difficult because they do not show important face parts, especially the nose and eyes. Furthermore, the quality of images used in real-world applications for masked face detection and recognition systems may be low, degrading the effectiveness of face recognition algorithms.

The problem statement addressed in this thesis involves the development of a new approach for detecting and recognizing masked faces that is robust against the challenges. We plan to use a combination of both traditional and modern machine learning and deep learning approaches to acquire highly accurate models for detection and recognition.

1.3 Aims and Objectives

To facilitate the development of a robust and reliable facial recognition system and facial recognition system, this article will:

- Review the current advances in the detection and recognition of faces with masks on.

- A new approach to identify the recognition of masked faces uses a combination of deep and machine learning techniques.
- Conduct significant experiments to evaluate the performance of the proposed method.

1.4 Structure of Thesis

This report is structured as follows:

Chapter 2 covers a review of the existing work done for masked face detection and recognition by researchers in past years.

Chapter 3 describes the proposed framework in detail. It has three main modules: Masked and unmasked face classification, masked face recognition, and a combined model that can perform both classifications.

Chapter 4 gives an overview of the databases and the performance metrics employed to assess the efficiency of the proposed approach. All the results are discussed in detail, including the required tables and figures.

Chapter 5 this chapter will give the detailed conclusion of this research and present the future scope of this thesis.

CHAPTER 2: LITERATURE REVIEW

The global pandemic of COVID-19 has resulted in the extensive utilization of facial coverings as a measure to prevent virus transmission and ensure personal safety. However, this practice has resulted in a significant challenge in face identification and verification as vital facial features are obscured. This study proposes a solution to the problem by suggesting using the techniques of deep learning and transfer learning models to recognize masked faces. R. K. Shukla et al. [1] proposed a methodology that employs a transfer learning strategy utilizing MobileNetV2 technology to handle the difficulty of recognizing faces hidden behind masks. The proposed model utilizes deep features to successfully recognize masked faces such as techniques like feature extraction and deep learning models.

The proposed model achieves a remarkable recognition accuracy of 99.82% with the proposed dataset, which was created using various authentic sources, including the internet and Kaggle datasets. The dataset comprises 8169 images, consisting of 4066 images of faces wearing masks and 4103 images of faces without masks. The dataset was developed to improve the effectiveness of current face detection and identification algorithms for facial recognition on masked faces during the COVID-19 outbreak. In this paper, facial images were cropped using annotations to acquire dimension labels and bounding boxes were utilized to increase the margins around the cropped faces. Comparative analysis shows that the presented model performs better than all four existing models in terms of results. They assess accuracy, depth, parameter count, and face size both with and without the face mask. The proposed methodology and dataset can provide valuable insights for researchers and practitioners developing effective solutions for face recognition and masked detection.

S. Nadeem et al. [2] presented an innovative framework named DeepMasknet for the detection of mask and no mask faces and recognizing masked and no masked faces. DeepMasknet utilizes a deep learning-based approach for both tasks which consist of two basic modules: a face mask detection module and a masked facial recognition module. The face mask detection module employes a pre-trained Faster R-CNN model to identify the presence of a face mask in an image, the pre-trained ResNet-50 model is used by the masked facial recognition module to identify features from the covered face region, which are then employed to recognize the masked face.

B. Huang et al. [4] presented a novel approach called Progressive Learning Loss (PLFace) that aims to achieve balanced performance in deep face recognition for both masked and mask-free faces by utilizing margin losses. In the suggested method, the significance of the masked and mask-free data is dynamically updated at various training phases. The training process emphasis the idea of reducing the anomalies in normal samples first, while masked samples are gathered later. The proposed approach comprises adaptability in the modification of the relative significance of masked and non-masked samples across various stages of training. The visualization of the deep features extracted by PLFace has been performed using t-SNE and contrasted with those obtained through CurricularFace. The visualization results demonstrate that the adoption of progressive learning leads to increased mixing of non-masked and masked samples in the feature space, thereby effectively eliminating the impact of mask bias.

The paper has used two datasets for training and two datasets for evaluation. The training datasets are MS1MV3 and MS1MV3+MA-X, which is MS1MV3 with a mask applied. The evaluation datasets are ICCV2021-MFR-MASK and RMFRD. About 5.1 million face photos from 9.3 thousand people make up MS1MV3, and "MA-X" refers to masked face augmentation achieved using a particular “X” probability. The experimental comparison of benchmarks for normal and masked faces shows that PLFace is successful in removing the mask bias in face recognition. PLFace considerably enhances the performance of masked face recognition in terms of accuracy along with preserving the model accuracy for the identification of normal face compared to its cutting-edge competitors.

B. Anil Kumar et al. [7] proposed a novel model of human face detection for both photo and real-time video images, irrespective of whether a mask is present or not. The presented methodology involves utilizing the model for the extraction of features and categorization of masked image called Caffe-MobileNetV2 (CMNV2). Specifically, the model focuses on extracting features around the forehead, eyes, nose and ears to accurately classify images. To enhance classification accuracy while minimizing training parameters, five additional layers are incorporated into the pre-trained MobileNetV2 architecture. The paper uses OpenCV for detecting human faces as well as a deep learning technique for recognizing a face’s region of interest (ROI). Experimental results show the performance of the proposed approach, as indicated by a outstanding accuracy for real-time video images and 99.64% accuracy for photo images. Additional metrics, such as 100% precision, 99.28% recall, 99.64% f1-score and 0.36% error rate show that it performs better than earlier models. The paper aims to enhance the detection performance of masked and unmasked faces, with a particular emphasis on facial masks.

Saurav Kumar et al. [8] presented a framework for developing a smart surveillance system that can recognize individuals putting on face masks in the aftermath of the COVID-19 pandemic. The proposed methodology employs the YOLOv3 algorithm for object detection and recognition of individuals behind face masks. Transfer learning is utilized through the framework of Darknet neural network to train the custom dataset. Additionally, the authors generated a custom dataset consisting of 900 images of individuals wearing face masks, captured using an iPhone XR camera in various institutional settings. The dataset is augmented using techniques such as rotation, flipping, zooming in/out, cropping, and variable brightness or contrast. Image augmentation techniques are employed to increase the dataset size and enhance model performance.

The outcome of this research revealed that the YOLOv3 algorithm attains a mean average precision (mAP) of 98.73% on their personalized dataset, surpassing YOLOv3-tiny by approximately 62%. The authors also provide a comparative analysis of detection time between YOLOv3 and YOLOv3-tiny, illustrating that YOLOv3 is slower but more robust in performance. Overall, the results illustrate the efficiency of the suggested methodology for the recognition of individuals wearing face masks.

Yiming Ge et al. [9] show Masked facial recognition (MFR) performs much better using the Convolutional Visual Self-Attention Network (CVSAN) algorithm than with other techniques. The suggested method leverages a self-attention feature map to convolution layer features, enforcing the enhancement of local features. The authors created the Masked VGGFace2 dataset using the face identification algorithm in order to train the CVSAN model. To verify the effectiveness of the algorithm, it is tested for large test sets with 7K masked face identities. The primary contribution of this paper is the incorporation of the Transformer structure into masked face recognition.

Pedro C. Neto et al. [11] provided an evaluation of various techniques for identifying faces that are either occluded or masked. The primary contribution of the paper is to explore whether solutions designed for the identification of masked face can improve the performance of facial recognition for general obstruction. The paper introduces numerous contributions to both the fields of masked and obstructed face recognition.

The results of the paper demonstrated the compatibility between MFR methods and OFR datasets. The analysis offered in the paper supports the idea that proposed solutions for general OFR, MFR can be done successfully, provided that the occlusions are of a reasonable size. The paper also emphasizes that the range of possible occlusions is increasing because of technological advances.

Iram Javed et al. [12] introduced a novel system aimed at detecting face masks and keeping track of social distancing in COVID-19 pandemic. This system is specifically designed to enforce standard operating procedures (SOPs), which include the wearing of face masks and the maintenance of social distancing in public areas. One of the key contributions of this study is the development of a large-scale outdoor dataset that comprises 10k images that are divided into two binary classes: face-masked and non-face-masked individuals. This dataset is expected to make it easier for automated identification of masked face and social distance estimation in public settings to develop in the future. They also suggested a whole pipeline for detecting masked face and analyzing social distancing in real-time in outdoor setting. The results demonstrate that the proposed pipeline gives enhanced performance compared to baseline version, providing a significant enhancement in accuracy by 5.3%.

Walid Hariri [13] proposed an approach for efficient recognition of masked faces in the context of the COVID-19 pandemic. The primary contribution of this research is to tackle the challenge of recognizing faces with masks on, which can prove to be a challenging task due to the occlusion of certain facial features. To address this problem, the proposed methodology involves the removal of the masked face region followed by the utilization of pre-trained CNN to reveal deep characteristics in the areas that aren't covered, such the forehead and eyes. The features obtained from the final convolutional layer are quantized providing a resilient representation. Finally, Multilayer Perceptron (MLP) is utilized for the classification process. To test the efficiency of the proposed approach, the Real-World-Masked-Face-Dataset is utilized for experimental results, which in contrast to other existing methods, show strong recognition performance.

Hoai Nam Vu et al. [14] presented a fresh method for recognizing masked faces using Local Binary Pattern (LBP) features with deep learning approaches. The proposed methodology integrates RetinaFace, a facial recognizer, as an encoder, and extracts LBP features from regions of the eye, forehead and eyebrow in images with masked face. Furthermore, this article introduces a new dataset, named COMASK20, which was gathered from 300 subjects at the authors' institution. The proposed technique was assessed on both the Essex dataset and the COMASK20 dataset, and it outperformed other previously proposed face recognition techniques, achieving an 98% F1-score on the Essex dataset 87% and on the COMASK20 dataset. The primary contribution of the paper lies in its proposed distance function, which has been found to be significantly more efficient in terms of both model computational time and evaluation compared to the cosine distance. Further, the proposed distance function has been observed to result in a substantial reduction in computational cost compared to Euclidean and Cosine with savings of 8% and 10.7%.

Hayat Al-Dmour et al. [15] presented a system grounded on deep learning for the purpose of detecting and recognizing masked faces by using multiple face dataset.

L. Queiroz et al. [16] focused on the effect of wearing masks on face recognition technology amid the pandemic. The authors conducted experiments utilizing two datasets, namely Flickr-Faces-HQ and SpeakingFaces. The findings indicate a degradation of 36.78% in recognition performance when faces were masked. However, this degradation was reduced to 1.79% with the application of advanced deep learning approaches in the cross-spectral domain. The proposed methodology involves the employment of the Cascade R-CNN model to locate and classify faces with or without masks, resulting in a relatively high performance of 0.879 mAP.

M. K. Kumcu et al.'s [17] proposed methodology involves the administration of the Beck Depression Inventory, a survey on face recognition issues brought on by mask wear during the pandemic a sample of 44 RRMS patients and 51 healthy controls. The findings of the study indicate that RRMS patients exhibited poorer performance in masked faces and needed to take off their masks more frequently than their healthy counterparts to identify faces. Additionally, RRMS patients demonstrated limited progress in identifying masked faces over the passage of time compared to healthy controls since the pandemic's onset. This study examines the new difficulties related to mask usage during the pandemic on MS patients' face recognition.

V. Sharma et al. [18] introduced a novel approach for Recognition of masked faces utilizing pre-existing recognition methods and publicly available dataset for masked faces. They used to retrain the FaceNet model using pre-trained ResNet v1 ResNet50 and Inception architectures achieving an exceptional accuracy of 99.98% on the training set. Furthermore, the paper outlines the difficulties that face recognition systems face due to face masks and variations in imaging conditions. The authors recommend creating a balanced masked face dataset to extend the model's generality to minority populations.

R. Golwalkar et al. [19] introduced a proposed system that utilizes deep metric learning and a self-developed deep learning network, named FaceMaskNet-21, for recognizing masked faces. This system boasts the ability to distinguish masked faces in live video streams, static images, and static video files with a less than 10 ms execution time and an accuracy of 88.92%. The suggested method has practical applications in various settings, including schools, colleges, malls, banks, and high-security zones. Its use can facilitate attendance and access authorization without necessitating the removal of masks.

FaceMaskNet-21 network uses a technique for deep metric learning that uses quadruplets to train the FaceMaskNet-21 network to produce a output feature vector of 128-d. Quadruplets improve masked face recognition performance compared to triplets. Following ReLU and max-pooling layers, the network has a third convolution layer with 13x13 dimensions and 384 filters. The output dense layer has 128 nodes, and it flattens the input values to plot them to output classes. The next layer is Softmax, after which the 128-d output encodings are prompted. The study utilized a dataset that included both children and adults wearing masks to test and train the proposed system. The overall accuracy achieved was 88.186%, with a slight reduction in accuracy when the dataset included the masked faces of children. This is because Children's exposed facial features are less noticeable when wearing masks.

The precision of the proposed system is similar to that of Hariri et al. (91.30%), who employed a pre-trained VGG-16 model and transfer learning. However, the use of VGG-16 makes it challenging to implement the model on portable embedded systems like ARM and mobile phones. FaceMaskNet-21, in comparison, is quicker and smaller, making it easier to implement on portable embedded systems.

G. Yuan et al. [20] introduced a face recognition method that uses a hierarchical segmentation-based mask learning technique to improve occlusion robustness. The feature masking (FM) operators produce multi-scale latent masks to get rid of occlusion-induced false responses and improve the contaminated facial features at different levels. The proposed MSML network is capable of accurately detecting and occlusions from feature representations at different levels are being removed while also integrating features from visible facial areas. The effectiveness and performance of the proposed method compared to previously proposed methods is demonstrated through experiments with synthetic or realistic occlusions for face verification and recognition. The proposed method can achieve prominent occlusion-robustness and does not compromise its performance on standard non-occluded face recognition tasks. The proposed MSML network is evaluated on the Multi-PIE, LFW, and MFV datasets, and the results indicate that MSML is robust to unseen realistic occlusions and achieves impressive performance. Specifically, the proposed MSML enhances accuracy by 3.5% and TAR (FAR=1e-3) by 10.83% compared to A18, thereby achieving the highest performance.

N. D. Kwak et al. [21] The present paper outlines a novel system that is capable of detecting faces that are obscured by masks, while also identifying the user via transfer learning via YOLOv5s and Facenet, respectively. The proposed system has been used to determine whether someone is wearing a mask as well as verify the identity of the person entering a given space. The YOLOv5s model has been employed for masked face detection, while Facenet has been used for masked face recognition. To achieve this, the authors have conducted transfer learning, altered the learning rate, epoch, and batch size, and then evaluated the results to select the optimum model. Face Mask Dataset (YOLO Format) and data collected from the web is used for detection masked and non-masked faces. A total of 2352 data points were utilized, with 1679 assigned to training data, 398 to verification data, and 280 to test data. The proposed system has demonstrated a strong ability to identify masked faces and recognize the user.

S. Guo et al. [22] introduced a novel thermal infrared face recognition approach that employs face masks by means of the generative deep learning method. The main contribution of this research is in addressing the issue of facial occlusion that arises due to masks in the context of visible light-based face recognition. The proposed approach utilizes thermal infrared technology to capture the heat that is emitted by a face through a mask, while the generative deep learning method is applied to eliminate the mask in an independent dataset.

The HUST-MIR Mask Face Databases are utilized in this study, comprising 86 volunteers and a total of 270,000 thermal infrared images containing both masked and unmasked faces. Out of these images, 50,000 are selected for the purpose of training the mask detection network, while the other 50,000 are used for testing. The findings show that the proposed method yields favorable recognition results, although the recognition rate is lower than that of the visible light solution with the mask. This can be attributed to the limited amount of data, given that there are only 86 volunteers, whereas visible light-based approaches involve datasets that contain millions of faces.

F. I. Eyiokur et al. [23] put forth a computer vision system that is based on deep learning techniques and aims to stop COVID-19 from spreading by detecting face masks, face-hand interactions, and measuring social distance. The proposed system employs two distinct face datasets, one of which is publicly available, to train and evaluate the models. The methodology utilized in this research involves the utilization of CNN methods and training procedures to devise the models. Furthermore, the paper compares the mask detection models with existing works and assesses the system's generalization capacity by utilizing publicly accessible datasets. This research reveals that the proposed method has remarkable accuracy in detecting face masks, face-hand interactions, and measuring social distance. To sum up, the study's principal contribution is the development of a computer vision system that can assist in stopping the spread of COVID-19 by identifying and monitoring the recommended protective measures.

D. Montero et al. [25] proposed a technique aimed at improving the accuracy of face recognition for masked faces, a fundamental requirement relating to the pandemic. The suggested model is predicated on the ArcFace framework and involves alterations to both the loss function and the backbone. The original face recognition database is modified by the authors by augmenting data and further merging both datasets for training. The ResNet-50-based neural network is selected and adjusted to yield the probability for the usage of masks with adding any computational cost. As for the loss function ArcFace loss and mask-usage loss is combined resulting in novel innovative Multi-Task ArcFace function. The experimental results show that the suggested method significantly improved the original model's performance for masked faces along with preserving the accuracy for non-masked datasets. Moreover, the model achieves 99.78% accuracy for the classification of mask-usage.

Deepalaxmi R. Shenvi et al. [26] advanced an artificial intelligence-driven system for preventing COVID-19 that includes temperature monitoring, auto-sanitization, and mask detection. Furthermore, the system incorporates facial recognition technology, which can maintain an accurate student database with temperature and auto-sanitization mechanisms for opening and closing doors. The proposed system is linked to a server that enables administrators to monitor the system's performance in real time from any location. The paper also outlines the challenges that traditional facial recognition technology confronts due to the widespread use of masks during the COVID-19 epidemic. In response to this issue, the paper introduces the Real-World Masked Face Recognition Dataset (RMFRD), which is currently the world's largest real-world masked face dataset. The paper proposes a deep learning multi-feature combination face recognition algorithm driven by large data sets to enhance the effectiveness of current facial recognition systems on faces with masks. Additionally, the paper also discusses several other AI-based systems proposed by different researchers aimed at detecting COVID-19 and predicting its global outbreak.

A. Cabinet et al. [27] proposed the Masked Face-Net dataset as a solution to the issue of identifying face masks that are either worn correctly or improperly in the COVID-19 context. The dataset is created using an image editing approach and comprises three distinct datasets for detecting masked faces, namely the Correctly Masked Face Dataset (CMFD), the Incorrectly Masked Face Dataset (IMFD), and their combined variation for global masked face detection. The dataset serves a dual purpose, that is, to detect individuals wearing face masks and those not wearing any, as well as correctly and incorrectly worn masks. Moreover, the study presents a mask-to-face deformable model that generates other masked face images, including those with unique masks. The dataset, which is publicly available on GitHub, contains 137,016 images. The proposed approach is based on deep learning models that effectively detect individuals wearing masks and those who are not. The results indicate that the proposed dataset is useful for training deep learning models designed to identify individuals wearing masks and those who are not and that it provides a level of classification granularity for mask-wearing analysis that is not available in other large datasets of masked faces.

J.-S. Yu [30] The paper presents a novel algorithm for detecting face masks and determining their standard wear based on an enhanced YOLO-v4 model. The proposed methodology relies on the deep learning object detection algorithm. Several evaluation metrics are employed to assess the algorithm's performance, demonstrating that the face mask recognition mAP can achieve a rate of 98.3% and that the frame rate is high, reaching 54.57 FPS, which is notably superior to the current algorithm.

To summarize, this paper introduces an algorithm for face mask recognition and standard wear detection that is both more accurate and efficient than existing algorithms. The proposed methodology uses deep learning and employs an enhanced CSPDarkNet53 model, an adaptive image scaling algorithm, and an improved PANet structure. The algorithm is evaluated using a face mask detection dataset and achieves high accuracy and frame rate.

S. Mishra et al. [31] introduced an innovative Indian Masked Faces in the Wild (IMFW) dataset, with the aim of addressing the challenges associated with unconstrained masked facial recognition in the Indian context. This dataset is composed of 200 subjects and features images taken in an unrestricted setting, both masked and unmasked. It is important to note that the proposed dataset incorporates different poses, illumination, resolution, and the various types of masks were worn by the subjects. The performance based on the proposed IMFW dataset, four pre-trained deep face recognition models are evaluated: VGGFace, ResNet50, LightCNN29, and ArcFace. The experimental results highlight the limits of present algorithms under a variety of scenarios. The main contribution of this study is the culturally diverse datasets that have been collected in unrestricted settings, which is not currently available in existing datasets. The proposed approach involves benchmarking existing facial recognition models on the proposed IMFW dataset.

This paper advances a solution to the issue of face recognition when masks are worn, an issue that has confronted society during the COVID-19 pandemic. These datasets are specifically designed to help the development of various applications involving masked faces. Additionally, the paper presents a multi-granularity masked face recognition model that outperforms industry-reported results, achieving an accuracy of 95%. The proposed methodology relies on deep learning and requires an extensive pool of facial samples. The paper's contribution is to resolve the challenge of face recognition with masks and offer datasets that can be applied by both industry and academia for the development of diverse applications.

S. Prasad et al. [33] presented an innovative form of semi-supervised learning, called maskedFaceNet, which aims to identify masked faces in real-time. To achieve this end, the proposed approach utilizes unlabeled data to enhance network performance while minimizing data annotation efforts. Additionally, the paper presents two empirical datasets, MASK-face-v1 and MASK-face-v2, which serve to benchmark the proposed method and guide future research. Results obtained through the proposed model exceed current state-of-the-art methods across several datasets, thereby testifying to the validity of the presented hypothesis across different dataset types. As part of its contribution, the paper also examines the suitability of objective loss functions for masked face detection. By implementing a smooth L1 loss, the paper demonstrates that this method proves less sensitive to outliers in the final detection process.

This article [34] provides a comprehensive overview of the Masked Face Recognition Competitions (MFR) conducted as part of the 2021 International Joint Conference on Biometrics (IJCB 2021). The primary objective of this competition was to improve the accuracy of face recognition for masked faces while considering the feasibility of deploying the proposed solutions, with due consideration given to the compactness of the face recognition models. Ten teams participated in the competition, submitting a total of 18 valid solutions. These teams belonged to diverse affiliations spread across nine different countries and included representatives from both academia and industry.

The submitted solutions were evaluated using a private dataset that simulated a collaborative, multisession, real masked capture scenario. With respect to model compactness, all solutions contained between 23M and 108M parameters, with the top three solutions having less than 87M parameters. The suggestion is that using a bigger and more complex deep-learning model does not always lead to better verification performance.

J. Xinbei et al. [35] presented a novel real-time approach for detecting face masks in public spaces, which is a crucial measure to aid in the mitigation of the transmission of COVID-19. Squeeze and Excitation YOLOv3 (SE-YOLOv3) is the proposed approach, which is founded on YOLOv3. The key contributions of the research are as follows: Firstly, the Properly Wearing Masked Face Detection Dataset (PWMFD) is introduced, consisting of mask-wearing samples from a total of 9205 images with three classifications present. Secondly, the SE-YOLOv3 mask detector is presented, which incorporates the attention mechanism by integrating the SE block into the Darknet53 architecture. This incorporation will facilitate the effective capture of inter-channel relations. This enables the network to allocate greater attention to critical features. Thirdly, the adoption of GIoU loss is implemented, which can more effectively depict the spatial variation between the forecasted and actual ground truth boxes in order to enhance the consistency of bounding box regression. Focal loss is employed to address the extreme class imbalance between foreground and background. Lastly, the utilization of the corresponding techniques for image augmentation is implemented in order to further strengthen the resilience of the model with regard to the specific undertaking.

The proposed technique surpasses the performance of YOLOv3 and other advanced detectors on PWMFD, attaining a greater 8.6% mAP in comparison to YOLOv3, while maintaining a comparable detection speed. In addition, the article analyzes the outcomes of diverse components in SE-YOLOv3 and investigates the influence of the attention module using feature map visualization.

T. Mare et al. [38] proposed a method to produce synthetic masks and superimpose them on human faces for the efficient performance of face recognition systems in identifying masked individuals, which has become a crucial issue based on the COVID-19 pandemic. The proposed method employs Spark AR Studio to generate nine masks with diverse characteristics, such as colors, shapes, and fabrics, and overlay them on the faces in the original images. Furthermore, the authors provide two enriched datasets that include 640,000 samples of masked faces. Subsequently, the authors conducted a human evaluation study to determine the authenticity of the produced masks and found that their method generates images that are significantly more realistic than other methods. The authors also assessed the effectiveness of advanced face recognition systems trained on the enhanced datasets and discovered that refining the systems on realistically generated faces improved their performance in recognizing masked individuals by more than 2%. In summary, this paper's contribution is an authentic approach to generating masked faces and enhancing face recognition systems to recognize masked individuals.

In the paper entitled "Deep Face Recognition: A Survey," M. L. Wang and W. Deng [39] presented a comprehensive analysis of recent advances in deep face recognition (FR) algorithms, databases, protocols, and application domains. The principal contribution of this paper is that it provides a summary of diverse network architectures and loss functions that have been introduced in the expeditious advancement of deep face recognition methodologies. The associated methodologies for facial processing have been categorized into two groups, namely "one-to-many augmentation" and "many-to-one normalization". The paper presents a summary and comparison of commonly used databases for both model training and evaluation, which include LFW, IJB-A/B/C, Megaface, and MS-Celeb-1M. The paper reviews various scenarios in deep face recognition, like cross-factor, heterogeneous, multiple-media, and industrial scenes. The paper emphasizes the technical challenges that are present as well as several promising directions for research that may be pursued in the future.

The proposed methodology involves the use of deep learning techniques characterized by a hierarchical architecture that combines pixels into invariant face representation, which has significantly improved state-of-the-art performance and enabled successful real-world applications.

G. Wu [40] The present paper introduces a novel algorithm for recognizing masked faces in contactless delivery cabinets based on the COVID-19 pandemic. The proposed approach is characterized by the incorporation of an attention mechanism neural network as well as locally constrained dictionary learning and dilated convolution techniques, all of which are designed to enhance the accuracy of recognition of masked face images. To evaluate the efficiency of the proposed methodology, the RMFRD and SMFRD databases of Wuhan University were selected for experimentation. The results demonstrate that the proposed algorithm yields a superior recognition rate compared to existing methods.

Z. Zhu et al. [41] introduced the Masked Face Recognition Challenge, which has been designed to extend the limits of practical masked face recognition (MFR) using the ultra-large-scale WebFace260M benchmark and its associated Face Recognition Under Inference Time Constraint (FRUITS) protocol. The aim of this task is to tackle the issue of face recognition systems failing to identify faces that are occluded by facial masks, a common occurrence during the COVID-19 pandemic. The article provides a detailed account of the training data, evaluation protocols, submission rules, test sets, and metrics used in the challenge. The challenge is evaluated based on both MFR and standard face recognition (SFR) metrics, and the primary evaluation metrics are designed to show a weighted sum that takes into account both masked and standard faces simultaneously. The article presents the preliminary competition results of the first phase of the challenge, in which 69 teams from academia and industry participated and 49 teams outperformed the baseline. Additionally, the article describes the extensive test set employed for MFR assessment, which includes 60,926 faces of 2,478 celebrities and is widely regarded as the largest real-world masked test set in the world. The test sets collection pipeline is illustrated, and the algorithms' performance is evaluated under Controlled-Masked, Wild-Masked, and All-Masked settings. Overall, the article contributes to the advancement of practical MFR and provides a benchmark for evaluating the performance of MFR algorithms.

A. Anwar et al. [42] presented a novel methodology for identifying masked faces through the utilization of already existing facial recognition systems with reliable accuracy. In order to achieve this objective, the authors have introduced an open-source tool named MaskTheFace, which is capable of masking faces, thereby leading to the creation of a large dataset of masked faces. This helps in generating the dataset that can be used to train an efficient facial recognition system that can achieve the desired level of accuracy for masked faces.

To train and test the proposed methodology, the authors of the paper have utilized the Facenet system. In addition, the authors have employed the VGGFace2 dataset, containing approximately three million images of 9131 identities, with an average of ∼ 362 images per identity that vary in terms of pose, age, ethnicity, and illumination. From this dataset, the authors have extracted a subset, named VGGFace2mini, by randomly selecting 42 images per identity. This particular dataset consists of unmasked images of the identities. Further, the authors have generated the VGGFace2-mini-SM dataset, which has been expanded by nearly twice its original size through the use of a random selection of masks (including surgical-green, surgical-blue, N95, and cloth) that have been applied to each respective image.

The authors of the paper have reported that the utilization of MaskTheFace has resulted in a 38% improvement in the true positive rate for both masked and non-masked faces in the current Facenet system. Additionally, the accuracy of the re-trained system was tested on a custom real-world dataset, MFR2, and was found to exhibit similar accuracy, thereby extending the applicability of the proposed methodology to real-life masked faces.

B. Mandal et al. [43] proposed a deep learning-based model that utilizes a pre-trained ResNet-50 architecture to identify masked faces. The authors carried out fine-tuning of the pre-trained model on their dataset and achieved an accuracy of 89%. The research builds on the existing pre-trained ResNet-50 architecture, which was initially trained on human faces, to address the challenge of recognizing an individual's identity while wearing a face mask. The proposed approach enables transfer learning to utilize the pre-trained model on their images, which comprise individuals without face masks. The authors implemented different architectural modifications and conducted hyperparameter tuning to enable the model to recognize individuals with a mask from images of the same individuals without a mask. The study used precision, recall, and F1-score as performance measures to evaluate the model's effectiveness. The authors compared their results with other relevant works in this area but with different datasets. The research concludes that the suggested framework possesses the capability of being incorporated into present face identification applications engineered with the intent of ensuring security verification.

C. Wang et al. [44] The present study proposes a novel tool for the purpose of generating masked faces from unmasked faces. Additionally, the study introduces a newly constructed database named Masked LFW (MLFW), which serves to assess the efficiency of facial recognition models on masked faces. The newly proposed tool effectively generates real and diverse masked faces, with the generated mask demonstrating substantial visual consistency with the original unmasked face. The MLFW database utilized in this study is based on the Cross-Age LFW (CALFW) database, which includes different mask templates covering the most common styles observed in daily life. The outcomes of the study indicate that the accuracy of facial recognition models decreases significantly when exposed to masked faces as a result of extensive experimentation. In particular, state-of-the-art models exhibited a decline in recognition accuracy ranging from 5% to 16% when tested on the MLFW database, as compared to their respective accuracy when tested on the original images.

Sachith Seneviratne et al. [45] presented a series of datasets and a benchmark to be utilized by researchers in the realm of masked identification through the utilization of contrastive representation learning. The proposed methodology is predicated upon the utilization of a pre-trained representation to construct a model on a singular dataset, subsequently fine-tuning this model on multiple datasets to generalize the feature embedding, and further fine-tuning that is based on the identification of negative pairs during the training phase. Moreover, the study provides a synthetic mask-generating code and a novel training approach that is specialized for masked versus unmasked face matching. The findings indicate that the specialized weights generated through the proposed methodology exhibit superior performance compared to conventional face recognition characteristics in the context of matching masked faces with unmasked ones.

L. Song et al. [46] introduced an approach for face recognition that is robust against occlusion. The method is the Pairwise Differential Siamese Network (PDSN), which establishes a correspondence between obscured facial segments and distorted characteristic elements. To eliminate these occluded elements from recognition, the proposed method generates a Feature Discarding Mask (FDM) by uniting relevant dictionary elements and then multiplying them with the original features. The present investigation conducts comprehensive experimentation on synthesized and authentic occluded facial datasets, and the findings decisively showcase that the suggested algorithm surpasses pre-existing cutting-edge systems. Furthermore, it demonstrates a remarkable ability to generalize effectively in the realm of general facial recognition tasks. The dataset that is employed as the probe set is Face scrub, and the evaluation of training data is conducted to support these findings.

B. Huang et al. [47] presented a newly introduced occlusion face recognition dataset, denoted as Webface-OCC which was designed to enhance the evaluation of occluded face recognition in real-world scenarios. The proposed methodology includes obscuring familiar occlusion types (e.g., glasses, masks, etc.) on the typical image of face via known facial key points. To enhance the diversity and authenticity of occlusion, the authors collected a number of occlusion templates from a variety of natural scenes. Additionally, a precise facial key point detection model is implemented to obtain key point information pertaining to the facial image. The Webface-OCC dataset encompasses 804,704 facial images, which have been obtained from a diverse set of 10,575 subjects with a wide range of occlusion types. Results from the experiments illustrate that the ArcFace, retrained by means of the Webface-OCC dataset, produces accuracy rates of 97.08% and 78.25% on both types of face datasets, simulated as well as real world, respectively, which is significantly higher than the corresponding counterparts. Furthermore, the retrained models surpass the original models (i.e., FaceNet and ArcFace) by a substantial margin. In particular, in comparison to the initial ArcFace model, the precision of the retrained model has experienced a notable enhancement across the four masked facial recognition datasets, with an increase of 36.22%, 29.14%, 27.04%, and 15.03%, respectively.

S. Yang et al. [48] presented a benchmark dataset for facial recognition named WIDER FACE, which is substantially more extensive than current datasets and features abundant annotations, such as poses, occlusions, and face bounding boxes. The dataset is exceedingly demanding due to considerable variations in scale, occlusion, and pose. The paper assesses multiple representative detection systems, thereby providing a comprehensive overview of their top performance, and proposes a solution to handle substantial scale variation. The proposed dataset is an effective training resource for facial recognition. The paper also discusses typical failure cases that are worth further exploration. The methodological framework adopted in the paper incorporates part-based methods, such as deformable part models (DPM) and aggregated channel features (ACF). The images in the WIDER FACE dataset were obtained using search engines like Google and Bing. The paper demonstrates that there is a discrepancy between the current performance of facial recognition technology and the practical requirements of the real world. The paper's contribution is to provide a new benchmark dataset that is more challenging and realistic than existing datasets, thereby facilitating future facial recognition research.

H. Du et al. [49] introduced a novel methodology for performing masked face recognition under the challenging conditions of a near-infrared-visible (NIR-VIS) domain gap, facial part occlusion, and a lack of masked face data. The proposed approach effectively addresses these challenges by utilizing a heterogeneous training approach that maximizes the mutual information of face representations between semi-siamese networks and allows for the integration of two distinct domains. Moreover, a technique that utilizes 3D facial restoration has been utilized to generate masked facial features using pre-existing NIR images. The resulting approach provides a face representation that remains consistent across different domains and is capable of withstanding occlusion caused by masks. The effectiveness of the proposed method is evaluated on three distinct NIR-VIS face datasets, comprised of Oulu-CASIA NIR-VIS, CASIA NIR-VIS 2.0, and BUAA-VisNir. The experimental results demonstrate that the proposed method outperforms existing methods and exhibits cross-dataset-generalization capacity. Moreover, the article showcases the average value and variability of the rank-1 accuracy and verification rate when the false acceptance rate is set at 0.1%. attained from complete 10-fold trials conducted on the CASIA NIR-VIS 2.0 dataset utilizing multiple training loss functions.

M. Geng et al. [50] presented a Masked Face Segmentation and Recognition (MFSR) dataset for the purpose of constructing and evaluating masked face recognition models. This dataset is divided into two portions. The first part of the dataset comprises of 9,742 masked face images that have undergone region segmentation annotations. The second part, on the other hand, contains 11,615 images of 1,004 individuals, each consisting of full-face masked images that manifest varied orientations, lighting conditions, and mask types. To resolve the issue of insufficient data, the authors have presented a novel Identity Aware Mask GAN (IAMGAN) that incorporates a multi-level identity preservation segment to synthetically produce masked face images. Furthermore, to intensify the discriminative power of masked face recognition models, the authors propose a new Domain Constrained Ranking (DCR) loss. The (DCR) mandates the convergence of masked facial attributes. The test findings on the MFSR dataset affirm the efficiency of the presented techniques.

P. L. Wilson et al. [51] presented a methodology utilizing Haar classifiers for swift and precise identification of facial characteristics. This approach involves the localization of the image analysis area for a facial feature to the most probable location for its existence. This technique aids in reducing false positives and elevating detection speed.

In this study, three classifiers were employed for the detection of the eyes, nose, and mouth. The classifiers' accuracy was assessed using images from the FERET database. The outcomes demonstrated a high level of precision for the eyes and nose, while the mouth's detection rate was lower due to the detection's minimum size requirement.

H. Al-Dmour et al. [52] introduced a methodology utilizing deep learning-based systems to detect and recognize masked faces, as well as determine their identity and whether they are appropriately masked. The given system performs the task of identifying faces and extracting their features from the input image, with extracted features being stored in the database of the system. From there, the system proceeds to conduct a comparison of image features against each, and every stored facial print contained within the database, thereby determining whether or not access should be authorized or denied.

The paper employs several face image datasets for the methodology, including the LFW, CelebA, and MaskedFace-Net datasets. LFW includes 13,233 face images of 5,749 persons, while CelebA has 202,599 face images of 10,177 individuals. MaskedFace-Net consists of 5,000 masked face images and 5,000 unmasked face images.

The proposed system has shown promising results, with the binary classification model achieving 99.77% accuracy and the multi-class model achieving 99.5% accuracy in classifying masked face images into three labels. The system also successfully recognized the person by achieving an average accuracy of 97.98%. Visualizing them confirms the system's ability to locate and match faces. The paper suggests utilizing feature selection along with deep learning for the identification of masked face and highlights the importance of utilizing artificial intelligence, particularly deep learning, in combating the COVID-19 pandemic.

A literature review Table 2.1 is provided below, presenting an overview of the current progress in the field of detecting and recognizing faces with masks.

Table 2 1: Summary of Existing Masked Face Detection and Recognition techniques

Author	Dataset	Features and Classifier	Results
Hayat Al-Dmour et al. [15]	RMFD, Masked-FaceNet dataset, Celebrities Face Recognition (CFR)	Haar feature-based cascade classifiers applied then pre-trained VGG-16 as a discriminative feature extractor	99.77% 99.5% 97.98%
Ashwan A. Abdulmunem et al. [3]	Face mask detection dataset from Kaggle	Deep-learning features passed to two models, first is MobileNetv2 and second model is a new CNN architecture	97.59%
Ambreen Sabha et al. [62]	"Face Mask Detection 12K Images Dataset" downloaded from the Kaggle	CoSumNet model applies transfer learning using a pre-trained ResNet-50 deep learning model, to extract high-level features from the processed frames	99.92% and loss 0.00077
Wen-Chang Cheng et al. [5]	MaskedFace-Net, TFM, MAFA, and RMFD VGGFace2_HQ_CROP	FaceNet approach combined with transfer learning and the cosine annealing mechanism	About 93%
Mohanad Azeez Joodi et al. [6]	MAFA dataset	Haar cascade as face detector, and then proposed CNN model is utilized as a classification model	97.55% to 98.43%

Ratnesh Kumar Shukla et al. [1]	Custom dataset contains 8169 images, including 4066 masked face images and 4103 non-masked face images	Transfer learning approach with MobileNet V2 based	99.82%
Saurav Kumar et al. [8]	Custom dataset of captured 900 images	The YOLOv3 algorithm along with transfer learning used to train a custom dataset	Mean average precision (mAP) of 98.73%, surpassing YOLOv3-tiny by approximately 62%
Yiming Ge et al. [9]	Generated a Masked VGGFace2 dataset and tested on MFR2	Convolutional Visual Self-Attention Network (CVSAN)	96.36%
Naeem Ullah et al. [2]	Custom unified mask detection and masked face recognition (MDMFR) dataset	DeepMasknet framework	Detection - 98.5% Recognition -96.7%
Busra Kocacinar et al. [10]	MadFare Dataset	Fine-tuned lightweight deep Convolutional Neural Networks (CNN)	90.40%

2.1 Research Gaps

From the existing literature, some of the research gaps have been identified as follows:

- Several researchers used custom datasets for classification purposes, so further research on that work can’t be carried out.

- There are limited unified models with higher accuracy that are also evaluated on diverse datasets that have been proposed.
- Some researchers have not explicitly explained how they addressed the insufficient samples with respect to each class problem in some of the available datasets for masked face recognition.
- Occlusion from face masks poses a significant hurdle for most existing methods for masked face detection and recognition, as these methods are not robust to such occlusion. This is especially problematic since identifying faces that are completely or partially covered by masks is becoming increasingly difficult.
- Variations in mask types can have a profound impact on the accuracy of masked face detection as well as recognition, with cloth masks being particularly challenging to detect compared to surgical masks.

CHAPTER 3: METHODOLOGY

In this research, two phases of the proposed methodologies are involved. The first phase is dataset preparation, followed by a pre-trained classification model that can be significantly used for both the automated detection of mask or no-mask-face images and the recognition of a masked-face individual. The proposed methodology for masked face detection comprised of two distinct phases. The initial phase involves data collection, data preprocessing and data splitting then the images passed to second phase that involves pre-trained VGG16 architecture with the binary classification layer at the end to detect either the person has put on the mask or not as presented in the given Figure 3.1.

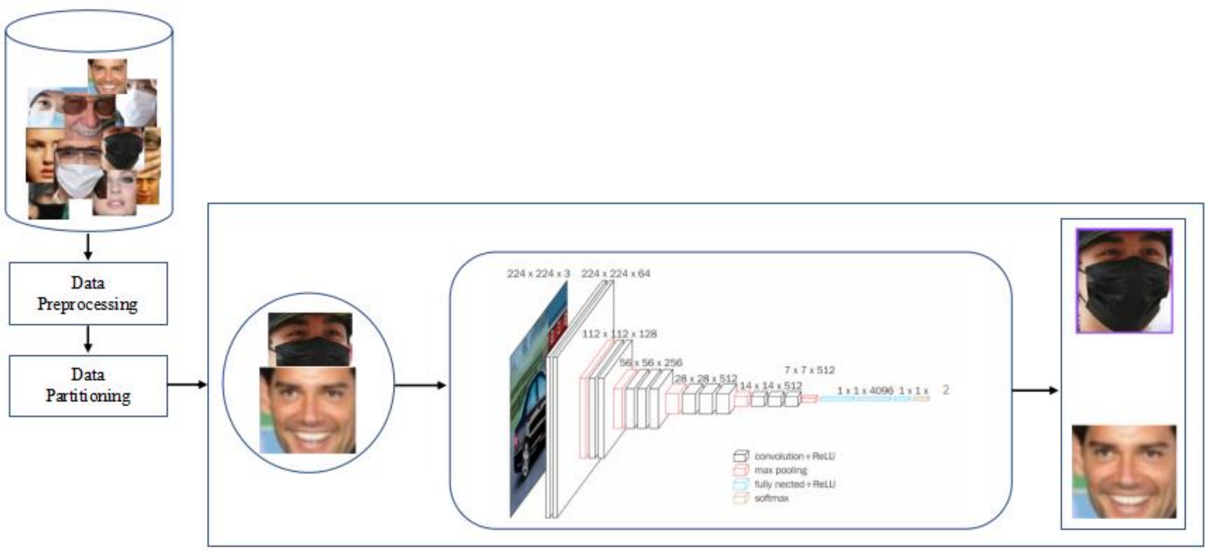


Figure 3. 1: Proposed Methodology for Masked Face Detection

The proposed methodology for masked face recognition comprised of two distinct phases. The initial phase involves data collection, data preprocessing and data splitting then the images passed to second phase that involves pre-trained VGG16 architecture with the multilabel classification layer at the end to recognize the person whether has put on the mask or not as presented in Figure 3.2.

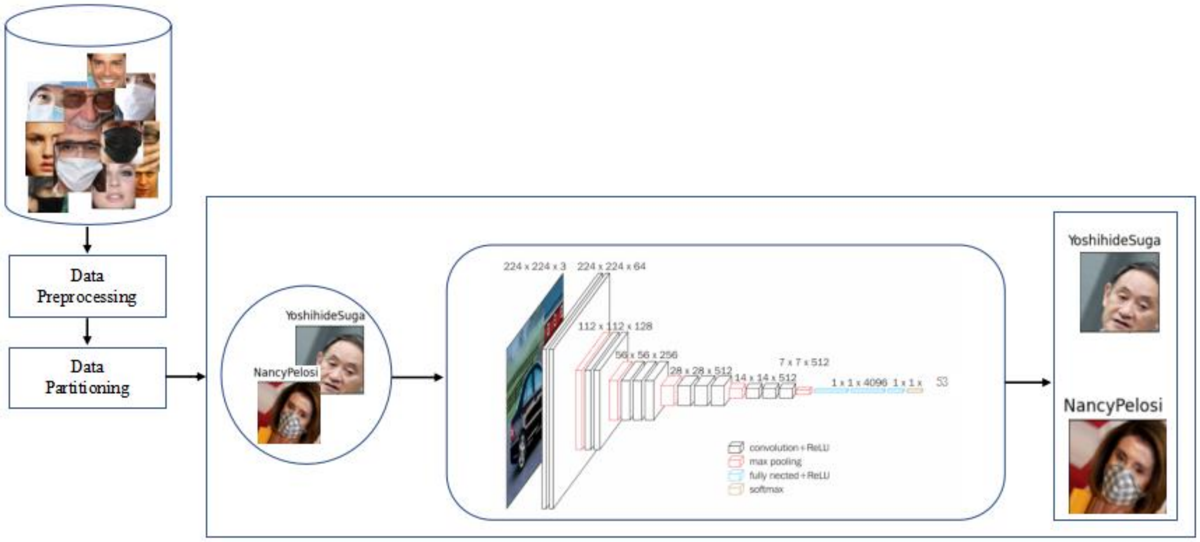


Figure 3. 2: Proposed Methodology for Masked Face Recognition

The proposed methodology for a unified model possesses the ability to detect and identify masked faces comprised of two distinct phases. The initial phase involves data collection, data preprocessing and data splitting then the images passed to second phase that involves pre-trained Inception V3 architecture with the binary as well as multilabel classification layers at the end to detect the masked faces as well as recognize either the person is wearing mask or not as presented in Figure 3.3.

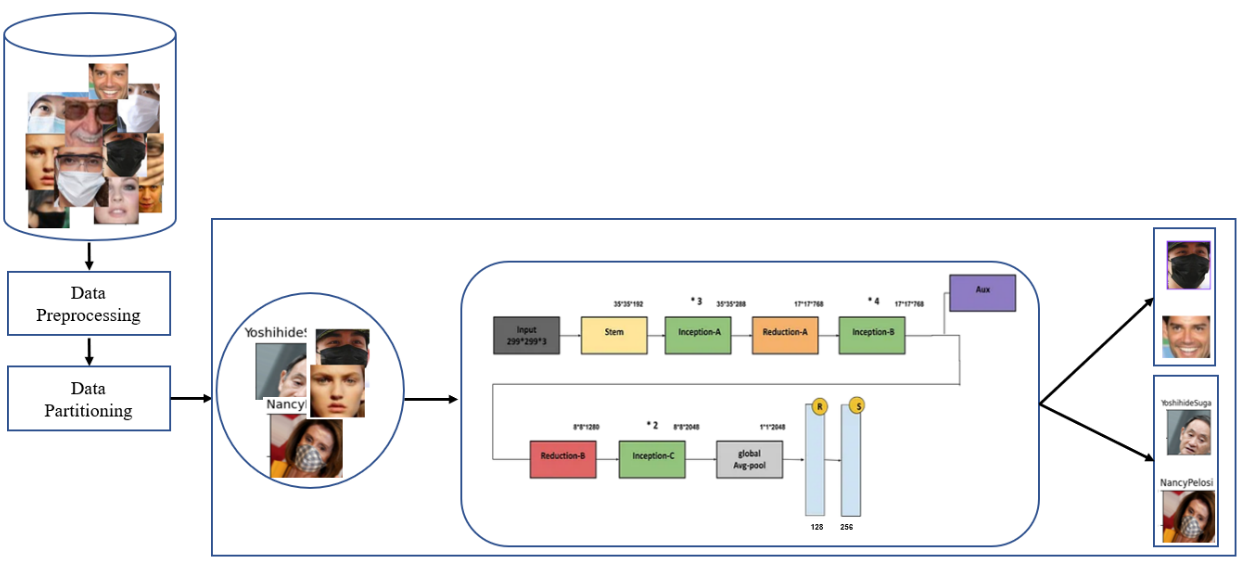


Figure 3. 3: Proposed Methodology for Unified Model

3.1 Dataset Preparation

The COVID-19 pandemic has resulted in an escalated requirement for masked face detection and recognition systems. These systems possess the ability to impose mask-wearing regulations, monitor virus outbreaks, and identify individuals who are non-compliant with the mask-wearing protocol. Nonetheless, developing precise systems for detecting and recognizing masked faces is a challenging task. The challenge of detecting and recognizing masked faces is primarily due to the difficulty in distinguishing between masked and non-masked faces as well as the significant changes in facial features caused by masks. The creation of a comprehensive dataset dedicated to detecting and recognizing masked faces would offer an invaluable resource for researchers and developers looking to confront the challenges of the COVID-19 pandemic.

The inaccessibility of a unified and balanced standard dataset for detecting masked faces and their recognition has posed a challenge in selecting from the limited publicly available datasets to evaluate the efficiency of masked face detection and recognition. It is important to note that these tasks require unique datasets. For masked face detection, a dataset with images of numerous individuals both with and without masks is needed, while for the recognition of masked face, multiple images of the same person wearing a mask are needed. In order to ensure accuracy in masked face detection and recognition, it is essential that our dataset be diverse in terms of gender, age, and race and contain both male and female images. The images should represent a wide range of variations in terms of mask type, lighting conditions, occlusion, face angle, environment, and format, among other factors.

3.1.1 Data Resizing

Dataset resizing is the process of scaling images to a uniform size in a dataset, which is widely implemented in machine learning to improve model performance. The resizing process involves standardizing the data by ensuring that images are uniformly sized, thus enabling the machine-learning model to learn more efficiently.

Implementing dataset resizing in masked face image classification has two primary benefits: Firstly, it can reduce the quantity of data that has to be processed by the machine learning model. This is particularly crucial for models that are trained on large datasets, as it can improve the training time and performance of the model. Secondly, resizing can increase the accuracy of the machine learning model by eliminating noise from the images, thereby simplifying the identification of facial features by the model.

Several notable machine learning models, such as ResNet and VGGNet, frequently use a prevalent size of 224x224 pixels, which is considered spacious enough to capture the essential facial characteristics and simultaneously small enough to enable quick processing by the machine learning model.

3.1.2 Data Augmentation

As discussed earlier, we have the problem of the unavailability of diverse multiple images of mask and non-mask faces of the same person. To solve this, a data-level approach has been taken so that the no. of images with respect to each class can be increased. Image Augmentation is performed to generate synthetic masked as well as non-masked face samples as per class.

Dataset augmentation is a method of artificially expanding a dataset by generating new data points from pre-existing ones. This can be achieved by implementing various modifications to the pre-existing data points, such as cropping, flipping, rotating, shifting, and introducing noise.

The augmentation of datasets is a crucial technique for enhancing the efficiency of machine learning models of classifying masked face images. This is due to its effectiveness in avoiding overfitting, which occurs when a machine learning model becomes excessively familiar with the training data and cannot thus generalize to new data.

Dataset MFR2 utilized for the recognition of masked faces undergo augmentation to tackle the insufficient masked images as well as non-masked ones with respect to each person (class) (Table 3.1). Implemented transformations are as follows:

- **Flipping:** The flipping technique involves a horizontal or vertical alteration of the image. Horizontal flip is set to ‘true’ to randomly flip images.
- **Rotating:** This involves a particular angle of rotation of images in the range of 0 to 180 degrees. Here, 30 is set as an angle to randomly rotate images.

- **Changing Brightness:** To modify the brightness of the images, 0.2 to 1.0 is set as the brightness range.
- **Shearing:** Setting the shear range results in stretching or compression in a single direction by a factor of 0.2.
- **Zooming:** To perform zooming in or out of the images, a factor of 0.2 is set.
- **Shifting:** Faces within an image shift horizontally and vertically by a fraction of 0.2 of total width and height.

Table 3. 1: Augmented MFR2 Dataset

MFR2 Dataset	Original Data	Augmented Data	No. of classes
Masked	170	3,555	53
Non-masked	98	2,054	53

Some masked and non-masked samples from the augmented data show multiple images of the same person shown in Figure 3.4.

Figure 3. 4: Masked and non-masked samples of Augmented MFR2 Dataset

3.1.3 Label Encoding

The procedure of transforming categorical data into numerical data is referred to as label encoding. This is applied to enhance the comprehensibility of data utilizing in machine learning models. In the context of masked face classification, the labels are the classes of images, such as ‘masked’ and ‘unmasked,’ while the additional label ‘person name’ is used as the class to recognize the masked face.

There are two main approaches to label encoding: One-hot encoding and Integer encoding. For this research, integer encoding has been implemented, which involves assigning each label a unique integer value starting from 0 to (no. of labels-1). The label "masked" has been assigned a value of 0, and the label unmasked has been assigned a value of 1. Similarly, for masked face recognition, an integer is assigned to each person's name as a label, which would be according to the given number of labels.

3.1.4 Data Splitting

The splitting of datasets into training, validation, and testing subsets is a familiar process in machine learning to avoid overfitting and make a well-generalized model for unseen data. In the context of masked face image classification, the training data is utilized for the model's training, the validation subset is employed to assess the model's performance during training, and the testing subset is employed to assess the model's final performance.

The training subset must be a representative sample of the entire dataset. For effective masked face image classification, it is important that the training data used for training the model include images of faces with diverse masks, poses, and lighting conditions. Similarly, the validation subset should also be a representative sample of the entire dataset. However, it should not be employed for model training since the model should not have access to the validation subset during training. This is done to enable the assess of model's performance on the validation subset to determine if the model is overfitting. The testing subset must be entirely new so that the model's final performance can be determined.

Various techniques are available for splitting datasets; some of the most commonly employed ones are Holdout splitting, K-fold cross-validation, Leave-one-out cross-validation, Stratified splitting, and Bootstrapping.

The optimal method of partitioning for the subsets assigned to training, validation, and testing is dependent on the specific machine learning model used and the dataset's size. For this research problem, the partitioning of 70/20/10 is preferred, implying that 70% of the dataset is allocated to training, 20% to validation, and 10% to testing. For masked face detection, holdout splitting is being done, which is considered the most simplistic approach to dataset splitting. Whereas for masked face recognition, stratified splitting is being implemented to ensure that each class (person name) with both a masked and non-masked face is equally represented in each subset. This is particularly vital for classification problems where the classes are not evenly distributed. The split ratio for this is set to75/15/10. The datasets are divided into training, validation, and testing sets, as illustrated in Table 3.2.

Table 3. 2: Datasets Splitting into Train, Validation and Test Sets

Dataset Split Sets		**MDMFR Dataset (11,852)**	**Face Mask Detect Dataset (7,553)**	**facedatahybrid (45,000)**	**MFR2 (5,609)**
Train Set	mask	5060	2,607	15,805	4,487
	non-mask	5000	2,679	15,694	
Validation Set	mask	400	745	4,515	533
	non-mask	400	765	4,484	
Test Set	mask	483	373	2,259	589
	non-mask	509	384	2,243	

3.2 Classification Stage

3.2.1 Fundamental Layers in a Convolutional Neural Network

Convolutional neural networks, also known as CNNs, represent a particular category within the realm of artificial neural networks extensively used for object detection, image classification, and segmentation. They are designed according to the functioning of the human visual cortex, which has a hierarchical arrangement of neurons that operate on image-based information in a manner similar to CNNs. The effectiveness of convolutional neural networks has been shown to be exceptionally high in different predictive tasks, particularly image recognition, and classification. The output of the CNN model is characterized by identifying distinctive features that can be detected anywhere across the input images. The architecture of a CNN consists of three layers, namely an input layer, multiple hidden layers, and an output layer. The hidden layers are commonly comprised of convolutional layers, activation layers, pooling layers, and fully connected (FC) layers. The initial three layers are iteratively applied in sequence, thereby enabling feature learning, while the final FC layer is utilized for classification purposes, as depicted in Figure 3.5.

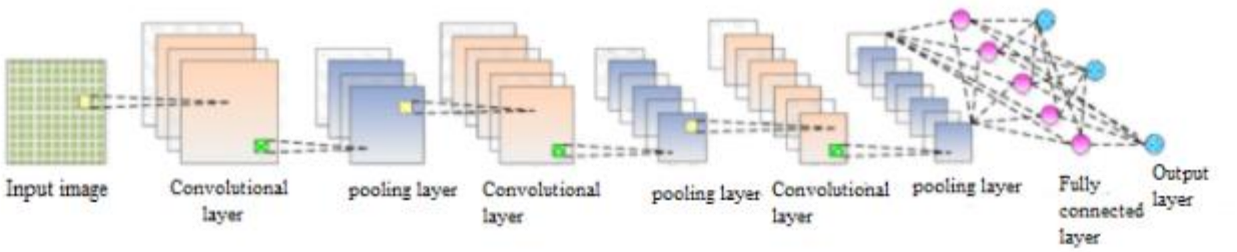


Figure 3. 5: Architecture of Convolutional Neural Network [53]

In a typical CNN, the primary components are as follows:

- **Input layer**

The input layer comprises an input image represented as a matrix constructed with pixel values that have the dimensions [height*width*depth], where depth represents the channels.

- **Convolutional layers**

The convolutional layer consists of an array of independent filters, or kernels, and each filter convolves separately with the image (Figure 3.6). In order to perform convolution, a filter is moved over the entire image, and the dot product is taken between sections of the image and the filter. All filters are initialized randomly, and these are the parameters that the network will subsequently learn. The initial layers search for fundamental patterns such as lines or corners. As we go deeper into the subsequent convolutional layers, the filters compute dot products on the previous convolutional layers. Therefore, the network extracts the fragmented pieces, or boundaries, and assembles them into larger entities.

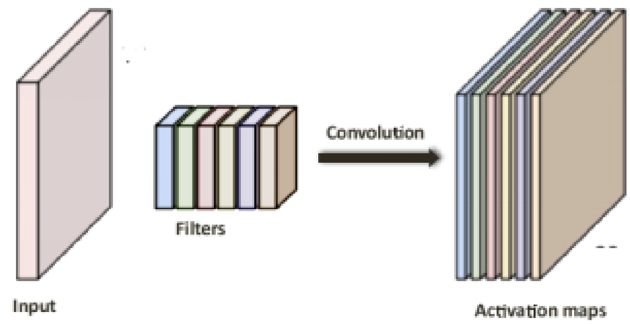


Figure 3. 6: Activation maps of a convolutional layer

- **ReLU (Rectified Linear Units) layers**

After every convolutional layer, it is common to employ an activation layer immediately thereafter. This layer aims to introduce nonlinearity in a system that already uses linear functions during the convolutional layers (element-wise multiplication and summation). ReLU layers facilitate network training without making significant compromises in accuracy. They also help to reduce the disappearing gradient problem; the phenomenon of slow training in the lower network layers arises due to a considerable reduction in gradient throughout these layers. ReLU uses the activation function presented as f(x) = max(0,x) for the input values. In other words, a layer merely converts all negative values to zero (Figure 3.7). The model's nonlinear properties and the network are amplified where the receptive fields of the convolutional layer remain unaffected.

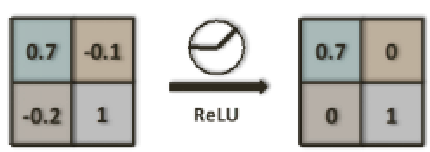


Figure 3. 7: ReLU operation

- **Pooling layers**

Pooling layers, also known as down sampling layers, have the primary objective of gradually diminishing the spatial dimensions of the representation, thus decreasing the computations and parameters in the network. Various non-linear functions, such as L2-norm pooling and average pooling, can be employed to implement pooling. However, max pooling is the most widely used approach. Max pooling divides the image into a set of non-overlapping blocks and selects the maximum value in each block, thus significantly reducing the spatial dimensions of the input volume. The fundamental concept behind pooling is that the precise location of a particular feature in the original input becomes less important than its location relative to other features. Pooling is mainly used for two purposes: reducing the number of parameters by 75%, which ultimately lowers the computational cost, and controlling overfitting. Overfitting is a phenomenon that arises when the model is excessively optimized on the training dataset but fails to generalize to the validation and testing datasets. Figure 3.8 illustrates the process of how max and average pooling are done.

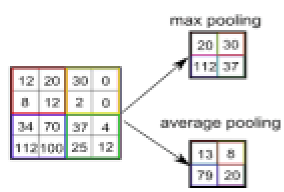


Figure 3. 8: Max and average pooling [54]

- **Fully connected (FC) layers**

After the convolutional, ReLU and pooling layers are applied repeatedly, the CNN's high-level reasoning is accomplished via fully connected (FC) layers. The high-level features presented in the image are effectively represented by the output obtained from the convolutional layers, and adding an FC layer allows a non-linear combination of those features. Although all features from convolutional layers may be good, combinations of those features may potentially yield superior outcomes. Neurons in an FC layer have complete connections to all activations in the preceding layer, similar to regular ANNs, and function similarly.

3.2.2 CNN-based Approaches

When implementing Convolutional Neural Networks (CNNs) for classification problems, two approaches can be employed, as represented in Figure 3.9. The first method, referred to as "Learning from Scratch", involves initializing the neuron weights with random values and relying entirely on the chosen dataset for network training. The second method, Transfer Learning, utilizes previously trained weights and parameters obtained from a different dataset to initialize the network training for the selected dataset. Fixed Feature Extraction and Fine Tuning are two distinct methods that can be used to implement Transfer Learning. Fixed feature extraction involves utilizing pre-trained weights without further adjustments, while fine-tuning involves the refinement of weights to improve performance.

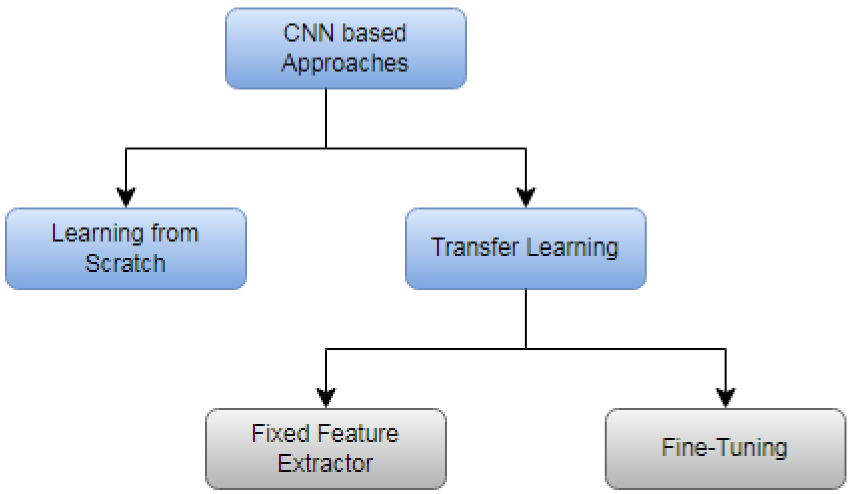


Figure 3. 9: CNN-based Approaches

- **Transfer Learning**
  - **Fixed Feature Extractor**

    The fixed feature extractor is a transfer learning approach where the weights and biases acquired from a vast dataset are used directly for the classification task of a given dataset. Therefore, there is no need for the network to be retrained on the selected/specified dataset. Only the feature extractor segment of the CNN architecture is implemented, and the features extracted via this mechanism can then be labeled using any classifier, such as a SoftMax classifier.

- **Fine-tuning**

  The process of fine-tuning involves the retraining of a segment of a network on a given dataset. During fine-tuning, a section of the network is retrained on a given dataset, which involves freezing weight values in a few initial layers and training only a select few on the task-specific dataset. The first step in this process is the establishment of the appropriate number of channels or neurons in the output layer, which should correspond to the total number of classes in the dataset. The weights and biases of the frozen layers are then set in accordance with the pre-trained architecture. Training parameters, such as total epochs, learning rates, batch sizes, and optimizers, are subsequently defined, and the unfrozen layers are trained to adjust the weights and biases of those layers with respect to the task-specific dataset.

  This fine-tuning approach to transfer learning has been implemented in our proposed classification phase of masked face detection and recognition.

3.2.3 Pre-trained Models

As previously mentioned, a pre-trained model is used with a transfer learning approach as well as for the classification of masked and non-masked images. The pre-trained models that have been implemented are VGG-16 and InceptionV3. The architecture of each model is explained below.

- **VGG-16**

The convolutional neural network VGG16, which was created by Karen Simonyan and Andrew Zisserman at the University of Oxford's Visual Geometry Group (VGG), has gained considerable attention in the field of computer vision. Their 2014 publication, titled "Very Deep Convolutional Networks for Large-Scale Image Recognition" [55], was the first to introduce this architecture. VGG16 was trained using the ImageNet dataset, which consists of more than 1.2 million images categorized into 1,000 classifications. VGG16's accuracy was evaluated using the ImageNet validation set, achieving an accuracy of 92.7% with a top-5 error rate of 7.3%.

VGG16 is a deep CNN that processes an input image of size 224x224x3. The first block of VGG16 comprises 16 layers of convolution and max pooling, with the first 13 layers being convolutional layers. The systematic arrangement of convolutional layers is done in blocks, with each block consisting of two convolutional layers followed by a max pooling layer. The convolutional layers use 3x3 filters, while the max pooling layers use 2x2 filters. The maximum kernel size in any convolutional layer is 3x3, and the filters slide over the image to extract features. Even the filter size in one of the convolutional layers is just 1x1, which is a linear transformation followed by non-linearity. The stride in the convolutional layers is fixed at 1, while the padding is set to preserve the spatial resolution of the image. The max pooling layer is present after most, but not all, of the convolutional layers, and the window size is 2x2 and the stride value is set as 2.

The final three layers of VGG16 are comprised of fully interconnected layers. The first two interconnected layers of VGG16 consist of 4096 neurons. However, the third fully connected layer consists of 1000 neurons, which is consistent with the number of categories in the ImageNet dataset employed in the ILSVRC competition. [56].

VGG16 has been utilized in a variety of computer vision tasks, such as image classification, object detection, and semantic segmentation. VGG16 has also been used as a pre-trained model for other CNNs. The detailed architecture of VGG-16 is presented in Figure 3.10.

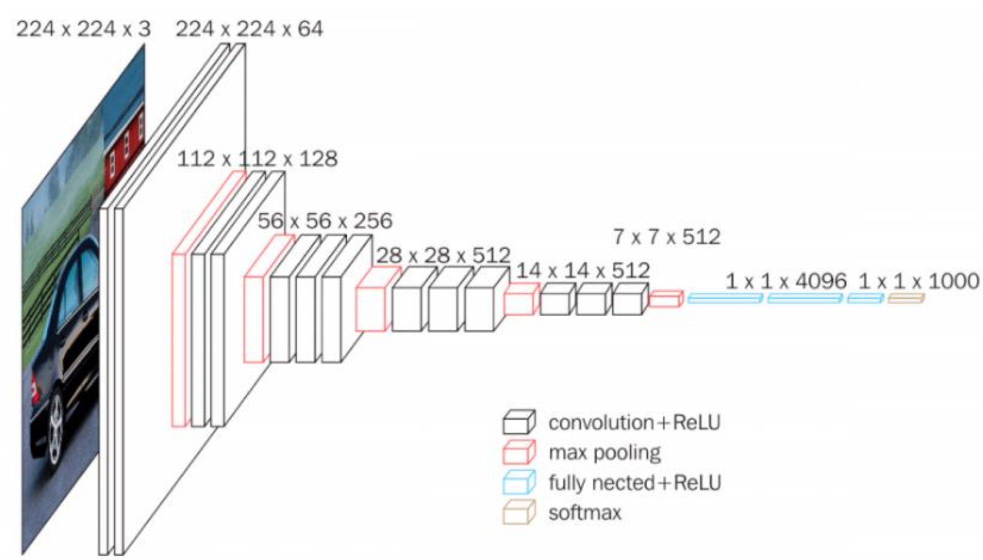


Figure 3. 10: VGG16 Architecture

- **InceptionV3**

Inception V3 was documented in a research paper titled "Rethinking the Inception Architecture for Computer Vision", [63] authored by Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, and Jonathon Shlens. This paper was made available to the public in the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) in the year 2015. The Inception V3 architecture, developed by Google in 2015, is a convolutional neural network (CNN) that represents a more advanced and efficient iteration of the previously introduced Inception V1 and V2 architectures. It incorporates several techniques to enhance its performance, such as factorized 7x7 convolutions, which involve breaking down large convolutions into smaller ones, thereby reducing the overall number of network parameters. Additionally, Inception modules are employed, which combine various filter sizes and pooling operations to extract image features across different scales. Furthermore, auxiliary classifiers are incorporated into intermediate layers of the network to aid in the regularization of the training process.

Extensive research has demonstrated the effectiveness of the Inception V3 architecture in image classification tasks. Notably, it achieved a top-5 error rate of 23.4% when tested on the ImageNet dataset, a widely recognized benchmark for image classification.

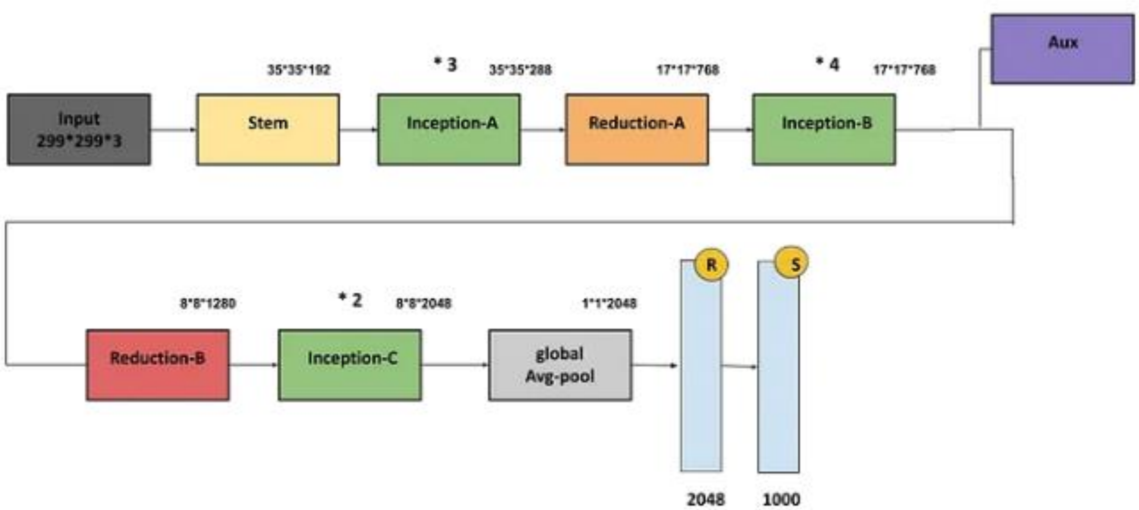


Figure 3. 11: InceptionV3 Architecture

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Datasets

In order to assess the performance of the proposed approach, it is evaluated on a set of four publicly available datasets, namely Kaggle fmdd, MDMFR, facedatahybrid, and LFW with Masks which include with-mask and without-mask face images. For masked face recognition, the dataset that has been evaluated is the MFR2 dataset.

This section provides a detailed explanation of the mentioned datasets.

4.1.1 Face Mask Detection Dataset (fmdd)

The dataset has been prepared, as specified in [57], and is composed of a total of 7553 RGB images, which are classified into two separate folders, namely with-mask and without-mask. Each image is assigned a label corresponding to its respective folder (Figure 4.1). It is noteworthy that there are 3725 images of faces with masks and 3828 images of faces without masks in this dataset.

Figure 4. 1: Samples of Face Mask Detection Dataset

4.1.2 MDMFR Dataset

The dataset is prepared by [58] and divided into train, validation, and test subsets. Each subset consists of 10,060, 800, and 992 images, respectively, divided into two subfolders, with mask and without mask, used as classes (Figure 4.2) for the detection of masked faces.

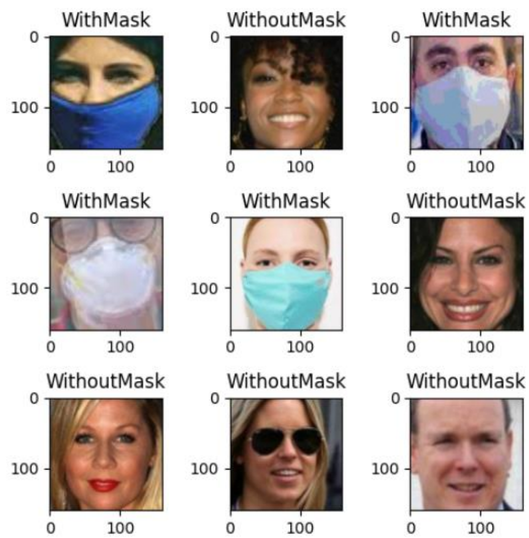


Figure 4. 2: Samples of MDMFR Dataset

4.1.3 facedatahybrid

The dataset collected by [59] consists of facial images of individuals that have been categorized into two groups, namely the mask and non-mask classes. The fundamental purpose of this dataset is to identify whether an individual is wearing a mask or not by analyzing facial images. where the data is comprised of 22579 and 22421 images of the respective classes.

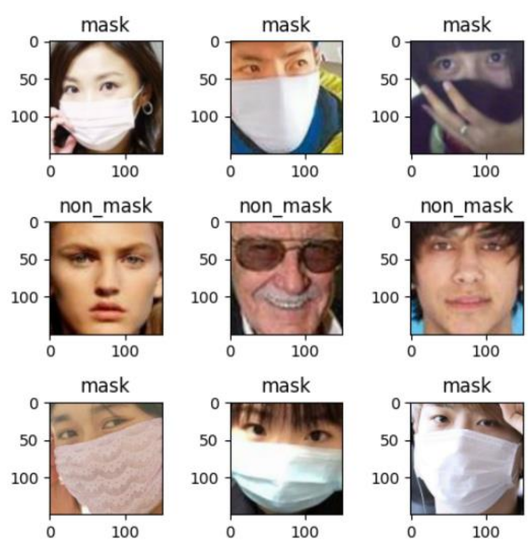


Figure 4. 3: Samples of facedatahybrid Dataset

4.1.4 LFW Dataset with Masks

The dataset [61] was created due to the surge in COVID-19 cases. The dataset is comprised of images of celebrities sourced from the Labeled Faces in the Wild Dataset, wherein each image is covered up with a face mask. This dataset serves as a base for novel tasks such as mask detection and masked face recognition. The dataset includes two distinct folders, Masked Faces and Unmasked Faces, each with images of 421 celebrities. Masks are applied in two colors, blue and black, and in three positions: low, mid, and high. These positions represent the various ways in which individuals wear masks.

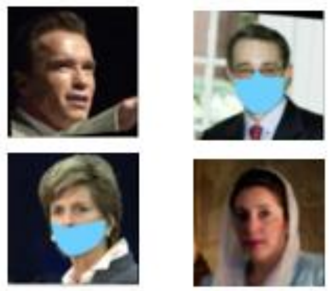


Figure 4. 4: Samples of LFW Dataset with Masks

4.1.5 MFR2

The dataset known as Masked Faces in the Real World for Face Recognition (MFR2) is a compact collection that features 53 prominent personalities from the realms of celebrity and politics [60]. A total of 268 images that have been aligned and preprocessed are included in this dataset. Each individual identity in the given dataset possesses an average of five distinct images. The dataset comprises both masked and unmasked faces of the respective identities. Further, the dataset has been preprocessed, involving image dimension and face alignment adjustments. Each of the images contained within the dataset has a standardized dimension of 160 x 160 pixels and a color depth of 3.

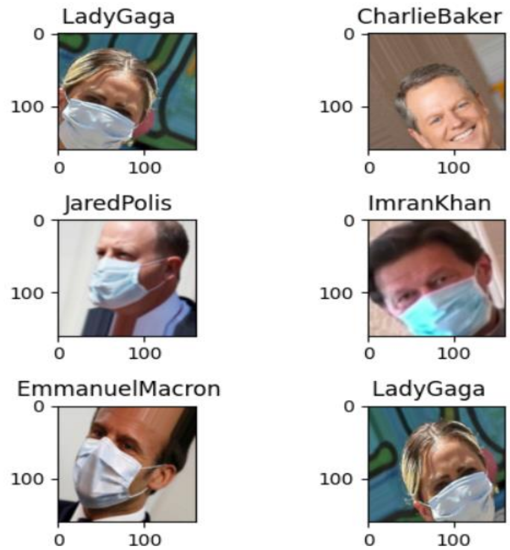


Figure 4. 5: Samples of MFR2 Dataset

4.2 Performance Measures

In order to achieve classification goals, the implementation of two tasks, namely masked face detection, and recognition, has been carried out. This section focuses on discussing the performance measures utilized to evaluate the proposed model, VGG16 and InceptionV3, which has been applied to the datasets referenced earlier. The evaluation of the model takes into account various parameters, including accuracy, sparse-categorical cross-entropy loss function, sensitivity, precision and F1-score which are presented in equations (1), (2), (3), (4) and (5), respectively.

$$Accuracy = \frac{TP + TN}{TP+TN+FP+FN} \ldots\ldots. (1)$$

Accuracy, which denotes the ratio of correct classifications to the total number of classifications, is a critical parameter in this context.

$$L(y, \hat{y}) = -\sum_{j=0}^{M}\sum_{i=0}^{N}(y_{ij} * log(\hat{y}_{ij})) \ \ldots\ldots. (2)$$

The Categorical Cross-Entropy loss function serves to compare the distribution of predictions with the true distribution. To attain this, the probability of the true class is set to 1 and 0 for all other classes, representing the actual and predicted probabilities, respectively. The loss function for both categorical cross entropy and sparse categorical cross entropy is identical, as mentioned above. The Yi format, which denotes the true labels, is the only factor that distinguishes it, with sparse categorical cross entropy being utilized for the integer encoding of the labels.

$$SEN = Recall = \frac{TP}{TP+FN} \ldots\ldots (3)$$

Sensitivity is calculated by dividing the sum of true positives across all classes by the sum of true positives and false negatives across all classes.

$$PRE = \frac{TP}{TP+FP} \ldots\ldots (4)$$

Precision is defined as the ratio of the product of precision and recall to the sum of precision and recall.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \ldots\ldots (5)$$

The F1-score is a performance metric that combines the precision and recall of a classifier into a single measure by calculating their harmonic mean. It is generally utilized to evaluate the performance of two classifiers. Let us assume that classifier A has a greater recall, and classifier B possesses a higher precision. In this context, the F1-scores of both classifiers can be employed to identify the one that produces better outcomes.

4.3 Experimental Results

The pre-trained VGG16 and InceptionV3 architectures are executed using the TensorFlow framework and trained on a single NVIDIA GeForce GTX 1060 6GB GPU. In this section, in this section, the results of this study are explained in two parts: the detection and recognition of masked faces using a pre-trained model with transfer learning and a unified model capable of both detecting and recognizing masked as well non-masked images.

4.3.1 Performance Evaluation on Masked Face Detection

The datasets used for masked face detection are MDMFR dataset, Face Mask Detection Dataset and facedatahybrid and LFW Dataset with Masks utilized as a blind test set.

For preprocessing, the face images are first adjusted to a resolution of 150 x 150. Labels encoded using integer encoding for each of two classes. Each pixel in the RGB images, which range from 0 to 255, is normalized by dividing it by 255. As for the fine-tuned optimal hyperparameters involved in models, the Stochastic Gradient Descent (SGD) optimizer is employed with a momentum of 0.9. The initial learning rate is set at 0.0001 and dropout value is 0.3. The training epochs are determined to be 50, where the batch size is set to 20. The dimensionality of the masked facial features of the models is set to 256, where the 'SoftMax' activation function is used for the output of two classes 'with_mask' as 0 or 'without_mask' as 1. During the training process, 70% of the dataset is utilized, with a 20% validation set. The fine-tuned VGG16 with transfer learning evaluated on fmdd, MDMFR and facedatahybrid datasets. Achieved results are presented in Table 4.1.

Table 4. 1: Detection Model Performance Measures on Training, Validation and Testing Datasets

Dataset	Training Accuracy	Val-Accuracy	Test-Accuracy	Train-Loss	Val-Loss	Test-Loss	Precision	Recall	F1-score
Face Mask Detection Dataset	98.47%	96.36%	96.83%	0.0442	0.0959	0.0748	0.97	0.97	0.97
MDMFR Dataset	99.65%	99.75%	99.6%	0.0132	0.0101	0.0178	100	100	100
facedatahybrid	97.56%	97.27%	97.05%	0.0687	0.0732	0.0847	0.97	0.97	0.97

The accuracy and loss trend of the training and validation over 50 epochs can be observed from the plot in figures 4.6, 4.7 and 4.8. The proposed model achieved its highest accuracy between 40-50 epochs with respect to minimum validation loss and reached an impressive accuracy, surpassing previous research findings.

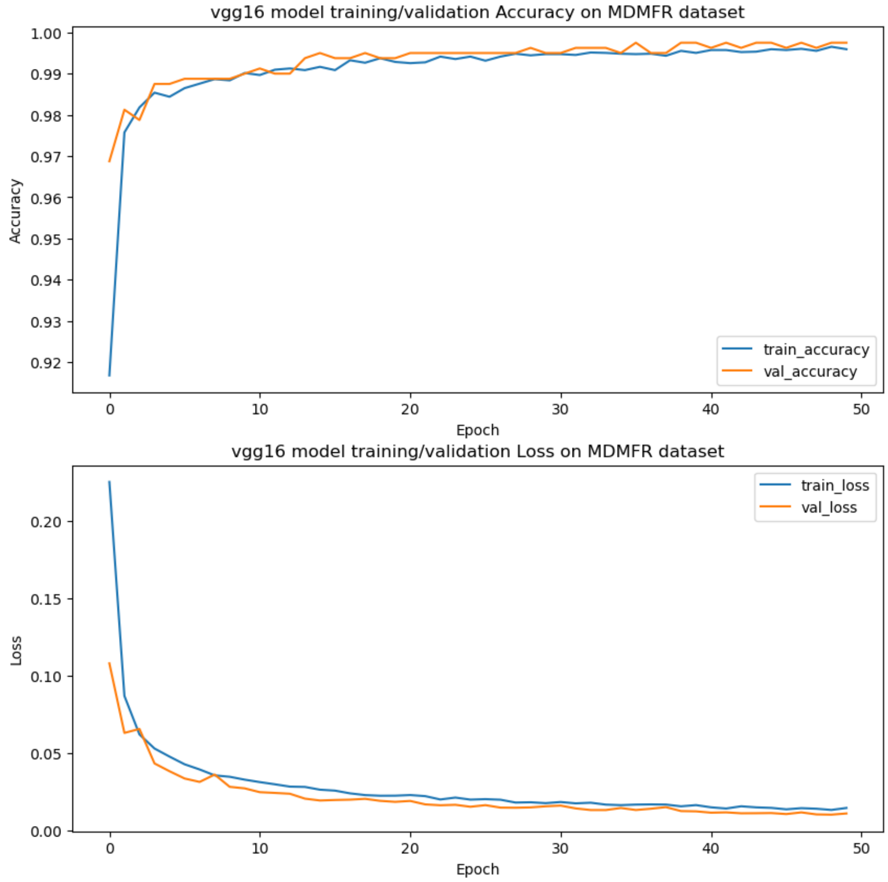


Figure 4. 6: Accuracy and Loss function graphs of MDMFR training and validation sets

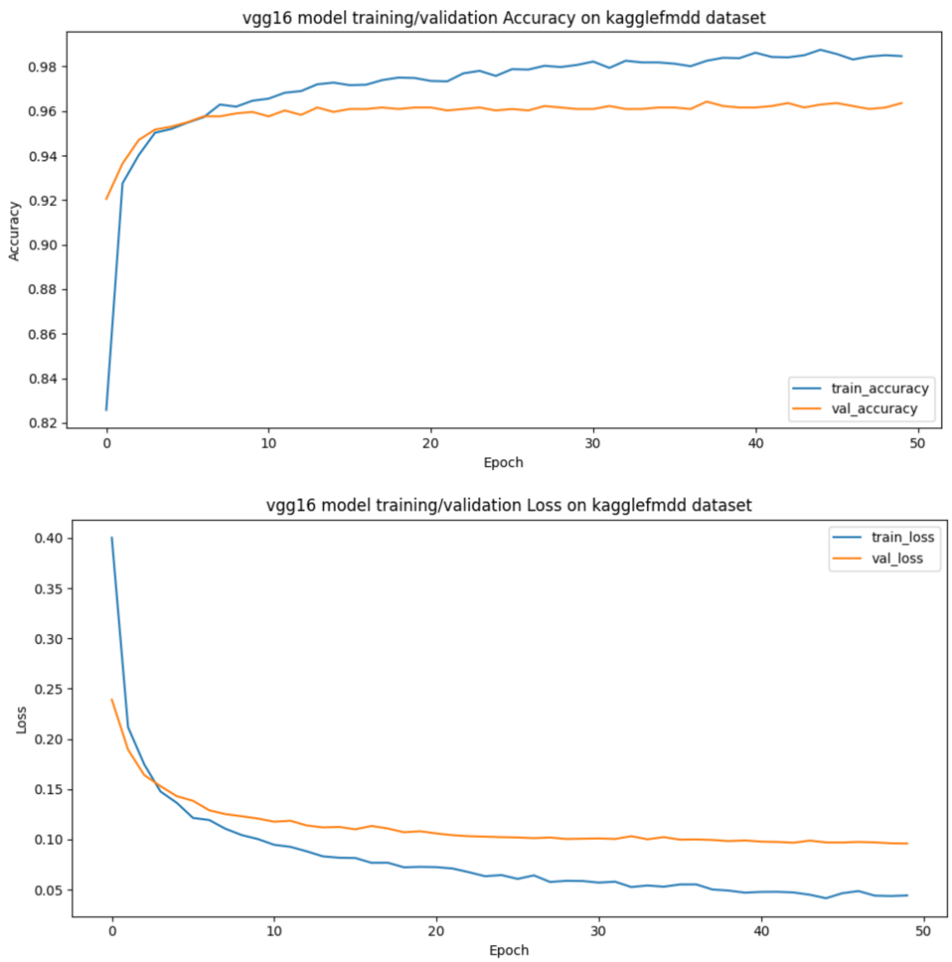
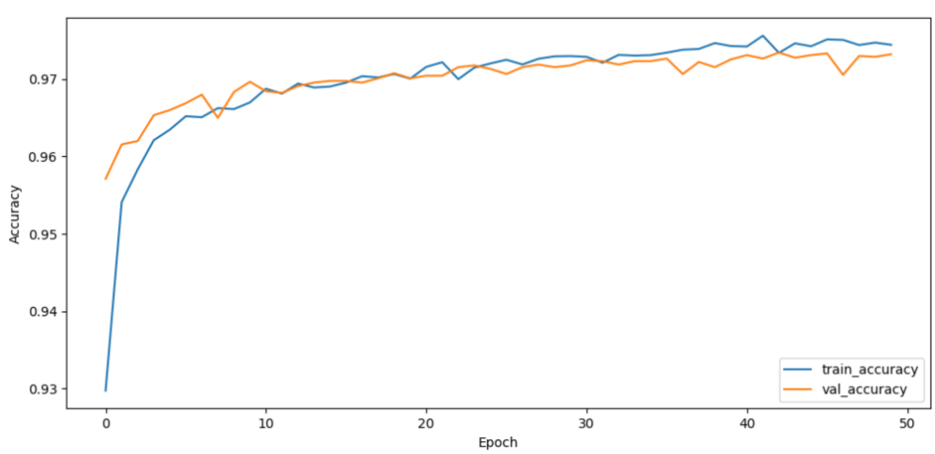


Figure 4. 7: Accuracy and Loss function graphs of fmdd training and validation sets



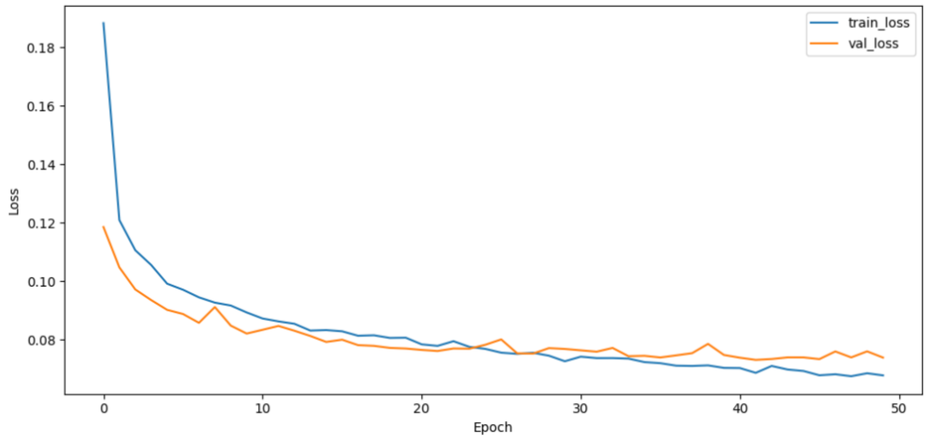


Figure 4. 8: Accuracy and Loss function graphs of facedatahybrid training and validation sets

From the results we can see that our proposed model for detection performed well on training as well as testing sets of detection datasets, giving accuracy of 96.83, 99.6% and 97.05% respectively. To visualize the accuracy of each class, confusion matrix has been utilized, which is a tabular representation of the model's predicted performance. Figures 4.9, 4.10 and 4.11 display the confusion matrix of each dataset, where the diagonal elements indicate the correct recognition accuracy for each labeled class. Specifically, '0' corresponds to with_mask, while '1' corresponds to without_mask.

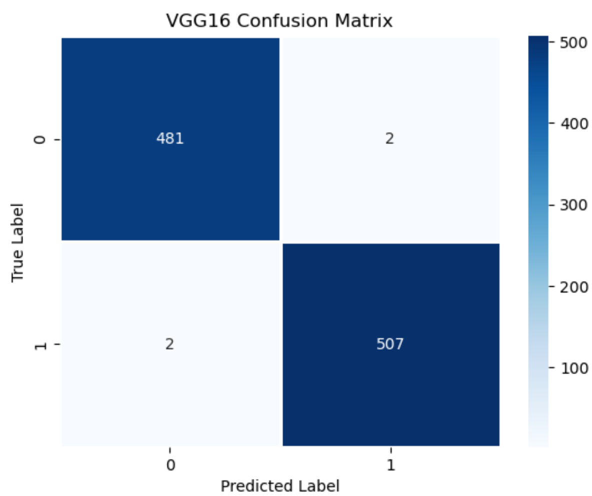


Figure 4. 9: Confusion Matrix of MDMFR testing data

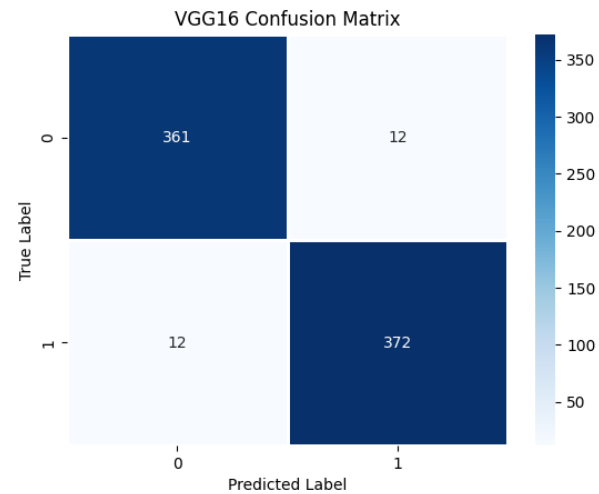


Figure 4. 10: Confusion Matrix of Face Mask Detection Dataset testing data

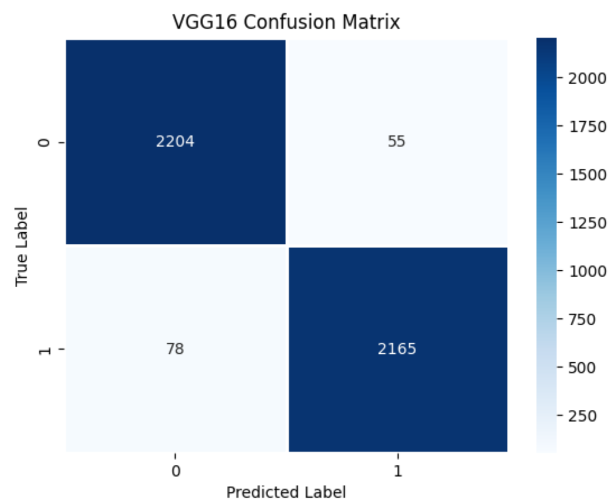


Figure 4. 11: Confusion Matrix of facedatahybrid testing data

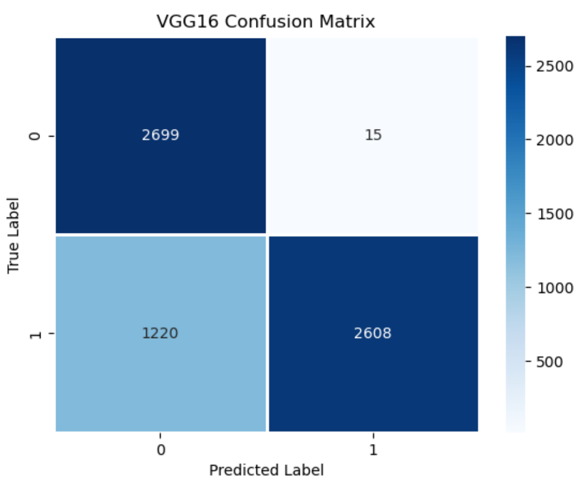
4.3.1.1 Cross-dataset Evaluation

Cross-dataset evaluation is a methodology that is utilized to evaluate the efficiency of a machine learning model on a dataset that differs from the one on which the model was originally trained. The purpose of this approach is to assess the model's capacity to generalize to new data. The trained weights that have been learned on the training datasets tested to the other datasets (Table 4.2) one by one to assess the performance measures. As it can be seen that fmdd dataset weights best fit on LFW dataset where MDMFR dataset weights performed well on facedatahybrid dataset and facedatahybrid dataset weights results high testing accuracy on MDMFR dataset. It also shows how well our pretrained VGG16 model with fine tuning has performed in an unknown environment as well.

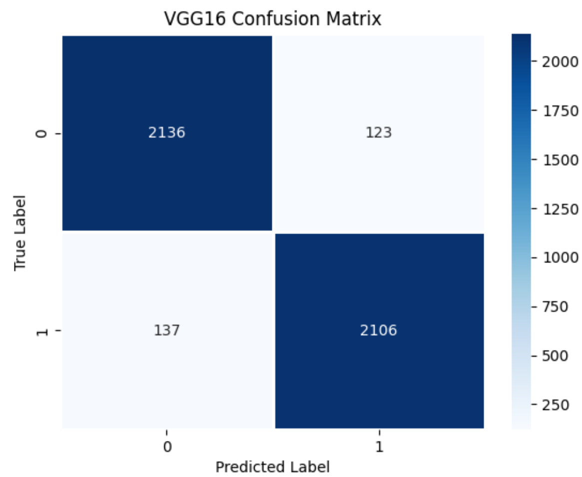
Table 4. 2: Detection Model Performance Measures on Cross-Datasets

Training Dataset	Testing Dataset	Accuracy	Loss	Precision	Sensitivity	F1-score
Face Mask Detection Dataset	MDMFR Dataset	94.25%	0.1374	0.95	0.94	0.94
	facedatahybrid	92.58%	0.2241	0.93	0.93	0.93
	LFW dataset with masks	99.05%	0.0253	0.99	0.99	0.99
MDMFR Dataset	Face Mask Detection Dataset	81.1%	0.783	0.87	0.81	0.81
	facedatahybrid	94.22%	0.1816	0.94	0.94	0.94
	LFW dataset with masks	75.13%	0.748	0.84	0.75	0.74
facedatahybrid	MDMFR Dataset	97.88%	0.069	0.98	0.98	0.98
	Face Mask Detection Dataset	81.95%	0.516	0.86	0.82	0.82
	LFW dataset with masks	75.27%	0.604	0.84	0.75	0.74

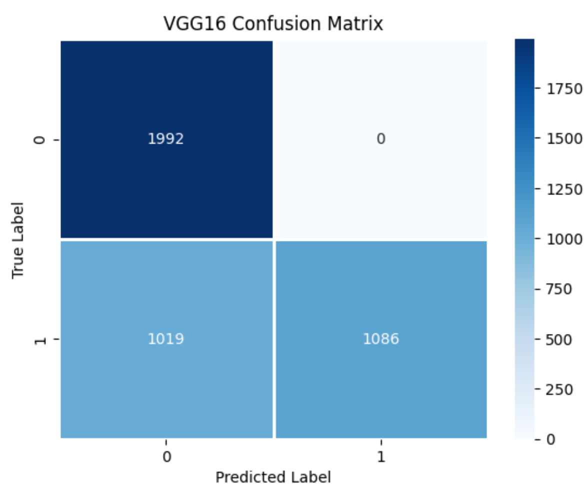
The performance of the proposed model can be visualized in the form of confusion matrix, which is a 2D matrix by the comparison of predicted and actual class. In figures 4.12, 4.13, 4.14 we can see how many times trained weights of each dataset predicted class were correct or incorrect by comparing it to the actual class.



(a)

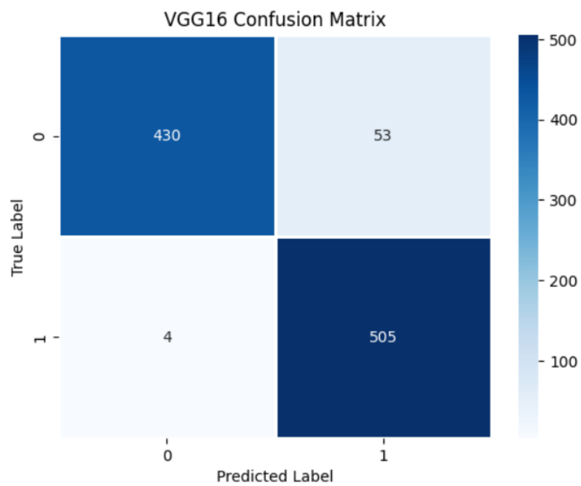


(b)

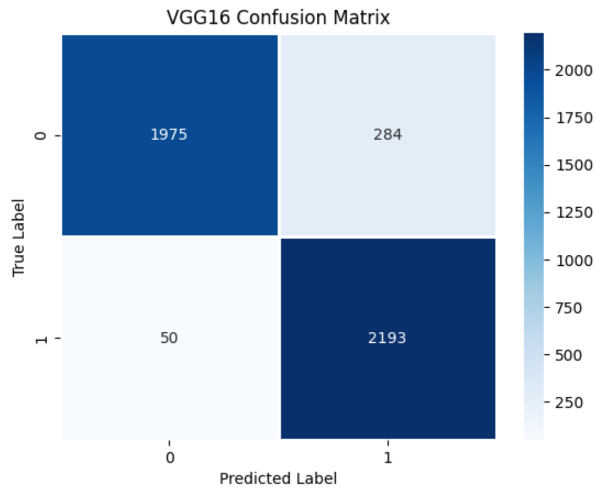


(c)

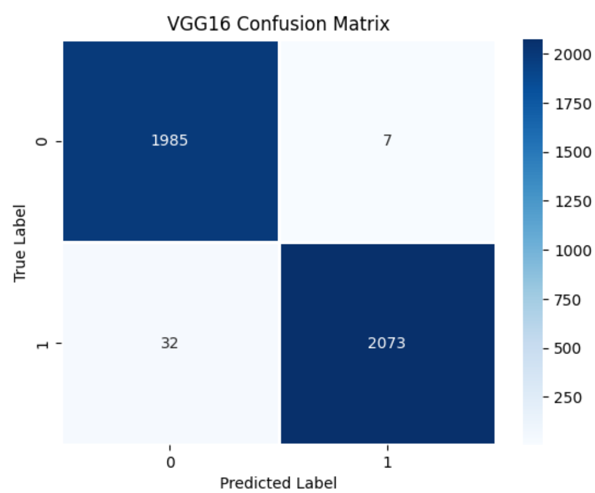
Figure 4. 12: Confusion Matrix of MDMFR Testing on (a) fmdd Dataset (b) facedatahybrid Dataset (c) LFW with Masks Dataset



(a)

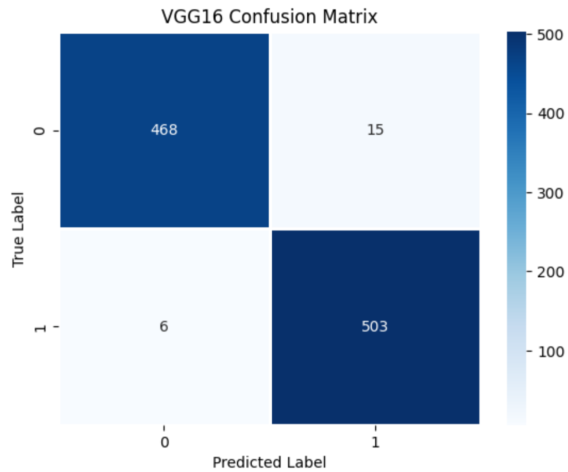


(b)

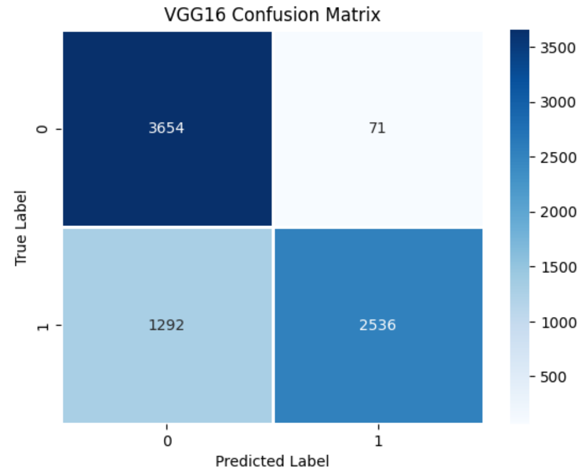


(c)

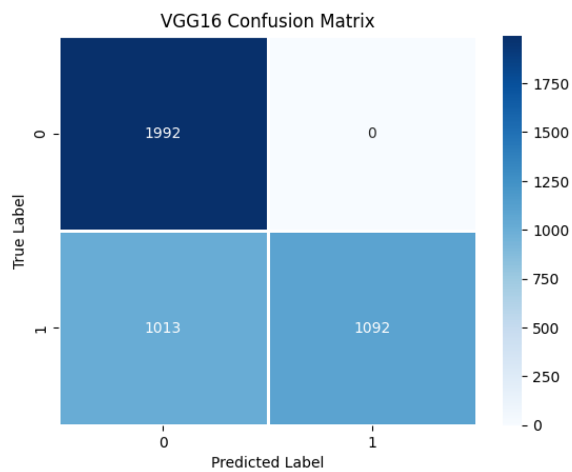
Figure 4. 13: Confusion Matrix of fmdd Testing on (a) MDMFR Dataset (b) facedatahybrid Dataset (c) LFW with Masks Dataset



(a)



(b)



(c)

Figure 4. 14: Confusion Matrix of face data hybrid Testing on (a) MDMFR Dataset (b) fmdd Dataset (c) LFW with Masks Dataset

4.3.1.2 Comparison of Different Methodologies

The comparison of our proposed pretrained model and other techniques from the existing work are shown in Table 4.3 below. We can see that our proposed pretrained network with transfer learning performed well for both the face mask detection dataset as well as for facedatahybrid with 96.83% and 97.05%.

Table 4. 3: Comparison of Detection Model with Existing Technique

Year	Author	Technique	Dataset	Accuracy
2023	Ambreen Sabha et al. [62]	CoSumNet model applied transfer learning using a pre-trained ResNet-50 deep learning model	Face mask detection dataset from Kaggle	97.73%
2023	**Proposed**	**Fine-tuned VGG16 with transfer learning**	**Face mask detection dataset from Kaggle**	**96.83%**
2023	Ambreen Sabha et al. [62]	CoSumNet model applied transfer learning using a pre-trained ResNet-50 deep learning model	facedatahybrid	97.10%
2023	**Proposed**	**Fine-tuned VGG16 with transfer learning**	**facedatahybrid**	**97.05%**

4.3.2 Performance Evaluation on Masked Face Recognition

There are two main approaches to the recognition of masked faces:

- **Face mask detection followed by face recognition:** The first approach involves the detection of a face mask within a given image, followed by the cropping of the face and its submission to a facial recognition system to facilitate identification.
- **Direct masked face recognition:** The second method, direct masked face recognition, is designed to recognize faces despite partial or complete obstruction by masks. This method presents a more difficult task but has the potential to deliver more accurate results than the two-step method.

For preprocessing, the face images are first adjusted to a resolution of 160 x 160. Each pixel in the RGB images, which range from 0 to 255, is normalized by dividing it by 255. The Adagrad optimizer is employed, with the learning rate set to 0.001 and dropout value of 0.3.

The training batch size is determined to be 5. The dimensionality of the masked facial features of the model is set to 256, where the 'SoftMax' activation function is used for the output of 53 classes representing each unique person's names given in the dataset.

During the training process, 75% of the dataset is utilized, with a 15% validation set. The pre-trained network VGG16 with transfer learning is applied to publicly available dataset MFR2, and the values of performance measures on Training, Validation, and Testing datasets are shown in Table 4.4.

Table 4. 4: Recognition Model Performance Measures on Training, Validation and Testing Sets

Dataset	Training Accuracy	Validation Accuracy	Testing Accuracy	Training Loss	Validation Loss	Precision	Sensitivity	F1-score
MFR2	100%	96.04%	95%	0.0739	0.2958	0.94	0.93	0.93

In Figure 4.15, we can see the correct labels predicted with respect to actual labels of both masked and non-masked faces in the test result along with the display of their respective images.

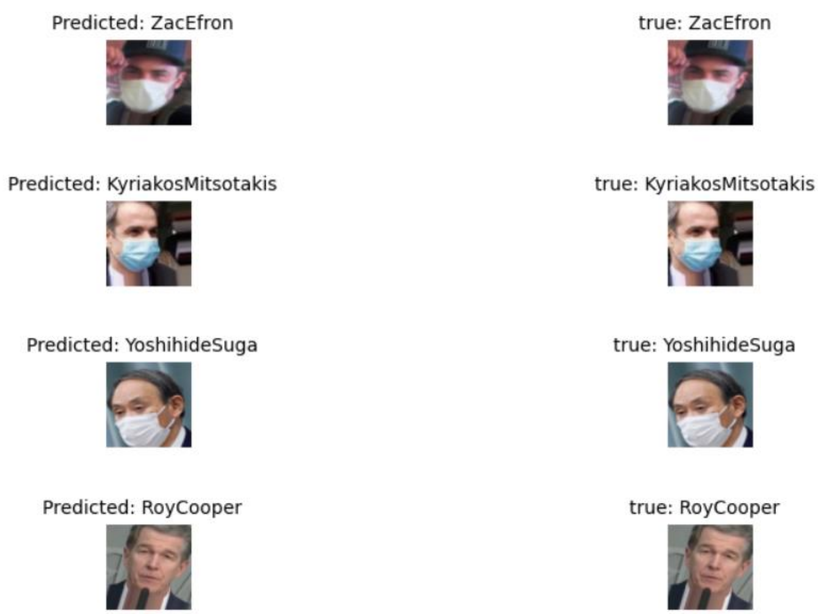


Figure 4. 15: Samples of Recognition Test Result

The graphical representation of the training and validation accuracies, as well as the loss trend over the course of 50 epochs, is depicted in Figure 4.16. It is noteworthy that maximal accuracy was achieved during the 10-20 epoch interval with respect to minimum validation loss, beyond which the model's performance remained constant. The analysis indicates that the training accuracy is at an impressive level of 93%, surpassing all previous research endeavors.

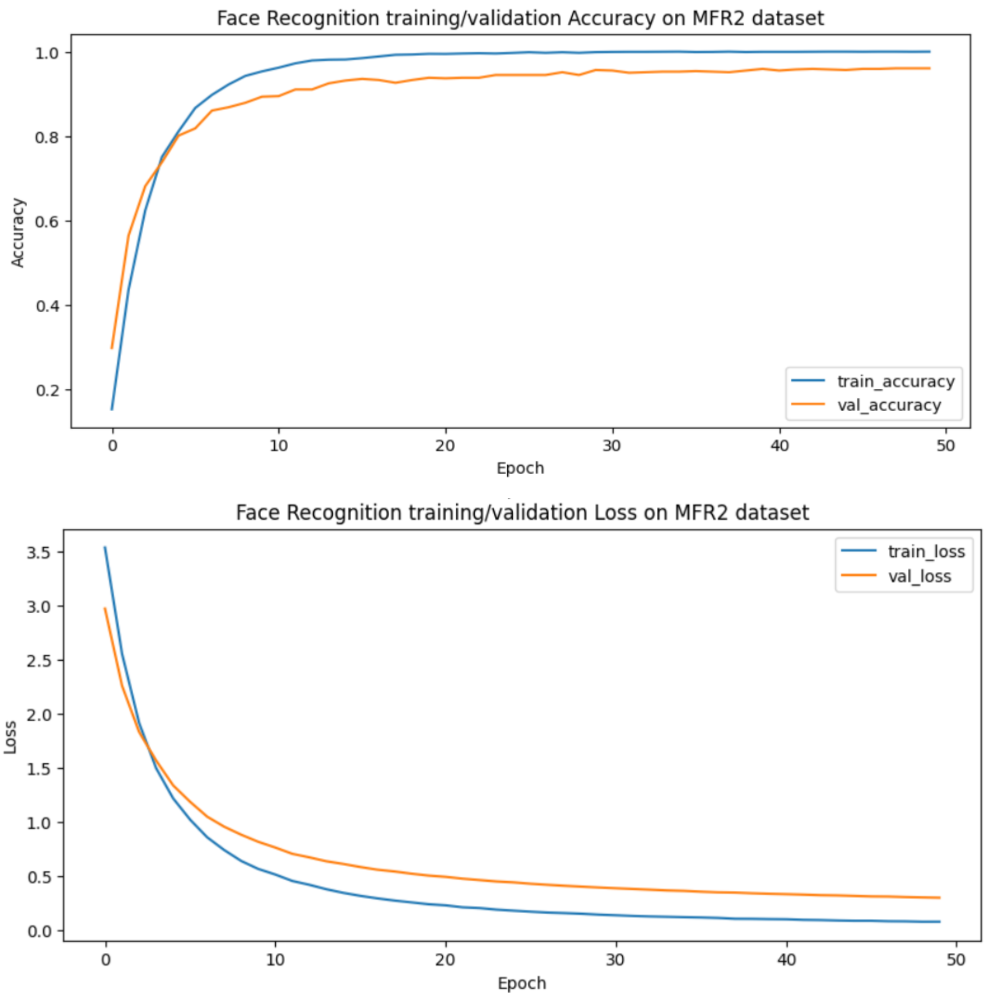


Figure 4. 16: Accuracy and Loss function graphs of MFR2 training and validation sets

4.3.2.1 Comparison of Different Methodology

The proposed model compared with the existing deep learning model and result mentioned in Table 4.5 presents that our proposed model achieved better accuracy.

Table 4. 5: Comparison of Recognition Model with Existing Technique

Year	Author	Technique	Accuracy
2023	Yiming Ge et al. [9]	Convolutional Visual Self-Attention Network	96.36%
2023	**Proposed Model**	**Fine-tuned VGG16 with transfer learning**	**95%**

4.3.3 Performance Evaluation of Unified Model

For the preprocessing of the proposed unified model, the face images are first adjusted to a resolution of 160 x 160. Labels encoded using integer encoding for each of two classes mask and non-mask and for 53 labels as person names. Each pixel in the RGB images, which range from 0 to 255, is normalized by dividing it by 255. As for the fine-tuned optimal hyperparameters involved in models, the Adagrad optimizer is employed with the initial learning rate set at 0.0001. The training epochs are determined to be 25, where the batch size is set to 5. The dimensionality of the masked facial features of the models is set to 128 and 256, where the two ‘SoftMax’ layers of activation function are used for each classification. One for the binary output of two classes ‘with_mask’ as 0 or ‘without_mask’ as 1 with the dropout value is 0.2 for the detection dense layer. and the other for the person recognition as multiclassification. During the training process, 80% of the dataset is utilized, with a 10% validation set. The fine-tuned InceptionV3 with transfer learning evaluated on MFR2 dataset for both masked face detection as well as recognition. Achieved results are presented in Table 4.6.

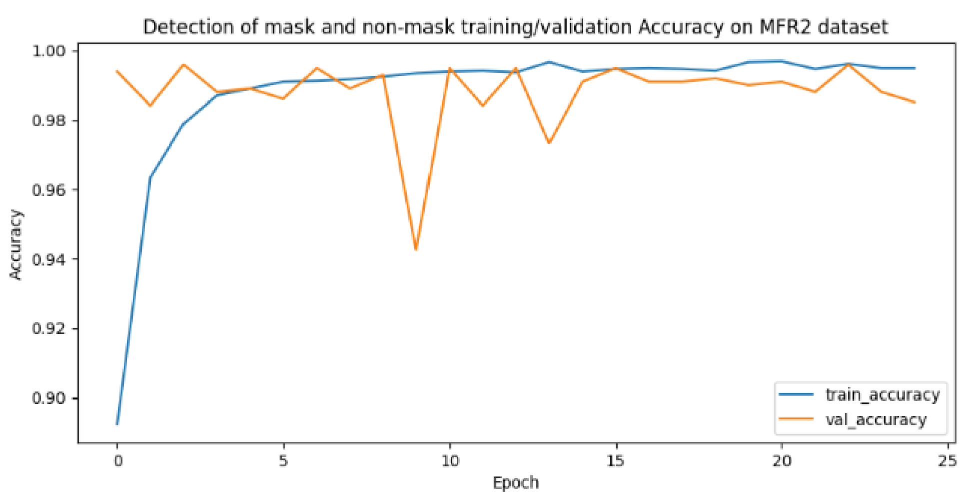
Table 4. 6: Unified Model Performance Measures on Training, Validation and Testing Datasets

Unified Model	Training Accuracy	Validation Accuracy	Testing Accuracy	Training Loss	Validation Loss	Precision	Sensitivity	F1-Score
Detection	99.5%	99.11%	99%	0.0150	0.0352	0.99	0.99	0.99
Recognition	97.57%	98.91%	98%	0.0874	0.0356	0.98	0.98	0.98

In Figure 4.17, we can see the correct detection and recognition labels predicted with respect to actual labels of both masked and non-masked faces in the test result along with the display of their respective images.

Figure 4. 17: Samples of Unified Model Test Result

The graphical representation of the training and validation accuracies for the masked face detection and recognition, as well as the loss trend over the course of 25 epochs, is depicted in Figure 4.18 and 4.19 respectively.



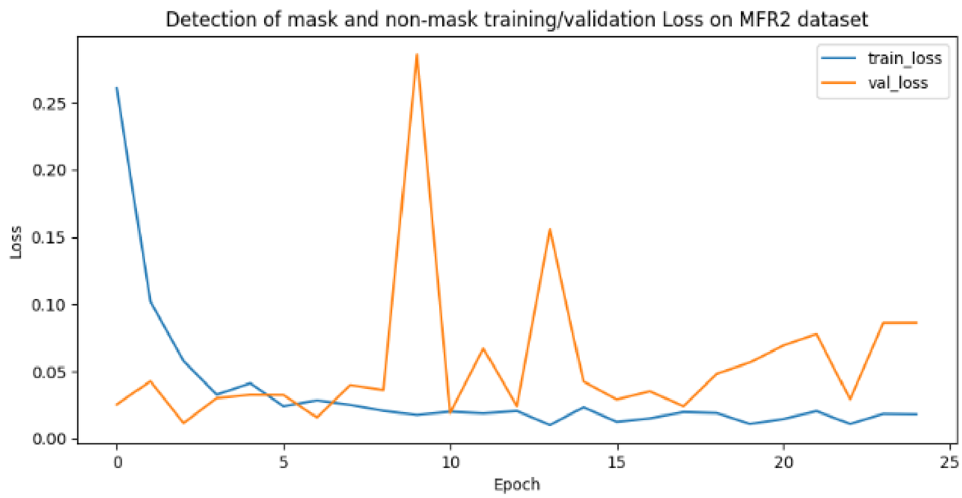
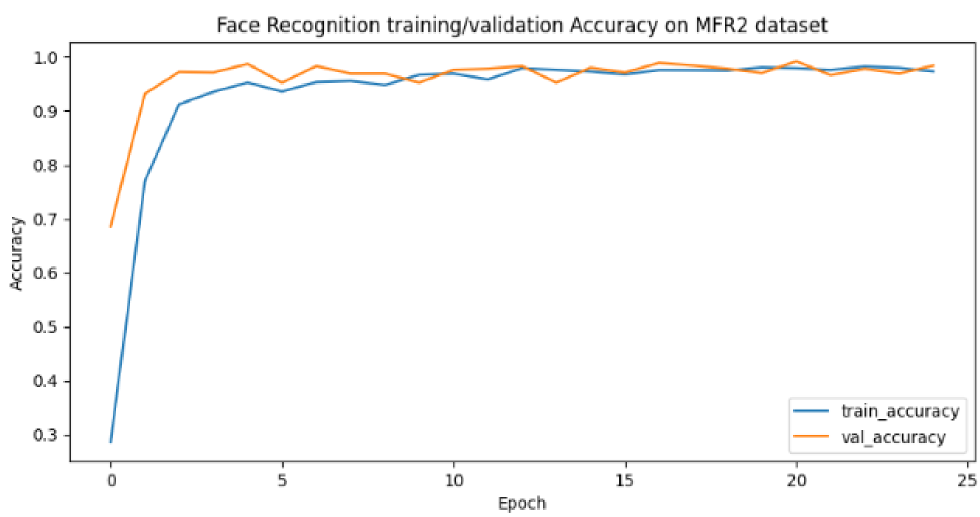


Figure 4. 18: Accuracy and Loss Function Graphs of Training/Validation Datasets for Masked Face Detection



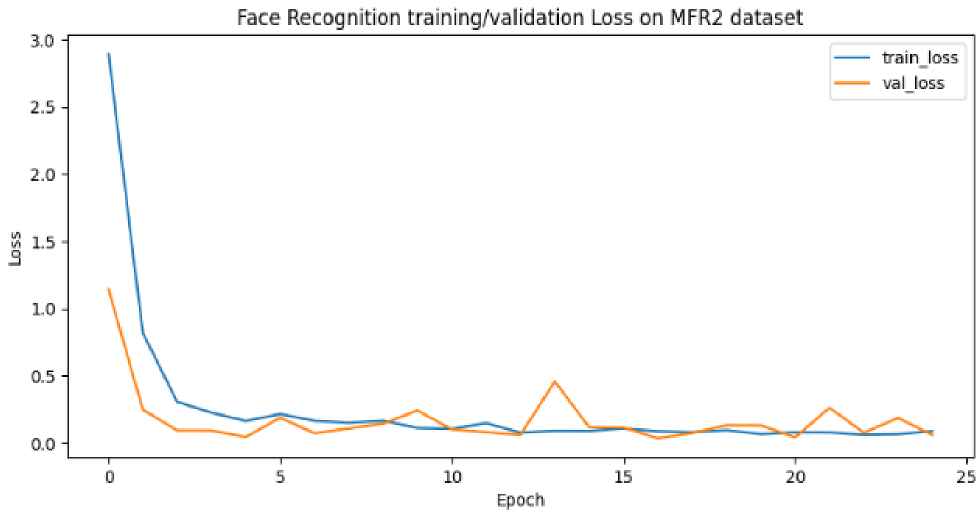


Figure 4. 19: Accuracy and Loss Function Graphs of Training/Validation Datasets for Masked Face Recognition

After using pretrained InceptionV3 model as a feature extractor, passing them through two dense layers for masked face detection and recognition each gives learned weights. At each epoch the loss from each classification learned weights then combined by giving some weightage e.g., alpha[det-loss] + beta[rec-loss] where alpha=1.0 and beta=2.0. Based on this calculated combined loss (Figure 4.20) the learned weights will propagate back to update and then be used for the testing.

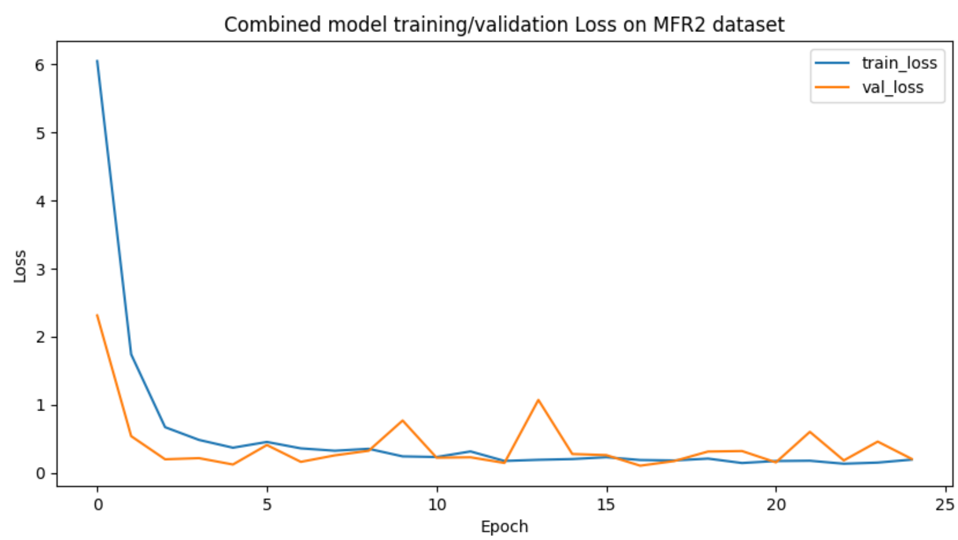


Figure 4. 20: Combined Training/Validation Loss Graph for the Unified Model

The result of the proposed unified model shows that it not only performed well for detection but also recognition by giving an overall accuracy of 99% and 98% respectively. To analyze the performance of model accuracy with respect to each class Confusion matrix is plotted which is the tabular form to visualize, where diagonal elements represents the correct accuracy of each class predicted by the model as shown in Figure 4.21 and 4.22, the correct detection accuracy for each class here label '0' represents masked and '1' represents non-masked, where for the correct recognition labels are 0 to 52 representing each unique person name.

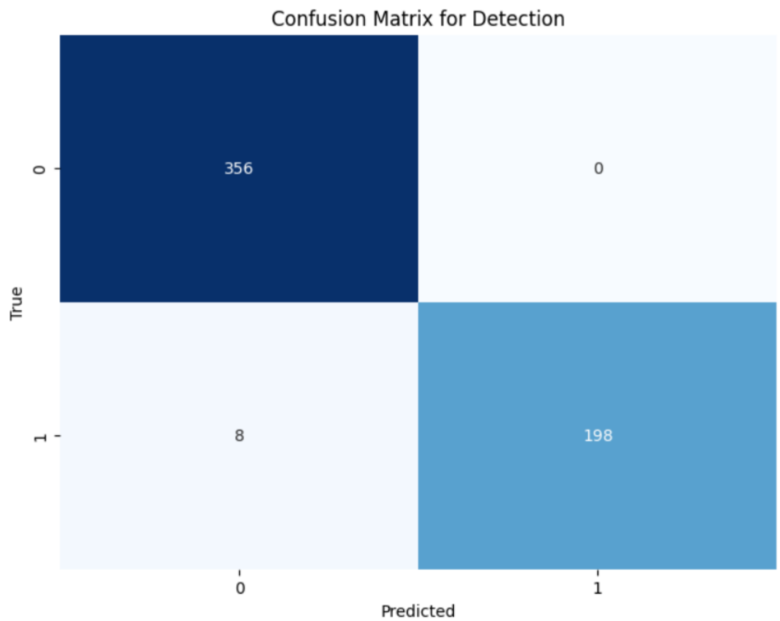


Figure 4. 21: Confusion Matrix of Unified Model for Detection

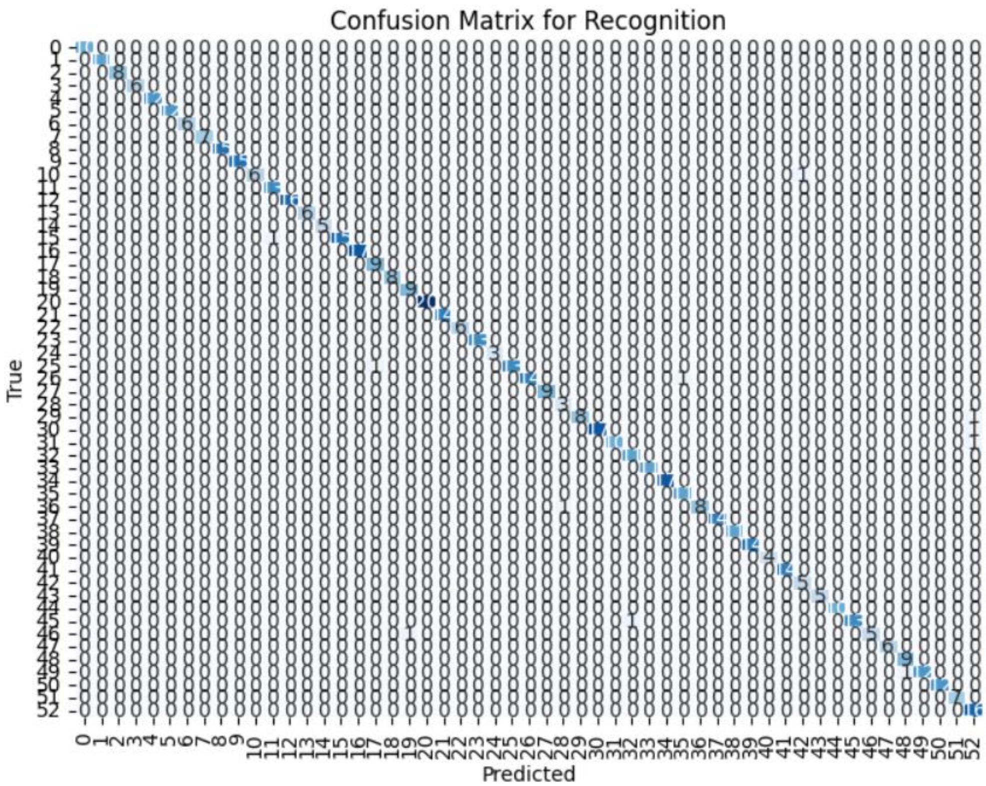


Figure 4. 22: Confusion Matrix of Unified Model for Recognition

The proposed unified model used for both masked face detection and recognition performed well giving high accuracy (Table 4.7) as compared to individual classification models. By giving 99% accuracy for masked face detection and 98% accuracy for masked face recognition it is proved that our unified model comprised of a pretrained InceptionV3 model with transfer learning can be used for masked face detection and recognition from images.

Table 4. 7: Comparison of Unified Model with Proposed individual Classification Models

Proposed Model	**Technique**	**Accuracy**
Detection Model	Pretrained VGG16 with Transfer Learning	99.6%
Recognition Model	Pretrained VGG16 with Transfer Learning	95%
Unified Model	**Pretrained InceptionV3 with Transfer Learning**	**99% (det)** **98% (rec)**

CHAPTER 5: CONCLUSION & FUTURE WORK

5.1 Conclusion

In conclusion, we have addressed the masked face recognition challenge that emerged due to the COVID-19 pandemic. We have developed a unified model for masked face detection and recognition that achieves high accuracy of 99% and 98% on challenging dataset. The model is robust to various occlusions and environmental conditions, making it suitable for real-world applications. Our work contributes to the growing frame of research on masked face recognition. We have explored diverse datasets, implemented different pretrained networks, and fine-tuned them to handle masked faces effectively. The proposed model demonstrates the potential of transfer learning for masked face detection and recognition.

The model can detect masked faces in various conditions, including different lighting conditions, backgrounds, and facial expressions. It can also recognize the identities of masked individuals, even when their facial features are partially obscured. This makes the model a valuable tool for applications such as security screening, access control, and surveillance. We believe that the proposed model represents a significant step forward in the field of masked face detection and recognition. The model has the potential to be used in a variety of applications where accurate masked face recognition is required. We encourage further research to improve the model's performance and extend its capabilities.

5.2 Contribution

- The present study involves the collection and pre-processing of diverse publicly available masked face images, including those with and without face masks, to facilitate their detection and recognition.
- The study further aims to develop a singular CNN-based framework, specifically VGG16 pre-trained, that can not only be used to classify masked/non-masked but also recognize various masked faces.
- Additionally, the study intends to develop a unified model of a pre-trained Network for the detection and recognition of masked face images.

- The performance of the model is evaluated based on various parameters, such as accuracy, precision, sensitivity, f1-score, and loss function.

5.3 Future Work

Previously suggested approaches have been limited using custom datasets or a lack of diverse image samples, which has had an impact on the quality of the classification outcomes. To address this challenge, we have proposed a model that utilizes a large set of images sourced from various online resources, resulting in improved quality and precision of classification outcomes. Unlike previous models that were designed to only recognize or detect masked images at a time, our proposed framework consists of a single pretrained-based model that can recognize multiple scenarios, thereby eliminating the need for multiple software or hardware resources. In the future, this framework may be expanded to address other classification challenges by incorporating additional classes, such as incorrectly masked faces or the inclusion of multiple masked faces within a single image.

REFERENCES

1. R. K. Shukla and A. K. Tiwari, “Masked Face Recognition Using MobileNet V2 with Transfer Learning,” *Computer Systems Science and Engineering*, vol. 45, no. 1, pp. 293–309, Jan. 2023, doi: 10.32604/csse.2023.027986.
2. S. Nadeem, A. Javed, M. A. Ghazanfar, A. Alsufyani, and S. Bourouis, “A novel DeepMaskNet model for face mask detection and masked facial recognition,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, pp. 9905–9914, Nov. 2022, doi: 10.1016/j.jksuci.2021.12.017.
3. A. A. Abdulmunem, N. D. Al-Shakarchy, and M. S. Safoq, “Deep learning based masked face recognition in the era of the COVID-19 pandemic,” *Ijece*, vol. 13, no. 2, p. 1550, Apr. 2023, doi: 10.11591/ijece.v13i2.pp1550-1559.
4. B. Huang *et al.*, “PLFace: Progressive Learning for Face Recognition with Mask Bias,” *Pattern Recognition*, vol. 135, p. 109142, Mar. 2023, doi: 10.1016/j.patcog.2022.109142.
5. W.-C. Cheng, H.-C. Hsiao, and L.-H. Li, “Deep Learning Mask Face Recognition with Annealing Mechanism,” *Applied Sciences*, vol. 13, no. 2, p. 732, Jan. 2023, doi: 10.3390/app13020732.
6. M. A. Joodi, “Increasing validation accuracy of a face mask detection by new deep learning model-based classification,” *Azeez Joodi | Indonesian Journal of Electrical Engineering and Computer Science*, Jan. 01, 2023. https://ijeecs.iaescore.com/index.php/IJEECS/article/view/28942/16951
7. B. A. Kumar and M. Bansal, “Face mask detection on photo and Real-Time video images using Caffe-MobileNetV2 transfer learning,” *Applied Sciences*, vol. 13, no. 2, p. 935, Jan. 2023, doi: 10.3390/app13020935.
8. S. Kumar, D. Yadav, H. Gupta, M. Kumar, and O. P. Verma, “Towards smart surveillance as an aftereffect of COVID-19 outbreak for recognition of face masked individuals using YOLOv3 algorithm,” *Multimedia Tools and Applications*, vol. 82, no. 6, pp. 8381–8403, Jul. 2022, doi: 10.1007/s11042-021-11560-1.
9. Y. Ge, H. Liu, J. Du, Z. Li, and Y. Wei, “Masked face recognition with convolutional visual self-attention network,” *Neurocomputing*, vol. 518, pp. 496–506, Jan. 2023, doi: 10.1016/j.neucom.2022.10.025.

10. “A Real-Time CNN-Based lightweight mobile masked face recognition system,” *IEEE Journals & Magazine | IEEE Xplore*, 2022. https://ieeexplore.ieee.org/document/9795021
11. “Beyond Masks: on the generalization of masked face recognition models to occluded face recognition,” *IEEE Journals & Magazine | IEEE Xplore*, 2022. https://ieeexplore.ieee.org/document/9857925
12. I. Javed *et al.*, “Face mask detection and social distance monitoring system for COVID-19 pandemic,” *Multimedia Tools and Applications*, vol. 82, no. 9, pp. 14135–14152, Sep. 2022, doi: 10.1007/s11042-022-13913-w.
13. W. Hariri, “Efficient masked face recognition method during the COVID-19 pandemic,” *Signal, Image and Video Processing*, vol. 16, no. 3, pp. 605–612, Nov. 2021, doi: 10.1007/s11760-021-02050-w.
14. H. N. D. Vu, M. T. T. Nguyen, and C. Pham, “Masked face recognition with convolutional neural networks and local binary patterns,” *Applied Intelligence*, vol. 52, no. 5, pp. 5497–5512, Aug. 2021, doi: 10.1007/s10489-021-02728-1.
15. H. Dmour, “Masked Face Detection and Recognition System based on deep learning algorithms,” Uts, May 2023, [Online]. Available: https://www.academia.edu/102036575/Masked_Face_Detection_and_Recognition_System_Based_on_Deep_Learning_Algorithms
16. L. Queiroz, “Biometrics in the Time of Pandemic: 40% Masked Face Recognition Degradation can be Reduced to 2%,” 2022. https://www.semanticscholar.org/paper/Biometrics-in-the-Time-of-Pandemic%3A-40-Masked-Face-Queiroz-Lai/595d16c5f2e41a739ad15a24bff9623a8febec92
17. M. K. Kumcu, S. Tezcan, B. Ölmez, F. N. D. Çelik, and C. Yücesan, “Masked face recognition in patients with relapsing–remitting multiple sclerosis during the ongoing COVID-19 pandemic,” *Neurological Sciences*, vol. 43, no. 3, pp. 1549–1556, Jan. 2022, doi: 10.1007/s10072-021-05797-9.
18. V. Sharma, “Masked face recognition,” 2021. https://www.semanticscholar.org/paper/Masked-Face-Recognition-Sharma-Gangaraju/b9c69124ecc6f2b92889c5d3ee4252b3da75c795

19. R. Golwalkar and N. Mehendale, “Masked-face recognition using deep metric learning and FaceMaskNet-21,” *Applied Intelligence*, vol. 52, no. 11, pp. 13268–13279, Feb. 2022, doi: 10.1007/s10489-021-03150-3.
20. G. Yuan, H. Zheng, and J. Dong, “MSML: Enhancing Occlusion-Robustness by Multi-Scale Segmentation-Based mask learning for face Recognition,” *Proceedings of the ... AAAI Conference on Artificial Intelligence*, vol. 36, no. 3, pp. 3197–3205, Jun. 2022, doi: 10.1609/aaai.v36i3.20228.
21. N. D. Kwak, “Study On Masked Face Detection And Recognition using transfer learning,” *koreascience.kr,* 2022, doi: 10.17703/IJACT.2022.10.1.294
22. S. Guo, T. Zhang, and T. Guo, *Thermal infrared face recognition with face mask based on the generative method*. 2022. doi: 10.1117/12.2622449.
23. F. I. Eyiokur, H. K. Ekenel, and A. Waibel, “Unconstrained face mask and face-hand interaction datasets: building a computer vision system to help prevent the transmission of COVID-19,” *Signal, Image and Video Processing*, Jul. 2022, doi: 10.1007/s11760-022-02308-x.
24. M. Loey, G. Manogaran, C. E. D. M. Taha, and N. E. M. Khalifa, “A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic,” *Measurement*, vol. 167, p. 108288, Jan. 2021, doi: 10.1016/j.measurement.2020.108288.
25. D. Montero, “Boosting Masked Face Recognition with Multi-Task ArcFace,” *arXiv.org*, Apr. 20, 2021. https://arxiv.org/abs/2104.0987
26. “CNN based COVID-19 Prevention system,” *IEEE Conference Publication | IEEE Xplore*, Mar. 25, 2021. https://ieeexplore.ieee.org/document/9396004
27. A. Cabani, K. Hammoudi, H. Benhabiles, and M. Melkemi, “MaskedFace-Net – A dataset of correctly/incorrectly masked face images in the context of COVID-19,” *Smart Health*, vol. 19, p. 100144, Mar. 2021, doi: 10.1016/j.smhl.2020.100144.
28. W. Hariri, “Efficient masked face recognition method during the COVID-19 pandemic,” *Signal, Image and Video Processing*, vol. 16, no. 3, pp. 605–612, Nov. 2021, doi: 10.1007/s11760-021-02050-w.

29. S. Sethi, M. Kathuria, and T. C. Kaushik, “Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread,” *Journal of Biomedical Informatics*, vol. 120, p. 103848, Aug. 2021, doi: 10.1016/j.jbi.2021.103848.
30. J.-S. Yu and W. Zhang, “Face Mask Wearing Detection algorithm based on improved YOLO-V4,” *Sensors*, vol. 21, no. 9, p. 3263, May 2021, doi: 10.3390/s21093263.
31. S. Mishra, “Indian masked faces in the wild dataset,” *arXiv.org*, Jun. 17, 2021. https://arxiv.org/abs/2106.09670
32. Z. Wang, “Masked Face Recognition Dataset and application,” *arXiv.org*, Mar. 20, 2020. https://arxiv.org/abs/2003.09093
33. “MaskedFaceNet: a Progressive Semi-Supervised Masked face detector,” *IEEE Conference Publication | IEEE Xplore*, Jan. 01, 2021. https://ieeexplore.ieee.org/document/9423368
34. F. Boutros, “MFR 2021: Masked Face Recognition Competition,” *arXiv.org*, Jun. 29, 2021. https://arxiv.org/abs/2106.15288
35. J. Xinbei, T. Gao, Z. Zhu, and Y. Zhao, “Real-Time face mask detection method based on YOLOV3,” *Electronics*, vol. 10, no. 7, p. 837, Apr. 2021, doi: 10.3390/electronics10070837.
36. P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria, and D. J. Hemanth, “SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2,” *Sustainable Cities and Society*, vol. 66, p. 102692, Mar. 2021, doi: 10.1016/j.scs.2020.102692.
37. K. Hammoudi, A. Cabani, H. Benhabiles, and M. Melkemi, “Validating the CorrectWearing of Protection Mask by taking a selfie: Design of a mobile application ‘CheckYourMask’ to limit the spread of COVID-19,” *Cmes-computer Modeling in Engineering & Sciences*, vol. 124, no. 3, pp. 1049–1059, Jan. 2020, doi: 10.32604/cmes.2020.011663.
38. T. Mare, “A realistic approach to generate masked faces applied on two novel masked face recognition data sets,” *arXiv.org*, Sep. 03, 2021. https://arxiv.org/abs/2109.01745
39. M. L. Wang and W. Deng, “Deep face recognition: A survey,” *Neurocomputing*, vol. 429, pp. 215–244, Mar. 2021, doi: 10.1016/j.neucom.2020.10.081.

40. G. Wu, "Masked face recognition algorithm for a contactless distribution cabinet," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–11, May 2021, doi: 10.1155/2021/5591020.
41. Z. Zhu, "Masked Face Recognition Challenge: The WebFace260M Track Report," *arXiv.org*, Aug. 16, 2021. https://arxiv.org/abs/2108.07189
42. A. Anwar, "Masked face recognition for secure authentication," *arXiv.org*, Aug. 25, 2020. https://arxiv.org/abs/2008.11104
43. B. Mandal, "Masked Face Recognition using ResNet-50," *arXiv.org*, Apr. 19, 2021. https://arxiv.org/abs/2104.08997
44. C. Wang, "MLFW: a database for face recognition on masked faces," *arXiv.org*, Sep. 13, 2021. https://arxiv.org/abs/2109.05804
45. "Multi-Dataset Benchmarks for Masked Identification using Contrastive Representation Learning," *IEEE Conference Publication | IEEE Xplore*, Nov. 01, 2021. https://ieeexplore.ieee.org/document/9647194
46. "Occlusion Robust face recognition based on mask learning with pairwise differential Siamese network," *IEEE Conference Publication | IEEE Xplore*, Oct. 01, 2019. https://ieeexplore.ieee.org/document/9009826
47. "When face Recognition meets occlusion: a new benchmark," *IEEE Conference Publication | IEEE Xplore*, Jun. 06, 2021. https://ieeexplore.ieee.org/document/9413893
48. S. Yang, "WIDER FACE: a face detection benchmark," *arXiv.org*, Nov. 20, 2015. https://arxiv.org/abs/1511.06523
49. H. Du, H. Shi, Y. Liu, D. Zeng, and T. Mei, "Towards NIR-VIS masked face recognition," *IEEE Signal Processing Letters*, vol. 28, pp. 768–772, Jan. 2021, doi: 10.1109/lsp.2021.3071663.
50. M. Geng, P. Peng, Y. Huang, and Y. Tian, *Masked Face Recognition with Generative Data Augmentation and Domain Constrained Ranking*. 2020. doi: 10.1145/3394171.3413723.
51. P. L. Wilson and J. D. Fernandez, "Facial feature detection using Haar classifiers," *Journal of Computing Sciences in Colleges*, vol. 21, no. 4, pp. 127–133, Apr. 2006, doi: 10.5555/1127389.1127416.
52. H. Al-Dmour, A. Tareef, A. M. Alkalbani, A. Hammouri, and B. Alrahmani, "Masked Face Detection and Recognition System based on deep learning algorithms," *Journal of*

Advances in Information Technology, vol. 14, no. 2, pp. 224–232, Jan. 2023, doi: 10.12720/jait.14.2.224-232

53. X. Zhang, Y. Wang , N. Zhang , D. Xu and B. Chen, "Research on Scene Classification Method of High-Resolution Remote Sensing Images Based on RFPNet," *Applied Sciences*, vol. 9, no. 10, p. 26, 2019.
54. T. R. D. [Online]. Available: https://medium.com/analytics vidhya/convolutional-neural-network-cnn-1bd606078aa5
55. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 1409.1556.
56. M. U. Hassan, "VGG16 – Convolutional Network for Classification and Detection," Neurohive / Neural Networks, May 2023, Available: https://neurohive.io/en/popular-networks/vgg16/
57. Face mask Detection Dataset," *Kaggle*, Jul. 31, 2020. https://www.kaggle.com/datasets/omkargurav/face-mask-dataset
58. "MDMFR Dataset," *Kaggle*, Apr. 13, 2023. https://www.kaggle.com/datasets/haleytriem/mdmfr-dataset
59. "facedatahybrid," *Kaggle*, Feb. 05, 2021. https://www.kaggle.com/datasets/mrviswamitrakaushik/facedatahybrid?select=Balanced+Facemask+Dataset
60. "Home." https://sites.google.com/view/masktheface/home
61. "LFW Dataset with Masks," *Kaggle*, Jul. 29, 2021. https://www.kaggle.com/datasets/akashguna/lfw-dataset-with-masks?select=LFW_without_Mask
62. A. Sabha and A. Selwal, "CoSumNet: A video summarization-based framework for COVID-19 monitoring in crowded scenes," *Artificial Intelligence in Medicine*, vol. 139, p. 102544, May 2023, doi: 10.1016/j.artmed.2023.102544.
63. C. Szegedy, "Rethinking the inception architecture for computer vision," *arXiv.org*, Dec. 02, 2015. https://arxiv.org/abs/1512.00567

APPENDICES

Experimental Results for Masked Face Recognition

Pre-trained Model	Learning-Rate	Dropout	Optimizer	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy	Testing Accuracy
VGG16	0.001	0.3	Adam	0.0481	98.53%	0.2996	92.7%	90%
	0.001	0.3	SGD	0.6779	78.83%	0.9458	74.41%	72%
InceptionV3	0.001	0.3	Adagrad	0.0089	100	0.2271	92.88	93%
	0.0001	0.1	Adagrad	0.1855	98.55	0.5523	87.43	86%
	0.0001	0.2	Adagrad	0.0083	99.98	0.2284	93.9	93%

Unified Model Results using Different Pre-trained Networks

Pre-trained Model	Training Accuracy	Validation Accuracy	Testing Accuracy	Training Loss	Validation Loss	Precision	Sensitivity	F1-score
VGG16	dect:0.9963 rec: 0.9720	dect:1.0000 rec: 0.8069	dect:100% rec: 81%	0.2217 dect: 0.0118 rec: 0.1049	1.5641 dect:0.0024 rec: 0.7808	0.81	0.80	0.79
VGG19	dect:0.9953 rec:0.9274	dect:0.9911 rec:0.7020	dect: 99% rec: 63%	0.4881 dect:0.0190 rec: 0.2346	2.6174 dect:0.0295 rec: 1.2940	0.66	0.61	0.61
ResNet50	dect:0.9955 rec: 0.9705	dect:0.9970 rec: 0.9673	dect: 99% rec: 91%	0.2307 dect:0.0157 rec: 0.1075	0.3353 dect: 0.0125 rec: 0.1614	0.93	0.91	0.90