Comparative Analysis of Deep Learning Models for COVID-19, Pneumonia, and Normal Chest X-ray Classification

Karim Hamdar
2092041

Davide Christian Mancosu Bustos
2089208

Abstract

The COVID-19 pandemic has put immense pressure on healthcare systems, emphasizing the need for rapid and reliable diagnostic tools. Chest X-ray imaging, as a non-invasive and widely accessible modality, plays a critical role in diagnosing respiratory conditions, including COVID-19 and pneumonia. Accurate classification of chest X-rays into COVID-19, pneumonia, and normal categories is essential for effective treatment, resource allocation, and management of patients. This study aims to evaluate and compare three deep learning architectures—ResNet, DenseNet, and custom CNNs—for their ability to classify chest X-rays into these three categories. We explore various preprocessing techniques, including resizing and padding, and assess the role of data augmentation in improving model robustness.

1. Introduction

1.1. Background and Motivation

The COVID-19 pandemic has underscored the need for effective diagnostic tools, with chest X-rays playing a pivotal role in diagnosing and managing not only COVID-19 but also other respiratory conditions such as pneumonia. Accurate classification of chest X-rays into categories such as COVID-19, pneumonia, and normal is essential for effective treatment and resource allocation.

1.2. Objectives

This study aims to develop and evaluate a robust classification system for chest X-ray images using advanced deep learning models. We compare state-of-the-art architectures, including DenseNet and ResNet, alongside custom Convolutional Neural Networks (CNNs). Our approach incorporates various preprocessing techniques, such as resizing and padding, to assess their impact on model performance and robustness against image variability.

1.3. Key Findings

ResNet models consistently delivered the highest accuracy and AUC scores, particularly when images were resized, highlighting their ability to generalize well across the dataset. DenseNet also demonstrated strong performance, especially with padded images, achieving high AUC scores. In contrast, the custom CNN models showed more variability, with the best results occurring when images were resized. Preprocessing techniques played a crucial role in model performance, with resizing generally proving more effective than padding. The impact of data augmentation was mixed, improving performance in some cases but slightly reducing it in others.

1.4. Organization of the Paper

The paper is structured as follows: Section 2 reviews related work in the field of chest X-ray classification using deep learning. Section 3 describes the dataset and preprocessing methods used. Section 4 details the architectures of the models implemented. Section 5 presents the experimental setup and results. Finally, Section 6 concludes the study and suggests directions for the Section 7 Future Work.

2. Related Work

The use of deep learning models for chest X-ray classification has seen a surge in interest due to the COVID-19 pandemic. Several studies have explored the application of deep learning architectures such as ResNet, DenseNet, and custom CNNs to classify X-ray images for the detection of COVID-19 and other respiratory conditions, particularly pneumonia.

One of the pioneering works in this area is by **Cohen et al. (2020)** [1], who developed a publicly available COVID-19 chest X-ray and CT image database. Their research emphasized the need for diverse datasets to build effective predictive models. This database has since been widely used in subsequent studies, including this one, to train deep learning models for fast, non-invasive, and cost-effective diagnostic tools. Cohen et al. [2] also explored the use of DenseNet with transfer learning to predict the severity of COVID-19-induced pneumonia, achieving a notable accuracy of 80.6. This inspired our approach of implementing DenseNet in our study to evaluate its effectiveness alongside other models.

In another key study, **Wang et al. (2020)** [3] introduced COVID-Net, a custom deep learning architecture specifically designed for COVID-19 detection from chest X-rays. Trained on a large dataset of 13,975 chest X-ray images, COVID-Net achieved a significant accuracy of 93.3, outperforming other popular architectures such as ResNet-50 and VGG-19. Wang et al.'s approach incorporated a lightweight model design tailored for clinical deployment, making their work particularly influential in optimizing the CNN architecture used in our experiments.

Our research builds upon these foundations by systematically comparing ResNet and DenseNet models with a custom CNN to assess their performance in classifying chest X-ray images into COVID-19, pneumonia, and normal categories. Unlike previous studies that primarily focused on a single architecture or dataset, our study evaluates multiple preprocessing techniques—such as resizing and padding—along with the effects of data augmentation. This comprehensive comparison aims to provide a deeper understanding of how different architectures and data preparation methods impact model performance, filling a gap in the literature regarding the effects of image preprocessing on deep learning models for chest X-ray classification.

By leveraging a balanced dataset from the Kaggle "COVID-19 Radiography Database" and extending the work of Cohen et al. and Wang et al., our study contributes valuable insights into the optimal combinations of deep learning architectures and preprocessing techniques for improving diagnostic accuracy in real-world medical settings. Furthermore, while previous studies have highlighted the importance of model accuracy, our research also places significant emphasis on computational efficiency, which is crucial for deploying these models in resource-limited environments.

3. Dataset and Preprocessing

The dataset used in this study is the publicly available *COVID-19 Radiography Database* from Kaggle, which was curated specifically for the classification of chest X-ray images into three distinct categories: COVID-19, pneumonia, and normal. The dataset comprises a total of 4,575 chest X-ray images, evenly distributed across the three classes, with 1,525 images in each category. This balanced distribution ensures that no class is overrepresented or underrepresented, reducing bias during the model training and evaluation phases.

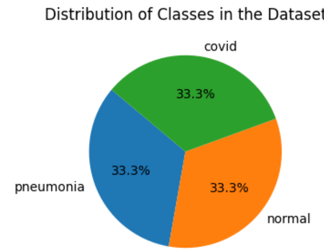

Figure 1. Balanced classes.

3.1. Dataset Composition

- **COVID-19 Images:** The dataset includes 1,525 chest X-ray images diagnosed as COVID-19. A significant portion of these images were collected from reputable repositories, including GitHub, Radiopaedia, The Cancer Imaging Archive (TCIA), and the Italian Society of Radiology (SIRM). Some images were augmented due to the limited availability of public COVID-19 images at the time of collection.

- **Pneumonia Images:** The pneumonia category consists of 1,525 chest X-ray images sourced from the Kaggle repository and the NIH Chest X-ray dataset.

- **Normal Images:** The remaining 1,525 images represent normal, healthy chest X-rays sourced from the Kaggle and NIH datasets.

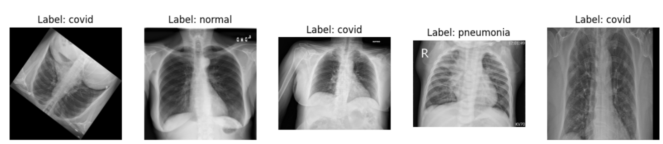

Figure 2. Images before preprocessing.

3.2. Preprocessing

Several preprocessing steps were applied to standardize and prepare the dataset for training deep learning models:

Image Resizing: All images were resized to 224x224 pixels to ensure compatibility with the deep learning models. This step was essential for reducing computational complexity and standardizing input dimensions.

Padding: Symmetrical padding was applied to non-square images to preserve the original aspect ratio, resulting in images of 224x224 pixels. This helped avoid distortions and ensured central image features remained intact.

Grayscale Conversion: All images were converted to grayscale to maintain consistency, given that some were originally in RGB format while others were grayscale.

Normalization: Pixel values were normalized to a range between 0 and 1, which helps improve model convergence during training by ensuring all inputs are on a comparable scale.

Data Augmentation: Various augmentation techniques, such as random rotations, horizontal and vertical flips, and scaling, were applied to the training set. These operations were implemented using TensorFlow's Keras `ImageDataGenerator` to increase data variability and improve model generalization.

Quality Check: Each image was visually inspected to ensure consistency and to eliminate any corrupted or irrelevant data. This quality control ensured that all images were ready for the training and testing processes.

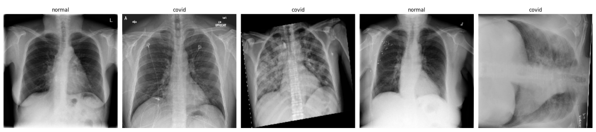

Figure 3. Resized images.

Figure 4. Padded images.

3.3. Dataset Splitting

The dataset was split into train and test in proportions 80 and 20 respectively. The train dataset was then split into train (80) and validation (20) using ImageDataGenerator from Keras library.Stratified splitting ensured that the proportion of each class was consistent across the training, validation, and test sets.

Stratified splitting ensured that the proportion of each class was consistent across the training, validation, and test sets.

3.4. Challenges and Considerations

Due to the visual similarities between COVID-19 and pneumonia, preserving image integrity through preprocessing was critical. Resizing and padding helped ensure the retention of important features, while grayscale conversion and normalization reduced computational complexity. Data augmentation further enhanced the robustness of the models by introducing variability, which helped the models generalize better.

These preprocessing steps were essential for ensuring data consistency and quality, allowing for more accurate training and evaluation of the models.

4. Methods

In this study, we implemented and compared three deep learning architectures—ResNet, DenseNet, and a custom Convolutional Neural Network (CNN)—for the task of chest X-ray image classification into three categories: COVID-19, pneumonia, and normal. The objective was to evaluate the effectiveness of these models and the impact of various preprocessing techniques and data augmentation strategies on model performance.

4.1. Model Architectures

4.1.1 Residual Network (ResNet)

ResNet is a state-of-the-art deep learning architecture known for its use of residual connections, which mitigate the vanishing gradient problem in deep networks by allowing gradients to flow more easily through the layers. In this study, we used the ResNet-50 architecture due to its demonstrated effectiveness in medical image classification tasks. ResNet's architecture consists of multiple convolutional layers interspersed with identity mappings (shortcut connections) to help preserve the information from earlier layers. This allows the model to efficiently handle the complex features in chest X-ray images.

4.1.2 DenseNet

DenseNet introduces a more compact architecture compared to ResNet, with its key innovation being dense connections between layers. Each layer in DenseNet receives the feature maps from all preceding layers, which enhances feature reuse and improves gradient flow across the network. For this study, we implemented DenseNet-121 due to its balance between accuracy and computational efficiency, and its effectiveness in capturing fine details in chest X-ray images.

4.1.3 Custom Convolutional Neural Network (CNN)

A custom CNN was designed specifically for this classification problem. It consists of several convolutional layers, followed by max-pooling and fully connected layers. We experimented with different filter sizes and dropout rates to prevent overfitting. The custom CNN is less complex than ResNet and DenseNet but provides computational efficiency. This architecture serves as a baseline to compare the performance of more advanced models.

4.2. Preprocessing and Data Augmentation

Preprocessing techniques, including image resizing, padding, grayscale conversion, and normalization, were applied to the chest X-ray images. Data augmentation techniques, such as random rotations, horizontal and vertical

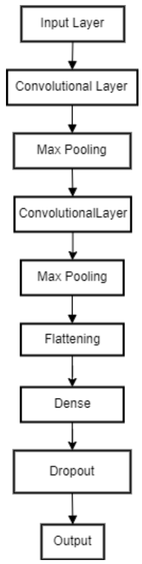

Figure 5. CNN.

flips, and scaling, were also employed to increase data variability and enhance model generalization. These augmentation methods were implemented using TensorFlow's Keras `ImageDataGenerator`.

Data augmentation was applied only to the training set to prevent overfitting, while the validation and test sets remained unaugmented to ensure consistent evaluation of model performance.

4.3. Training Setup and Hyperparameters

Each model was trained using the Adam optimizer, with a learning rate initially set to 1e-5. Categorical cross-entropy was used as the loss function. Early stopping was applied based on validation loss, and all models were trained for up to 100 epochs.

4.4. Evaluation Metrics

We evaluated model performance using several metrics:

- **Accuracy:** Proportion of correctly classified images.
- **Precision:** Proportion of true positive predictions among all positive predictions.
- **Recall:** Proportion of true positives identified out of all actual positives.
- **F1-Score:** Harmonic mean of precision and recall.
- **Area Under the Curve (AUC):** Evaluates the model's ability to distinguish between the three classes.

4.5. Rationale for Model Selection and Alternative Approaches

ResNet and DenseNet were selected due to their success in previous medical imaging tasks, particularly for their ability to capture complex features in chest X-rays. The custom CNN was included to provide a computationally efficient alternative and serve as a baseline for comparison. We considered other architectures, such as VGGNet and Inception, but opted for ResNet and DenseNet due to their balance between accuracy and computational cost. Although we trained models from scratch, transfer learning was considered as an alternative approach. Transfer learning involves fine-tuning pre-trained models on large datasets, such as ImageNet, for specific tasks. However, for this study, training from scratch allowed us to evaluate the impact of preprocessing and data augmentation on model performance directly.

5. Experiments

In this section, we describe the experiments conducted to evaluate the performance of three deep learning architectures—ResNet, DenseNet, and a custom Convolutional Neural Network (CNN)—on the classification of chest X-ray images into three categories: COVID-19, pneumonia, and normal. Our experiments were designed to assess the impact of different preprocessing techniques, data augmentation strategies, and model architectures. Additionally, an ablation study was conducted to determine the significance of various preprocessing techniques and architectural components.

5.1. Experimental Setup

The experiments were conducted on the Kaggle COVID-19 Radiography Database, which contains 4,575 X-ray images evenly distributed across the three classes (COVID-19, pneumonia, and normal). All models were trained from scratch for 100 epochs using the Adam optimizer with a learning rate of 1e-5, categorical cross-entropy as the loss function, and early stopping based on validation loss. TensorFlow's Keras framework was used to implement and train the models.

5.2. Experiment 1: Impact of Preprocessing Techniques

Preprocessing is crucial for ensuring that the models can learn from standardized and consistent data. In this experiment, we evaluated two preprocessing techniques: image resizing and padding. All images were resized to 224x224 pixels, which matches the input size requirements for ResNet, DenseNet, and the custom CNN models. Padding was also applied to preserve the aspect ratio of non-square images.

- **Resizing:** The resized images ensured uniformity in the input dimensions, simplifying the learning process for the models.

- **Padding:** Padding was applied symmetrically to retain the original aspect ratio, followed by resizing to 224x224 pixels.

The results showed that resizing the images led to better overall performance across all three models. Specifically, ResNet achieved the highest accuracy and AUC when trained on resized images, while DenseNet and the custom CNN also demonstrated improved performance. Padding, while preserving the image aspect ratio, led to a slight drop in accuracy due to the introduction of irrelevant information in the padded regions.

5.3. Experiment 2: Effect of Data Augmentation

Data augmentation was applied during training to increase the variability of the input data and prevent overfitting. The augmentation techniques used included random rotations (up to 30 degrees), horizontal and vertical flips, and scaling (between 0.8 and 1.2 times the original size). The validation and test sets were not augmented to provide a consistent evaluation of model performance.

Two sets of experiments were conducted:

- **With Data Augmentation:** The models were trained with data augmentation applied to the training set.
- **Without Data Augmentation:** The models were trained without any augmentation.

The results were mixed. Data augmentation significantly improved the robustness of ResNet and DenseNet models, particularly in terms of recall and AUC. However, the custom CNN exhibited a slight decrease in performance when data augmentation was applied. Overall, the use of data augmentation was more beneficial for ResNet and DenseNet, improving their ability to generalize.

5.4. Experiment 3: Model Comparisons

We compared the performance of ResNet, DenseNet, and the custom CNN across various configurations, including:

- Preprocessing techniques (resizing and padding)
- Data augmentation (with and without augmentation)

The performance of the models was evaluated using accuracy, precision, recall, F1-score, and AUC. Table 1 summarizes the results for each configuration.

The ResNet model with resized images and no data augmentation achieved the best overall performance, with an accuracy of 93.1% and an AUC of 0.9838. DenseNet also performed well, particularly when padded images were used. The custom CNN showed more variable performance, with the best results achieved with resized images without augmentation.

Model	Preprocessing	Precision	Recall	Accuracy	AUC	Loss
CNN	Augmented, Resized	0.8453	0.7649	0.8217	0.9387	0.4912
CNN	Augmented, Padded	0.6939	0.4826	0.5256	0.7594	0.9002
CNN	Resized, No Augmentation	0.9239	0.9191	0.9208	0.9799	0.2709
CNN	Padded, No Augmentation	0.8431	0.8168	0.8282	0.9380	0.4951
DenseNet	Augmented, Resized	0.8592	0.8237	0.8411	0.9524	0.4201
DenseNet	Augmented, Padded	0.8970	0.8734	0.8884	0.9623	0.3718
DenseNet	Resized, No Augmentation	0.9121	0.9013	0.9047	0.9770	0.2814
DenseNet	Padded, No Augmentation	0.9229	0.9153	0.9162	0.9805	0.2592
ResNet	Augmented, Resized	0.8853	0.8527	0.8727	0.9731	0.3148
ResNet	Augmented, Padded	0.9067	0.8951	0.8975	0.9772	0.2978
ResNet	Resized, No Augmentation	0.9311	0.9278	0.9310	0.9838	0.2741
ResNet	Padded, No Augmentation	0.9272	0.9272	0.9272	0.9822	0.2831

Table 1. Model performance metrics for different preprocessing and augmentation configurations.

5.5. Ablation Study: Impact of Preprocessing and Augmentation

We conducted an ablation study to investigate the significance of different preprocessing techniques and data augmentation. We systematically removed certain components (e.g., resizing, augmentation) to observe their impact on model performance. The results showed that:

- **Resizing:** Removing resizing led to a significant drop in performance across all models. For example, the accuracy of ResNet dropped from 93.1% to 89.7%.
- **Padding:** Padding introduced unnecessary information in some cases, leading to lower accuracy, particularly for CNN models.
- **Data Augmentation:** Data augmentation improved generalization for ResNet and DenseNet but reduced performance for the custom CNN.

The ablation study confirmed that resizing and data augmentation were critical for improving the robustness of ResNet and DenseNet.

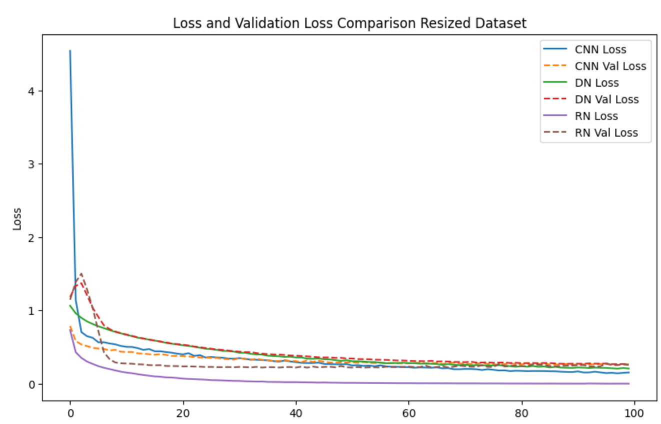

Figure 6. Loss comparison of Resized

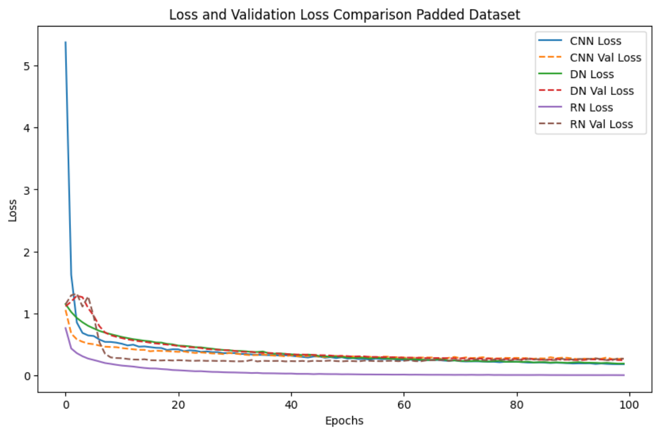

Figure 7. Loss comparison of Padded

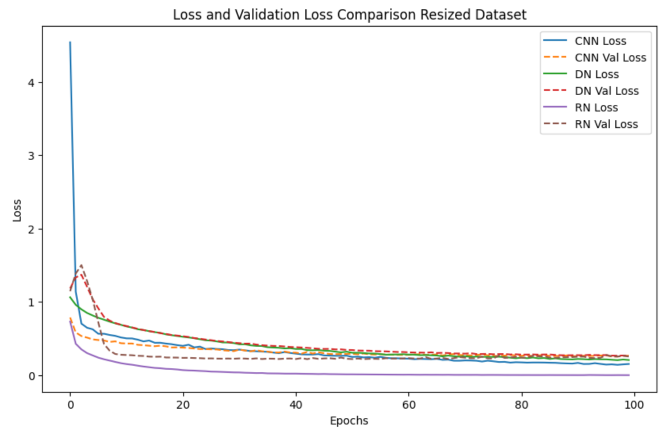

Figure 10. AUC comparison of Padded

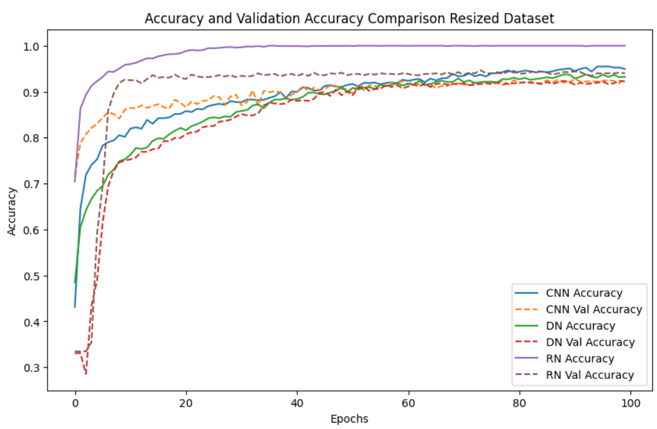

Figure 8. Accuracy comparison of Resized

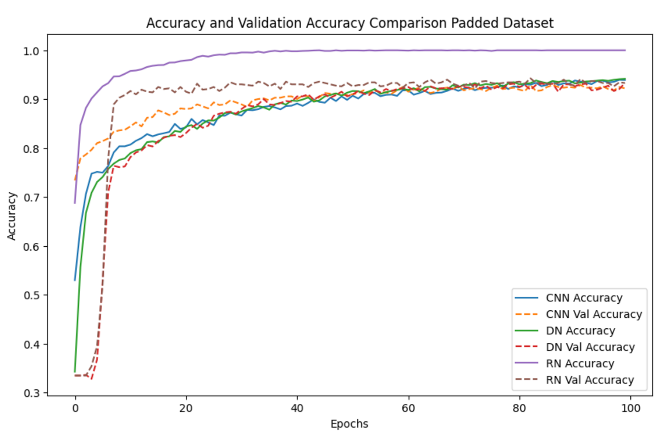

Figure 9. Accuracy comparison of Padded

5.6. Discussion and Comparison with Prior Work

Our experimental results are consistent with prior studies, such as those by Cohen et al. and Wang et al., which demonstrated the superior performance of ResNet and DenseNet in chest X-ray classification tasks. Wang et al.'s COVID-Net achieved an accuracy of 93.3%, comparable to the 93.1% accuracy obtained by our ResNet model. Our study expands on previous work by systematically comparing preprocessing techniques and data augmentation strategies across multiple architectures.

6. Conclusion

This study aimed to develop and evaluate deep learning models for the classification of chest X-ray images into three categories: COVID-19, pneumonia, and normal. We implemented and compared three architectures—ResNet, DenseNet, and a custom Convolutional Neural Network (CNN)—and assessed the impact of preprocessing techniques and data augmentation strategies on model performance.

The key results of our experiments indicate that **ResNet consistently outperformed** the other models across all metrics, achieving the highest accuracy (93.1%) and AUC (0.9838) when trained on resized images without augmentation. **DenseNet** also demonstrated strong performance, particularly when padded images were used, showing that its densely connected layers helped capture critical features, especially in more challenging image configurations. The custom CNN, while computationally efficient, exhibited **variable performance**, with the best results achieved when trained on resized images without augmentation.

One of the most important findings from this study is the significant influence of **preprocessing techniques** on model performance. Image resizing generally led to better results across all models compared to padding, which occasionally introduced unnecessary information. **Data augmentation** had a mixed impact: it significantly improved generalization and robustness in the more complex models (ResNet and DenseNet) but slightly reduced performance in the simpler custom CNN.

7. Future Work

Future research should focus on expanding the dataset to include a larger variety of chest X-ray images, particularly for rare cases, which could further improve the generalization of the models. Additionally, exploring **transfer learning**—leveraging pre-trained models on large-scale medical datasets—could potentially enhance classification accuracy with fewer labeled images. Finally, experimenting with **more advanced augmentation techniques** and newer deep learning architectures, such as EfficientNet or Transformer-based models, may lead to improvements in both accuracy and computational efficiency, making these models more applicable for real-time clinical settings.

References

[1] JP Cohen, L Dao, K Roth, et al. Predicting covid-19 pneumonia severity on chest x-ray with deep learning. *Cureus*, 12(7):e9448, July 2020.

[2] Joseph Paul Cohen et al. Covid-19 image data collection: Prospective predictions are the future. *arXiv preprint arXiv:2006.11988*, 2020.

[3] L Wang et al. Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports*, 10:19549, 2020.