**COVID-19 CLASSIFICATION IN X-RAYS**

**Entry**

In December 2019, a new infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first identified in Wuhan, China. This disease, now known as COVID-19, rapidly spread throughout the world, resulting in a global pandemic that has affected millions of people and caused widespread mortality.

One of the most common methods for diagnosing COVID-19 is the use of X-radiation (X-ray). X-rays can provide detailed images of the chest and can be used to identify characteristic features associated with COVID-19, such as ground-glass opacities [A] and consolidation [B]. However, the interpretation of X-ray can be challenging, particularly in cases where the disease is mild or where there is overlap with other chest diseases.

To address these challenges, researchers have turned to machine learning techniques to develop algorithms that can automatically classify X-ray as either COVID-19 positive or negative. One such technique is the use of convolutional neural networks (CNNs), which have shown promise in a range of computer vision applications, including medical image analysis.

**Localization and relevance of the topic**

The use of image classification techniques and CNN models for COVID-19 prediction from X-ray falls under the larger field of medical image analysis. This field involves the development of algorithms and techniques for the analysis and interpretation of medical images, including X-rays, MRI scans, and CT scans.

The COVID-19 pandemic has highlighted the urgent need for accurate and efficient methods of COVID-19 diagnosis, particularly as the disease can be asymptomatic [C] or present with mild symptoms. X-ray have been shown to be a valuable tool in the diagnosis of COVID-19, but their interpretation can be challenging and time-consuming. By developing machine learning algorithms that can accurately classify X-ray as COVID-19 positive or negative, we can improve the efficiency and accuracy of COVID-19 diagnosis, allowing for earlier detection and treatment of the disease.

The insights that this work could bring are significant, including the potential for earlier detection of COVID-19 and improved accuracy in the interpretation of X-rays. This could have important implications for patient outcomes, as early detection and treatment of COVID-19 can be critical in preventing the progression of the disease.

**Working hypothesis**

The aim of this project is to develop and evaluate an image classification model based on Convolutional Neural Networks (CNNs) for COVID-19 prediction in X-rays. We will use a CNN to classify X-ray as COVID-19 positive or negative with certain accuracy.

COVID-19 infection causes characteristic changes in the chest that can be detected on X-ray. By training a CNN model on a dataset of X-rays, we aim to develop an accurate and efficient method for COVID-19 prediction in X-rays.

**Methodology**

To develop our image classification model for COVID-19 prediction in X-rays, we have selected deep learning approach based on Convolutional Neural Networks (CNN). CNNs have been shown to be highly effective for image classification tasks, making them a natural choice for analyzing medical images.

Our methodology consists of the following steps:

1. Data Collection: We collected an open-source dataset of chest X-rays from COVID-19 positive and negative patients from Kaggle.
2. Splitting data into train, test, and validation sets: We divided the dataset into three subsets: a training set to train the model, a validation set to tune the model hyper parameters [D], and a test set to evaluate the performance of the model.
3. Data Augmentation: We augment the training set by applying various transformations such as rotation, flipping, and cropping. This helps to increase the size of the training set and improve the model's ability to generalize.
4. Data Transformation: We transform all the sets by resizing the images to a fixed size and normalizing the pixel values.
5. Training the pre-trained ResNet18 architecture model: We fine-tuned a pre-trained ResNet18 CNN model on our training and validation data using transfer learning [E]. This approach allows us to leverage the pre-trained model's weights and optimize the model for X-rays dataset.
6. Evaluating predictions on test data with a confusion matrix: We used a confusion matrix to evaluate the model's performance on the test set. The confusion matrix helps us determine how well the model has learned to distinguish between COVID-19 positive and negative X-rays.

**State of research**

The field of COVID-19 diagnosis using deep learning techniques has seen a surge in research since 2020. Several studies have reported high accuracy rates for COVID-19 prediction using deep learning models trained on CT scans and X-rays of chests and lungs.

One of the earliest studies on COVID-19 prediction using deep learning was conducted by Wang et al. in March 2020, where they used a deep learning model to classify COVID-19 patients from non-COVID-19 patients using CT scans. They reported an accuracy of 79.3% on test dataset. [2]

Subsequent studies have explored the use of deep learning models for COVID-19 prediction using both CT scans and X-rays. A study by Apostolopoulos and Mpesiana in April 2020 explored the use of CNNs for COVID-19 prediction using chest X-rays and reported an accuracy of 97.82% [3]

Another study by Li et al. in May 2020 used transfer learning to develop a deep learning model for COVID-19 prediction using CT scans. They reported an accuracy of 96%. [4]

These studies, along with several others, demonstrate the potential of deep learning models for COVID-19 prediction using CT scans and X-rays. However, there is still a need for further research to improve the accuracy and efficiency of these models, especially in real-world clinical settings.

**References**

[1] Prashant Pandey (2021). Chest X-Ray COVID-19, Pneumonia. Retrieved March 27, 2023, from Kaggle: <https://www.kaggle.com/datasets/prashant268/chest-xray-covid19-pneumonia>

[2] Wang, S., et al. (2020). A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). medRxiv.

[3] Apostolopoulos, I. D., & Mpesiana, T. A. (2020). Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Physical and Engineering Sciences in Medicine, 43(2), 635-640.

[4] Li, L., et al. (2020). Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. Radiology, 296(2), E65-E71.

**Definition of terms**

[A] Ground-glass opacities: Radiological finding in which the lungs appear hazy or cloudy on medical imaging, such as chest X-rays.

[B] Consolidation: Radiological finding in which a region of the X-ray appears denser or more opaque than normal.

[C] Asymptomatic: Refers to a medical condition in which a person does not experience any symptoms of a disease or condition, despite being infected or having the condition.

[D] Hyper parameters: Adjustable parameters that are set before the training of a deep learning model, and are not learned during the training process. These parameters determine the structure and complexity of the model

[E] Transfer learning: A machine learning technique in which a pre-trained model is used as the starting point for a new task, rather than training a new model from scratch.

**Structure of the work/Results**

The results of the study show that the CNN model developed using the ResNet-18 architecture was able to achieve an accuracy of 86% on the test set. The test set consisted of 50 normal and 50 infected X-ray images. The confusion matrix for normal X-rays showed that the model correctly classified 46 out of 50 normal images, which is a relatively high rate of accuracy. Similarly, the confusion matrix for infected X-rays showed that the model correctly classified 46 out of 50 infected images. This suggests that the model has a good ability to distinguish between normal and infected X-rays, and has potential for use in diagnosing COVID-19 in patients. However, it is important to note that the study was limited to a specific dataset and further validation is needed on larger and more diverse datasets to determine the generalizability of the model.