eISSN 2255-0569

ORIGINAL

A Simplified Approach to Covid-19 Pneumonia Classification

Un enfoque simplificado para la clasificación de neumonía por Covid-19

Ngonidzashe Mathew Kanyangarara , Rajendra Kumar

Department of Computer Science & Engineering, School of Engineering & Technology, Sharda University, Greater Noida, India.

Corresponding author
Rajendra Kumar
E-mail: rajendra04@gmail.com

Received: 9 - IV - 2023
Accepted: 5 - V - 2023

doi: 10.3306/AJHS.2023.38.05.117

Abstract

Introduction: The outbreak of Covid-19 has triggered a worldwide problem, especially in Asia and America. The World Health Organization (WHO) declared the sickness a pandemic on March 20, 2020. It arrived in waves, and most countries around the world have now experienced two waves and are on the approach of experiencing the third. The goal of this study is to build up and certify a Computer-Aided Diagnosis (CADx) system for distinguishing between COVID-19 positive patients and Non-Covid Patients people.
Methods: Chest X-ray (CXR) images are used to accomplish Covid-19 Pneumonia Classification. From public datasets GitHub 2295 CXR images were obtained which include 712 COVID-19 positive and 1583 normal cases. The proposed CADx system utilized a Conventional Neural Network (CNN) model for data argumentation and CNN is built, compiled and trained with help of Tensor flow and Keras. For the sake of appraisal, the dataset is estranged into three categories: Train, Test and Validation.
Results: The three sets accuracy was evaluated and the results for Training, Validation and Test were observed as 97.77%, 97.81% and 97.72%, respectively.
Conclusion: The presented study can create a precise Computer-Aided Diagnosis system for the two categories of classification.

Keywords: Computer-Aided Diagnosis (CADx), Conventional Neural Network (CNN), Keras and Tensorflow, Chest X-ray (CXR).

Resumen

Introducción: El brote de Covid-19 ha desencadenado un problema mundial, especialmente en Asia y América. La Organización Mundial de la Salud (OMS) declaró la enfermedad pandémica el 20 de marzo de 2020. Llegó en oleadas, y la mayoría de los países del mundo ya han experimentado dos oleadas y están a punto de experimentar la tercera. El objetivo de este estudio es construir y certificar un sistema de Diagnóstico Asistido por Ordenador (CADx) para distinguir entre pacientes COVID-19 positivos y personas No-Covid.
Métodos: Se utilizan imágenes de radiografía de tórax (CXR) para realizar la clasificación de la neumonía Covid-19. De los conjuntos de datos públicos GitHub se obtuvieron 2295 imágenes CXR que incluyen 712 COVID-19 positivos y 1583 casos normales. El sistema CADx propuesto utiliza un modelo de Red Neuronal Convencional (CNN) para la argumentación de datos y la CNN se construye, compila y entrena con la ayuda de Tensor flow y Keras. Para la evaluación, el conjunto de datos se divide en tres categorías: Entrenamiento, Prueba y Validación.
Resultados: Se evaluó la precisión de los tres conjuntos y los resultados de Entrenamiento, Validación y Prueba fueron del 97,77%, 97,81% y 97,72%, respectivamente.
Conclusiones: El estudio presentado puede crear un sistema preciso de Diagnóstico Asistido por Ordenador para las dos categorías de clasificación.

Palabras clave: Diagnóstico asistido por ordenador (CADx), Red neuronal convencional (CNN), Keras y Tensorflow, Radiografía de tórax (CXR).

Introduction

Pneumonia is a pervasive ailment caused by viruses, microorganisms, and fungus, among other microbes. The term "pneumonia" is derived from the Greek word "pneumon", which means "lung." As a result, the word pneumonia is linked to lung disease[14]. Pneumonia arises when a bacterial or viral contagion in the lungs produces noteworthy harm and irritation. The coronavirus is responsible for the damage to the lungs in COVID pneumonia[2]. When COVID pneumonia builds up, it causes extra signs, such as:

- Shortness of breath
- Increased heart rate
- Low blood pressure

AI-based computer-aided diagnostic tools are used to detect and treat different types of cancer, such as breast cancer and brain tumors. These tools are commonly referred to as deep learning (DL). Deep learning (DL) is a subset of machine learning (ML) stimulated as a result of the structure of the human brain. This is a subfield of machine learning that uses neural networks to interpret data related to the biology of a being's brain[3]. CNNs are the most widely used of these tools because they can analyze and interpret complex data. The Objectives of this research are:

- Have a model that is going accelerate prediction processes and to assist medical professionals.
- Create a CADx which will give an accurate classification between Covid-19 patients and Normal patients using CXR image.
- Make it easy even for leman to be able to check whether he/she is Covid-19 positive or negative, by just uploading the CXR image onto the CADx.

Related work

The use of deep learning (DL) in the healthcare sector and other image related tasks has increased exponentially over the last decade[12]. DL models have been shown to help classify Computed Tomography (CT) scans of pneumonia and Tuberculosis (TB), malignant cancer pictures, diabetic retinopathy, microbiological slide images, and many pattern/object recognition problems in various studies8,10. Pathologists, computer scientists, and radiologists work together in the field of pathology to diagnose diseases such as cancer, pneumonia, and tuberculosis using computer-assisted diagnostics.

Recent developments in a number of medical image-processing tasks involved algorithms that have outperformed using deep learning models and on enormous datasets. Some examples include the Classification of skin cancer5, the detection of diabetic retinopathy[15], etc. Automatic diagnosis is gaining popularity through chest radiography[7]. These algorithms are increasingly being used to detect pulmonary nodules and classify pulmonary tuberculosis. Using the published Open dataset, researchers have discovered that the same deep convolutional network design does not work for all anomalies; ensemble models outperform single models in classification accuracy, and deep learning improves classification accuracy dramatically. A method presented in is superior to the rule-based methods[13]. Many other applications also used training-based system for recognition[8,9] and classification purpose with the help of machine learning algorithms like artificial neural network (ANN)[6], convolution neural network (CNN). The CNN are used for optimizing the performance to enhance the models by changing hyperparameters and hidden layers. The problem of dealing with overfitting and model normalization during model training are generally done using dropout and batch normalization respectively[10,11].

Materials and methods

The results of extensive tests and assessments carried out to determine the viability of the suggested approach are reported in this study. To design and train the Convolutional Neural Network Model, we have used Keras, a deep learning framework that is open-source with a Tensor Flow API. All experiments are performed on a typical personal computer (Windows 10, 64-bit, CPU) equipped with graphics processing unit (GPU) and a Google Colab Notebook.

Datasets

The COVID-19 chest X-ray dataset[4] containing 2295 images is used in the study[1]. The original dataset is divided into two main folders (training and testing) and two subfolders containing Covid-19 Positive and Non-Covid chest X-ray images, respectively 2295 CXR images were obtained which included 712 COVID-19 positive and 1583 normal. For experimental purposes, images are split as 80% for training and 20% for validation. Each of the training set validation set and the testing set contains the subdirectories as COVID-19 and normal. The COVID-19 folder contains the COVID X-ray images, while the normal directory contains the negative X-ray images.

Pre-processing and Augmentations

To ensure that the datasets grew in size and quality, we used a variety of data augmentation techniques. This helped in reducing overfitting tribulations and enhances our model's overview aptitude while training. **Table I** presents the configurations used for image augmentation.

Firstly, well rescale of the data is done by a factor of 1/255. That will help us to do the normalization. We then divided our training into 20% of the validation set. In order to do that, we used the validation split constraint.

Tabla I: Image augmentation parameters

Parameter	Value
Rescale	1/255
Validation_split	0.2
Zoom Range	0.2
Horizontal Flip	True

The Model

Input Image

The accumulated input image is set for pre-processing to make the model's execution effective. To maximize the datasets, data augmentation techniques are used. Physical data collection is complex due to the COVID19 global pandemic[6]. The augmented data is sent to the preceding convolution to retrieve essential elements.

Convolution steps

Convolutions extract local aspects from large input data sets and multiply the resulting NxN matrices. Conv2D uses filters, kernel size, input shapes and activation functions for image classification1. In this study domain, the values of the variables are already stated in **table I**.

Max Pooling Layer

In advanced learning algorithms, max-pooling reduces dimensionality and extracts maximum features. The pooling layer reduces the number of variables and regularizes overfitting by finding the average of the provided elements[6].

Fully Connected Layer

It plots pooling layers, flattened and fed into the subsequent layer. The fully connected layer is essential for classification in CNN. The following classification outcome is performed with the aid of the activation function (sigmoid, Dense). The summary of layers of the proposed network architecture (CNN model) is shown in **Figure 1**.

For the mining of features from the input image, we used a modus operandi identified as convolution. The input is fed to the conv 2d block, after which ReLU (Rectified Linear Unit) is applied, followed by MaxPooling and lastly the dropout regularization. After this, the convolution is repeated but with more filters i.e., flattening them using the final layer into a one-dimensional array, which is heading to a fully connected layer.

Finally, an opaque (dense) layer with a sigmoid activation utility is deployed. So, in short, we feed the image into the conv2d for feature extraction. And lastly, that information is fed to the tightly connected artificial neural network. Supply forward neural architecture's output is presented in **figure 2**.

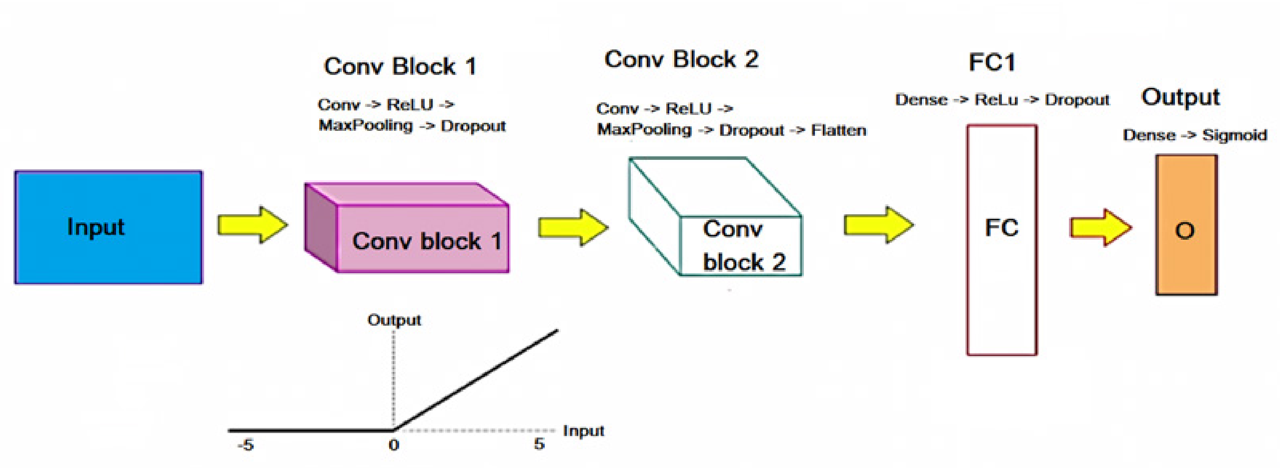
Figure 2: Training of the model.

Model" "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 150, 150, 32)	2432
max_pooling2d	(None, 75, 75, 32)	0
dropout (dropout)	(None, 75, 75, 32)	0
conv2d_1 (Conv2D)	(None, 75, 75, 64)	51264
max_pooling2d_1	(None, 37, 37, 64)	0
dropout_1 (dropout)	(None, 37, 37, 64)	0
flatten (flatten)	(None, 87616)	0
dense (dense)	(None, 256)	22429952
dropout_2 (dropout)	(None, 256)	0
dense_1 (dense)	(None, 1)	257

Total params: 22,483,905
Trainable params: 22,483,905
Non-trainable params: 0

Figure 1: The architecture of the proposed CNN.



Results

For the compilation of the model, we used Adam Optimizer with learning at a pace of 0.001. After that, we had to train our model for 30 epochs on all of the training images that we had. So, when we trained the model, four values per epoch were lost. The precision and loss of training data are very useful in assessing the effectiveness of training.

Figure 3: Training and validation loss.

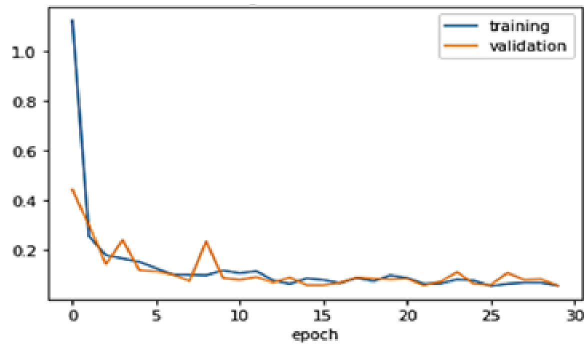


Figure 4: Training and validation accuracy.

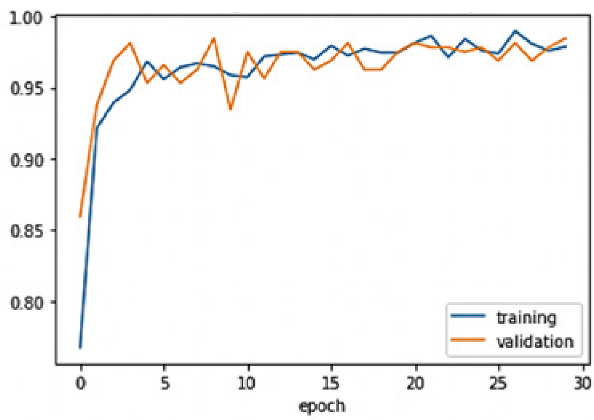
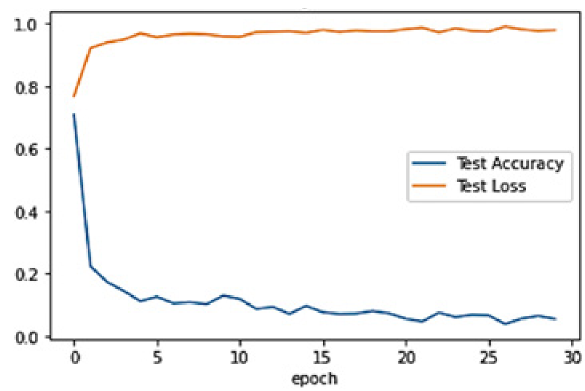


Figure 5: Test accuracy and test loss.



The accuracy, on the other hand, is the percentage of right guesses, and the validation accuracy is the measurement using data that has not been used in training. After training, we need to know how the model performed during the training phase. We plotted two graphs, one for loss, and the other one for accuracy (as shown in **figures 3, 4** and **5**). The accuracy of our final training observed was 97%. The validation accuracy was around 98%, while the training loss was only 0.0677%. And the validation was 0.0372. Therefore, a good accuracy of the test of roughly 97% was achieved and the test loss was around 0.7071%.

Discussion

From the chest X-ray pictures acquired from front views, we build a model to detect and categorize pneumonia for Covid-19. This technique starts by shrinking the chest X-ray pictures to a fraction of their original size. The photos are then enhanced by a convolutional neural network framework, which extracts and classifies characteristics from the images. When compared to other approaches, our model's validation accuracy was marginally greater.

We had to redo the model's training process numerous times before getting the same results each time. This will benefit developing countries with a doctor shortage, such as the majority of African countries. Significant improvements could be made if we had access to patient and non-patient statistics from around the world; however, our system is constrained because of short of information.

Conclusion

From a series of X-ray scans, the study demonstrated how to distinguish between positive and negative Covid-19 patients. The proposed model is implemented from the ground up, which distinguishes it from existing systems that heavily rely on transfer learning. The model is light CNN as it uses less number of layers and resulting faster training over any graphics processing unit (GPU). The accuracy of model is also very good to predict the classification accurately. The presented research is going to be expanded in the near future to perceive and classify X-ray images of Retinal Image Analysis, Skin Cancer Detection and more like Brain analysis.

Interests Conflict

The authors declare no conflict of interest.

References

1. Abdullahi U I, Mehmet O, Sertan S, Fadi A, Polycarp S Y. Pneumonia Classification Using Deep Learning from Chest X-ray Images During COVID-19. Cognitive Computing 2021; 11-3.

2. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan. Lancet, 2020; 395(10223): 507-13.

3. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface, 2018; 15(141): 20170387.

4. Cohen J. COVID-19 Chest X-ray Dataset. https://github.com/ieee8023/covid-chestxray-dataset

5. Esteva A, Kuprel B, Roberto AN, Justin K, Susan MS, Helen MB, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 2017; 542(7639): 115-8.

6. Gupta J K, Kumar R. An efficient ANN Based approach for Latent Fingerprint Matching. International Journal of Computer Application, 2010; 7(10): 18-21.

7. Kallianos K, Mongan J, Antani S, Henry T, Taylor A, Abuya J, et al. How far have we come? Artificial intelligence for chest radiograph interpretation. Clin Radiol, 2019; 74(5): 338-45.

8. Kalembo V S, Ngonidzashe M K, Prince T P, Gupta V, Kumar R.. A Deep Learning Model for Face Recognition in Presence of Mask. Acta Informatica Malaysia, 2022; 6(2): 38-41.

9. Kumar R, Singh R C, Khokher R. A systematic review of palm and dorsal hand vein recognition techniques. Academic Journal of Health Sciences, 2022; 37(1): 100-9.

10. Kumar R, Singh R C, Kant S. Dorsal Hand Vein Recognition Using Very Deep Learning. Macromolecular Symposia, 2021; 397(2000244): 1-13.

11. Kumar R, Singh R C, Kant S. Dorsal Hand Vein-Biometric Recognition using Convolution Neural Network. Adv. in Intell. Sys. and Comp. DOI: 10.1007/978-981-15-5113-0_92.

12. Mohammad T I, Md A A, Ahmed T M, Khalid A. Abnormality detection and localization in chest x-rays using deep convolutional neural networks. 2017; http://arxiv.org/abs/1705.09850.

13. Okeke S, Mangal S, Uchenna J M, Do-Un J. An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare. Journal of Healthcare Engineering, 2019; 4180949: 7.

14. Peng H, Seyoun P, Rongkai Y, Junghoon L, Linda C C, Cheng T L, et al. Added value of computer-aided CT image features for early lung cancer diagnosis with small pulmonary nodules: a matched case-control study. Radiology. 2018; 286(1):286-95.

15. Varun G, Lily P, Marc C, Martin C S, Derek W, Arunachalam N, et al W. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, JAMA, 2016; 316(22): 2402-10.