**Comparative Analysis of Different Convolutional fine-tuned models for ensuring Mask Usage in real-time applications.**

*Shaunak Nagvenkar*

*21BCE0296*

**1. Abstract**

This paper focuses on the comparative analysis of various deep learning models in predicting and thus implementing stricter mask usage. The importance of masks spreads across multiple domains, such as healthcare and public safety, and has become particularly crucial during the COVID-19 pandemic. Developing technologies to detect and educate the public on correct use of masks is essential. In this paper, we examine various existing deep learning models including VGG16, VGG19, ResNet50, Xception, and InceptionV3, to assess their effectiveness in detecting mask usage, using a dataset that categorizes comprising of images of human faces belonging to three categories which are masks worn properly, masks worn improperly, and no mask. The pre-trained version of these models chosen were used first with that is trained on the ‘ImageNet’ weights. They are later fine-tuned according to the specific task before being evaluated. The accuracy, precision, recall, and F1-score were used to measure the performance of the models and the results display the strengths of these models in terms of classification accuracy and computational efficiency. This analysis provides valuable insights into the trade-offs between accuracy and processing time, guiding the selection of suitable models for real-time mask detection applications.

**2. Introduction**

**2.1 Background Information:**

Surgical masks are typically protective masks that act as barriers to the nose and mouth blocking the viruses and bacteria and also other foreign entities from entering the human body as well limiting the harmful germs emitted by the wearer from spreading into the environment. They consist of multiple layers of nonwoven fabric, usually polypropylene, with an outer fluid-repellent layer, a middle melt-blown filter for trapping bacteria and viruses, and an inner absorbent layer to capture moisture from breathing. Is helps in preventing inter person spread of diseases which is why they are usually seen in medical field. Although surgical face masks were widely in use in fields such as medicine and industrial applications the wide spread use of surgical masks among people as a personal safety equipment has only increased since the COVID-19 pandemic in limiting the spread of infectious diseases.

**2.2 Motivation behind the proposed work:**

The COVID-19 pandemic highlighted the importance of respiratory masks, which can help prevent the spread of diseases. They can also help minimize the transmission of infections by protecting people from respiratory droplets. The ability to accurately identify mask usage in various settings, such as schools, hospitals, and workplaces, has become a vital part of any healthcare facility's efforts to prevent the spread of infections. Traditional methods of detecting mask use usually relies on human intervention, which can be inefficient and erroneous. Integrating Artificial Intelligence techniques in automating this task can help significantly in reducing the errors and be more efficient in real time use. In the era of computer vision, vast convolutional neural networks (CNNs) like VGG16, ResNet50, and InceptionV3 have been developed. This has helped us to create image model that can perform mask classifications. These models have been widely used and also successful in various image classification tasks. Most of these models are readily made available and have a pre-trained version of its which makes using them very easy as there no need for training them. Benefiting from their pre-trained weights on large datasets like ImageNet on which they are trained but their effectiveness in various specific tasks.

**2.3 Focus of the proposed work:**

The main goal of this study is to generate a complete analysis of the currently used deep learning convolutional neural network models used in detecting the proper usage of masks in real time scenario. The efficiency of these models to classify a person based on whether he or she is wearing a mask is determined and the based on these results the most significant algorithm is determined.

**2.4 Proposed work contribution:**

Face masks played a crucial role in limiting spread of the COVID-19 virus along with many other measures implemented. It essentially acted as the primary line of defense during that time. In many cities masks are regularly used to tackle pollution as it filters the air we breathe which, makes a huge difference on our lungs in the long run. In order to maintain this discipline in such times or situations, making the role of detecting face mask very necessary for public safety. This paper wants to find the best models that can be used in this specific task by fine-tuning them to fit the purpose.

**3. Literature Review**

**3.1 Existing work in the context of proposed work:**

The study ***"Comparative Study of Model Optimization Techniques in Fine-Tuned CNN Models"(1)*** by Ramaprasad Poojary and Akul Pai shows how optimizers like SGD, Adam, and RMSProp compare with each other on a cat vs. dog dataset for classification purpose. SGD for ResNet50 model achieves nearly 99% accuracy with a 0.001 learning rate across 15 epochs while RMSProp optimizer shows slightly better results for InceptionV3 model. The conclusion depicts that among all the models studied SGD was found to be the best optimizer.

Authors Amit Chavda, Ankit Damani, and Jason Dsouza in their paper ***"Multi-Stage CNN Architecture for Face Mask Detection"(2)*** created a model that classifies faces as masked or unmasked. It is trained on a diverse set of datasets and was found to be more efficient that traditional methods in terms of accuracy.

In ***"Face Mask Detection System using CNN," (3)*** Sneha Sakshi and team created a MobileNetV2-based system with over 99% accuracy for mask detection in both static images and real-time videos, ensuring high performance in public spaces during the COVID-19 pandemic.

The paper ***"Comparative Analysis of CNN Models to Diagnose Breast Cancer" (4)*** by Ketulkumar Govindbhai Chaudhari leverages machine learning models like SVM, RCNN, and Faster RCNN for breast cancer detection using medical imaging. The dataset used is the BreakHis dataset and concludes that Faster RCNN is more effective than the other models.

The paper ***“Fine-tuning CNN Image Retrieval with No Human Annotation” (5)*** *by* Filip Radenovic, Giorgos Tolias and Ondrej Chum shows a self-supervised approach for image retrieval with CNNs, that will allow to learn from unlabeled data. It performs very well compared to traditional methods, specifying the strength of supervised learning.

The paper ***“Detection of a facemask in real-time using deep learning methods: Prevention of Covid 19”(6)*** discusses a CNN for real-time face mask detection, classifying images into three categories. The model achieves training and testing accuracies of 98.9% and 98.74%, effectively handling varied lighting conditions.

The paper ***“Effect of Data Augmentation on Fine-tuned CNN Model Performance”*** (7) examines how data augmentation impacts fine-tuned CNNs, specifically VGG16 and ResNet50. Results show improved accuracies of 93.5% and 95% for training, highlighting the importance of augmentation in reducing overfitting.

The paper ***"A Comparative Analysis of Hybrid Deep Learning Models for Human Activity Recognition" (8)*** assesses hybrid models combining CNNs with RNNs for activity classification using the PAMAP2 dataset. Hybrid models outperform standalone models, with CNN-BiLSTM achieving the best performance.

The paper ***“Automatic Face Mask Detection System in Public Transportation in Smart Cities Using IoT and Deep Learning” (9)*** introduces a real-time mask detection system using VGG16 on Raspberry Pi, achieving over 99% accuracy. Its low latency makes it suitable for crowded environments.

The paper ***"Fine-tuning Convolutional Neural Networks for Fine Art Classification”*** (10) focusses on utilization of CNNs in fine art classification. The dataset used in this study was the WikiArt dataset. It was concluded that retaining of layers helps improved performance.

**3.2 Restrictions on the Current Work:**

Furthermore, as mentioned in "Automatic Face Mask Detection System in Public Transportation in Smart Cities Using IoT and Deep Learning," while high accuracy is achieved in studies like "Face Mask Detection System using CNN," there is little discussion of the computational costs and latency challenges when deploying such models in real-time systems, especially in resource-constrained environments like public transportation. Last but not least, a gap exists in fully optimizing model performance for a variety of applications because many existing works do not include a comparative examination of hybrid models or architectures that can potentially give greater performance in real-time circumstances.

There are still a few drawbacks with CNN-based models for detection tasks, notwithstanding their improvements. Numerous studies, like the one in "Comparative Study of Model Optimization Techniques in Fine-Tuned CNN Models," use optimization approaches that are primarily focused on a small number of optimizers and may not fully address other optimization strategies that could enhance performance. The emphasis in research such as "Multi-Stage CNN Architecture for Face Mask Detection" is mostly on controlled surroundings or static images, and the models might not operate as well in real-world scenarios with large variations in illumination or occlusions.

**3.3 Research Deficits from Previous Studies:**

The corpus of current research offers a solid basis for CNN-based model construction and optimization for applications like image classification, medical diagnosis, and face mask recognition. The existing work in this field has provided a solid basis the construction of a CNN model for specific use cases like healthcare. Nonetheless, there exist notable gaps in various studies. In the paper "A Comparative Analysis of Hybrid Deep Learning Models for Human Activity Recognition," appropriate analysis of hybrid models which combined CNNs and RNNs techniques, to improve both spatial and temporal feature extraction. More detailed research is required to completely understand the effects of data augmentation approaches, especially when dealing with highly imbalanced datasets, in "Effect of data-augmentation on fine-tuned CNN model performance". Furthermore, not enough attention has been paid to how well-tuned models function in real-time applications in a variety of environmental scenarios (dim lighting, shifting angles, or congested environments). As noted in "Automatic Face Mask Detection System in Public Transportation in Smart Cities Using IoT and Deep Learning," the difficulty of effectively implementing these models in IoT systems creates further research need regarding scalability and optimization for low-latency situations.

**3.4 Objectives of the proposed work:**

A comparative analysis of optimized and fine-tuned CNN models in order to find methods that improve efficiency and performance in image classification tasks is performed in this study. This work will use datasets related to face mask identification to assess the effects of different training fine-tuning on model accuracy and computational simplicity. It also employs concepts of data augmentation that will help create a more robust model to be used in real-time scenarios.

**4.Proposed Methodology**

**4.1 Methods and Approaches:**

This research employs a structured methodology to compare various deep learning models for detecting mask usage in images. The following steps outline the key methods and approaches utilized in the study.

**Dataset Preparation**: The **Face Mask Dataset** contains 14,535 images classified into three categories: **Incorrect Mask**, **With Mask**, and **Without Mask**. Designed to support face mask detection model development, the dataset includes diverse orientations, mask types, and occlusions to ensure real-world applicability. An openly accessible **Mask Face Dataset (No, Mask, Improper)** dataset consisting of images of people wearing surgical masks was categorized into three classes Improper Mask (signifying mask not worn properly), Mask (signifying worn improperly) and No mask (mask not worn properly). This dataset is utilized to train and validate the models for the specified classification task. The images it contains were manually classified into the categories they belong to before training.

**Preprocessing**: The images will undergo preprocessing steps, including resizing to a uniform dimension, normalization and data augmentation techniques to enhance the diversity of the training set and reduce overfitting. The models used require images of different images of uniform pixel dimensions, for which the images that will be fed into the model will be first reshaped to the required size. Finally, to create robust models, the images will be augmented with various techniques.

**Model Choice:** The models chosen for this study were VGG16, VGG19, ResNet50, Xception, and InceptionV3. The models were used with pre-trained ‘ImageNet’ weights on which they are trained. For the purpose of mask identification, each model will be having different fully connected neural network layer that helps it to understand the features effectively and the entire model will be finally fine-tuned to ensure that it performs efficiently on the particular dataset.

**Model Training and Fine-Tuning**:

All the models will be fed with the same data going through same pre-processing stages. Owing to the fact that different models have varying architectures and so need to be trained of different parameters. Adam optimizer is used as an optimizing technique for all the models. Callback and class-weight as well were used in the training phase to tackle the issues of overfitting and data imbalance respectively. The models use the pre-trained ‘ImageNet’ weights and is initially trained only for the fully connected neural networks by freezing all the convolutional layers. Further the entire model is fine-tuned to fit the specific use case of face mask detection.

**4.2 Data and Measurements**

**Dataset Description**

The dataset used for the process of testing the trained models and obtains efficiency is publicly available on Kaggle uploaded by Larxel, that can be utilized for classification model training. This dataset comprises images of individuals in various settings, annotated to indicate mask usage with their names suggesting their status. It consists of 529 improper mask, 992 mask, and 554 no mask face images. They were first classified into sub folders as explained below.

The modified dataset consists of three distinct classes:

* **Mask**: Images where individuals are wearing masks correctly, covering both the nose and mouth.
* **Improper Mask**: Images where individuals are wearing masks incorrectly, such as under the chin or covering only the mouth.
* **No Mask**: Images where individuals are not wearing a mask at all.

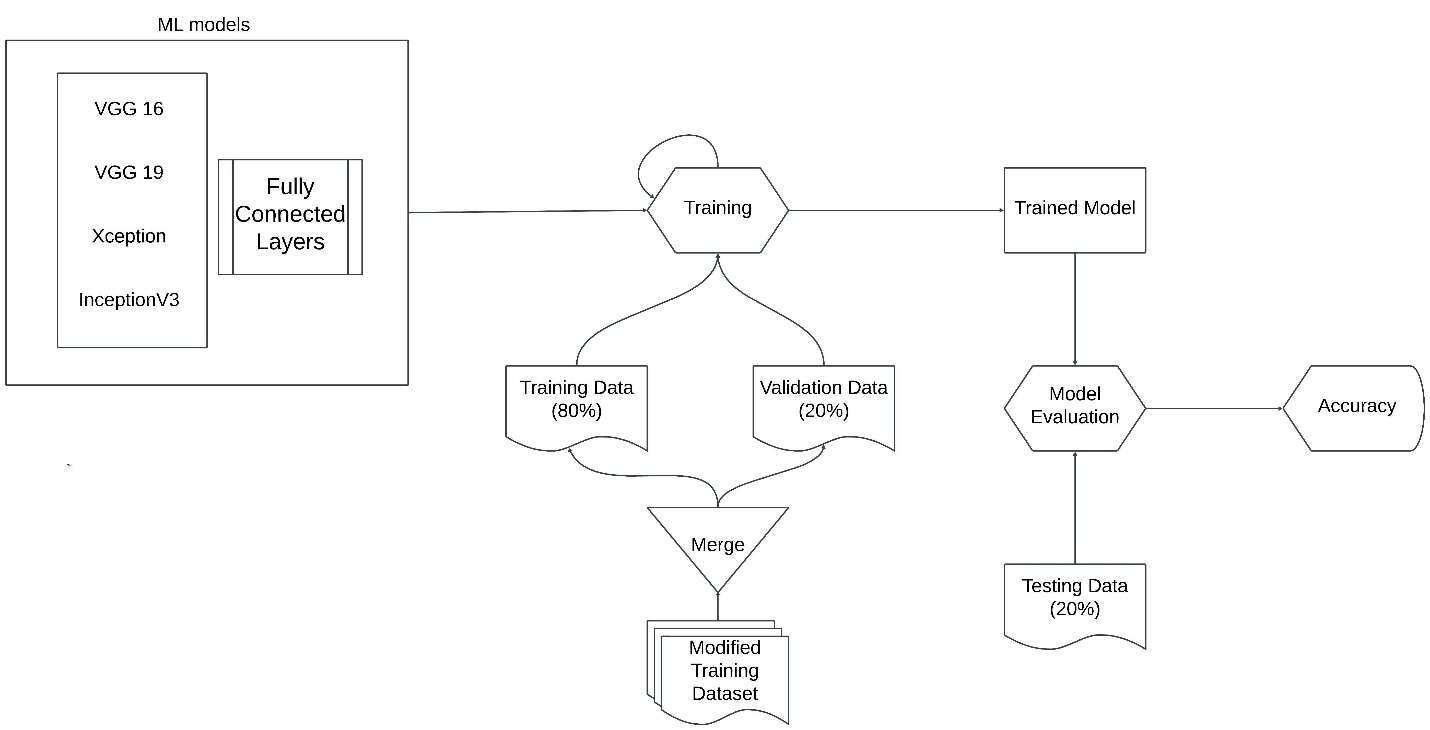
**Measurements:**

To assess the performance of each deep learning model in predicting mask usage, the following evaluation metrics will be employed:

* **Accuracy**
* **Precision**
* **F1-score**
* **Inference Time**
* **Confusion Matrix**
* **Receiver Operating Characteristic Curve**

**5. System Architecture**

**5.1 Architecture Diagram**



**5.2 Algorithms of the Proposed Work**

The proposed work employs a wide range of convolutional algorithms for effective face mask detection. The following models were used:

* **VGG16:**
* **VGG19**
* **ResNet50**
* **InceptionV3**
* **Exception**

**5.3 Description of the Proposed Work**

The proposed work aims to draw comparison between the convolutional algorithms specified above to determine the which would be the best in the specific use case of predicting face mask usage detection. The models were trained with various strategies and some unique strategies specific to the models. These models initially took up the pre -trained ‘ImageNet’ weights and were subsequently fine-tuned to the specific task so as to provide the most optimal output. These models can be used in multiple real-time applications such mask screening in airports and hospitals which are accessed by multiple people at the same time.

**5.4 Algorithm Improvement & Justification of the Proposed Methodology**

The new fine-tuned algorithms that are already trained on pre-existing ‘ImageNet’ trained to fit a classification model for face masks can help in increasing the efficiency detecting defaulters, the people who are not wearing masks in this case creating alerts and thus maintain public safety. The various features implemented can help in can help in making the model more efficient with depicted by the model evaluation scores.

**6. Simulations (Experiments)**

**6.1 Evaluation Criteria:**

As specified in the previous sections, the analysis will take into consideration the processing time and the accuracy of classifications for all the model into account and hence with this data make a determination of the most suitable algorithm for this purpose. The metrices used in this process were as follows:

* **Accuracy**
* **Precision**
* **Recall**
* **F1-score**
* **Inference time per image**
* **Confusion Matrix**
* **Receiver Operating Characteristic Curve**
  1. **Application Interface & Experimental Setup**

The following hardware and software setup was used for the study.

* **Tools/Frameworks**:

**Development environment**: VS Code

**Programming Language**: Python 3.11

**Machine Learning libraries used**: TensorFlow 2.17

* **Hardware**:

**GPU**: AMD Ryzen 7 4800H with Radeon Graphics

**RAM**: 16 GB

* **Parameter settings**:

**Learning Rate**: 1 for initial training and 1e-5 from fine-tuning phase using Adam optimizer

**Train Validation Split Ratio**: 8:2

**6.3 Results Obtained through Proposed Approach**

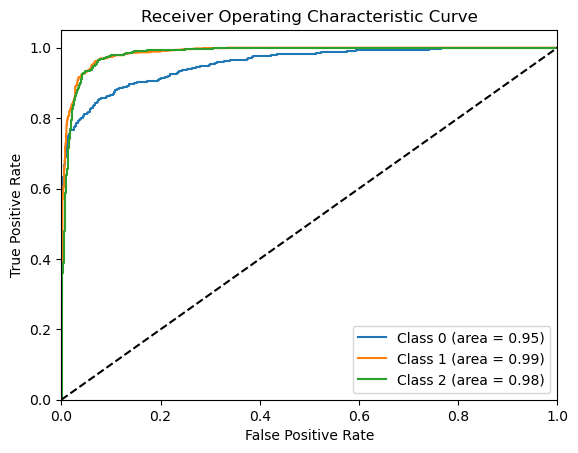
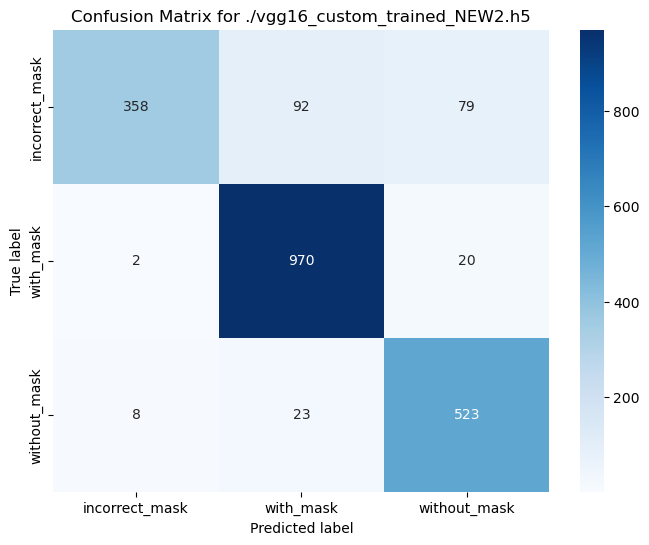
On performing above mentioned analytical operation the following results were obtained:

**Testing Phase**

| **Model** | **Accuracy** | **Inference Time (s/image)** |
| --- | --- | --- |
| **VGG16** | 89% | 0.0151 |
| **VGG19** | 87% | 0.0167 |
| **ResNet50** | 85% | 0.0148 |
| **Xception** | 82% | 0.0179 |
| **InceptionV3** | 73% | 0.0218 |

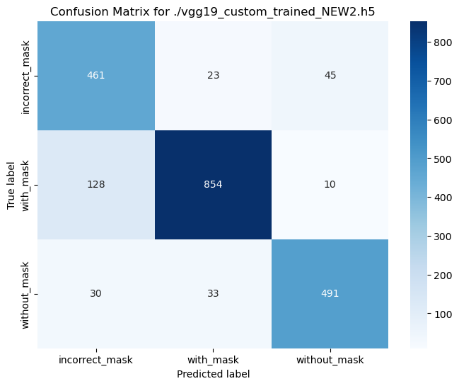
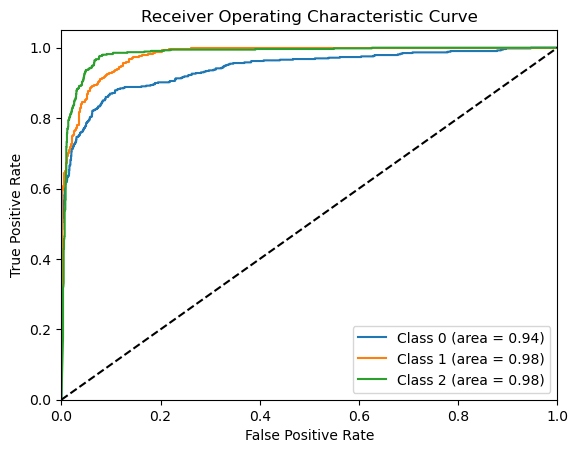
**VGG16**

|  | **Precision** | **Recall** | **F1-Score** |
| --- | --- | --- | --- |
| **Improper Mask** | 0.97 | 0.68 | 0.80 |
| **Mask** | 0.89 | 0.98 | 0.93 |
| **No Mask** | 0.84 | 0.94 | 0.89 |

****

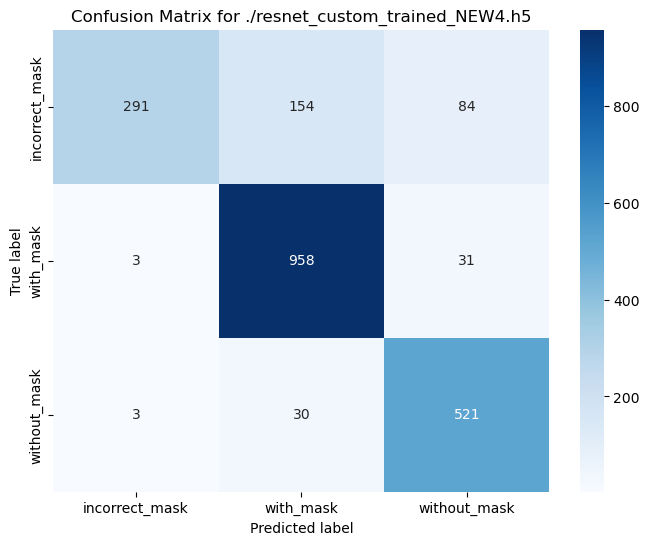
**VGG19**

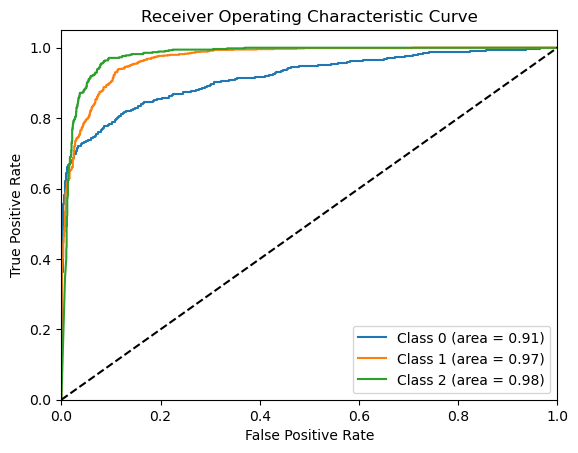
|  | **Precision** | **Recall** | **F1-Score** |
| --- | --- | --- | --- |
| **Improper Mask** | 0.74 | 0.87 | 0.80 |
| **Mask** | 0.94 | 0.86 | 0.90 |
| **No Mask** | 0.90 | 0.89 | 0.89 |

**  
**

**ResNet50**

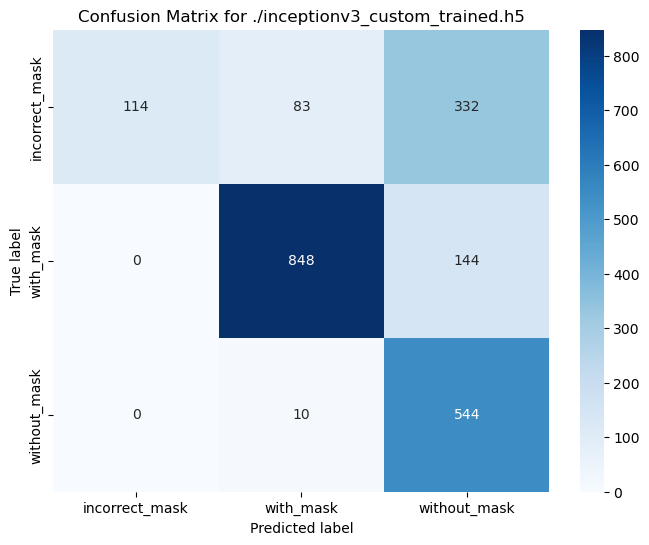
|  | **Precision** | **Recall** | **F1-Score** |
| --- | --- | --- | --- |
| **Improper Mask** | 0.98 | 0.55 | 0.70 |
| **Mask** | 0.84 | 0.97 | 0.90 |
| **No Mask** | 0.82 | 0.94 | 0.88 |

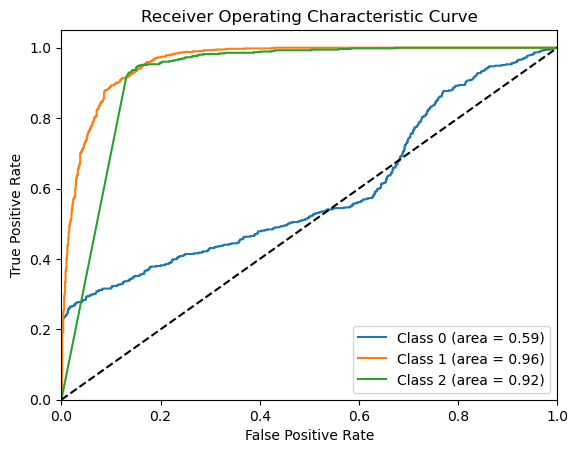
****

****

**InceptionV3**

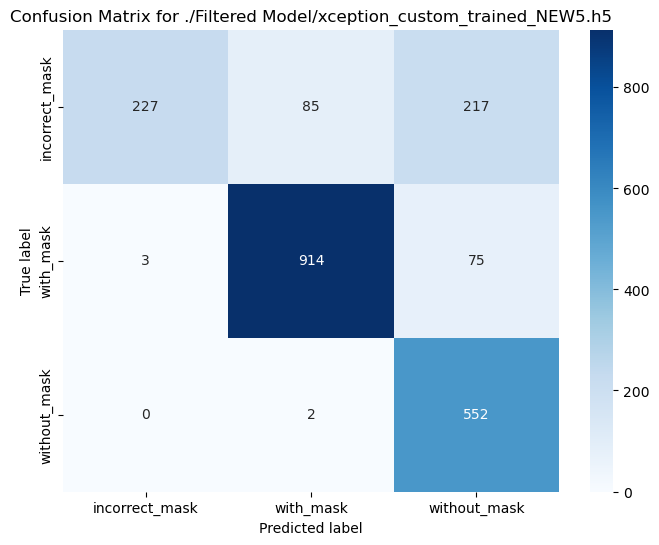
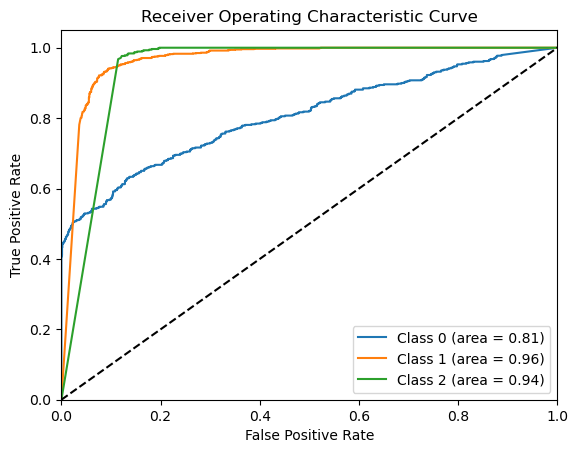
|  | **Precision** | **Recall** | **F1-Score** |
| --- | --- | --- | --- |
| **Improper Mask** | 1.00 | 0.22 | 0.35 |
| **Mask** | 0.90 | 0.85 | 0.88 |
| **No Mask** | 0.53 | 0.98 | 0.69 |





**Xception**

|  | **Precision** | **Recall** | **F1-Score** |
| --- | --- | --- | --- |
| **Improper Mask** | 0.99 | 0.43 | 0.60 |
| **Mask** | 0.91 | 0.92 | 0.92 |
| **No Mask** | 0.65 | 1.00 | 0.79 |

****

**6.4 Description of the Results Obtained through Proposed Approach**

The performance evaluations provide a detailed comparison of each model in terms of accuracy and inference time per image. The VGG16 model demonstrated the best trade-off between accuracy and processing time, achieving an accuracy score of 89% with an inference time of only 0.0151 seconds per image. Similarly, the VGG19 model achieved 87% accuracy, though its processing time was slightly higher at 0.0167 seconds per image. ResNet50 showed an accuracy of 85% and the fastest inference time of 0.0148 seconds per image, making it the most time-efficient model.

The Xception model achieved an accuracy of 82% with a processing time of 0.0179 seconds per image, while InceptionV3 had the lowest accuracy at 73% and the highest inference time of 0.0218 seconds per image. Overall, VGG16 emerged as the most balanced model in terms of both performance and efficiency, while Xception showed promise for improvement in accuracy at a slight cost of processing time.

**6.5 Comparison with Existing Studies and Methods**

In this section, the performance of the proposed models is compared with existing studies and methodologies. The analysis focuses on metrics such as accuracy, precision, recall, and F1-score, as well as inference time. Compared to related works discussed in the literature review, our approach offers competitive accuracy and processing efficiency. For instance:

* + **VGG16** Achieved an F1-score of 0.93 for detecting masks, indicating high precision and recall, with a balanced trade-off between speed and accuracy.
  + **VGG19** Performed slightly lower in terms of accuracy but still achieved a high F1-score of 0.90 for mask detection, making it a reliable choice.
  + **ResNet50** Excelled in time efficiency, with a very low inference time of 0.0148 seconds per image, while maintaining competitive F1-scores across categories.
  + **Xception** Showed potential with the highest precision for detecting masks but required improvements in recall for categories like improper masks.
  + **InceptionV3** Demonstrated limitations in accuracy and inference time compared to other models, suggesting it may not be ideal for real-time applications.

The results highlight trade-offs between speed and accuracy, with VGG16 and Xception offering the most practical options depending on the application’s priorities.

**7. Conclusion, Limitation and Future Scope of Work**

The evaluation of different models highlights VGG16 as the most efficient in terms of both accuracy and inference time, achieving an accuracy of 89% and processing images in just 0.0151 seconds. Xception, while slightly slower at 0.0179 seconds per image, achieved excellent classification performance, especially for mask detection. Most models demonstrated inference times under 0.02 seconds, ensuring near-instantaneous classification—a critical factor for real-time applications.

However, certain limitations exist:

1. Dataset Bias The models were tested on a custom dataset, which may not fully represent real-world diversity.

2. Improper Mask Detection Models like Xception struggled with recall in categories such as improper mask detection.

3. Model Complexity Higher accuracy often came at the cost of increased inference time, as seen with Xception and InceptionV3.

Future work can focus on:

* Fine-tuning model architectures to improve feature extraction and classification accuracy.
* Exploring ensemble methods to combine the strengths of multiple models.
* Extending the dataset to include a wider variety of images for better generalization.
* Investigating lightweight architectures for deployment on resource-constrained devices.

In conclusion, while VGG16 is the most practical model for real-time applications due to its speed and reliability, Xception offers higher classification accuracy and warrants further optimization to address its recall challenges. Both models hold significant potential for improving mask detection systems in real-world scenarios.

**8. References**

1. Poojary, R, Pai, A. 2019. Comparative Study of Model Optimization Techniques in Fine-Tuned CNN Models. *International Conference on Electrical and Computing Technologies (ICECTA)*.
2. Chavda, A., Damani, A., Dsouza, J., Badgujar, S. 2021. Multi-Stage CNN Architecture for Face Mask Detection. *IEEE International Conference for Convergence in Technology (I2CT)*.
3. Sakshi, S., Yadav, S. S., Gupta, A. K, Kumar, U. 2021. Face Mask Detection System using CNN. *International Conference on Advance Computing Innovative Technologies in Engineering (ICACITE).*
4. Chaudhari, K. G. 2018. Comparative Analysis of CNN Models to Diagnose Breast Cancer. *International Journal of Innovative Research in Science, Engineering and Technology*.
5. Radenovic, F. Tolias, G. Chum, O. 2019. Fine-tuning CNN Image Retrieval with No Human Annotation. IEEE Transaction on Pattern Analysis and Machine Intelligence.
6. Kashyap, G, S, Soholt, J. Siddiqui, A, Siddiqui, R, Malik, K, Wazir, S, Brownlee, A. E. I. 2024. Detection of a Facemask in Real-Time using Deep Learning Methods: Prevention of COVID-19.
7. Poojary, R, Raina R, Mondal A. K. 2021. Effect of Data Augmentation on Fine-tuned CNN Model Performance. (IAES) *International Journal of Artificial Intelligence*.
8. Abbaspour, S, Fotouhi, F, Sedagatbaf, A, Fotouhi, F, Vahabi, M, Linden M. 2020. Comparative Analysis of Hybrid Deep Learning Models for Human Activity Recognition.
9. Kumar, T. A, Rajmohan, R, Pavithra, M, Ajagbe, S. A, Hodhod R, Gaber, T. 2022. Automatic Face Mask Detection System in Public Transportation in Smart Cities Using IoT and Deep Learning.
10. Cetinic, E, Lipic, T, Grgic, S. 2018. Fine-tuning Convolutional Neural Networks for Fine Art Classification.

**9. Code**

Python Jupyter Notebook:

https://drive.google.com/file/d/17AIDF87EvEjjfeBTIjOHhlHn\_1mt4HIR/view?usp=sharing

**10. Dataset Links:**

**Mask Face Dataset (No, Mask, Improper):**

https://www.kaggle.com/datasets/emrahaydemr/mask-face-dataset-no-individual-improper