

Assessment of State-of-the-Art Deep Neural Networks for the Detection of COVID-19 via Chest X-Rays



Azzam Shaikh
M.S. in Robotics
ashaikh2@wpi.edu



Anoushka Baidya
M.S. in Robotics
abaidya@wpi.edu



Chaitanya Gokule
M.S. in Robotics
cagokule@wpi.edu



Cristobal Rincon Rogers
M.S. in Robotics
crinconrogers@wpi.edu

May 1, 2024

GitHub repo: github.com/azzamshaikh/CS_534_AI_Assessment_of_SOTA_COVID-19_CXR_Models

Abstract

At the end of 2019, Wuhan, China, emerged as the ground zero for the outbreak of COVID-19, a disease caused by the SARS-CoV-2 virus, leading to a worldwide health emergency declared by the World Health Organization. With the virus affecting 216 countries and causing extensive morbidity and mortality, accurately differentiating COVID-19 from similar respiratory diseases became imperative. Leveraging artificial intelligence, specifically state-of-the-art convolutional neural networks, offers a promising solution for enhancing the accuracy and speed of diagnosing COVID-19 through chest X-rays. Given the varied state-of-the-art convolutional neural networks (CNNs) explored for COVID-19 detection in chest X-rays, this paper synthesizes these approaches to present a comprehensive analysis. This project delves into the effectiveness, adaptability, and efficiency of pretrained state-of-the-art models like ResNet-50, VGG-19, and InceptionV3. Furthermore, fine-tuned state-of-the-art models, such as COVID-ResNet, DarkCovidNet, and DenseNet169 and XGBoost, are evaluated to determine if fine-tuned models provide significant benefits over renown, pretrained models, assessing their diagnostic accuracy and operational benefits. Through a rigorous evaluation framework involving accuracy, sensitivity, specificity, and computational efficiency, this project aims to identify the most promising CNN architectures for real-world application, offering a path forward in enhancing rapid and reliable COVID-19 screening methods. Based on the results of the testing accomplished, it was determined that the DenseNet169 and XGBoost model developed by Nasiri & Hasani provides the best results, based on MCC score, to determine COVID-19 in chest x-rays.

Keywords: Neural Networks; CNN; SOTA; COVID-19; Deep Learning; Chest X-rays; ResNet, VGG, Inception, Darknet, DenseNet

1 Introduction

1.1 Motivational Background

In December of 2019, in Wuhan, China, the first official case of COVID-19 was identified (Page, Hinshaw, & McKay, 2021). COVID-19, an abbreviation for coronavirus disease 2019, is a contagious disease that is a strain of coronavirus, also known as SARS-CoV-2, or ‘Severe Acute Respiratory Syndrome coronavirus 2’ (B. Hu, Guo, Zhou, & Shi, 2020). Coronavirus is described as ”a diverse group of viruses infecting many different animals, and can cause mild to severe respiratory infections in humans” (B. Hu et al., 2020). While most cases of coronaviruses are known to occur in livestock, human coronaviruses that were discovered in the 1960s which mainly caused a common cold, and thus were not a priority to study (King, 2020). However, in 2002, an epidemic of SARS occurred resulting in the deaths of ”one in 10 of the 8000 plus people it infected” (King, 2020). This coronavirus was referred to as SARS-CoV-1. After this outbreak, in 2012, another epidemic occurred when a new coronavirus strain appeared in Saudi Arabia; due to the location of the disease, it was called the Middle East respiratory syndrome (MERS), and thus the virus was referred to as MERS-CoV (King, 2020). According to King (2020), in all these cases, the origin of these viruses started from bats. Bats live with ”a cornucopia of coronaviruses without falling ill”. Through bats, the coronaviruses were transmitted to other animals and eventually to humans. The SARS-CoV and MERS-CoV viruses were responsible for the deaths

of 774 (Pasley, 2020) and 858 (Organization, n.d.) individuals, respectively. Based on initial analysis of the SARS-CoV-2, there were similarities in the sequence identity between itself and SARS-CoV and MERS-CoV (T. Hu et al., 2020) as well as their capability to transfer via human-human transmission (T. Hu et al., 2020). At the time, it was difficult to predict the severity of this coronavirus strain. However, the ability for SARS-CoV-2 to rapidly transfer the disease between individuals led to the World Health Organization to declare the outbreak a Public Health Emergency of International Concern (T. Hu et al., 2020). Within 6 months of the initial discovery, 216 countries had declared an exposure to COVID-19 (T. Hu et al., 2020). As of March 10, 2023, a total of 676,609,955 cases had been detected and 6,881,995 deaths had been declared (University, n.d.).

The symptoms of COVID-19 can vary from patient to patient. A variety of factors such as age, weight, underlying health conditions, and disabilities can impact the body's response. Generally, COVID-19 causes respiratory symptoms that are similar to a cold, a flu, or pneumonia (for Disease Control & Prevention, 2023). The symptoms tend to appear 2-14 days after exposure (for Disease Control & Prevention, 2022). Some individuals may show no symptoms, and some may have severe reactions (for Disease Control & Prevention, 2023). Typical symptoms include fevers, loss of taste or smell, coughs, sore throats, difficulty breathing, fatigue, body aches, congestion, and others (for Disease Control & Prevention, 2023). Due to the similarities in symptoms between a cold, a flu, and COVID-19, it is important to undergo diagnosis to determine a patient's sickness and provide appropriate treatment and potential medication. In today's age, testing for COVID-19 can be done at a clinic or at home via viral tests such as polymerase chain reaction tests and antigen tests (CDC, 2020). However, early on during the COVID-19 pandemic, the global supply of viral test kits were not readily available. Thus, providers attempted to use imaging, such as CT scans or chest x-rays, to determine if a patient had COVID-19 (Krishnaraj & Matsumoto, 2020). Patients who have COVID-19 can exhibit signs in the images that may point to COVID-19, but in many cases, these signs also appear for other lung related problems (Krishnaraj & Matsumoto, 2020). According to Xu et. al (2020), a patient with COVID-19 shows "multiple patchy shadows in both lungs" (Xu et al., 2020). Figure 1 visualizes the described patchy shadows. Figure 2 depicts a typical chest x-ray as a reference. There is clear visual distinction between a normal chest x-ray and one with COVID-19.

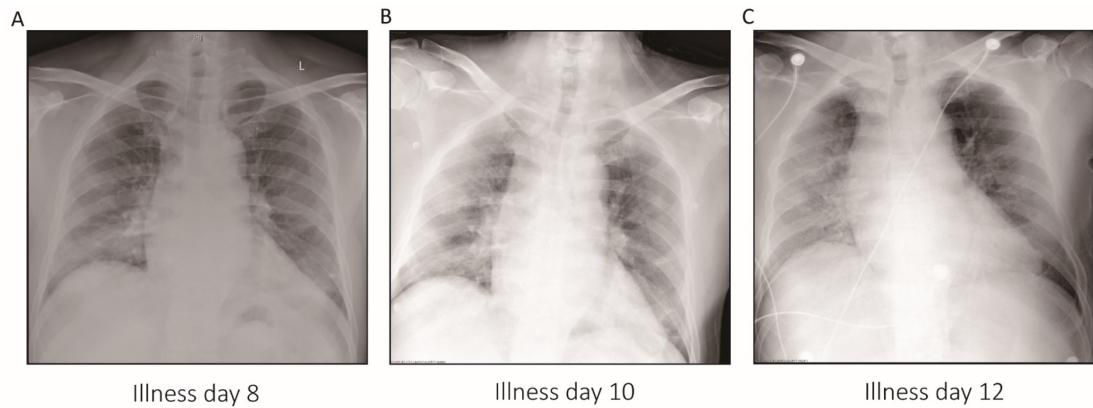


Figure 1: Chest x-ray of COVID-19 positive patient over the span of a few days (Xu et al., 2020).

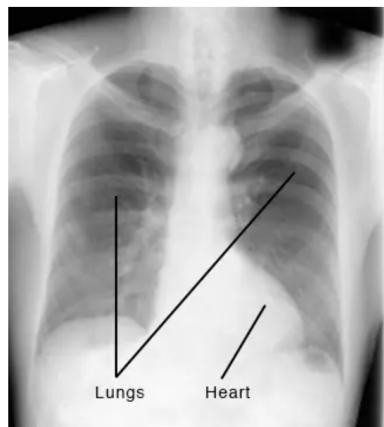


Figure 2: Typical chest x-ray (Clinic, n.d.).

Challenges occur when comparing a COVID-19 positive x-ray versus a non-COVID, viral pneumonia. According to Durrani et. al (2022), in a study comparing RT-PCR positive and negative suspected patients, the visual similarities in the chest x-rays in the middle and lower lobes to be statistically significant (Durrani et al., 2021). Figure 3 visualizes a COVID-19 negative suspect patient x-ray to a COVID-19 positive patient. Both images include multiple patchy shadows making it difficult to distinguish the results to the plain eye. Durrani et. al also suggested using detailed clinical evaluations and radiological workup to distinguish the results (Durrani et al., 2021).

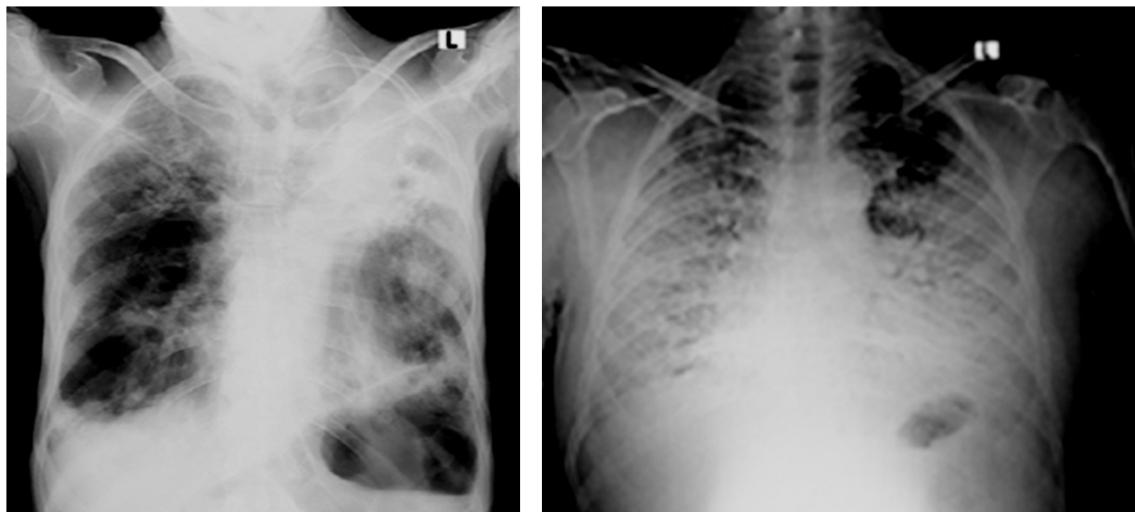


Figure 3: Chest x-ray of COVID-19 negative suspected patient (left) versus a COVID-19 positive patient (right) (Durrani et al., 2021).

According to the American College of Radiology, CT scans and X-rays are generally not recommended as the primary tool for diagnosis due to similarities in images between COVID-19 and other respiratory infections, the high percentage of COVID-19 positive patients with normal X-rays, and due to the close proximity required between provider and patient when conducting tests (ACR, 2020). It is possible to reduce and minimize the impact of proximity between a provider and patient during testing via PPE and appropriate hygienic practices. However, due to the visual inspection required by radiologists, it is not trivial to have clear distinction in the images between COVID-19 and other respiratory diseases as well as reducing false positive readings of COVID-19 positive patients.

To improve the capabilities of interpreting diagnosis and classification of COVID-19 versus other respiratory diseases and against false positives, artificial intelligence can be leveraged to classify chest x-rays and assist in the diagnosis process. This approach would help automate the diagnosis process and deliver faster results as compared to the conventional process.

1.2 Current State-of-the-Art Solutions

Various state-of-the-art methods have been developed for the diagnosis of COVID-19 via chest x-rays.

Nayak et. al (Nayak, Nayak, Sinha, Arora, & Pachori, 2021) proposed a deep learning (DL) based method for early diagnosis of COVID-19 infections. It evaluates the performance of eight pre-trained convolutional neural networks (CNN) models: AlexNet, VGG-16, GoogleNet, MobileNet-V2, SqueezeNet, ResNet-34, ResNet-50, and Inception-V3. The study focuses on classifying COVID-19 cases from normal cases, with a comprehensive analysis of factors such as batch size, learning rate, number of epochs, and type of optimizers to determine the best-suited model. ResNet-34 achieved the highest performance with an accuracy of 98.33%. The paper also compares its results with state-of-the-art methods, highlighting the potential of CNN models for assisting in early COVID-19 detection through medical imaging.

Motamed et. al (Motamed, Rogalla, & Khalvati, 2020) proposed a semi-unsupervised generative adversarial network (GAN), called RANDGAN, for detecting COVID-19 without requiring labeled data. Utilizing the largest publicly available COVID-19 chest X-ray dataset, COVIDx, the study employs transfer learning for lung segmentation to enhance the classification accuracy. RANDGAN outperformed conventional GANs, improving anomaly detection in medical images and increasing the area under the ROC curve from 0.71 to 0.77. The research emphasizes the significance of lung segmentation and proposes a novel approach to COVID-19 detection that could potentially speed up testing processes in healthcare systems.

Darapaneni et. al (Darapaneni et al., 2022) studied the prediction of COVID-19 by leveraging deep learning models, specifically VGG-16, ResNet-50, and Inception-V3 to classify images as COVID-19 positive or negative. The paper demonstrates a methodology that includes image preprocessing, data augmentation, and the use of transfer learning to fine-tune pre-trained models on a dataset of 21,025 chest X-ray exams. The VGG-16 model achieved the highest accuracy among the models tested, indicating its effectiveness in distinguishing between COVID-19 and normal X-ray images. The study underscores the potential of deep learning to assist in rapid and accurate COVID-19 diagnosis, aiming to alleviate the burden on healthcare systems and improve patient management during the pandemic.

Cohen et. al (Cohen et al., 2020) focuses on predicting the severity of COVID-19 pneumonia with deep learning techniques. The authors developed a model that uses pre-trained neural networks on non-COVID-19 chest X-ray datasets to predict two severity scores: the extent of lung involvement and the degree of opacity. The model, which leverages DenseNet architecture, achieved a mean absolute error (MAE) of 1.14 for the extent score and 0.78 for the opacity score. The study demonstrates the potential of deep learning models in assessing COVID-19 severity from chest X-rays, proposing a valuable tool for patient management and treatment efficacy monitoring, especially in ICU settings.

Signoroni et. al (Signoroni et al., 2021) develops an end-to-end deep learning architecture named BS-Net for predicting the severity of COVID-19. This architecture integrates multiple tasks, including lung segmentation, spatial alignment, and multi-regional score estimation. The scoring system used, known as the Brixia score, provides a semi-quantitative assessment of lung compromise. Utilizing a clinical dataset of almost 5,000 annotated images, the study demonstrates the model's high accuracy across all processing stages. It emphasizes the model's ability to outperform human annotators in rating accuracy and consistency, suggesting its utility in computer-assisted monitoring of COVID-19 patients. The research also highlights the generation of highly detailed explainability maps to visualize the network's focus on lung areas, enhancing the understanding of its decision-making process.

Tartaglione et. al (Tartaglione, Barbano, Berzovini, Calandri, & Grangetto, 2020) investigates the challenges of using deep learning to classify COVID-19 due to the small size of available datasets. It emphasizes the potential biases introduced by transfer learning from non-COVID to COVID datasets and explores the impact of dataset composition and model complexity on classification performance. The study utilizes various datasets, including a novel dataset (CORDA) collected from a hospital in Northern Italy, and tests different CNN architectures, highlighting the importance of careful dataset selection and model tuning to avoid overfitting and ensure generalizability. The paper's extensive experimental evaluation offers insights into the complexities of developing effective deep learning models for COVID-19 detection from X-rays, underlining the need for larger, diverse datasets to improve model accuracy and reliability.

Wang, Lin, and Wong (Wang & Wong, 2020) proposed a tailored, deep learning model for the detection of COVID-19 in x-ray images with a model called COVID-Net. This open-source model was developed using a human-machine collaborative design strategy, leveraging both human-driven principled network design and machine-driven design exploration. The model is trained on the COVIDx dataset, a large, open-access benchmark dataset compiled from several public data repositories, making it one of the largest of its kind for COVID-19 positive cases. COVID-Net's performance is evaluated using explainability methods to ensure

it makes predictions based on relevant information from the images. The study highlights the potential of COVID-Net to aid clinicians in screening and offers a foundation for further research and development in the field of medical image analysis for COVID-19 detection.

Li, Li, and Zhu (Li, Li, & Zhu, 2020) introduce COVID-MobileXpert, an on-device, lightweight deep neural network (DNN) mobile application for COVID-19 patient triage and follow-up using chest X-rays. The system employs a novel three-player knowledge transfer and distillation (KTD) framework comprising a pre-trained attending physician (AP) network, a fine-tuned resident fellow (RF) network, and a trained lightweight medical student (MS) network. This framework enables the MS network to perform on-device COVID-19 patient triage and follow-up efficiently. Experiments demonstrate the potential of COVID-MobileXpert for rapid deployment with various architecture and tuning parameter settings. The paper emphasizes novel loss functions and training schemes to ensure robust feature learning, highlighting the importance of on-device machine learning models for rapid, effective, and privacy-preserving assessment of COVID-19 patients.

Farooq and Hafeez (Farooq & Hafeez, 2020) proposed a deep learning model to screen COVID-19 from radiographs with a model called COVID-ResNet. The framework utilizes a fine-tuned ResNet-50 architecture, employing state-of-the-art training techniques such as progressive resizing, cyclical learning rate finding, and discriminative learning rates. The model achieved a state-of-the-art accuracy of 96.23% on the COVIDx dataset within only 41 epochs. This paper showcases a computationally efficient and highly accurate model for multi-class classification of COVID-19 and other types of infections from radiographs, highlighting its potential to aid in early screening and reduce the healthcare system's burden.

Hemdan, Shouman and Karar (Hemdan, Shouman, & Karar, 2020) proposed a deep learning framework to diagnose COVID-19 in x-ray images with a model called COVIDX-Net. It validates the framework on a dataset of 50 chest X-ray images, including 25 confirmed positive COVID-19 cases, using seven different convolutional neural network (CNN) models such as VGG19 and MobileNetV2. The models showed promising results in automated COVID-19 classification, with VGG19 and DenseNet models achieving the best performance. The study highlights the potential of deep learning in assisting the early diagnosis of COVID-19, providing a valuable tool for radiologists.

Ozturk et al. (Ozturk et al., 2020) proposed a deep learning model to automate the detection of COVID-19 cases with a model called DarkCovidNet. The model, designed for binary (COVID vs. No-Findings) and multi-class (COVID vs. No-Findings vs. Pneumonia) classifications, achieved an accuracy of 98.08% for binary classes and 87.02% for multi-class scenarios. DarkCovidNet is built upon the DarkNet model, modified for COVID-19 detection with 17 convolutional layers. This approach facilitates rapid and accurate diagnosis, potentially assisting radiologists, especially in regions with a scarcity of specialists.

Nasiri and Hasani (Nasiri & Hasani, 2022) proposes a novel method for diagnosing COVID-19 from chest X-ray images using DenseNet169 to extract features, which are then classified by the XGBoost algorithm. This approach was tested against a dataset including COVID-19, pneumonia, and no-findings labels, achieving notable accuracy improvements over existing methods. The method outperformed others in both speed and accuracy, demonstrating the effectiveness of combining deep neural networks with gradient boosting for medical image analysis.

State-of-the-art (SOTA) methods in diagnosing COVID-19 from chest X-rays leverage deep and machine learning for significant benefits, including high diagnostic accuracy (via models like ResNet, VGG, COVID-Net), fast processing (e.g., COVID-MobileXpert), adaptability across datasets (COVIDX-Net, DarkCovidNet), diagnostic transparency (DeltaNet), enhanced training efficiency (COVID-ResNet), and improved precision through hybrid models (DenseNet169 with XGBoost). These approaches provide scalable, precise, and efficient diagnostic solutions, supporting rapid screening and relieving healthcare system pressures.

1.3 Proposed Work

This project aims to provide an assessment of various state-of-the-art models and their effectiveness in detecting COVID-19 through chest X-rays. Four state-of-the-art approaches are selected and implemented for this assessment.

First, to create a benchmark for the analysis, the methodology proposed by Darapaneni et. al (Darapaneni et al., 2022) will be developed. The proposed methodology was based on the usage of state-of-the-art pre-trained models, including VGG-16, ResNet-50, and Inception-V3, to detect COVID-19 through chest X-rays. These models are renowned for their high accuracy and performance across various image classification tasks. Within the testing completed by Darapaneni, the three listed models had testing accuracy scores of 97%, 94%, and 94%, respectively. Thus, the initial focus of the project will be on leveraging these pretrained architectures to create a performance benchmark and ensure a thorough analysis.

The remaining three models will focus on developing models that have been enhanced and fine-tuned specifically for COVID-19.

For the first fine tuned model, the model developed by Farooq & Hafeez, COVID-ResNet, will be implemented. This model presents a 3-stage training approach to fine tune a pretrained ResNet-50 architecture. This model achieved a testing accuracy of 96.23% with only 41 epochs (Farooq & Hafeez, 2020).

For the second fine tuned model, the model developed by Ozturk et. al, DarkCovidNet, will be implemented. This model presents a modified DarkNet architecture with fewer layers and filters while gradually increasing the number of filters. This model achieved a testing accuracy of 98.08% for binary classification and 87.02% for multi-class classification (Ozturk et al., 2020)

For the third fine tuned model, the model developed by Nasiri & Hasani will be implemented. This model presents a two step classification process by utilizing a pretrained model, a DenseNet169, to extract image features and then passing those features to an XGBoost algorithm to perform the classification. This model achieved a testing accuracy of 98.23% for binary classification and 89.70% for mutli-class classificaiton (Nasiri & Hasani, 2022).

With these six models, an analysis can be performed to compare the results of using 3 pretrained models and 3 fine-tuned models to determine the best method for detecting COVID-19 in chest X-rays.

1.4 Dataset, Experimental Setup, and Performance Analysis

Three commonly used datasets are currently available for training and testing models. One dataset was curated by Cohen et. al (Cohen et al., 2020) which contains over 600 X-ray images of patients who are positive and or suspected of COVID-19 and/or other viral/bacterial pneumonias. Another commonly used dataset was curated by Lin et. al (Wang & Wong, 2020) which pulls X-ray images from eight various datasets, including the dataset created by Cohen et. al. This dataset contains over 84,000 images from 45,000 patients. The third dataset, winner of the COVID-19 Dataset Award by the Kaggle Community, was curated by Chowdhury and Rahman et. al (Chowdhury et al., 2020; Rahman et al., 2021). This dataset includes chest x-ray images of 3,616 COVID-19 positive cases, 10,192 normal cases, and 6,012 viral pneumonia cases. Due to varying input layers that each model may have, over the course of the project development, an appropriate dataset from one or multiple datasets will be finalized. The final dataset curated will be one that provides the capability for multiclass (COVID vs. No-Findings vs. Pneumonia) classification of chest x-rays. Each dataset will undergo any data augmentation and/or normalization as required per the selected approach.

The experiments to train and test the models will be completed using the appropriate frameworks as proposed by approach. Initial collaboration will occur on Google Colab due to convenience and remote GPU access.

With the state-of-the-art models described, to evaluate the effectiveness of these approaches, the performance of each model will be compared. This analysis will aid in identifying the best state-of-the-art model.

Six metrics will be utilized to compare the results of the models. First, for each model, a confusion matrix will be generated based on the result of the testing. The confusion matrix will identify four key aspects of the results by comparing the predicted output against the actual output. The matrix is shown in Figure 4. The true positive value indicates correctly identified images with a characteristic, while the true negative indicates correctly identified images without a characteristic. A false positive indicates that an image has been classified to contain a certain characteristic when in fact it does not, while a false negative is the opposite of this.

		Predicted	
		Negative (N) -	Positive (P) +
Actual	Negative -	True Negative (TN)	False Positive (FP) Type I Error
	Positive +	False Negative (FN) Type II Error	True Positive (TP)

Figure 4: Example confusion matrix for a binary classification problem (Suresh, 2020).

Using this confusion matrix, the accuracy of the model can be identified. The accuracy describes the

model's ability to make correct predictions. The calculation for accuracy is described in Equation 1 (Suresh, 2020).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{\text{Correct Predictions}}{\text{Total Predictions}} \quad (1)$$

The precision of the model can also be described. The precision describes how correct the true prediction is. The calculation for precision is described in Equation 2 (Suresh, 2020).

$$Precision = \frac{FP}{TP + FP} = \frac{\text{Predictions Actually Positive}}{\text{Total Predicted Positive}} \quad (2)$$

The sensitivity, or recall, can be described. The sensitivity describes how many actual observations are predicted correctly. The calculation for sensitivity is described in Equation 3 (Suresh, 2020).

$$Recall = \frac{TP}{TP + FN} = \frac{\text{Predictions Actually Positive}}{\text{Total Actual Positive}} \quad (3)$$

Using precision and sensitivity, the F1-score can also be described. This metric is the "harmonic mean of precision and recall" (Suresh, 2020) which provides details on the overall accuracy of the model. The calculation for the F1-score is described in Equation 4 (Suresh, 2020).

$$F1 Score = 2 * \frac{Recall * Precision}{Recall + Precision} \quad (4)$$

An Area Under the Receiver Operating Characteristic Curve (AUC-ROC) will also be created for each model. This will measure the model's ability to distinguish between classes (COVID-19 positive vs. negative) across various threshold settings, indicating diagnostic ability.

Additionally, the Matthew's correlation coefficient (MCC) will be calculated for each model. An MCC is used to "measure the difference between the predicted values and actual values" and is the "best single-value classification metric to summarizes a confusion matrix" (Voxco, 2021). The calculation for the MCC is described in Equation 5 (Voxco, 2021).

$$MCC = \frac{TN * TP - FN * FP}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (5)$$

The primary metric for selecting the most effective approach is the MCC score, as it is the best single value classification value. However, the additional five metrics will be taken into consideration for evaluating models with equivalent or similar results. Using these six metrics, the overall best model can be identified.

With the foundation of the project defined, the remainder of the report will dive deep into the selected models for this assessment, propose the high-level solution and process, discuss the experimental results of the testing, and conclude with the lessons learned and future work.

2 State-of-the-Art Methods Literature Review

As mentioned in the Proposed Work, out of the 12 researched SOTA solutions, four specific methods will be analyzed for the project. This section will describe each approach in more detail.

2.1 Darapaneni et. al - Pretrained Models

As described in Section 1.2, Darapaneni et al. (Darapaneni et al., 2022) introduced a model utilizing pretrained neural networks to detect COVID-19 from chest X-ray images with notable efficiency. By leveraging deep learning architectures such as ResNet-50, VGG19 and InceptionV3, their model capitalizes on the strength of these networks in image recognition tasks. The core of their methodology lies in the use of transfer learning, where these models, initially trained on the vast ImageNet database, are fine-tuned on a specialized dataset of chest X-ray images representing COVID-19 cases alongside other pulmonary conditions.

This approach is underscored by the application of image pre-processing and data augmentation techniques including rotation, flipping, and scaling to simulate a wider variety of X-ray images. Such strategies are crucial for overcoming the challenge of limited datasets typically available in medical imaging, thus enhancing the model's capability to generalize across diverse cases. The choice of ResNet-50 and InceptionV3

is particularly strategic; ResNet-50's deep residual learning framework helps mitigate the vanishing gradient problem, allowing for the training of deep networks. In contrast, InceptionV3's architecture is designed to capture image features at various scales effectively, proving essential in identifying the subtle yet critical indicators of COVID-19 infection visible in chest X-rays.

Darapaneni et al. (Darapaneni et al., 2022)'s model also demonstrates the practical application of transfer learning in rapidly deploying diagnostic models in response to emerging health crises. This strategy not only speeds up the development process by utilizing pre-trained weights but also maintains high accuracy levels, which is crucial for diagnostic tools. However, they also acknowledge the challenges inherent in this approach, including the potential for model bias and the necessity for datasets that are both high-quality and representative. Moreover, the authors emphasize the importance of these AI-driven models serving as supportive tools to augment the expertise of medical professionals rather than replacing them, highlighting the need for transparency and explainability in AI solutions to facilitate their adoption in clinical settings.

In conclusion, Darapaneni et al.'s (Darapaneni et al., 2022) work exemplifies the potential of leveraging pretrained models through transfer learning to address the urgent need for accurate COVID-19 diagnostics. Their comprehensive approach not only showcases the efficacy of deep learning models in medical imaging but also underscores the importance of dataset quality and the collaborative role of AI in enhancing clinical decision-making.

2.2 Farooq & Hafeez - COVID-ResNet

As described in Section 1.2, Farooq and Hafeez (Farooq & Hafeez, 2020) developed a model called COVID-ResNet that achieved a accuracy of 96.23% on the COVIDx dataset within only 41 epochs. The model utilizes a pretrained ResNet-50 architecture and fine tunes the training process via data augmentation and three stage training approach.

ResNet50 is a deep convolutional neural network developed by Microsoft Research in 2015. It is a variant based on the residual network (ResNet) architecture. This specific model contains 50 layers. Residual networks employ deep residual learning that allows the "network to learn a set of residual functions that map the input to the desired output. These residual connections enable the network to learn much deeper architectures than was previously possible, without suffering from the problem of vanishing gradients" (Kundu, 2023). This model has shown successful performance in varying, complex visual recognition tasks.

Farooq and Hafeez expand this network by implementing a custom, fine turned training approach that allows the "ability to feed images of sizes other than which they are trained with" (Farooq & Hafeez, 2020).

The dataset used for this work was acquired from the COVIDx dataset - revision 5 - prepared by (Wang & Wong, 2020). It contains 5941 chest x-ray images that are classified into four categories: normal, bacterial pneumonia, viral pneumonia (non-COVID-19), and COVID-19. The dataset is already split into training and test datasets into unknown values.

For the data augmentation, the transformations utilized include vertical flips, random rotation of images up to 15 degrees, and varying lighting conditions. Only the training data was augmented.

For the network model, as mentioned previously, a pretrained ResNet-50 model is utilized. No adjustments are made to the overall model. The main component for COVID-ResNet is the unique three-stage training process. Progressive resizing is used throughout the training stages by adjusting the scale of the input images for each stage.

For stage one, the input images are resized to 128x128x3 pixels and network under goes two training phases. First, "only the newly added head of the network is trained while preserving the [pretrained] weights for the rest of the body with a learning rate of 1e-3 for 3 epochs" (Farooq & Hafeez, 2020). For the second phase, "the whole network is fine-tuned (both the body and the head of the model) using discriminative learning rate for 5 epochs" (Farooq & Hafeez, 2020).

For stage two, a similar two training phases occur. First, the head of the model from stage one is tuned with resized images scaled to 224x224x224x3 with a learning rate of 1e-4 for 3 epochs. For the second phase, the whole network is tuned for 5 epoch using a discriminative learning rate.

For the last stage, the whole network is tuned with with resized images scaled to 229x229x229x3 for 25 epochs. For this layer, "discriminative learning rates [are used] where the earliest layer was trained with a learning rate of 1e-6 and the last layer was trained with a learning rate of 1e-4. All the layers in between were trained with equidistance learning rates between these two values" (Farooq & Hafeez, 2020).

For all stages, an Adam optimizer with batch size of 32 is used in training.

2.3 Ozturk et. al - DarkCovidNet

As introduced in section 1.2, Ozturk et al. (Ozturk et al., 2020) developed the DarkCOVIDNet model based on the Darknet-19 architecture, which forms the backbone of the YOLO (You Only Look Once) real-time object detection system. Their model achieved notable accuracies of 98.08% for binary classification and 87.02% for multi-class classification.

DarkCOVIDNet is a modified version of Darknet-19, featuring fewer layers and filters. Darknet-19 consists of 19 convolutional layers and five pooling layers using Maxpool, each followed by BatchNorm and LeakyReLU operations (Ozturk et al., 2020). Batch normalization standardizes inputs, reducing training times and increasing model stability. Pooling layers, such as Maxpool, are essential for reducing the spatial dimensions of the input volume, thereby reducing computational complexity and controlling overfitting. Batch normalization further aids in model convergence. Lastly, Leaky ReLU, an activation function used in neural networks, allows a small, non-zero gradient for negative inputs, preventing the "dying ReLU" problem where neurons output zero for negative inputs, potentially leading to inactive neurons.

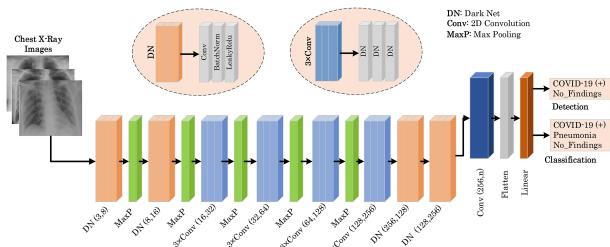


Figure 5: DarkCovidNet architecture (Ozturk et al., 2020)

Ozturk et al. adapted the Darknet-19 model to create DarkCOVIDNet, comprising 17 convolutional layers (Ozturk et al., 2020). The model architecture can be seen in 5. Each DarkNet layer includes a single convolutional layer followed by BatchNorm and LeakyReLU operations. In specific layers of the architecture, single convolutional layers are applied three times. Ozturk et al. referred to these as "3x Conv Layers". The model ends with Avgpool and Softmax layers. For training, they employed the Adam optimizer with a cross-entropy loss function and a learning rate set to 3e-3. Ozturk et al. utilized an 80-20 dataset split, with 80% for training and 20% for validation, training their model for 100 epochs.

In their study, Ozturk et al. utilized X-ray images from two sources for COVID-19 diagnosis. They employed a COVID-19 X-ray image database curated by Cohen JP, comprising 127 X-ray images diagnosed with COVID-19 (Ozturk et al., 2020). This database included 43 female and 82 male positive cases, with incomplete metadata. Additionally, they used the ChestX-ray8 database provided by Wang et al. for normal and pneumonia images (Ozturk et al., 2020). To address data imbalance, Ozturk et al. randomly selected 500 no-findings and 500 pneumonia class frontal chest X-ray images from this database.

2.4 Nasiri & Hasani - DenseNet169 and XGBoost

As outlined in 1.2, Nasiri and Hasani (Nasiri & Hasani, 2022) introduced a model named DenseNet169. This model adopts a pre-trained network which is designed and trained using transfer learning. It is utilized in way that allows for feature extraction followed by classification via XGBoost algorithm.

DenseNet169 is a deep convolutional neural network (DCNN) architecture. It is designed to alleviate the vanishing-gradient problem encountered in very deep networks. This architecture connects each layer to every other layer in a feed-forward fashion, making it more efficient in feature propagation and information flow. DenseNets consist of dense blocks, where each layer is connected to every other layer in a feed-forward manner.

DenseNet169 is a pre-trained model, it has been trained on a large dataset (typically ImageNet) for a general vision task like image classification, as proposed by (Nasiri & Hasani, 2022). As a result, it has learned to extract generic features from images. In this research, the weights of DenseNet169 were not re-trained, but rather the pre-trained weights were used directly to extract features from chest X-ray images. Furthermore, according to (Nasiri & Hasani, 2022) this proposed approach exhibited superior speed and reduced computational complexity compared to the method introduced by (Ozturk et al., 2020). This advantage stems from the

absence of DNN training requirement in the proposed approach. As a result, the proposed method exclusively involves training the XGBoost algorithm.

The extracted features from DenseNet169 are then used as input to the Extreme Gradient Boosting (XGBoost) algorithm. XGBoost is a powerful ensemble learning technique that is adept at handling high-dimensional data and performing classification tasks efficiently. By providing the features extracted by DenseNet169 as input to XGBoost, the research leverages the complementary strengths of both deep learning and gradient boosting techniques for improved COVID-19 detection from X-ray images.

3 Proposed High-level Solution and Process

As described, this project aims to provide an assessment of various state-of-the-art models and their effectiveness in detecting COVID-19 through chest X-rays. Four, supervised, state-of-the-art approaches will be implemented, trained, and tested.

Out of the three datasets listed in Section 1.4, the dataset that has been selected is the Kaggle COVID-19 Radiology Database. This dataset was a commonly used dataset within various papers that includes chest x-ray images of 3,616 COVID-19 positive cases, 10,192 normal cases, and 1,345 viral pneumonia cases.

Figure 8 shows the dataset split by their classifications.

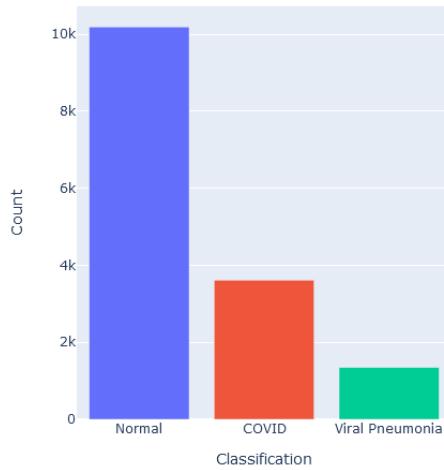


Figure 6: Kaggle Dataset split by classification.

Figure 7 shows a sample of the dataset by classification.

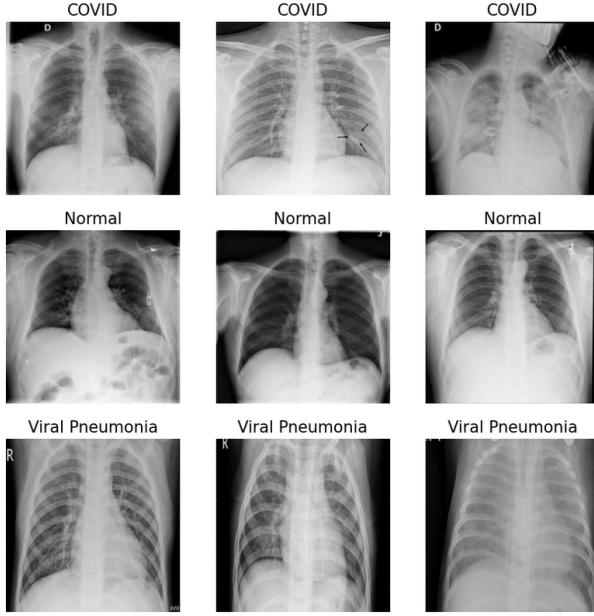


Figure 7: Visualization of a sample of the dataset.

Since the original dataset was heavily skewed, a portion of the dataset was selected to have similar numbers of images per classification. This will assist in training models accurately. Thus, the portion of the dataset can be seen in Figure 8. The dataset contains 1,143 COVID x-rays, 1,341 normal x-rays, and 1,345 viral pneumonia x-rays. It can be seen that this dataset is not skewed as compared to the full dataset.

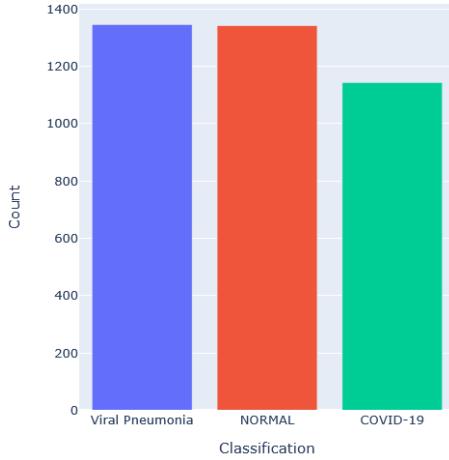


Figure 8: Smaller portion of Kaggle dataset.

Since this dataset is currently split by classification, the data will need to be split into training, validation, and testing datasets. The dataset will be split by using 70% of the dataset for training. For the remaining 30%, that dataset will be further split into 50% divisions for the validation and testing sets. This results in a total split of 70% for training, 15% for validation, and 15% for testing. Various approaches can be used to generate these datasets and are generally platform dependent. Thus, the following subsections will describe the workflow used to load the data into the different datasets.

After the datasets are split, the different models - as described in Section 2 - can be developed and trained with the training and validation datasets. Once trained, the final model can be used to make predictions with the test dataset and an evaluation of the model can occur. Section 4 discusses the results of the work based on the output of the testing.

3.1 Darapaneni et. al - Pretrained Models

It explores the integration of Convolutional Neural Networks (CNNs) for the diagnosis of COVID-19 using chest X-ray images, focusing on the application of ResNet50, VGG19, and InceptionV3 models.

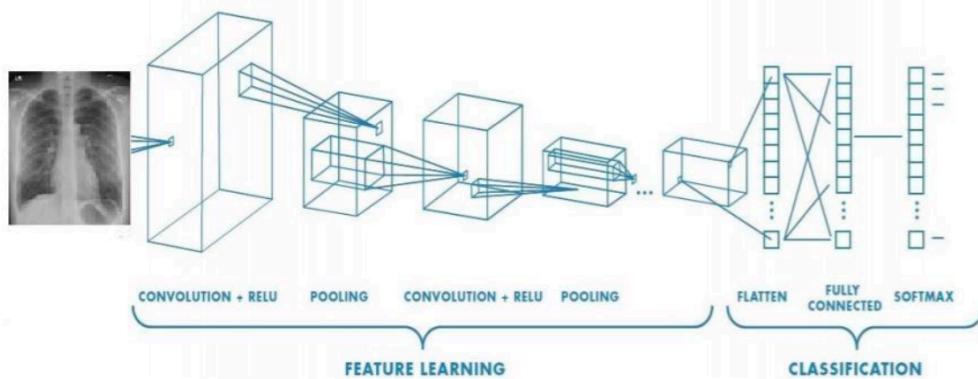


Figure 9: System Architecture.

The Figure 9 (Darapaneni et al., 2022) encapsulates the systematic approach employed in deploying State-of-the-Art (SOTA) Convolutional Neural Networks (CNNs) such as ResNet50, VGG19, and InceptionV3 to analyze chest X-ray images for the detection of COVID-19. This streamlined workflow initiates with feature learning—a convolutional process that iteratively applies filters to the input image to create a feature map, highlighted by ReLU (Rectified Linear Unit) activation functions that introduce non-linearity to the model’s learning capabilities. Pooling layers intermittently reduce the dimensionality of these feature maps, distilling the essential information while retaining spatial hierarchy.

The transition from convolutional layers to a fully connected network marks the shift from feature learning to classification. The feature maps are ‘flattened’ into a one-dimensional vector that serves as the input for the fully connected layers, which act as a high-level reasoning in the neural network. The culmination of this intricate architecture is the softmax layer, which translates the neural network outputs into probabilistic predictions for each class—Normal, Viral Pneumonia, and COVID-19.

The implementation uses TensorFlow framework for data handling and model training within a PyTorch framework. It proceeds to mount Google Drive containing the dataset and establishes a custom dataset class, `ChestXRayDataset`, for efficient data loading and preprocessing. This class is designed to handle image transformations required for model inputs, notably resizing and normalizing images to fit the input dimensions of the respective CNN architectures.

The dataset, initially located in a specified directory, is divided into training, validation, and testing subsets using a `random split` methodology, ensuring a balanced representation of the three classes: Normal, Viral Pneumonia, and COVID-19. Data loaders are then instantiated to facilitate batch processing during the training and evaluation phases, adopting a batch size of six for manageable memory consumption.

For model preparation, `get_pretrained_model` function is leveraged to obtain instances of the pre-defined CNN architectures with weights pre-trained on ImageNet. These models are adapted for the task by modifying their final layers to output three classes corresponding to the dataset labels. The training process is encapsulated in the `train_and_save_model` function, where models undergo training over a specified number of epochs with real-time validation accuracy monitoring for performance tracking. The function `evaluate_model_accuracy` is subsequently used to quantify model performance on the test set, providing a direct assessment of each model’s ability to classify unseen data accurately.

Finally, it emphasizes the visualization of predictions and the calculation of statistical metrics to interpret the models’ diagnostic capabilities thoroughly. This approach ensures a comprehensive evaluation of the CNN models’ performance in detecting COVID-19 from chest X-ray images, highlighting the potential of deep learning applications in augmenting medical diagnostics.

The code for the developed models and supporting data loading processes are detailed in Appendix A.

3.2 Farooq & Hafeez - COVID-ResNet

For the COVID-ResNet model, initially a Keras and Tensorflow based model was developed. However, significant issues resulted with misclassifications with the results. It was unclear if the model was generating the errors or if the dataloaders were creating issues. Thus, a PyTorch implementation was developed. In order to load and split the dataset, a [ChestXRayDataset](#) class was developed to load the dataset. Once the object was created, the [random_split](#) function was used to divide the dataset into training, validation, and testing datasets. A [DataLoader](#) object was then created for each dataset; these would be used to load the data into the various models. While the described approach does not use a validation dataset, a validation set was used as part of the training process.

After creating the data loaders, the models could be trained and tested. The code for the developed models and supporting data loading processes are detailed in Appendix B.

3.3 Ozturk et. al - DarkCovidNet

The DarkCovidNet model was developed and tested using the [Fastai](#) Python module, a high-level deep learning library built on [PyTorch](#). To organize the data, a directory structure specified by the Fastai data loader was used for splitting images into training, validation, and test sets. These sets were populated by randomly sampling images from each class and assigning them to their respective directories, with a split of 70% for training, 15% for validation, and 15% for testing. The data was then loaded using a [DataBlock](#) instance from the [fastai.vision.data](#) library, which utilized a [GrandparentSplitter](#), a utility class in the Fastai library used for splitting a dataset into training and validation sets based on the names of the grandparent directories, to separate the dataset into 'train' and 'valid'. Additionally, the images were resized to 256x256.

The code for the developed models and supporting data loading processes are detailed in Appendix C.

3.4 Nasiri & Hasani - DenseNet169 and XGBoost

For the DesneNet169 model, Keras framework was utilized for implementing the model. To facilitate loading and splitting of the dataset, a data frame was constructed to store the folder paths and classifications of each image. Utilizing the [train_test_split](#) function from the [sklearn.model_selection](#) library, the dataset was partitioned into train and test sets according to predefined splits. Once these sets were created, they were subsequently loaded into [ImageDataGenerators](#) to prepare them for utilization within the model. the pre-trained DenseNet169 model, accessed through the Keras library's applications module, is employed to extract high-level features from the preprocessed images. The extracted features from the training set are then utilized to train an XGBoost classifier, a gradient boosting algorithm for classification tasks.

The code for the developed models and supporting data loading processes are detailed in Appendix D.

4 Experimental Results and Discussions

This section will cover the results for the four developed models and a discussion.

4.1 Darapaneni et. al - Pretrained Models

Upon the completion of training and testing on the VGG19, Inception V3, and ResNet50 models, an assessment of the model could be completed using the metrics defined in Section 1.4.

The confusion matrices for each model are presented in Figures 10, 11, and 12. These matrices display a high degree of accurate classification for COVID-19 and viral pneumonia across all models, with exceptional performance on normal X-ray classifications by the Inception V3 model. The detailed confusion matrices can be observed in the respective figures.

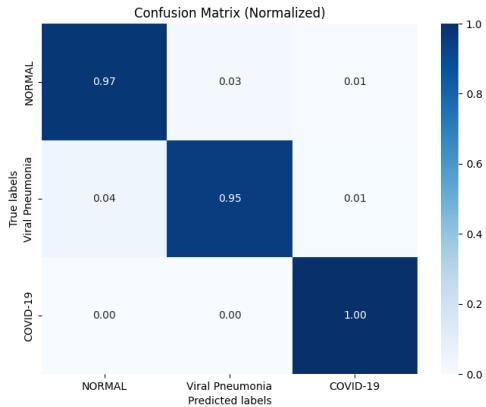


Figure 10: Confusion matrix for the Resnet50 model.

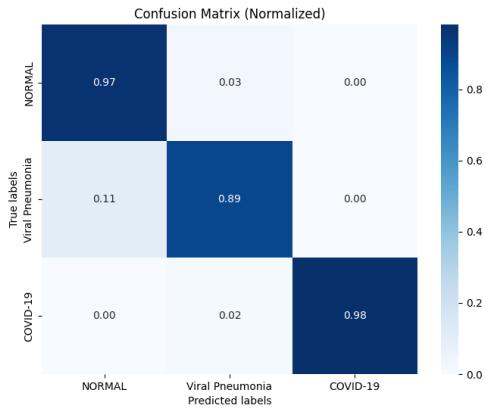


Figure 11: Confusion matrix for the VGG19 model.

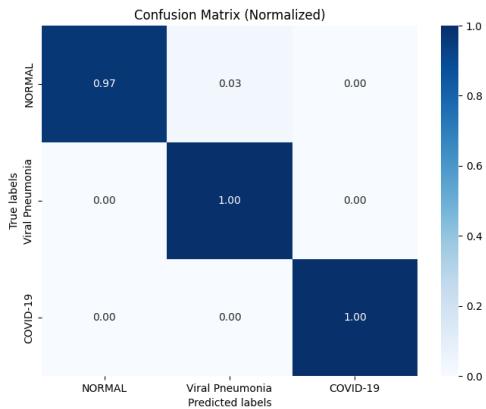


Figure 12: Confusion matrix for the Inception model.

The classification reports for VGG19, Inception V3, and ResNet50 models indicate high precision, recall, and F1-score, with the Inception V3 model achieving the highest Matthew's correlation coefficient (MCC) of 0.9845, as shown in Tables 1, 2, and 3.

Table 1: Performance Analysis for ResNet50 data

Metric	Classification			Overall
	COVID-19	Viral Pneumonia	Normal	
Precision	0.98	0.98	0.95	
Recall	1.00	0.95	0.97	
F1-Score	0.99	0.96	0.96	
Accuracy				0.97
MCC				0.9557

Table 2: Performance Analysis for VGG19

Metric	Classification			Overall
	COVID-19	Viral Pneumonia	Normal	
Precision	1.00	0.95	0.90	
Recall	0.98	0.89	0.97	
F1-Score	0.99	0.92	0.93	
Accuracy				0.95
MCC				0.9198

Table 3: Performance Analysis for Inception Model

Metric	Classification			Overall
	COVID-19	Viral Pneumonia	Normal	
Precision	1.00	0.97	1.00	
Recall	1.00	1.00	0.97	
F1-Score	1.00	0.98	0.99	
Accuracy				0.99
MCC				0.9845

The comparative analysis of the models reveals distinct performance characteristics. The ResNet50 model demonstrates commendable performance with precision, recall, and F1-scores closely distributed around 0.97, indicating a balanced ability to classify all three classes—normal, viral pneumonia, and COVID-19—with high accuracy. The VGG19 model follows closely, with its precision and recall metrics averaging around 0.95, suggesting a slightly less, yet still robust, capability in differentiating between the classes.

Notably, the Inception V3 model outperforms the other two with an impressive precision and recall, both averaging at 0.99, and a Matthew's Correlation Coefficient of 0.9845, reflecting its superior performance in this multi-class classification task. This metric, in particular, underscores the Inception V3 model's strength in producing consistent and reliable predictions across all classes. It's also worth highlighting that while the ResNet50 and VGG19 models show some minor misclassifications, particularly in differentiating between normal and viral pneumonia, the Inception V3 model achieves perfect recall for COVID-19.

The Receiver Operating Characteristic (ROC) curves mirror these findings, with Inception V3 demonstrating a closer curve to the top-left corner of the ROC space, indicating a higher True Positive rate with a low False Positive rate. These findings underscore the potential of Inception V3 in medical diagnostics, suggesting that its architecture may be particularly suited for tasks requiring fine-grained differentiation between highly similar classes as it can be seen in Figures 13, 14, and 15.

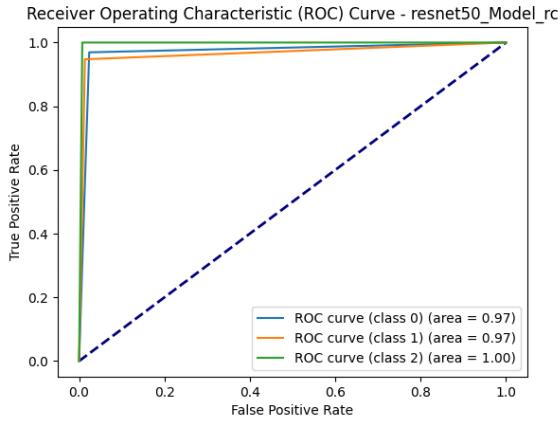


Figure 13: Multiclass AUC-ROC curve for the Resnet50 model.

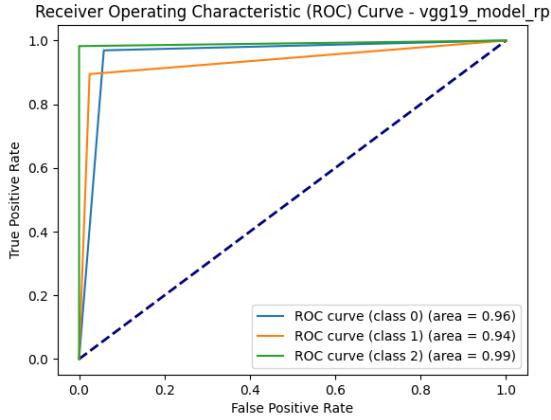


Figure 14: Multiclass AUC-ROC curve for the Vgg19 model.

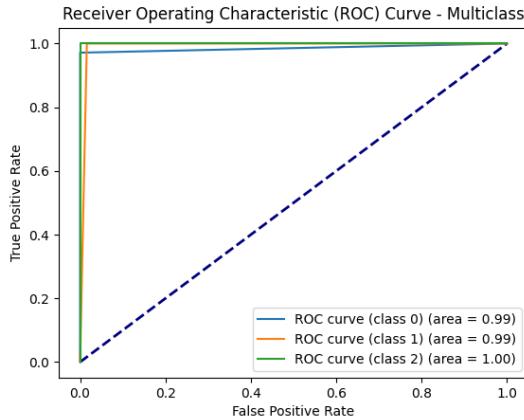


Figure 15: Multiclass AUC-ROC curve for the Inception model.

Figure 13, 14, and 15 shows the AUC-ROC curve for the COVID-ResNet Model.

4.2 Farooq & Hafeez - COVID-ResNet

After training and testing the models, an assessment of the model could be completed using the metrics defined in Section 1.4.

To begin, the confusion matrix can be presented, as shown in Figure 16.

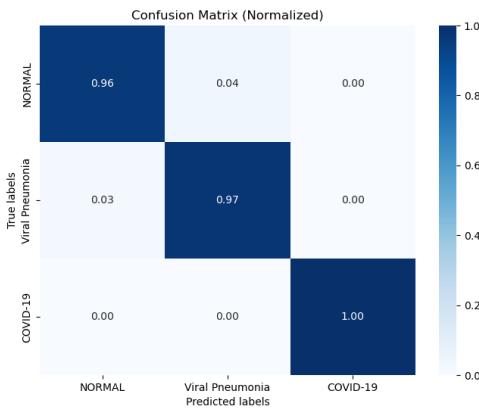


Figure 16: Confusion matrix for the COVID-ResNet model.

Based on the results of the confusion matrix, it can be seen that a significant portion of the predicted labels were correctly classified for each class. This indicates high accuracy and minimal outliers. When computing the model metrics, an accuracy score of 97.9% was achieved and a Matthew's correlation coefficient of 0.9687 was achieved. Table 4 tabulates the various results achieved as described per the performance analysis of the models.

Figure 21 shows the AUC-ROC curve for the COVID-ResNet Model.

Table 4: Performance Analysis for COVID-ResNet.

Metric	Classification			Overall
	Normal	Viral Pneumonia	COVID-19	
Precision	0.97	0.96	1.00	
Recall	0.96	0.97	1.00	
F1-Score	0.97	0.97	1.00	
Accuracy				0.979
MCC				0.9687

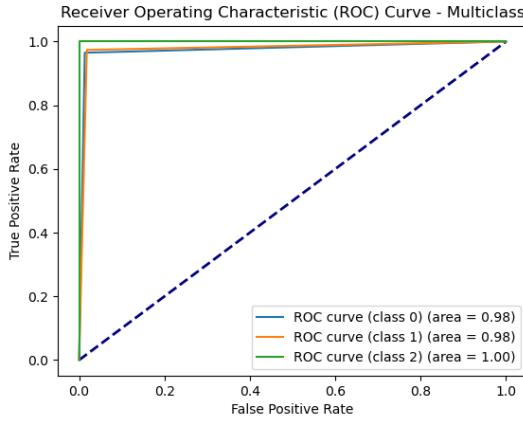


Figure 17: Multiclass AUC-ROC curve of the COVID-ResNet model.

Overall, the model results are highly successful. Compared to the original study, where a multiclass classification accuracy of 96% was achieved, this model slightly improved the results, achieving an accuracy of 97.9%. While there were some misclassifications between Normal x-rays and viral pneumonia x-rays, all COVID-19 x-rays were successfully predicted.

4.3 Ozturk et. al - DarkCovidNet

After training and testing the models, an assessment of the model could be completed using the metrics defined in Section 1.4.

To begin, the confusion matrix can be presented, as shown in Figure 18.

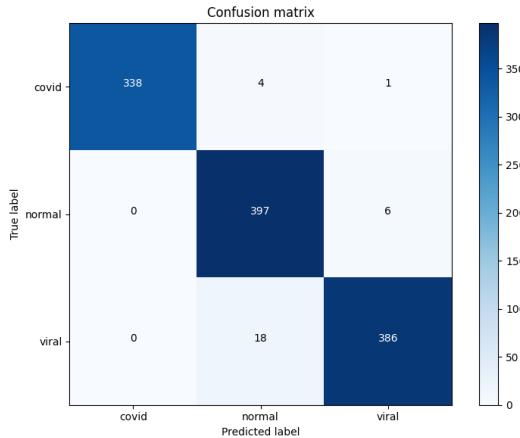


Figure 18: Confusion matrix for the DarkCovidNet model.

Based on the confusion matrix results, a significant portion of the predicted labels for all classes were correctly classified, indicating high accuracy. The matrix demonstrates a majority of labels were correctly identified, with few outliers. The model achieved an accuracy score of 97% and a Matthew's correlation coefficient of 0.9831 when computed. Table 5 tabulate the various results achieved per the models' performance analysis.

Table 5: Performance Analysis for DarkCovidNet.

Metric	Classification			
	Normal	Viral Pneumonia	COVID-19	Overall
Precision	0.95	0.98	1.0	
Recall	0.99	0.96	0.99	
F1-Score	0.97	0.97	0.99	
Accuracy				0.9817
MCC				0.9831

Figure 19 shows the AUC-ROC curve for the DarkCovidNet Model.

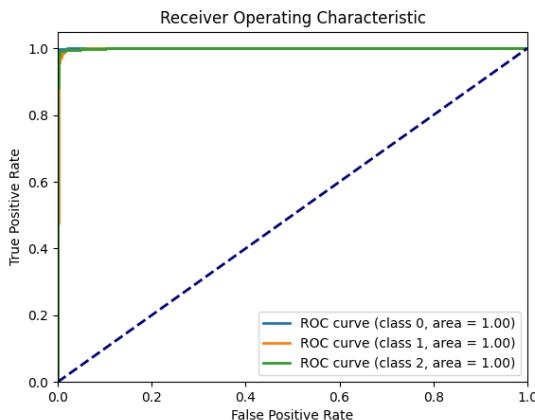


Figure 19: Multiclass AUC-ROC curve of the DarkCovidNet model.

The model's results are highly satisfactory. Upon closer examination of the dataset, the validation set comprised 343 COVID-19 X-ray images, 403 normal x-ray images, and 404 viral pneumonia x-ray images, indicating a relatively balanced and unbiased dataset. Compared to the original study, where a multiclass classification accuracy of 87% was achieved, this model has significantly improved, achieving an accuracy of 97%.

4.4 Nasiri & Hasani - DenseNet169 and XGBoost

After implementing and testing the model and classifier, the results obtained allowed for a detailed assessment of the model performance. The confusion matrix reveals that a substantial number of predicted labels were accurately classified across all classes, suggesting both high accuracy and few outliers. The calculated model metrics show an impressive accuracy rate of 97.9% and a Matthew's correlation coefficient of 0.9882. These findings are summarized in Table 6, which presents the various outcomes as per the performance analysis of the models.

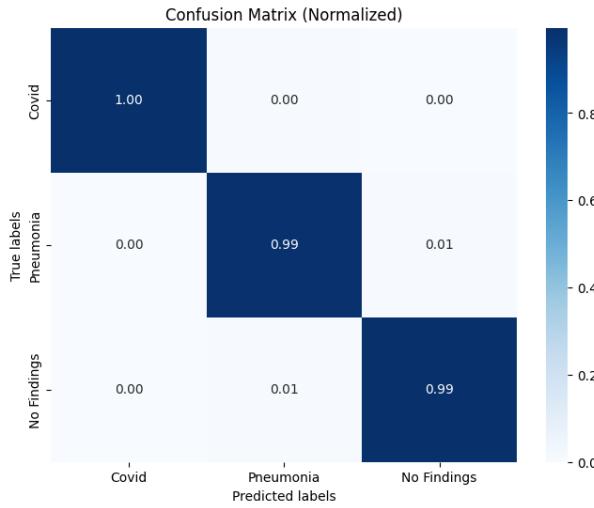


Figure 20: Confusion matrix for the DenseNet169 model.

Table 6: Performance Analysis for DensNet169.

Metric	Classification			Overall
	COVID-19	Viral Pneumonia	Normal	
Precision	1.00	0.99	1.00	
Recall	1.00	1.00	0.99	
F1-Score	1.00	0.99	0.99	
Accuracy				0.9922
MCC				0.9882

Figure 21 shows the AUC-ROC curve for the COVID-DenseNet169 Model.

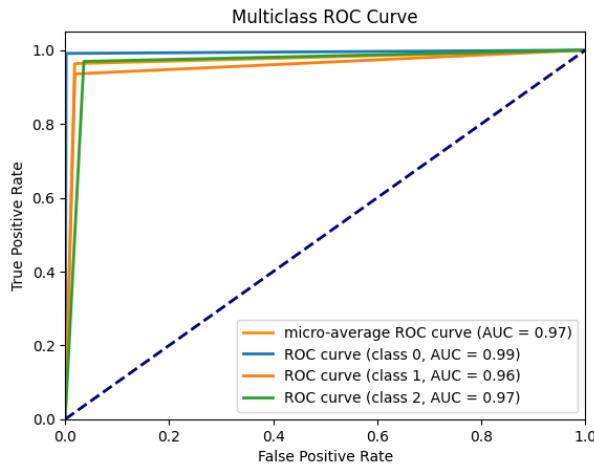


Figure 21: Multiclass AUC-ROC curve of the DenseNet169 model.

Overall, the model results are exceptionally positive. In comparison to the original study, which reported a multiclass classification accuracy of 89.70%, this model demonstrated a decent improvement, achieving an accuracy of 98%. With a very small number of incorrect classifications between Normal xrays and those indicating viral pneumonia, the model successfully identified all x-rays associated with COVID-19.

4.5 Discussion

As mentioned in the Performance Analysis section (1.4), the MCC score will be used as the primary evaluator for the best model. Table 7 contains the MCC scores of the 6 models evaluated for this assessment.

Table 7: MCC Scores for All Evaluated Models.

Model	MCC Score
VGG-19	0.9198
Inception V3	0.9845
ResNet-50	0.9557
COVID-Resnet	0.9687
DarkCovidNet	0.9831
DenseNet169 and XGBoost	0.9882

Table 8 summarizes the additional performance parameters obtained.

Based on the results of the overall testing, the best models can be ordered as shown in Table 9.

The most effective model for detecting COVID-19, based on the MCC score, is the DenseNet169 and XGBoost model that was developed by Nasiri & Hasani. The usage of the pretrained DenseNet169 model combined with the XGBoost algorithm for classification presented the best, state-of-the-art approach for detecting COVID-19.

The second most effective model is the Inception V3 pretrained model. It is impressive for a pretrained model to provide such successful results.

The third most effective model is the DarkCovidNet. The modified Darknet-19, developed for COVID-19 CXR detection proved successful for this application.

The fourth most effective model is the COVID-ResNet. By employing a unique training methodology, the COVID-ResNet provided excellent results with only 41 epochs. The accuracy even succeeded the base, pretrained ResNet-50, which is the fifth most effective model.

The sixth effective model was the pretrained VGG-19 model which had the lowest MCC and accuracy scores.

5 Lessons Learned

In this project, the team ventured into the realm of deep learning, focusing on image classification using various deep learning libraries. Through hands-on learning and problem-solving, several existing codebases were implemented and modified, addressing various aspects of model development and evaluation. Due to the mixed skill set with deep learning implementations, the teams overall proficiency in Python for deep learning has improved, along with skills in data preprocessing and manipulation, and understanding of model training, evaluation, and interpretation. The implementation of certain libraries posed challenges, with the fastai library offering substantial abstractions for deep learning model implementation.

Furthermore, the team encountered unexpected behavior during model training, highlighting the non-guaranteed nature of success in deep learning. For example, the COVID-ResNet model misclassified normal x-ray images, prompting further investigation into the cause. Conversely, the DarkCovidNet performed well on the first attempt. In order to improve the models, attention must be paid when adjusting hyper parameters and conduct additional testing.

Additionally, the team explored the impact of learning rate, batch size, and optimization algorithm on training performance. These factors are crucial in determining the efficiency and effectiveness of training neural networks. The learning rate governs the size of optimization steps, influencing convergence speed and the risk of convergence issues. Batch size affects training dynamics by influencing update stability and resource requirements. The choice of optimization algorithm, such as SGD, Adam, or RMSprop, dictates how model weights are adjusted, with each algorithm offering unique characteristics that impact convergence and performance. Properly tuning these hyperparameters based on dataset and model architecture is essential for achieving optimal training outcomes.

Table 8: Performance Analysis Values of Evaluated Models.

Model	Accuracy
VGG-19	0.950
Inception V3	0.990
ResNet-50	0.970
COVID-Resnet	0.9790
DarkCovidNet	0.9817
DenseNet169 and XGBoost	0.9922

Model	Precision		
	Normal	Viral Pneumonia	COVID-19
VGG-19	0.90	0.95	1.00
Inception V3	1.00	0.97	1.00
ResNet-50	0.95	0.98	0.98
COVID-Resnet	0.97	0.96	1.00
DarkCovidNet	0.95	0.98	1.00
DenseNet169 and XGBoost	1.00	0.99	1.00

Model	Recall		
	Normal	Viral Pneumonia	COVID-19
VGG-19	0.97	0.89	0.98
Inception V3	0.97	1.00	1.00
ResNet-50	0.97	0.95	1.00
COVID-Resnet	0.96	0.97	1.00
DarkCovidNet	0.99	0.96	0.99
DenseNet169 and XGBoost	0.99	1.00	1.00

Model	F1-Score		
	Normal	Viral Pneumonia	COVID-19
VGG-19	0.93	0.92	0.99
Inception V3	0.99	0.98	1.00
ResNet-50	0.96	0.96	0.99
COVID-Resnet	0.97	0.97	1.00
DarkCovidNet	0.97	0.97	0.99
DenseNet169 and XGBoost	0.99	0.99	1.00

Table 9: Best Models in order per MCC Score.

Rank	Model	MCC Score
1	DenseNet169 and XGBoost	0.9882
2	Inception V3	0.9845
3	DarkCovidNet	0.9831
4	COVID-Resnet	0.9687
5	ResNet-50	0.9557
6	VGG-19	0.9198

6 Conclusions and Future Work

6.1 Conclusion

This project aimed to provide an assessment of various state-of-the-art models and their effectiveness in detecting COVID-19 through chest X-rays. Four state-of-the-art approaches are selected and implemented for this assessment.

First, to create a benchmark for the analysis, the methodology proposed by Darapaneni et. al (Darapaneni et al., 2022) was developed. The proposed methodology was based on the usage of state-of-the-art pre-trained models, including VGG-16, ResNet-50, and Inception-V3, to detect COVID-19 through chest X-rays. The second model, developed by Farooq & Hafeez (Farooq & Hafeez, 2020), COVID-ResNet, was implemented and presents a unique, 3-stage training approach to fine tune a pretrained ResNet-50 architecture. The third model, developed by Ozturk et. al (Ozturk et al., 2020), DarkCovidNet, was implemented and presents a modified DarkNet architecture with fewer layers and filters while gradually increasing the number of filters. The final model, developed by Nasiri & Hasani (Nasiri & Hasani, 2022) was implemented and presents a two step classification process by utilizing a pretrained model, a DenseNet169, to extract image features and then passing those features to an XGBoost algorithm to perform the classification.

Using these six models and the Kaggle COVID-19 Radiography Database, pipelines were developed to train and test the models. Based on the results from the testing, the model developed by Nasiri & Hasani (Nasiri & Hasani, 2022), which is based on a pretrained DenseNet169 model and a XGBoost classifier, provided the best results based on MCC score. This model is best suited for the application of detecting COVID-19 in chest x-rays.

Based on the timeline shown in Figure 22, all milestones defined for the project were successfully accomplished.

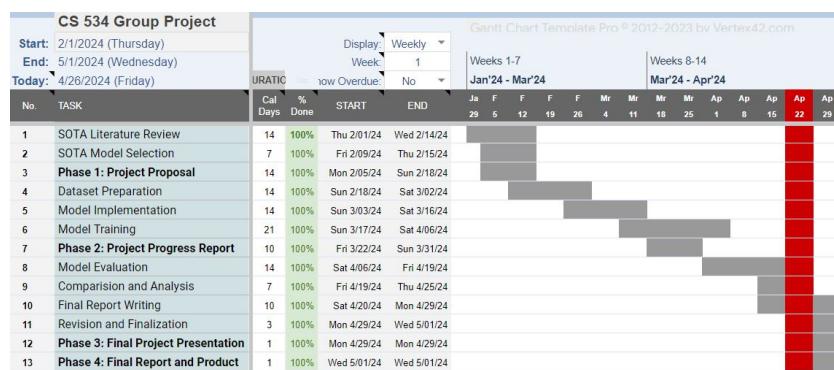


Figure 22: Gantt Chart depicting current progress of the project.

6.2 Future Work

The purpose of this report was to provide an assessment of the current SOTA methodologies that have been developed thus far. Within the literature review that was completed, a 1.2, a total of 12 approaches were studied. Eight approaches currently remain to be considered within this assessment. Future work for this project will include the training and testing of the remaining models described in this report to provide a fully complete and comprehensive approach for detecting COVID-19. Furthermore, using the lessons learned within this study, an entirely new, custom model can be proposed to tackle this issue.

References

- ACR. (2020, 03). *Acr recommendations for the use of chest radiography and computed tomography (ct) for suspected covid-19 infection*. Retrieved from <https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection>
- CDC. (2020, 02). *Coronavirus disease 2019 (covid-19)*. Retrieved from <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/testing.html>

- Chowdhury, M. E. H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. B., ... Islam, M. T. (2020). Can ai help in screening viral and covid-19 pneumonia? *IEEE Access*, 8, 132665-132676. doi: 10.1109/ACCESS.2020.3010287
- Clinic, M. (n.d.). *Chest x-ray*. Retrieved from <https://www.mayoclinic.org/tests-procedures/chest-x-rays/multimedia/chest-x-ray/img-20006961>
- Cohen, J. P., Dao, L., Morrison, P., Roth, K., Bengio, Y., Shen, B., ... Duong, T. Q. (2020). *Predicting covid-19 pneumonia severity on chest x-ray with deep learning*. Retrieved from <https://paperswithcode.com/paper/predicting-covid-19-pneumonia-severity-on> doi: <https://doi.org/10.48550/arXiv.2005.11856>
- Darapaneni, N., Maram, S., Singh, H., Subhani, S., Kour, M., Nagam, S., & Paduri, A. R. (2022). *Prediction of covid-19 using chest x-ray images*. Retrieved from <https://paperswithcode.com/paper/prediction-of-covid-19-using-chest-x-ray> doi: <https://doi.org/10.48550/arXiv.2204.03849>
- Durrani, M., Shahid, A., Tahir, U. K., Haq, I.-u., Yousaf, A., & Naveed, S. (2021, November). Comparison of chest x-rays findings in covid-19 suspected and confirmed cases at a university teaching hospital: A retrospective comparative study. *Pakistan Journal of Medical Sciences*, 38(1). Retrieved from <http://dx.doi.org/10.12669/pjms.38.1.4624> doi: 10.12669/pjms.38.1.4624
- Farooq, M., & Hafeez, A. (2020). *Covid-resnet: A deep learning framework for screening of covid19 from radiographs*. Retrieved from <https://paperswithcode.com/paper/covid-resnet-a-deep-learning-framework-for> doi: <https://doi.org/10.48550/arXiv.2003.14395>
- for Disease Control, C., & Prevention. (2022, 10). *Coronavirus disease 2019 (covid-19) – symptoms*. CDC. Retrieved from <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>
- for Disease Control, C., & Prevention. (2023, 07). *About covid-19*. Retrieved from <https://www.cdc.gov/coronavirus/2019-ncov/your-health/about-covid-19.html>
- Hemdan, E. E.-D., Shouman, M. A., & Karar, M. E. (2020). *Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images*. Retrieved from <https://paperswithcode.com/paper/covidx-net-a-framework-of-deep-learning> doi: <https://doi.org/10.48550/arXiv.2003.11055>
- Hu, B., Guo, H., Zhou, P., & Shi, Z.-L. (2020, 10). Characteristics of sars-cov-2 and covid-19. *Nature Reviews Microbiology*, 19, 141-154. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7537588/> doi: 10.1038/s41579-020-00459-7
- Hu, T., Liu, Y., Zhao, M., Zhuang, Q., Xu, L., & He, Q. (2020, 08). A comparison of covid-19, sars and mers. *PeerJ*, 8. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7443081/> doi: 10.7717/peerj.9725
- King, A. (2020, 05). An uncommon cold. *New Scientist*, 246, 32-35. doi: 10.1016/s0262-4079(20)30862-9
- Krishnaraj, A., & Matsumoto, A. (2020, 09). *Covid-19 and imaging: Why ct scans and x-rays have a limited role in diagnosing coronavirus*. Retrieved from <https://blog.radiology.virginia.edu/covid-19-and-imaging/>
- Kundu, N. (2023, 01 23). *Exploring resnet50: An in-depth look at the model architecture and code implementation*. Retrieved from <https://medium.com/@nitishkundu1993/exploring-resnet50-an-in-depth-look-at-the-model-architecture-and-code-implementation-d8d8fa67e46f>
- Li, X., Li, C., & Zhu, D. (2020). *Covid-mobileexpert: On-device covid-19 patient triage and follow-up using chest x-rays*. Retrieved from <https://paperswithcode.com/paper/covid-xpert-an-ai-powered-population> doi: <https://doi.org/10.48550/arXiv.2004.03042>
- Motamed, S., Rogalla, P., & Khalvati, F. (2020). *Randgan: Randomized generative adversarial network for detection of covid-19 in chest x-ray*. Retrieved from <https://paperswithcode.com/paper/randgan-randomized-generative-adversarial> doi: <https://doi.org/10.48550/arXiv.2010.06418>
- Nasiri, H., & Hasani, S. (2022, August). Automated detection of covid-19 cases from chest x-ray images using deep neural network and xgboost. *Radiography*, 28(3), 732–738. Retrieved from <http://dx.doi.org/10.1016/j.radi.2022.03.011> doi: 10.1016/j.radi.2022.03.011
- Nayak, S. R., Nayak, D. R., Sinha, U., Arora, V., & Pachori, R. B. (2021). Application of deep learning techniques for detection of covid-19 cases using chest x-ray images: A comprehensive study. *Biomedical Signal Processing and Control*, 64, 102365. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1746809420304717> doi: <https://doi.org/10.1016/j.bspc.2020.102365>
- Organization, W. H. (n.d.). *Middle east respiratory syndrome coronavirus (mers-cov)*. Retrieved from https://www.who.int/health-topics/middle-east-respiratory-syndrome-coronavirus-mers#tab=tab_1
- Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Rajendra Acharya, U. (2020). Automated

- detection of covid-19 cases using deep neural networks with x-ray images. *Computers in Biology and Medicine*, 121, 103792. Retrieved from <https://paperswithcode.com/paper/automated-detection-of-covid-19-cases-using> doi: <https://doi.org/10.1016/j.combiomed.2020.103792>
- Page, J., Hinshaw, D., & McKay, B. (2021, 02). In hunt for covid-19 origin, patient zero points to second wuhan market. *Wall Street Journal*. Retrieved from <https://www.wsj.com/articles/in-hunt-for-covid-19-origin-patient-zero-points-to-second-wuhan-market-11614335404>
- Pasley, J. (2020, 02). *How sars terrified the world in 2003, infecting more than 8,000 people and killing 774*. Retrieved from <https://www.businessinsider.com/deadly-sars-virus-history-2003-in-photos-2020-2>
- Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Abul Kashem, S. B., ... Chowdhury, M. E. (2021). Exploring the effect of image enhancement techniques on covid-19 detection using chest x-ray images. *Computers in Biology and Medicine*, 132, 104319. Retrieved from <https://www.sciencedirect.com/science/article/pii/S001048252100113X> doi: <https://doi.org/10.1016/j.combiomed.2021.104319>
- Signoroni, A., Savardi, M., Benini, S., Adami, N., Leonardi, R., Gibellini, P., ... Farina, D. (2021, July). Bs-net: Learning covid-19 pneumonia severity on a large chest x-ray dataset. *Medical Image Analysis*, 71, 102046. Retrieved from <http://dx.doi.org/10.1016/j.media.2021.102046> doi: 10.1016/j.media.2021.102046
- Suresh, A. (2020, 11). *What is a confusion matrix?* Retrieved from <https://medium.com/Analytics-vidhya/what-is-a-confusion-matrix-d1c0f8fed45>
- Tartaglione, E., Barbano, C. A., Berzovini, C., Calandri, M., & Grangetto, M. (2020, September). Unveiling covid-19 from chest x-ray with deep learning: A hurdles race with small data. *International Journal of Environmental Research and Public Health*, 17(18), 6933. Retrieved from <http://dx.doi.org/10.3390/ijerph17186933> doi: 10.3390/ijerph17186933
- University, J. H. (n.d.). *Covid-19 dashboard*. Retrieved from <https://gisanddata.maps.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6>
- Voxco. (2021, 12). *Matthews's correlation coefficient: Definition, formula and advantages - voxco*. Retrieved from <https://www.voxco.com/blog/matthewss-correlation-coefficient-definition-formula-and-advantages/>
- Wang, L., & Wong, A. (2020). *Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images*. Retrieved from <https://paperswithcode.com/paper/covid-net-a-tailored-deep-convolutional> doi: <https://doi.org/10.48550/arXiv.2003.09871>
- Xu, Z., Shi, L., Wang, Y., Zhang, J., Huang, L., Zhang, C., ... Wang, F.-S. (2020, April). Pathological findings of covid-19 associated with acute respiratory distress syndrome. *The Lancet Respiratory Medicine*, 8(4), 420–422. Retrieved from [http://dx.doi.org/10.1016/S2213-2600\(20\)30076-X](http://dx.doi.org/10.1016/S2213-2600(20)30076-X) doi: 10.1016/S2213-2600(20)30076-x

A Appendix A: Developed Code for Pretrained Models

The code developed for the Pretrained models can be found here.

B Appendix B: Developed Code for COVID-ResNet Model

The code developed for the COVID-ResNet model can be found here.

C Appendix C: Developed Code for DarkCovidNet Model

The code developed for the DarkCovidNet model can be found here.

D Appendix D: Developed Code for DenseNet169 and XGBoost Model

The code developed for the DenseNet169 and XGBoost model can be found here.